

Proceedings of the 43rd International Symposium on Automation and Robotics in Construction

Singapore, June 22-26, 2026

Zero-Shot Instance Segmentation of Shield Tunnel Linings via Adaptive SAM and Generative Inpainting

Chengxin Shi[1,2], Jiepeng Liu[1,2], Dongsheng Li[1,2], Pengkun Liu[1,2,3*], Yongjing Wang[4], Shan Huang[4]

[1]School of Civil Engineering, Chongqing University, Chongqing 400045, China
[2]State Key Laboratory of Safety and Resilience of Civil Engineering in Mountain Area, Chongqing 400045, China
[3]Hong Kong Center for Construction Robotics, The Hong Kong University of Science and Technology, Hong Kong, China
[4]China MCC5 Group Corporation Ltd., Chengdu 610000, China

scx@stu.cqu.edu.cn, liujp@cqu.edu.cn, lds@cqu.edu.cn, pengkunliu@cqu.edu.cn, 892238201@qq.com, 475347170@qq.com

Abstract -

Precise segmentation of shield tunnel linings is a prerequisite for structural health monitoring and digital twin construction. However, existing methods are heavily relying on extensive annotated datasets, which are scarce in tunneling scenarios. To address this, we propose a novel zero-shot framework for the adaptive recognition and instance segmentation of tunnel linings. The proposed method starts by unfolding tunnel point cloud data to an intensity image, enabling the use of a Vision-Language Model (VLM) and a prompt-based Segment Anything Model (SAM). To enhance performance, we employ (i) an adaptive hybrid prompt generation mechanism, integrating VLM with pixel intensity priors to automatically generate bounding box and point prompts for SAM and (ii) a Stable Diffusion inpainting module to restore missing regions caused by occlusions, enabling SAM's zero-shot capability to achieve precise instance-level segmentation of tunnel linings. Experimental results in two different tunnel datasets, compared to the original baseline SAM without prompts, demonstrate the robustness of the framework, achieving a mean Intersection over Union (mIoU) of 90.58% in standard circular tunnels. This approach effectively overcomes data scarcity issues and provides an efficient, largely automated solution for intelligent tunnel maintenance.

Keywords -

Shield Tunnel; Point Cloud; Zero-Shot Segmentation; Segment Anything Model; Vision-Language Models

1 Introduction

With the acceleration and deepening of urbanization, the effective utilization of urban underground space has become a key way to address issues such as traffic congestion, population density, and resource constraints [1] . However, such facilities typically exhibit characteristics that include diverse facility types, complex environments, and extensive coverage areas. Consequently, the challenges to ensuring tunnel operational safety and quality inspection remain worthy of attention.

Currently, conventional operational maintenance is largely based on manual inspections and fragmented digital recording. This approach mainly involves maintenance personnel performing periodic assessments on-site of facility conditions, with reports compiled manually [2, 3, 4]. Moreover, the tunnel environment is characterized by poor implementation, high humidity, and hazardous conditions, making manual inspection labor-intensive and risky [5]. To address these challenges, Yi et al. [6]extract segments from point clouds to construct the corresponding three-dimensional segment models, thereby obtaining precise structural models for subsequent digital twin-based tunnel operations and maintenance. Feng et al. [7] achieve a precise segmentation of the segments to allow detailed monitoring, such as calculating the position and orientation of the lining ring and analyzing joint deformation. These techniques offer valuable information for the comprehensive management of tunnels throughout their lifecycle.

However, both constructing digital twin geometric models and monitoring deformation fundamentally require precise segmentation of structural components. Existing tunnel lining segmentation methods primarily involve direct point cloud segmentation and projection-based segmentation. However, both approaches require extensive datasets for training, imposing significant limitations in data-scarce scenarios like tunnels.

The Segment Anything Model (SAM) proposed by Meta AI [8], demonstrates unprecedented zero-shot segmentation capabilities. It can precisely segment any object within an image based on simple user prompts (e.g., points, boxes, rough boundaries). Although this model exhibits strong generalizability, its performance may vary across specific scenarios. Preliminary research has explored the application of SAM's to remote sensing image analysis, medical image segmentation, and industrial defect detection[9], achieving favorable results due to SAM's robust pre-training foundation in image segmentation.

This paper proposes a novel point cloud segmentation method for tunnel linings, leveraging SAM's zero-shot segmentation capability to achieve high-precision automated segmentation. We employ (i) Vision-Language Model (VLM) with pixel intensity prior information to automatically generate bounding boxes and point prompts, aiming to enhance SAM's segmentation capabilities. (ii) We introduce the Stable Diffusion inpainting module to complete missing image portions, thereby improving the quality of images input to SAM. The final results demonstrate that our method performs effectively.

2 Related Work

Numerous research achievements have emerged in the field of point cloud segmentation, laying a solid foundation for technological advancement. Within the domain of

tunnel point cloud segmentation, approaches primarily fall into two categories.

The first category comprises direct segmentation methods based on point clouds, exemplified by Zhou et al.[10], whose research validated the application of deep learning in point cloud segmentation. The second category involves projection-based segmentation approaches, which reduce the dimensionality of three-dimensional data to two-dimensional images for processing. This reduction in dimensionality simplifies the segmentation challenge. For example, Zhang et al. [11] utilise traditional image processing techniques for segmentation, suitable for straightforward point cloud data.

However, current deep learning-based segmentation in tunneling still faces specific limitations, primarily in relation to scarce training data. Consequently, SAM's flexible prompting mechanism has been extended to specific engineering scenarios. Jing et al. [12] proposed a segmentation framework that integrates depth map information with visual SAM features, significantly improving the accuracy of the location of the edge for engineering components. Ye et al. [13] generated template prompts centered on initial points based on lining design parameters, supplemented by coarse mask prompts using polygonal envelopes, to achieve instance segmentation in shield-tunnel construction. Xu et al. [14], focusing on the implementation of engineering automation, further optimized the efficiency of inference and the compatibility of the deployment of large visual models. Their lightweight segmentation architecture achieved a real-time balance between component segmentation and condition monitoring in construction scenarios, providing technical support for intelligent supervision of construction processes.

Existing research indicates that segmentation methods based on large visual models are progressively transitioning from general-purpose segmentation to engineering-specific segmentation. However, existing SAM-based methods mainly rely on standard prompts and do not explicitly address the challenges of missing data and structural discontinuities in tunnel inspection images. To address these limitations, This study introduces an adaptive prompt generation module and an image restoration model, offering an enhanced solution for tunnel lining segmentation and advancing progress in this field.

3 Methodology

The proposed method is shown in Figure 1, consisting of four core steps: data preprocessing, PCD image to intensity and completion, SAM integration and optimization, and reprojection. Specifically, we preprocess the raw point cloud and perform a center axis fitting. Subsequently, the dimensionality reduction is achieved through cylindrical unfolding to generate intensity images. Finally, adaptive bounding boxes and point prompts are generated using VLMs and gray-value-driven techniques, which guide the SAM to perform zero-shot segmentation.

3.1 Data Preprocessing

Raw Point Cloud Data (PCD) from shield tunnels often contains non-structural artifacts such as equipment mounts and pipelines. To ensure data quality, we employ a statistical outlier removal filter combined with manual refinement using CloudCompare to eliminate noise while preserving the tunnel geometry.

Precise axis estimation is critical for subsequent unfolding. We propose an iterative least-squares fitting algorithm. First, Principal Component Analysis (PCA) provides an initial axis L_0. In each iteration i, the tunnel is sliced into normal planes S_i at equal intervals along L_i. For each plane, the cross-section center (a_i, b_i) and radius R_i are estimated by minimizing the geometric distance:

$$E(a_i, b_i, R_i) = \sum_{j=1}^{N_i} \left(\sqrt{(x_j - a_i)^2 + (y_j - b_i)^2} - R_i \right)^2 \tag{1}$$

The optimal centers are then fitted to form an updated spatial curve L_{i+1}. This process repeats until the deviation $d(L_{i+1}, L_i)$ falls below a convergence threshold ε, yielding the final spatial axis L and radius parameters.

3.2 PCD to Intensity Image and Completion

To take advantage of 2D foundation models, the 3D tunnel surface is projected onto a 2D plane. Based on the fitted axis L, we construct a local orthogonal basis $\{\mathbf{U}, \mathbf{E}, \mathbf{N}\}$ for each point p_i. The 3D coordinates are transformed into longitudinal distance u_i and azimuth angle θ_i:

$$u_i = (p_i - \mathbf{c}) \cdot \mathbf{U}, \quad \theta_i = \text{atan2}(\delta_i \cdot \mathbf{N}, \delta_i \cdot \mathbf{E}) \tag{2}$$

where $\mathbf{c}$ is the reference center, and δ_i is the radial vector.

The unfolded point set $P' = \{(u_i, \theta_i, I_i)\}$ is then rasterized in a 2D grid. The image width W and height H are determined by the angular resolution $\Delta\theta$ and the axial resolution Δu. To handle non-uniform point density, we employ a mean aggregation strategy where the pixel value $I_{\text{final}}[y, x]$ is calculated as the average intensity of all points falling within that grid cell:

$$I_{\text{final}}[y, x] = \frac{\sum I_i}{C[y, x]} \quad \text{if } C[y, x] > 0 \tag{3}$$

However, the generated intensity image inevitably contains void regions due to scanner blind spots, equipment occlusions, or extreme sparsity in the point cloud. These discontinuities disrupt the geometric integrity of the segments and can hinder the recognition performance of SAM.

To address this, we employ a generative inpainting approach based on the Stable Diffusion model. Specifically, we first generate a binary mask that identifies the missing regions. The binary mask for missing regions is automatically generated by identifying pixels with zero intensity (or below a sparsity threshold) resulting from the unfolding process. The incomplete intensity image and the mask are then fed into the Stable Diffusion Inpainting module. Using its powerful latent generative priors, the model synthesizes realistic textures to fill the masked areas based on contextual information from valid surrounding pixels. This process effectively restores the tunnel surface, providing a complete and high-quality input for the downstream segmentation task.

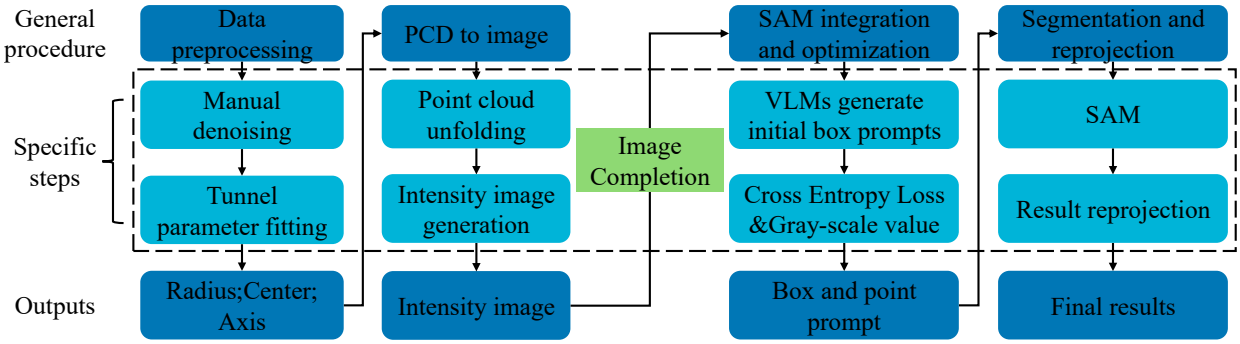


Figure 1. The framework of the proposed adaptive segmentation.

3.3 SAM integration and optimization

To enable precise zero-shot segmentation of shield tunnel segments, we propose an adaptive hybrid prompt generation module. Unlike standard SAM approaches that rely solely on generic visual foundation models, our method integrates open-vocabulary detection with domain-specific intensity priors to construct a robust prompt set:

$$P = \{V_{\text{box}}, V_{\text{point}}\} \tag{4}$$

for the SAM encoder. The specific model framework is shown in Figure 2.

3.3.1 Bounding Box Generation and Optimization

For the bounding box prompts V_{box}, we utilize the DINO-X model, which accepts the unfolded intensity image I and a text prompt T_k = "rectangles" as input. The model encodes image features $E_I(I)$ and text features $E_T(T_k)$, fusing them through a cross-modality decoder to predict initial candidate boxes. However, raw outputs often suffer from fragmentation. To address this, we introduce an optimization strategy based on cross-entropy scoring.

Specifically, we extract the classification logits from the detector head, which correspond to the cross-entropy loss used during pre-training. These logits serve as confidence scores S_{conf} for each candidate box. We first filter candidates based on a confidence threshold, and then apply an iterative merging algorithm. Boxes with high spatial overlap (IoU) are merged, prioritized by their cross-entropy scores, to form a purified set of bounding boxes V_{box}. These optimized boxes provide the coarse localization for the target segments.

3.3.2 Intensity-Driven Point Prompt Sampling

Although bounding boxes provide spatial limits, relying solely on them for segmentation often results in the background being segmented, affecting the segmentation accuracy. We formulate a heuristic sampling strategy to generate point prompts V_{point}. For each optimized bounding box, we analyze the pixel intensity distribution. Pixels falling within the background value range are sampled as negative prompts to suppress background noise. In contrast, pixels within the median intensity range of the segment are sampled as positive prompts. We randomly sample N positive points and M negative points per instance. Finally, the optimized boxes V_{box} and the intensity-driven points V_{point} are concatenated and fed into SAM to generate the final high-quality segmentation masks.

3.4 Reprojection

Upon leveraging the zero-shot capabilities of SAM to obtain high-quality 2D instance segmentation masks, the final step is to transfer these labels back to the original 3D point cloud space. This process closes the loop of our framework, allowing for the direct segmentation of the raw tunnel data.

Since our unfolding process (described in Section 3.2) maintains a deterministic mapping between the 3D points and the 2D grid pixels, the reprojection is formulated as a direct index lookup operation.

For every point p_i in the original point cloud P, we retrieve its corresponding pixel coordinates (x_i, y_i) calculated during the rasterization stage. The semantic label L_i for the 3D point p_i is then assigned based on the value of the corresponding pixel in the 2D mask:

$$L_i = \mathcal{M}_{SAM}(y_i, x_i) \tag{5}$$

By iterating through all points in P, we obtain a fully segmented 3D point cloud where each point carries an instance label consistent with the result of the 2D segmentation. This approach ensures that the precise boundaries delineated by SAM in the image domain are accurately preserved in the 3D geometric space.

4 Experiments and Results

4.1 Dataset and Setup

To evaluate the precision and robustness of the proposed method, we selected experimental data from two distinct tunnels. As shown in Figure 3, Tunnel A is a circular shield tunnel in Chongqing, while Tunnel B originates from the publicly available Seg2Tunnel T1 dataset [15], representing a horseshoe-shaped tunnel. For specific parameters, see Table 1.

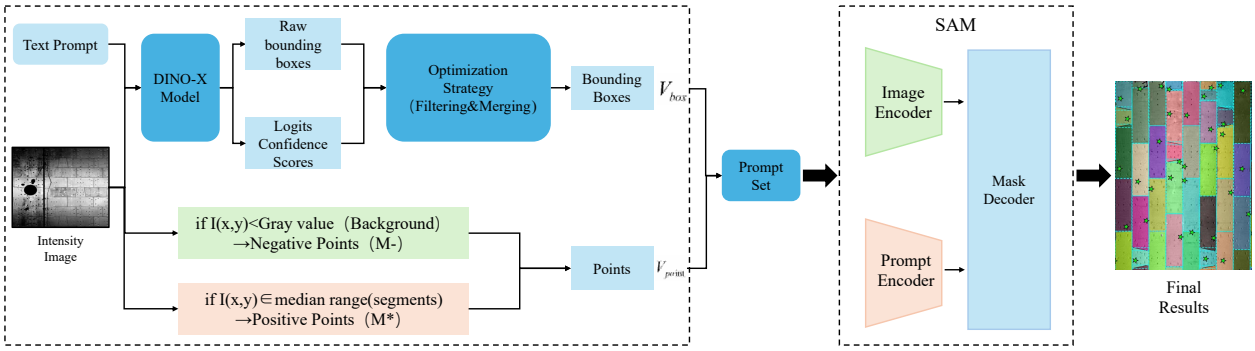


Figure 2. Optimized SAM Framework

Table 1. Key parameters of the shield tunnel datasets

Parameter	Tunnel A	Tunnel B
Design inner diameter (m)	7.5	5.5
Segments per ring	7	6
Segment width (m)	1.6	1.2
Total point count	16 million	14 million

All experiments were conducted with an NVIDIA GeForce RTX 3090 GPU (24GB VRAM). The model is implemented with the PyTorch 2.0 framework running on CUDA 11.8.

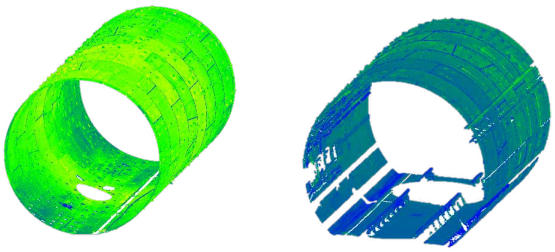


Figure 3. Overview of the experimental datasets.

4.2 Evaluation method

We offer Precision, Recall, Intersection over Union (IoU), and Mean Intersection over Union (mIoU) to quantify segmentation performance. The specific calculation formulas are as follows:

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{6}$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{7}$$

$$\text{IoU} = \frac{|M_{\text{GT}} \cap M_{\text{Pred}}|}{|M_{\text{GT}} \cup M_{\text{Pred}}|} \tag{8}$$

where:

- TP_i denotes the true positives (TP) for class i,
- FN_i denotes the false negatives (FN) for class i,
- FP_i denotes the false positives (FP) for class i,
- M_{GT} denotes the ground truth mask of the segment instance,
- M_{Pred} denotes the predicted mask of the segment instance.

To evaluate the overall quality of the segmentation, we calculate the average value of IoU in all annotated segment instances, i.e., mIoU:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \text{IoU}_i \tag{9}$$

where:

- N is the total number of annotated segment instances,
- IoU_i is the IoU value of the i-th segment instance.

For instance-level evaluation, predicted instances are matched to ground-truth instances using a one-to-one IoU-based strategy. A prediction is considered a true positive if its IoU with the best-matching ground-truth instance exceeds 0.5. Otherwise, it is counted as a false positive. Precision and Recall measure instance detection performance, whereas IoU evaluates the mask quality of matched instances.

4.3 Results and Analysis

For qualitative results, as illustrated in Figure4, we select Tunnel A as an example, presenting the unfolded intensity image, the segmentation predictions of the instance at the pixel-level, and the predicted point cloud reprojected. Among these, we evaluate the original SAM as a baseline to quantify . In this baseline, we employ the official model to perform fully automatic instance segmentation on the unfolded intensity images, without using any external prompts or inpainting modules. The ground truth for Tunnel A was manually annotated using the open-source software LabelMe, with labels numbered sequentially according to the number of segments, where 0 denotes background.

Table 2 summarizes the quantitative performance of our method in two datasets.

For circular Tunnel A, the method achieves a high mIoU of 90.58, with a precision of 87.95 and a recall of 97.33. The high values for precision and recall indicate that the model effectively identified most tunnel segment instances

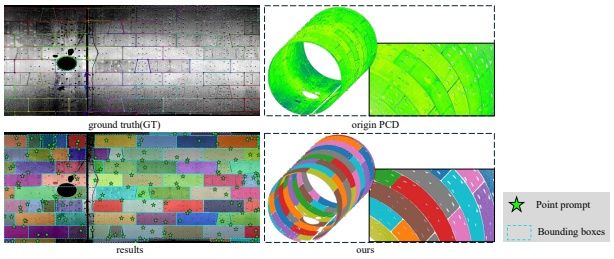


Figure 4. The segmentation results from tunnel A

(TP=73), with minimal false positives (FP=10) or false negatives (FN=2).

In contrast, performance for the horseshoe-shaped Tunnel B declined, achieving an mIoU of 75.59. While recall remained robust at 89.58, indicating the model successfully located most segments, precision dropped significantly to 75.44. This directly reflects a higher number of false positives (FP=14).

As shown in Figure5, the primary cause of this precision degradation is the extensive occlusion introduced by the equipment and cables attached to the tunnel lining. Unlike Tunnel A, these obstructions in Tunnel B result in large-scale data voids within the original point cloud.

Although our repair module mitigated this issue to some extent, the sheer size of the missing regions and the complex visual texture of the utility lines themselves posed significant challenges. Our method may misclassify missing background regions in intensity images as potential segmentation instances, thereby increasing the false positive (FP) rate. Nevertheless, recall remains robust at 89.58, demonstrating that even under severe occlusion, our method can still correctly segment a large number of tunnel segments, exhibiting significant potential for generalization.

Compared with the prompt-free SAM baseline, the proposed framework shows clear quantitative gains on both tunnels. The baseline only achieves 34.17% precision, 90.67% recall, and 80.31% mIoU on Tunnel A, and 31.48% precision, 70.83% recall, and 62.22% mIoU on Tunnel B, indicating that directly applying SAM tends to produce many spurious instances, whereas our method substantially improves precision and overall segmentation quality without sacrificing recall.

To analyze the impact of the inpainting module, Fig.6 presents a qualitative comparison of the unfolded intensity images before and after diffusion-based completion. While the inpainting process improves the visual continuity by filling missing regions, it may occasionally introduce synthetic textures that resemble structural boundaries. These pseudo-structures can mislead SAM and lead to false-positive detections, especially when large void regions must be synthesized with limited contextual information.

The proposed framework integrates three key components to achieve robust tunnel segment segmentation. The diffusion-based inpainting module improves the continuity of the unfolded intensity images by restoring missing regions caused by occlusions or sparse scanning. The box prompts generated by the vision-language model provide coarse instance localization, guiding SAM to focus on candidate segment regions. In addition, intensity-driven point

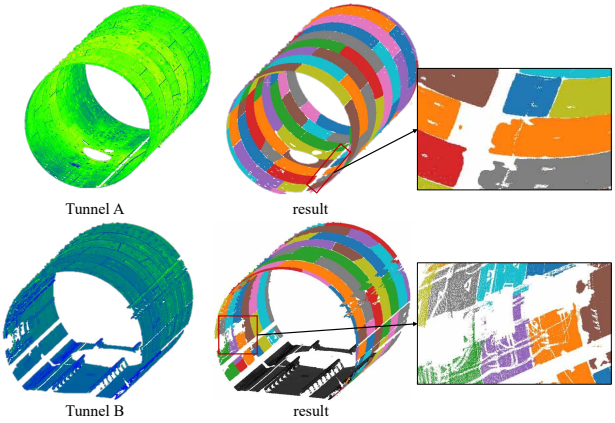


Figure 5. View of Tunnel Segmentation Results

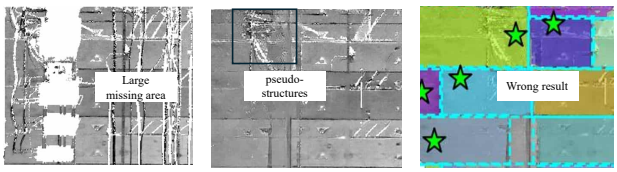


Figure 6. Error analysis of Tunnel B using the diffusion-based inpainting module.

prompts refine the segmentation boundaries by leveraging intensity variations along segment joints. These components jointly contribute to the overall segmentation performance of the framework.

Table3 presents the processing time of Tunnel A in the pipeline. It contains nine rings and containing 10 million points.

In summary, this paper proposes an automated zero-shot framework for instance-level segmentation of shield tunnel linings in 3D point clouds. Quantitative results in Table 2 show that the framework achieves high-quality instance segmentation for multi-type segments, accurately localizing most instances even under unfavorable conditions and validating its strong generalization potential.

5 Discussion

This paper constructs a complete closed-loop system. Experimental results demonstrate that this method preserves the continuity and topological consistency of segmented annular structures within circular tunnels, thereby reducing the reliance on complex three-dimensional models and extensive training datasets.

The comparison with the SAM baseline further confirms that directly applying a generic segment-everything model is insufficient in the presence of geometric distortions and severe occlusions: SAM tends to generate fragmented and spurious masks, particularly around repaired or missing regions. This approach relies entirely on pre-trained models and automated prompt-generating workflows, enabling zero-shot instance segmentation within minutes to

Table 2. Results of the Tunnel A and Tunnel B datasets

Tunnel	Method	Prec(%)	Rec(%)	mIoU(%)	TP	FP	FN
A	SAM (baseline)	34.17	90.67	80.31	68	131	7
A	**Ours**	87.95	97.33	90.58	73	10	2
B	SAM (baseline)	31.48	70.83	62.22	34	74	14
B	**Ours**	75.44	89.58	75.59	43	14	5

Table 3. Runtime analysis of the proposed framework

Module	Time (s)
Data preprocessing	120
Unfolding and Inpainting	102
SAM integration and optimization	40
Reprojection	37
Total	299

meet the efficiency demands of engineering inspection and maintenance.

However, when geometric structures deviate from circularity, unfolded images may exhibit distortion, which can affect boundary recognition in SAM model. Although SAM effectively addresses boundary adaptation issues, such distortions and occlusions still diminish segmentation accuracy, particularly in tunnels with complex geometries.

However, due to the generative nature of diffusion models, the restored textures may not perfectly correspond to the true tunnel geometry. Although the inpainting process improves the visual continuity of the unfolded intensity images, it may also introduce synthetic patterns that resemble structural edges or segment boundaries. These pseudo-structures can interfere with SAM's boundary recognition and occasionally lead to false-positive detections. This phenomenon becomes particularly pronounced in tunnels with heavy equipment occlusions, where large missing regions must be synthesized with limited contextual cues. This behavior is consistent with the experimental observations in Tunnel B, where complex occlusions increase the likelihood of misclassification. Despite these limitations, the inpainting module significantly improves the overall structural continuity of the input images, contributing to more stable and reliable segmentation results in most scenarios.

This approach relies entirely on pre-trained models and automated workflows, enabling zero-shot instance segmentation within minutes to meet the efficiency demands of engineering inspection and maintenance.

6 Conclusion and Future Work

This paper addresses the typical engineering scenario of segment segmentation in shield-tunnel construction by proposing a zero-shot adaptive prompt-generation segmentation framework. By establishing a closed-loop workflow 3D - 2D - 3D, it converts unstructured tunnel point clouds into two-dimensional intensity unfolded images, effectively bridging the gap between 3D engineering data and 2D foundation models. Using a general-purpose visual model and gray values for adaptive prompt generation, it achieves high-precision segmentation (90.58% mIoU) without requiring task-specific custom training. Experimental results provide preliminary evidence of the applicability of the proposed method across different tunnel geometries and missing areas, particularly in maintaining the continuity and topological consistency of ring-shaped structures within standard circular tunnels.

These findings validate the feasibility of the proposed approach, offering a practical technical path to reduce annotation costs and improve the efficiency of model deployment. Concurrently, it establishes a foundational data set for subsequent tunnel deformation monitoring and the construction of digital twin models.

However, this method also has certain limitations. Future research should focus on two primary areas:

- Fine-tuning image restoration models: Current general models exhibit insufficient learning of texture features in intensity images, resulting in poor performance when faced with large-scale voids. Subsequent work will address this issue to achieve enhanced robustness.
- Hierarchical prompting: Explore a hierarchical approach where a primary model identifies larger structures (e.g., tunnel rings), while a secondary, fine-tuned VLM targets detailed components (e.g., structural elements), potentially using 3D bounding boxes derived directly from fitted axes as more robust prompts.

7 Acknowledgments

This study was financially supported by the National Key Research and Development Program of China (Grant No. 2023YFC3805901), Postdoctoral Fellowship Program of CPSF (Grant No. GZC20242128), and China Postdoctoral Science Foundation (Grant No. 2024M763870) to which the authors are very grateful. Pengkun Liu extends sincere appreciation to The Hong Kong Center for Construction Robotics (InnoHK center supported by the Hong Kong ITC, InnoHK-HKCRC) for their invaluable support.

References

[1] Jian Zhou, Hongning Qi, Kang Peng, Yulin Zhang, and Manoj Khandelwal. Comprehensive review and future perspectives on prediction and mitigation of tunnel-induced ground settlement: A bibliometric analysis and methodological overview (2002–2022). *Tunnelling and Underground Space Technology*, 154:106081, 2024. ISSN 0886-7798. doi:https://doi.org/10.1016/j.tust.2024.106081. URL `https://www.sciencedirect.com/science/article/pii/S0886779824004991`.

[2] J.A. Richards. Inspection, maintenance and repair of tunnels: International lessons and practice. *Tunnelling and Underground Space*

Technology, 13(4):369–375, October 1998. ISSN 0886-7798. doi:10.1016/s0886-7798(98)00079-0. URL http://dx.doi.org/10.1016/s0886-7798(98)00079-0.

[3] X. Jiang, Y. Yuan, and X. Liu. *Performance-based Predictive Maintenance of Shield Tunnels*, page 1007–1016. CRC Press, October 2014. ISBN 9780429227196. doi:10.1201/b17618-147. URL http://dx.doi.org/10.1201/b17618-147.

[4] A. Haack, J. Schreyer, and G. Jackel. Report to ita working group on maintenance and repair of underground structures: State-of-the-art of non-destructive testing methods for determining the state of a tunnel lining. *Tunnelling and Underground Space Technology*, 10(4):413–431, October 1995. ISSN 0886-7798. doi:10.1016/0886-7798(95)00030-3. URL http://dx.doi.org/10.1016/0886-7798(95)00030-3.

[5] Leanne Attard, Carl James Debono, Gianluca Valentino, and Mario Di Castro. Tunnel inspection using photogrammetric techniques and image processing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:180–188, October 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2018.07.010. URL http://dx.doi.org/10.1016/j.isprsjprs.2018.07.010.

[6] Cheng Yi, Dening Lu, Qian Xie, Shuya Liu, Hu Li, Mingqiang Wei, and Jun Wang. Hierarchical tunnel modeling from 3d raw lidar point cloud. *Computer-Aided Design*, 114: 143–154, September 2019. ISSN 0010-4485. doi:10.1016/j.cad.2019.05.033. URL http://dx.doi.org/10.1016/j.cad.2019.05.033.

[7] Yong Feng, Xiao-Lei Zhang, Shi-Jin Feng, Yong Zhao, Qing-Zhao Kong, and Shu-Lin Zhu. Metro tunnel deformation detection based on laser scanning robot and point cloud semantic segmentation. *Tunnelling and Underground Space Technology*, 166:106966, December 2025. ISSN 0886-7798. doi:10.1016/j.tust.2025.106966. URL http://dx.doi.org/10.1016/j.tust.2025.106966.

[8] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, page 3992–4003. IEEE, October 2023. doi:10.1109/iccv51070.2023.00371. URL http://dx.doi.org/10.1109/iccv51070.2023.00371.

[9] Jay N. Paranjape, Shameema Sikder, S. Swaroop Vedula, and Vishal M. Patel. *S-SAM: SVD-Based Fine-Tuning of Segment Anything Model for Medical Image Segmentation*, page 720–730. Springer Nature Switzerland, 2024. ISBN 9783031723902. doi:10.1007/978-3-031-72390-2_67. URL http://dx.doi.org/10.1007/978-3-031-72390-2_67.

[10] Yunxiang Zhou, Ankang Ji, Limao Zhang, and Xiaolong Xue. Attention-enhanced sampling point cloud network (aspcnet) for efficient 3d tunnel semantic segmentation. *Automation in Construction*, 146:104667, February 2023. ISSN 0926-5805. doi:10.1016/j.autcon.2022.104667. URL http://dx.doi.org/10.1016/j.autcon.2022.104667.

[11] Yuxian Zhang, Xuhua Ren, Jixun Zhang, and Zichang Ma. A method for deformation detection and reconstruction of shield tunnel based on point cloud. *Journal of Construction Engineering and Management*, 150(3), March 2024. ISSN 1943-7862. doi:10.1061/jcemd4.coeng-14225. URL http://dx.doi.org/10.1061/jcemd4.coeng-14225.

[12] Yixiong Jing, Cheng Zhang, Haibing Wu, Guangming Wang, Olaf Wysocki, and Brian Sheil. Infradiffusion: zero-shot depth map restoration with diffusion models and prompted segmentation from sparse infrastructure point clouds, 2025. URL https://arxiv.org/abs/2509.03324.

[13] Zehao Ye, Wei Lin, Asaad Faramarzi, Xiongyao Xie, and Jelena Ninić. Sam4tun: No-training model for tunnel lining point cloud component segmentation. *Tunnelling and Underground Space Technology*, 158:106401, 2025. ISSN 0886-7798. doi:https://doi.org/10.1016/j.tust.2025.106401. URL https://www.sciencedirect.com/science/article/pii/S0886779825000392.

[14] Gang Xu, Yingshui Zhang, Qingrui Yue, and Xiaogang Liu. Automated detection and quantification of structural component dimensions using segment anything model (sam)-based segmentation. *Automation in Construction*, 176:106304, August 2025. ISSN 0926-5805. doi:10.1016/j.autcon.2025.106304. URL http://dx.doi.org/10.1016/j.autcon.2025.106304.

[15] Wei Lin, Brian Sheil, Pin Zhang, Biao Zhou, Cheng Wang, and Xiongyao Xie. Seg2tunnel: A hierarchical point cloud dataset and benchmarks for segmentation of segmental tunnel linings. *Tunnelling and Underground Space Technology*, 147:105735, May 2024. ISSN 0886-7798. doi:10.1016/j.tust.2024.105735. URL http://dx.doi.org/10.1016/j.tust.2024.105735.

Development of Feature Extraction Using High-Resolution RGB Images and Homography Transformation to Inspect RC Walls Rebar Works

Young-Gun Baek[1], Rong-Lu Hong[1], Il-Gyo Choi[2], Kyung-Ho Lee[1], Ju-Hyung Kim[1]

[1] Department of Architectural Engineering, Hanyang University, Seoul, Republic of Korea
[2]ON Construction Co., Ltd., Yeongdeungpo-gu, Seoul, Republic of Korea
byggun456@naver.com, abraham0814@hanyang.ac.kr, jupi0515@hanmail.net, mbl00@naver.com, kcr97jhk@hanyang.ac.kr

Abstract –

Reinforcing bars (rebars) are essential for resisting tensile forces and ensuring structural safety in reinforced concrete systems. However, current rebar inspection practices rely on manual and sampling-based measurements, leading to human errors and limiting inspection coverage. Although automated 3D and LiDAR-based inspection systems have been explored, their adoption rate remains low due to high cost and workflow complexity. To address these challenges, this study presents a cost-efficient vision-based inspection framework that combines high-resolution image analysis and marker-assisted geometric correction. The proposed method detects rebar and marker regions in low-resolution images, refines intersection keypoints in high resolution, performs marker-based rectification, and computes rebar spacing, diameter, and quantity from the corrected geometry. Experiments under varied distances and viewing angles demonstrated accurate and reliable performance, achieving 100% accuracy in counting vertical and horizontal rebars, mean absolute errors of 0.59 mm for diameter and 4.05 mm for spacing. These results confirm practical, low-cost automated rebar measurements for construction inspection.

Keywords –

Reinforcing bar; Convolutional Neural Network; Homography Transformation

1 Introduction

Quality control in the construction industry is a fundamental process for securing the structural safety of buildings. In particular, reinforcing bars (rebars) are primary load-bearing components that resist tensile forces in reinforced concrete members; therefore, improper placement of rebar can lead to serious defects such as cracking or corrosion [1,2]. Accordingly, accurate inspection of the number, diameter, spacing, and position of rebars before concrete casting is critical to ensure structural safety and construction quality [3]. However, conventional rebar inspection still depends on manual measurement and visual inspection, which are time-consuming and prone to human errors [4]. Specifically, of wall reinforcement, the bars are densely arranged and extend across a large area, which increases the amount of inspection work and makes the process highly labor-intensive. For these reasons, automation and systematization of rebar inspection processes is urgently needed to improve inspection efficiency and reliability[4].

Previous studies primarily rely on LiDAR-based 3D point cloud models to measure rebar diameter, quantity, and spacing [5–6]. Although laser-scanning-based approaches provide high accuracy, their applicability to real-time field inspection is limited due to the high equipment cost and the heavy computational burden [4,7]. As a cost-effective alternative to laser scanners, research has progressed to inspect rebars based on image processing and deep learning on RGB images [8,9]. This is mainly due to the widespread availability, low hardware cost, and ease of deployment of RGB cameras. However, these methods require an orthogonal camera angle, thereby limiting their applicability in dynamic construction environments. Photogrammetry-based 3D reconstruction has also been explored, but this approach typically requires a large number of images for a single rebar segment and involves long processing times, limiting its feasibility for on-site use [7]. Depth cameras have been frequently employed in rebar inspection studies, providing guaranteed measurement accuracy as laser scanners [1,10]. However, the accuracy of depth cameras decreases as the distance from the rebar increases and is further reduced as the camera angle deviates from the orthogonal condition to the rebar mat [5,10]. While these limitations may not be significant for small areas or short measurement ranges, they cause the

inspection process to become inefficient when applied to extensive wall reinforcement. Furthermore, the spatial complexity and tight schedules of construction sites hinder the ability to maintain a fixed camera angle.

To overcome these challenges, this study introduces a low-cost rebar placement quality inspection method based on a high-resolution DSLR (digital single-lens reflect) camera with marker-based calibration [11]. The high-resolution DSLR camera ensures the acquisition of necessary pixel density for the precise extraction of fine rebar features, even from long distances. Compared to LiDAR systems, the high-resolution DSLR camera is highly cost-effective and easy to maintain, thus securing high field applicability. Moreover, markers enable the compensation of perspective distortion through homography transformation and serve as a reference for converting pixel units into actual distances [12].

However, even with high-resolution images, extracting fine rebar features from long distances remains challenging. Traditional image processing techniques are susceptible to lighting variations, shadows, complex backgrounds, and noise, which hinder reliable detection in real-world construction environments [10]. Meanwhile, convolutional neural network (CNN)–based methods provide strong detection performance across various environments, but the computational cost increases significantly when processing high-resolution images, which limits real-time applicability [13]. Most CNN-based studies mitigate this computational burden using downsampling, but this leads to the loss of fine visual details and degrades measurement accuracy [8,9,12]. Because rebars are thin, small objects, their pixel footprint becomes extremely limited at long distances, and downsampling further reduces available information, increasing the likelihood of detection and measurement failure.

To address these limitations, the proposed method does not directly apply a CNN to the entire high-resolution image [11]. Instead, it restructures the role of CNNs by decomposing the analysis into two resolution-specific stages. At the low-resolution stage, a CNN is employed solely for global scene understanding and candidate localization, identifying rebar intersection regions and markers with minimal computational cost. These coarse predictions are not used for measurement but only to determine where high-resolution analysis is necessary. Subsequently, only the selected regions of interest are reprojected into the original high-resolution image and processed by a dedicated high-resolution network for metric-level localization. Through this resolution-aware role separation, the proposed method preserves fine geometric details required for accurate rebar measurement while avoiding the prohibitive computational cost of processing the entire image at high resolution [11].

The proposed method detects rebar intersection regions and markers from low-resolution images through a Dual-head Faster R-CNN and then processes the corresponding high-resolution patches with a modified High-Resolution Network (HRNet) to extract precise coordinates of rebar intersections. Subsequently, homography transformation is performed with the marker corner points as the reference frame to provide a perspective-corrected image and to compute the rebar count, diameter, and spacing. Through this process, the proposed method preserves detailed high-resolution information while minimizing computational cost, enabling efficient and accurate rebar inspection even at long distances and under various camera angles.

2 The Proposed Method

2.1 Overview of the Proposed Framework

This study proposes a method for the automated extraction of rebar count, diameter, and spacing in dense and extensive wall reinforcement structures using high-resolution images and fiducial markers. The overall procedure consists of four stages [11]. First, the entire input image is processed by a Dual-head Faster R-CNN to detect markers and rebar cross-section regions. For computational efficiency, the input image is scaled down to 600×400 pixels. The model is designed with a single backbone and two separate heads: one to predict markers in the form of bounding boxes and the other to predict rebar cross-section regions in the form of keypoints. This architecture enables the efficient extraction of necessary Regions of Interest (ROIs) even from noisy field images. Second, the detected rebar cross-section ROIs are projected onto the original high-resolution image, cropped to 256×256 pixels, and input into modified HRNet. Since modified HRNet maintains high-resolution representations for feature extraction, it can reliably estimate the four coordinates of the rebar corner points, even when captured from a long distance or against a complex background. Third, a Homography transformation is performed based on the marker's corner points to generate a perspective-corrected image. This transformation achieves perspective correction in non-frontal shooting environments and computes the pixel-to-real distance scale factor using the marker's known dimensions. Finally, the rebar diameter is calculated from the distances between the four corner points, while the rebar spacing is derived from the distances between centroids, defined as the intersection points of the diagonals formed by those corner points. The overview of the proposed method is shown in Figure 1.

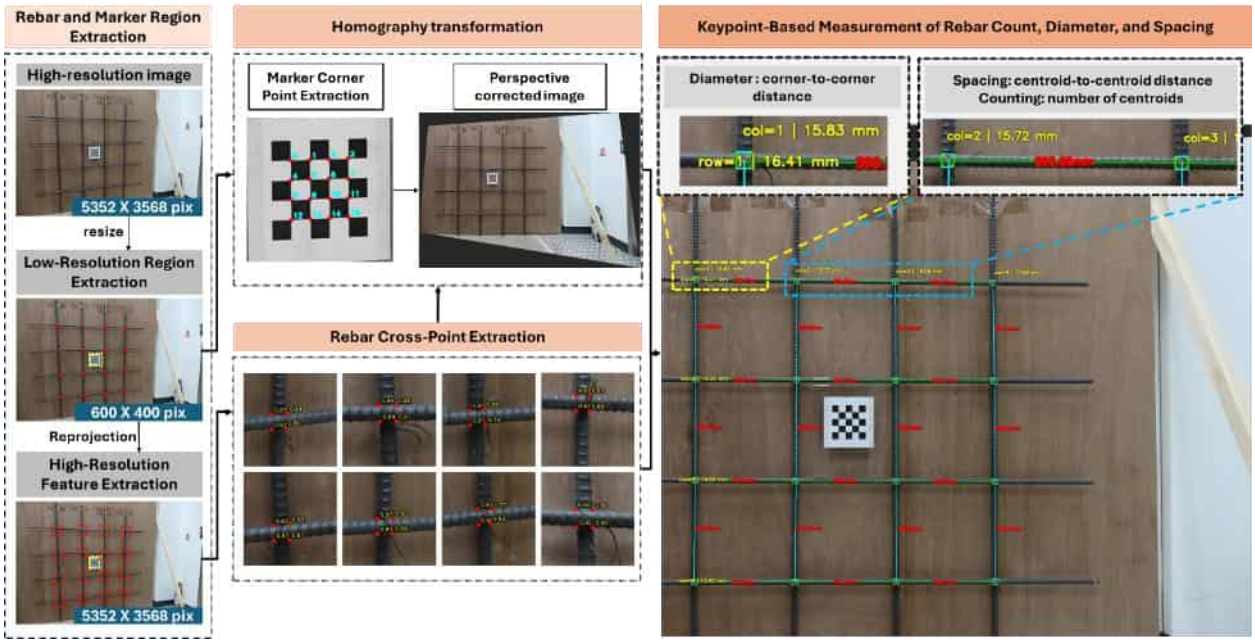


Figure 1. Overview of the proposed method

2.2 Dual-head Faster R-CNN for Rebar and Marker Detection

Since directly processing high-resolution images incurs a significant computational and memory burden, the proposed method first detects ROIs at a low-resolution stage and then performs precise keypoint extraction on the corresponding high-resolution patches. We designed a Dual-head Faster R-CNN with a single backbone and two heads to detect the ROIs from the low-resolution input image [11]. The model architecture shown in Figure 2. The Detector Head, which maintains the structure of the conventional Faster R-CNN, is designed to detect markers as bounding boxes, thereby filtering background and noise when the corner points are subsequently extracted. The Point Heatmap Head is an added component that performs convolutional operations directly on the feature maps extracted from the Feature Pyramid Network (FPN) to generate a heatmap, ensuring stable detection of the rebar keypoints. This architecture is designed to efficiently extract only the necessary ROIs even from noisy field images.

2.2.1 Experimental Settings of Dual-head Faster R-CNN

A dataset of 500 high-resolution rebar and marker images, resized to 600×400 pixels, was collected in a laboratory setting. Data augmentation (flips, blurring,

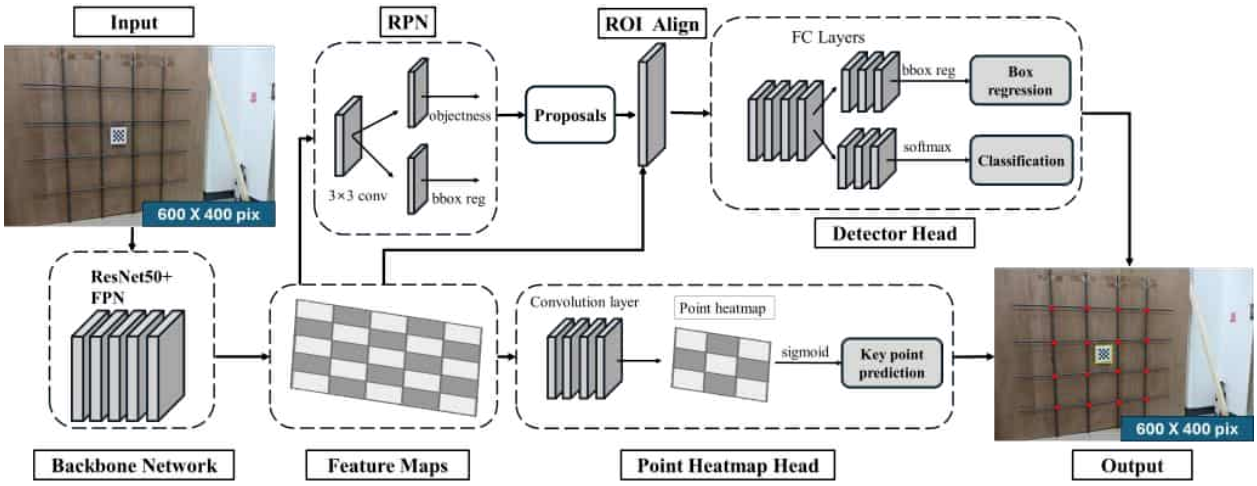


Figure 2. Architecture of the Dual-head Faster R-CNN

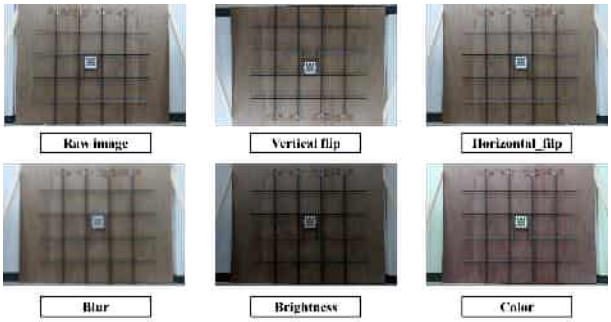


Figure 3 . Examples of data augmentation for Dual-head Faster R-CNN

brightness, color adjustments) expanded the final dataset training, 10% for validation, and 10% for testing. Examples of data augmentation are illustrated in Figure 3. Label consistency was preserved by applying identical geometric transformations to both images and annotations. The model introduced a Dual-head Faster R-CNN with a ResNet-50 FPN backbone, initialized using ImageNet pre-trained weights to leverage transfer learning for fast and stable convergence. Optimization was performed using SGD with a Cosine Annealing with Warm Restarts scheduler, a batch size of 2, and 70 epochs with early stopping. The total loss is the weighted sum of the standard Detection Loss (L_{det}) and a BCEWithLogits-based Heatmap Loss (L_{hm}), defined as:

$$L_{total} = L_{det} + w \cdot L_{hm} \quad (1)$$

Where w is the weight coefficient for the Heatmap Loss which was set to 1.8. The evaluation of bounding boxes was performed using standard metrics, including AP50, AP75, and mIoU. The evaluation of keypoint estimation was performed using the Percentage of Correct Keypoints (PCK) metric with pixel radius thresholds of 2, 5, and 8 (PCK@2, PCK@5, and PCK@8). which is computed as:

$$PCK@\rho = \frac{N_{correct}(\rho)}{N_{gt}} \quad (2)$$

$N_{\text{correct}}(\rho)$denotes the number of predicted keypoints within a pixel distance threshold ρ from the ground truth, and N_{gt} represents the total number of ground-truth keypoints.

2.3 Modified HRNet for Keypoint Extraction

Although initial rebar intersection points are detected at low resolution using a Dual-head Faster R-CNN, their direct reprojection to the high-resolution image is insufficient for precise localization due to pixel-level inaccuracies. Therefore, this study employs a high-resolution patch-based refinement approach using HRNet [11]. Accordingly, high-resolution patches of 256×256 pixels are cropped around the intersection point coordinates predicted by the dual-head Faster R-CNN, and the four corner vertices of the intersection region are precisely estimated using the modified HRNet. Figure 4 presents the architecture of the modified HRNet. To reduce computational overhead, deeper stages of the original HRNet that are unnecessary for precise keypoint estimation on small image patches are omitted. In addition, HRNet integrates a high-resolution feature stream with parallel lower-resolution branches through multi-resolution fusion. Furthermore, adopting a keypoint-based approach simplifies the model structure and reduces computational costs, as it estimates only the minimum required coordinates without predicting the entire pixel area, unlike bounding-box- or mask-based methods. Through this patch-based refinement process, high-resolution rebar features can be extracted with substantially reduced computational cost, as the network operates only on localized regions instead of processing the entire high-resolution image.

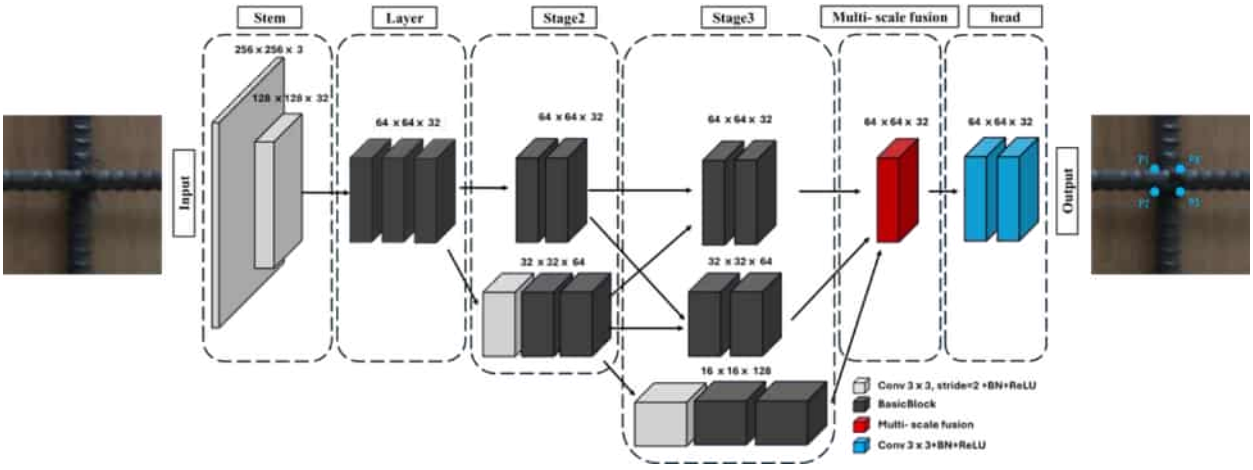


Figure 4. Architecture of the modified HRNet

2.3.1 Experimental Settings of Modified HRNet

The dataset for training modified HRNet was constructed from 880 image patches of size 256×256 pixels, cropped around the rebar intersection centers predicted by the dual-head Faster R-CNN. Extensive data augmentation was applied, including horizontal flips, rotation, blurring, and brightness and color adjustments. Examples of data augmentation are illustrated in Figure 5. The dataset was split into 80% for training, 10% for validation, and 10% for testing. Modified HRNet was trained using the AdamW optimizer, and a batch size of 5, for a maximum of 100 epochs. Early stopping was applied with a patience of 15 epochs, and no transfer learning was employed. The loss function was defined as the mean squared error (MSE) between the predicted heatmaps and the corresponding ground-truth heatmaps, computed using the batch-averaged loss to improve training stability. Keypoint estimation performance was evaluated using the PCK metric at pixel radius thresholds of 2, 5, and 8 (PCK@2, PCK@5, and PCK@8).

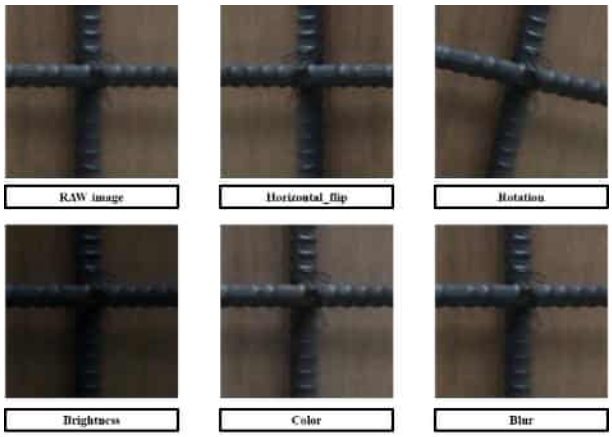


Figure 5. Examples of data augmentation for Modified HRNet

2.4 Geometric Feature Based Rebar Inspection and Quantification

2.4.1 Homography Transformation

In this study, rebar diameter and spacing were quantitatively extracted by combining high-resolution rebar intersection point coordinates with chessboard marker-based geometric correction. To ensure stable corner detection under long-distance imaging conditions, 5×5 chessboard markers with a side length of 20 mm were employed. The marker bounding boxes, detected by the Dual-head Faster R-CNN, were mapped onto the original high-resolution image, and their corner points were precisely extracted using the findChessboardCorners function of OpenCV. The detected marker corner points were matched with their

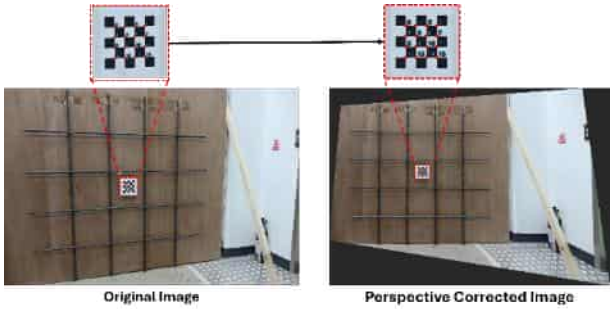


Figure 6. Perspective-corrected image using marker-based Homography Transformation

corresponding reference points defined on a marker-aligned virtual planar grid to estimate the homography matrix H. The estimated transformation was then applied to the input image as well as to the rebar corner coordinates estimated by HRNet, enabling rebar measurements to be conducted in a common, perspective-corrected system. Figure 6 illustrates an example of the perspective-corrected image generated using the marker-based homography.

2.4.2 Rebar Diameter Calculation

In this study, the rebar diameter is estimated from the distances between four corner coordinates predicted by HRNet on a perspective-corrected image obtained via marker-based homography. The pixel distances are subsequently converted to real-world units using a scale factor derived from the known geometry of the marker. To ensure consistent point ordering across samples, the top-right point in the image coordinate system is selected as the reference point ($P1$), and the remaining points are arranged in a counterclockwise order as $P2$, $P3$, and $P4$. Based on the resulting quadrilateral, opposing sides are paired and their average lengths are computed. Here, Length A corresponds to the vertical rebar direction, while Length B represents the horizontal rebar direction. Figure 7, rebar diameter estimation is performed on the perspective-corrected image by measuring the distances between HRNet-detected corner points and converting them into real-world units using the marker-based scale factor. Finally, the estimated diameters for the vertical and horizontal rebars are stored separately according to the row and column indices defined during the spacing extraction process.

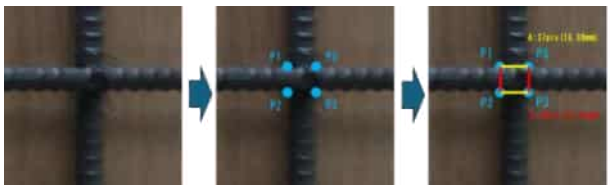


Figure 7. Rebar diameter estimation

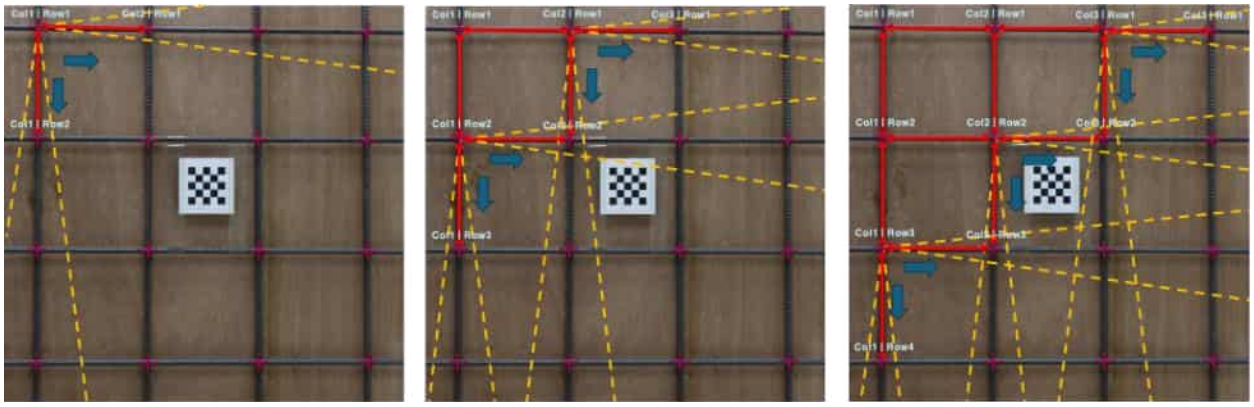


Figure 8. Grid recognition and spacing extraction based on rebar centroids

2.4.3 Calculation of Rebar Spacing and Count

In this study, the centroid of each rebar cross-section is defined as the intersection point of the diagonals $\overline{P1P3}$ and $\overline{P2P4}$ computed from the four corner coordinates estimated by HRNet. These centroids collectively form a grid-like pattern across the image, and their relative spatial arrangement is used to extract rebar spacing. To establish the grid structure, the centroid located at the top-left position in the image coordinate system is selected as the reference point. From this reference, the spacing search is sequentially expanded in the horizontal and vertical directions. An angular tolerance of 30° is applied during the search to accommodate geometric distortions caused by construction tolerances and variations in camera viewpoint. Each detected centroid is then assigned a unique column (col) and row (row) index based on its relative horizontal and vertical displacement from the reference centroid. During this indexing process, the spacing between adjacent centroids is simultaneously computed. By enforcing consistent directionality and angular tolerance, this grid-based strategy is designed to remain effective even under non-uniform rebar layouts or partially missing intersections, which are commonly encountered in real construction environments. Figure 8 illustrates the grid recognition procedure. The yellow lines indicate the 30° angular tolerance used during centroid association, while the red arrows represent the recognized spacing between adjacent centroids. The figure also visualizes the sequential assignment of row and column indices as each centroid is identified.

3 Results and Discussion

3.1 Experimental Setup

Image data were acquired using an 18-megapixel Canon EOS 100D DSLR camera, and a total of 500 images containing rebars and fiducial markers were collected in a controlled laboratory environment. The experiment was conducted in a controlled laboratory environment designed to reflect practical wall reinforcement conditions. Four D16 rebars, each 1.5 m in length, were arranged on a veneer plywood in a 3 × 3 grid pattern with a spacing of 300 mm, and fiducial markers were installed parallel to the rebars. The ground-truth rebar spacing was measured using a tape measure for quantitative comparison with the estimated results.

For image acquisition, the shooting distance was varied from 2.0 m to 4.0 m at intervals of 0.5 m to evaluate the effect of increasing distance and associated focus degradation on measurement accuracy. At each distance, more than 30 images were captured under free-view (non-frontal) perspectives. The marker-based configuration also enabled the estimation of camera extrinsic parameters, allowing the influence of viewpoint variations on image projection and measurement accuracy to be analyzed. Model performance was evaluated using the Mean Absolute Error (MAE) and the Mean Relative Error (MRE).

3.2 Detection and Refinement Model Evaluation

3.2.1 Performance of Dual-head Faster R-CNN

The Dual-head Faster R-CNN achieved optimal performance at the 68th epoch, yielding AP50 = 1.0000, AP75 = 1.0000, mIoU = 0.9633, and PCK@8/5/2 of 0.9990/0.9901/0.5315. Detailed results are summarized in Table 1.

Table 1. Evaluation metrics of Dual-head Faster R-CNN

Marker Detection		Rebar Detection	
AP50	1.0000	PCK@8	0.9992
AP75	1.0000	PCK@5	0.9901
mIoU	0.9633	PCK@2	0.5315

3.2.2 Performance of Mmodified HRNet

In the training of modified HRNet, early stopping with a patience of 15 epochs was applied and triggered at epoch 71. Evaluation on an independent test dataset yielded PCK@8/5/2 of 0.9980/0.9957/0.8930. Table 2 summarizes the prediction performance of the HRNet model.

Table 2 Evaluation metrics of modified HRNet

Evaluation index	Result
PCK@8	0.9980
PCK@5	0.9957
PCK@2	0.8930

3.3 Rebar Quantification Results

3.3.1 Rebar Counting Results

The proposed method was evaluated under varying distances (2.0–4.0 m) and viewpoints. Under all tested conditions, the numbers of both vertical and horizontal rebars were correctly estimated with 100% accuracy. These results demonstrate the robustness and practical applicability of the proposed method in real-world field environments.

3.3.2 Rebar Diameter Measurement Results

The results of rebar diameter estimation show that the MAE remained within 0.50–0.59 mm (3.1–3.7%) up to 3.5 m, increasing to 0.77 mm at 4.0 m. This trend is consistent with the reduction in effective pixel resolution as the camera-to-object distance increases, which limits localization precision. The results are summarized in Table 3.

Table 3. Rebar diameter accuracy at different distances

Distance	MAE (mm)	MRE (%)
2.0 m	0.5859	3.7045
2.5 m	0.5100	3.3702
3.0 m	0.5009	3.1310
3.5 m	0.5788	3.5556
4.0 m	0.7747	4.8427

The effect of the viewing angle on rebar spacing estimation was evaluated. The MAE increased from 0.51 mm to 0.86 mm, while the corresponding error rate rose from 3.20% to 5.40% as the viewing angle increased. This behaviour is explained by the fact that, although homography-based planar correction mitigates perspective distortion, residual errors arise at larger view angles due to the three-dimensional geometry of the rebars. A summary of the results is provided in Table 4.

Table 4. Rebar diameter accuracy at different angles

Angle	MAE (mm)	MRE (%)
0-10°	0.5127	3.2045
10-20°	0.5788	3.6178
20-30°	0.6209	3.8808
≤ 30°	0.8642	5.4012

3.3.3 Rebar Spacing Measurement Results

The results of rebar spacing estimation show that the MAE remained within 3.6–4.1 mm (1.2–1.4%) up to 3.5 m, increasing to 4.65 mm (1.55%) at 4.0 m. This trend is consistent with the reduction in effective image resolution at longer distances, which limits the precision of marker feature point localization. The quantitative results are summarized in Table 5.

Table 5. Rebar spacing accuracy at different distances

Distance	MAE (mm)	MRE (%)
2.0 m	3.9208	1.3060
2.5 m	3.6027	1.2061
3.0 m	4.0948	1.3641
3.5 m	3.9787	1.3253
4.0 m	4.6484	1.5487

The effect of the viewing angle on rebar spacing estimation was analyzed, with the MAE ranging from 3.8 mm to 4.5 mm and the corresponding error rates between 1.2% and 1.5%, showing a gradual increase as the viewing angle increased. In particular, an increase in error was observed for viewing angles exceeding 30°. This behavior is attributed to the fact that, although marker-based planar correction mitigates perspective distortion, residual errors arise at larger view angles due to the three-dimensional geometry of the rebars, which is consistent with the trends observed in diameter estimation. The results are summarized in Table 6.

Table 6. Rebar spacing accuracy at different angles

Angle	MAE (mm)	MRE (%)
0-10°	3.8600	1.2858
10-20°	4.1303	1.3759
20-30°	3.9335	1.3103
≤ 30°	4.4560	1.4843

4 Conclusion

This study presents an image-based metrology framework for efficient quality inspection of wide-area rebar placement structures, such as concrete walls. Unlike conventional inspection practices that rely on manual measurements, the proposed framework adopts a two-stage inspection pipeline integrating low-resolution

region detection and high-resolution keypoint refinement using a Faster R-CNN, HRNet, and marker-based homography transformation. To support this framework, the Faster R-CNN detector was extended with a dual-head structure to jointly detect rebars and markers. A lightweight HRNet configuration was adopted for efficient high-resolution keypoint detection.

Experimental validation under varying distances and viewpoints demonstrated robust performance. The proposed method achieved 100% accuracy in identifying both vertical and horizontal rebars under all tested distances and viewing angles. The mean absolute error was 0.59 mm (3.53%) for diameter estimation and 4.05 mm (1.35%) for spacing estimation. These results fall within the permissible tolerance for rebar diameter estimation (0.8 mm) [9] and the commonly accepted error rate for rebar spacing estimation (10%) [10]. These results fall within permissible tolerances confirming stable and practical measurement performance.

Despite these promising results, several limitations remain. First, the experiments were conducted only in a controlled laboratory setting, while real construction sites may introduce complex visual conditions such as uneven lighting, shadows, and background clutter, affecting keypoint detection robustness. Second, the proposed framework requires manual installation of markers, which may introduce additional preparation time. Third, measurement errors may increase when the viewing angle exceeds approximately 30° due to the three-dimensional geometry of rebars, requiring careful camera positioning. Fourth, measurement accuracy may degrade at longer distances due to reduced image resolution.

Future work will address these limitations. First, real construction site data will be collected to evaluate the robustness and generalizability of the proposed method under diverse environmental conditions. Second, rebar-attached markers will be developed to reduce the time and complexity of marker installation. Third, marker-based algorithms will be developed to estimate viewing angles and provide real-time guidance for proper camera positioning in field conditions. Finally, approaches to improve effective pixel resolution at long distances will be investigated to enhance measurement accuracy.

Nevertheless, the proposed approach demonstrates the potential of a low-cost, image-based solution to support practical and scalable rebar quality inspection in the construction industry.

Acknowledgement

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (grant number RS202400356697).

References

[1] Y. Kardovskyi and S. Moon. Artificial intelligence quality inspection of steel bar installation by integrating Mask R-CNN and stereo vision. *Automation in Construction*, 130:103850, 2021.

[2] M. S. Shetty and A. K. Jain. *Concrete Technology: Theory and Practice*. S. Chand Publishing, New Delhi, India, 2019.

[3] Concrete Reinforcing Steel Institute (CRSI). *Placing Reinforcing Bars: Recommended Practices*, 9th ed. Concrete Reinforcing Steel Institute, Schaumburg, IL, USA, 2011.

[4] J. Song, H. Zhang, Y. Li, and J. Wang. An automatic rebar spacing measuring method based on the YOLOv8-GB model. *Measurement*, 242:116278, 2025

[5] M. K. Kim, J. P. P. Thedja, and Q. Wang. Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3D point cloud data. *Automation in Construction*, 112:103077, 2020.

[6] X. Yuan, A. Smith, R. Sarlo, C. D. Lippitt, and F. Moreu. Automatic evaluation of rebar spacing and quality using LiDAR data: Field application for bridge structural assessment. *Automation in Construction*, 146:104708, 2023.

[7] A. H. Qureshi, W. S. Alaloul, A. Murtiyoso, S. J. Hussain, S. Saad, and M. A. Musarat. Smart rebar progress monitoring using 3D point cloud models. *Expert Systems with Applications*, 249:123562, 2024.

[8] X. Zhang, Y. Liu, J. Li, and H. Wang. A high-precision quality inspection system for steel bars based on machine vision. *Sensors*, 18(8):2732, 2018.

[9] W. Chen, X. Fu, W. Chen, and Z. Peng. Intelligent detection of rebar size and position using improved DeepLabV3+. *Applied Sciences*, 13(19):11094, 2023.

[10] U. Woo, M. Lee, T. Lee, H. Choi, S. M. Kang, and K. K. Choi. Real-time rebar spacing measurement system for quality control in construction. *Automation in Construction*, 171:106012, 2025.

[11] Y. G. Baek. Rebar inspection method using high-resolution image-based feature extraction and homography transformation. Master's Thesis, Hanyang University, Seoul, Republic of Korea. 2026.

[12] I. E. Bal, D. Dais, E. Smyrou, and V. Sarhosis. Novel invisible markers for monitoring cracks on masonry structures. *Construction and Building Materials*, 300:124013, 2021.

[13] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis. Efficient high-resolution deep learning: A survey. *ACM Computing Surveys*, 56(7):1–38, 2024.

FIM-Based Task and Motion Planning for Robotic Bricklaying

Martin Slepicka[1,2], Canberk Yalçıklı[1,2], Mohammed Reza Kolani[1,2], Stavros Nousias[1,2] and André Borrmann[1,2]

[1]Chair of Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2]Georg Nemetschek Institute, Technical University of Munich, Germany

mohammadreza.kolani@tum.de, martin.slepicka@tum.de

Abstract -

Construction robotics has a significant potential to automate a wide range of construction tasks, but it requires a structured pipeline that links high-level design representations to concrete robotic actions. Robotic bricklaying is particularly well-suited to automation because it consists of repetitive, predefined steps. However, it also introduces process-planning challenges that must be resolved at the level of individual components. Fabrication Information Modeling (FIM) has been introduced as an intermediate layer between design and fabrication for construction-scale additive manufacturing. This study extends the Fabrication Information Modeling framework to enable the transformation of high-level architectural design data into detailed, machine-interpretable specifications suitable for robotic bricklaying operations. Specifically, it proposes an integrated FIM-based data and process representation that supports the systematic planning and execution of automated bricklaying tasks. The framework's utility is validated through detailed bricklaying simulations and a successful real-world application within a controlled laboratory environment.

Keywords -

Fabrication Information Modeling; Industry Foundation Classes; Automation in Construction; Robotized Bricklaying

1 Introduction

Robotics in construction is increasingly adopted to automate repetitive, rule-based tasks in domains such as masonry, welding, and structural steelwork. Through their inherent precision and reliability, robotic systems enhance both occupational safety and operational productivity in hazardous, labor-intensive construction environments [1]. Brick wall assembly, characterized by standardized operations, is a prime candidate for robotic automation [2]. However, achieving a fully automated workflow requires addressing the inherent geometric complexities of masonry [3] and establishing a digital system that bridges building design with automated construction processes [4].

Building Information Modeling (BIM) serves as the foundation, providing a semantically rich digital representation of the built asset [5, 6]. Despite the advancements in digital modeling, a significant gap persists between BIM-based design data and robotic task execution. Robots remain unable to autonomously interpret the high-level, human-centric abstractions used to define BIM entities into the precise sequences required for physical construction [7]. While BIM provides a robust geometric and semantic foundation for Industry 4.0 integration, bridging this gap requires a dedicated methodology that extends BIM data into digital manufacturing workflows, transforming static information into actionable robotic instructions [8].

Fabrication Information Modeling (FIM) acts as a BIM-based digital manufacturing framework that transforms high-level design data into detailed, robot-interpretable instructions. By establishing a cyber-physical link between digital models, processes, and the physical asset, FIM enables integration across the design, fabrication, and assembly phases [4].

This study proposes an FIM-based task-planning framework for automated robotic brick wall assembly. The framework enriches the BIM model by incorporating a fabrication information model, which adds semantic fabrication attributes at the brick level and enables robotic execution of the assembly plan. This process translates high-level entities (e.g., wall definitions) into granular data that supports masonry task planning and execution. The proposed approach was integrated into a robotic simulation framework and validated through real-world robotic bricklaying.

The remainder of this paper is organized as follows: Section 2 provides background on robotic bricklaying, fabrication information modeling in various applications, and FIM-based planning and execution, and FIM; Section 3 details the proposed methodology; Section 4 describes the system implementation in simulation and laboratory environments; Section 5 presents the results; and finally, Section 6 concludes the paper.

2 Background

Automation in construction has become a central research focus, as digital technologies and robotic systems are increasingly being investigated for their potential to enhance productivity, safety, and efficiency on construction sites[7]. Within this broader research context, robotic bricklaying remains a promising yet relatively understudied area of construction automation. It provides a valuable basis for investigating task and motion planning, execution strategies, and information requirements in automated construction workflows.

2.1 Robotic bricklaying

Several studies have explored the integration of robotics into bricklaying workflows at different levels of automation. Early methods demonstrated that human–robot collaboration, often facilitated through mobile or augmented systems, can adapt to on-site conditions, thereby enhancing placement accuracy and reducing material waste [9]. Moving beyond collaborative scenarios, the automation of cable-driven robotic systems has been studied using BIM-driven simulation frameworks that optimize robot kinematics, process physics, and site logistics, resulting in collision-free and energy-aware trajectories [10]. Higher levels of automation have been achieved in a robotic masonry context; while Zhou et al. [11] embedded viscoelastic mortar behavior into a force–position feedback controller, enabling physics-informed adjustment of trajectories, velocities, and bond thickness. Complementing these, data-centric pipelines such as the IFC-to-Robot Operating System (ROS) workflow [7] have begun to translate BIM/IFC wall models into brick-level Cartesian pick-and-place tasks and validate them in a simulation environment. Furthermore, [12] presents a multi-robot coordination system for bricklaying that optimizes robot placement, material location, and task decomposition to reduce overall construction time.

Taken together, existing studies remain project- or platform-specific and only partly address full automation of generating robot-interpretable task data from digital design entities. This leaves robotic brick-wall assembly and construction in need of a framework that can systematically handle robotic process planning down to individual brick components, while resolving discrete process-planning challenges, such as junction assembly scenarios involving T- and L-junctions of connected brick walls within a workspace and an arm robot's reachability.

2.2 Fabrication Information Modeling

Fabrication Information Modelling (FIM) is a BIM-centric digital fabrication methodology introduced to the construction sector by [4, 8]. It is designed to structure

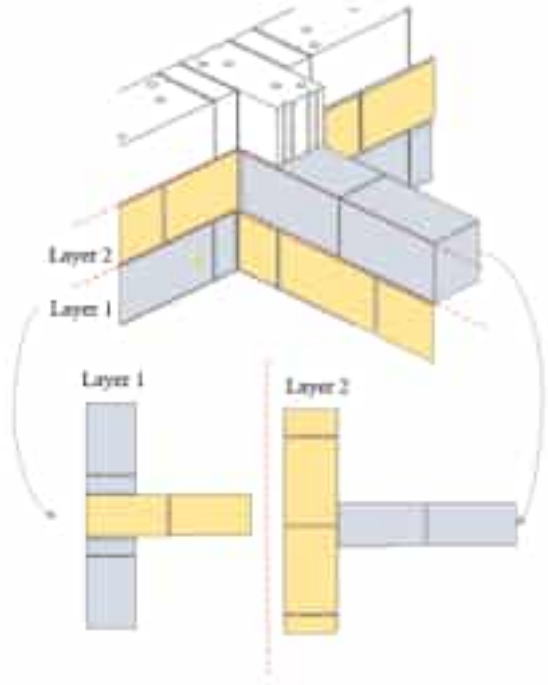


Figure 1. Illustrates the layer-by-layer changing priority wall status between walls and its relevance on the connection point brick placement adaptation [13]

and automate the data flow between high-level BIM geometry and robot-interpretable execution information. Unlike BIM, which primarily describes design intent, constructable element properties, or abstract spatial representations, FIM explicitly models how a structure will be fabricated and processed by deriving fine-grained fabrication data, toolpaths, material-deposition coordinates, sequencing, and machine-control parameters. A key step in FIM's geometry processing is component slicing, where wall or fabrication elements are processed to ordered layers that expose intermediate contours for planning and manipulation [4], enabling FIM to reference fabrication-relevant geometry at multiple levels of granularity. The framework further enables designers to interact with intermediate fabrication representations such as slice contours and path curves, forming a foundational stage for integrating feedback, optimisation, and robotic parameter control functionalities [8]. While representing fabrication and process information via slicing, the Industry Foundation Classes (IFC) schema enables aggregation, allowing multiple geometric representations to co-exist simultaneously [4]. Aggregation capability supports both automation, real-time monitoring and recalibration mechanisms such as RTDE (Real-Time Data Exchange)[14], which integrate material behaviour, nozzle pressure, movement dynamics, and environmental sensor feedback to enable closed-loop

process control during extrusion-based Additive Manufacturing [15]. The proposed framework addresses the research gap by approaching the FIM framework from a BIM-to-FIM perspective and extending FIM's established methodology and IFC-based information base for robotic bricklaying [15].

Despite these advances, robotic brick wall assembly still presents multi-scale challenges, particularly because assembly plans often lack the detailed fabrication and execution information required for robotic implementation, especially at geometrically complex conditions such as wall junctions.

2.3 FIM-based robotic execution

Recent work on construction robotics has increasingly explored methods for linking BIM-based representations with robotic task and motion planning. Initial system-oriented studies demonstrated that IFC models can be converted into robot-compatible planning environments, enabling motion planning, localization, and simulation-based task definitions [16]. Parallel graph-based approaches further showed that BIM's semantic and geometric information can be structured in machine-queryable formats to support robotic navigation and information processing [17]. While these efforts contribute important translation mechanisms, they primarily operate at the project or system level, where the focus remains on robot control, simulation environments, or data conversion pipelines, rather than component-scale fabrication. To improve component awareness, smart-component-oriented workflows segment IFC building entities into metadata-rich representations that carry individual states and robot-communication interfaces, enabling automated IFC–SDF conversion and robotic simulation execution [18]. Similarly, RDF-based middle-layer models organize BIM entities into semantic knowledge graphs that support ontology-based querying and planning of robotic tasks [19]. Although these methods improve data standardization and component-level communication, they do not yet provide a fabrication-focused structure that models building elements according to their manufacturing process to be executed in a single manufacturing process.

Overall, existing studies demonstrate multiple methodologies for establishing BIM–ROS bridges across different scales of robotic construction task planning and execution. However, current studies partly address the automation of generating robot-interpretable task data from design entities. This leaves robotic brick-wall assembly in need of a framework that can handle process planning at the individual brick component level, while resolving discrete process-planning challenges of brick walls.

To address these challenges, we developed a FIM framework that is structured as a pipeline that relies on the IFC data structure to manage and utilize fabrication data for robotic brick wall assembly. Within this pipeline, planning utilizes geometric and semantic fabrication information derived from individual wall elements, aggregating them into a structure-level FIM for the composite brick wall. Fabrication information is represented using a Labeled Property Graph (LPG) in the FIM Information Graph, which is suitable for representing IFC data structures while enabling efficient retrieval and updates of task data. The planning loop maps back the fabrication task states and geometric parameters, enabling component-level task validation, adaptation, and feedback exchange with robotic control. The framework is validated in a simulation environment, followed by robotic execution in a real-world experiment, confirming its applicability beyond digital planning.

3 Methodology

The workflow presented in this work builds on the research of M. Slepicka and A. Borrmann [15], which addressed the transformation of high-level design information into detailed, robot-interpretable data for additive manufacturing. In this context, continuous material deposition in additive manufacturing can be conceptualized as an "infinite brick," whereas the notion of a "finite brick" represents its discretized counterpart, i.e., robotic bricklaying. To support "finite brick", the FIM data structure is extended to represent individual bricks and their corresponding tasks for robotized assembly, as illustrated in Fig. 2.

Similar to 3D printing, a bricklaying fabrication model can be defined by discretizing the component into individual layers, where each layer is described by the brick placement surface (slice) – usually a plane – and the path along which the bricks are placed (toolpath). For "finite brick", the toolpath (`IfcCurve`) is additionally discretized into a list of individual points (`IfcPointOnCurve`), each representing the center of placement for a brick. For regularly shaped bricks, slices are aligned with the top surface of each brick, toolpaths follow their longitudinal centerline, and the discretized points are positioned at the center of their top surface. Thus, compared with additive manufacturing, an additional discretization step is introduced to provide discrete brick placement positions rather than a continuous toolpath.

For data storage, the IFC schema supports multiple IfcShapeRepresentations definitions per layer (cf. Section 2.2). This approach enables the systematic organization of fabrication information into multiple levels of detail, enhancing the model's adaptability and supporting efficient data retrieval. To fully describe the placement of a cuboid (brick) in three-dimensional space, both position and orientation are required. Rather than defining these

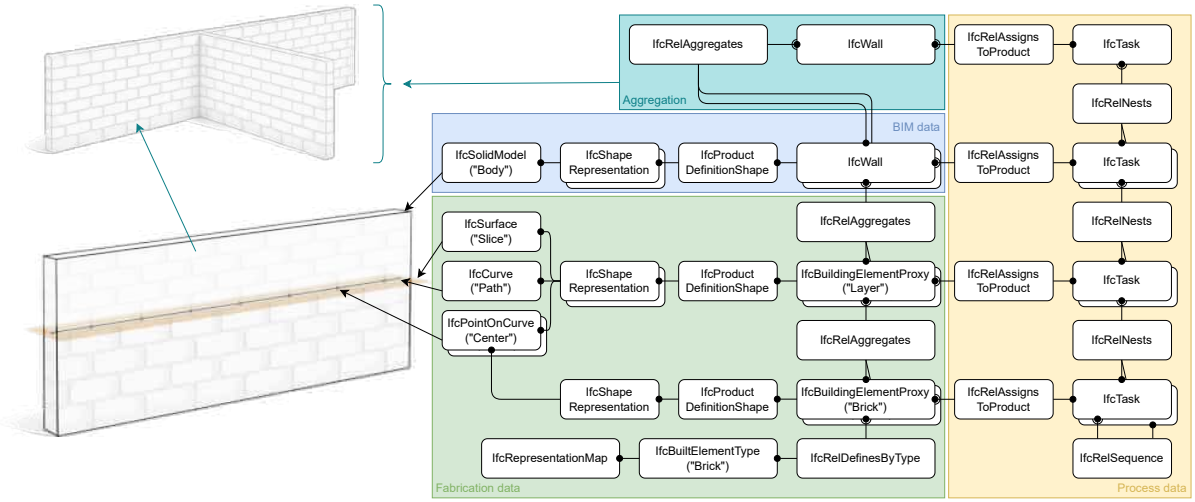


Figure 2. FIM data structure for describing brick walls composed of multiple segments. In the data structure, geometric (left), semantic (middle), and process information (right) are represented.

explicitly, they are captured parametrically through three representations for each layer: the slicing surface (slice) provides the normal orientation, the path provides the lateral orientation, and the point list provides the coordinates. At the same time, this parametric formulation establishes clear rules for discretization from volume to slice, from slice to path, and ultimately to the list of points. For each point, a semantic entity that defines the individual bricks is modeled using `IfcBuildingElementProxy`.

Overall, as defined in [15], a FIM model describes a contiguous component and its corresponding process. For brick walls, we define a contiguous component as the aggregation of wall segments for which all bricks can be placed without violating any placement condition (i.e., each brick must be fully supported). For example, a T-junction of two wall segments built with interlocking bricks at the joint can be regarded as a single contiguous wall. That means, the BIM model is subdivided into multiple FIM models, each representing one contiguous component.

For each FIM model, a new aggregated wall entity is generated that groups all BIM elements belonging to the same fabrication process (see Fig. 2, top). The adaptive planning algorithm identifies how individual wall entities are connected via the relationship `IfcRelConnectsElements` (see Ifc 4.3.2 Documentation [20, Section 5.4.3.48]). With this relationship, it is possible to define the connection type of adjacent wall segments. If the relationship is not present in the model, the adjacent wall segments are regarded as separate components.

In this study, the positional planning algorithm follows an iterative approach. After extracting the relevant BIM data, each wall segment is discretized into individual brick positions, as described in the preceding sections. Brick positions located near junctions are then modified to interlock adjacent segments in an alternating pattern, as illustrated in Fig. 1. Once all brick positions are defined, the bricklaying sequence is determined; here, a straightforward layer-by-layer strategy is adopted. More advanced sequencing algorithms could be applied in future work to optimize construction speed and minimize robot motion.

Aligned with the overall component breakdown structure (wall, wall segment, layer, and individual brick), the process can be modeled in the same hierarchical way. Although the semantic elements can implicitly suggest a task sequence, FIM describes the manufacturing process explicitly using `IfcTask`. As illustrated in Fig. 2 (right), each modeled brick is assigned an individual task representing its placement action. These tasks are organized hierarchically, with brick-level tasks nested within layer tasks, which in turn are nested within component-level tasks and higher-level groupings. In addition, individual tasks are connected through sequencing relationships (`IfcRelSequence`). These relationships can be planned and modeled in a separate planning stage to explicitly define the placement order.

To enable FIM-based robotic bricklaying in ROS, the individual placement tasks must be translated into executable robot instructions. Since each task is linked to the corresponding “Brick” IFC entity, the target pose can be retrieved directly from the FIM model; combined with the known source pose of the next brick, this allows a trajectory to be generated using a standard motion planner. To

support visualization and simulation within the ROS environment, SDF files are additionally required to represent the brick entities (cf. Section 2.3). In the FIM approach, however, SDF models are not generated through a one-to-one BIM-to-SDF conversion for every individual brick, but rather once for each brick type occurring in the model. For subsequent use, the corresponding SDF file can also be linked directly within the FIM model by means of an external reference (`IfcExternalReference`).

4 Implementation

This section presents a proof-of-concept implementation. In the implemented pipeline, BIM data is accessed using the open-source library IfcOpenShell, geometric modeling is performed with PythonOCC, and the resulting fabrication information is stored in a Neo4j graph database via Cypher queries. The implementation's validity is subsequently evaluated using a small-scale robotic setup comprising a UR10e robotic arm equipped with a Robotiq 2F-85 two-finger gripper, controlled via ROS2.

BIM-to-FIM pipeline initialization. As a starting point for the iterative pipeline, a BIM model must be provided. Using IfcOpenShell, the model is first filtered for brick walls. All brick walls that are connected via an `IfcRelConnects` relation are then grouped and aggregated. Then, for each aggregated wall, a separate FIM model is created by setting up a Neo4j database and populating it with all relevant IFC data for that wall (cf. Fig. 2, "Aggregation" and "BIM data"). As a subsequent step, the geometric information of all wall segments is derived to enable slicing operations analogous to the 3D printing methodology described in Section 2.

Segment-wise wall discretization. Using PythonOCC, the volumetric model – typically represented as a BRep or an extruded surface – is subdivided into horizontal slices, referred to as "fabrication layers", with a spacing defined by the brick height. For each layer, a centerline is computed as an alignment curve along which brick positions can be planned. These alignment curves are then discretized into a set of points representing the positions of individual bricks. The discretization is performed according to the selected brick-laying pattern, for example, the pattern shown in Fig. 1, in which every second layer is shifted by half a brick length along the centerline to create an interlocking arrangement. The spacing between adjacent brick positions is determined by the brick length plus a 1 cm clearance space. In this study, the brick dimensions were set to $12 \times 4.8 \times 3.6$ cm. In addition, bricklaying was simplified to dry stacking to focus on the pick-and-place workflow, so no additional allowance for mortar application was required. In this planning stage, the discretization is performed naively, placing only full bricks even when they exceed segment boundaries, providing a straightforward discretization of all individual wall segments.

Brick placement. In accordance with the discretization, semantic entities are instantiated for each layer and for each discretized point using the `IfcBuildingElementProxy` class. The layer entities are designated as `IfcFabricationLayer`, while the brick entities are labeled `IfcBrick`. For all instances, properties can be defined to capture the most relevant characteristics, such as identifiers or static parameters. The geometric representations created in the previous steps are attached to the semantic entities as illustrated in Fig. 2. And, to describe the individual bricks efficiently, type elements (`IfcBuiltElementType`) are instantiated for each type of brick including a representation map that defines the brick geometry.

Rule-based adaptations of brick positions and shapes. In a second iteration, all brick positions located near a segment boundary are reexamined. Two cases are distinguished: either the brick is located near a junction, indicated by an `IfcRelConnects` relationship, or it lies close to the boundary (end) of the wall segment. If the position is near the wall end, it is evaluated whether the corresponding brick extends beyond the wall's boundary representation. When this occurs and an adjustment of the brick length is required, the brick entity is reassigned to a different type entity with suitable dimensions, creating a new type entity if necessary. If the resulting brick length falls below a predefined threshold, a minimum length is enforced and the adjacent brick is shortened accordingly. This adaptation is constrained by the requirement that the modified arrangement must still satisfy the interlocking brick pattern, thereby ensuring proper interlocking across layers. For each updated brick, the position of the corresponding center point is then recomputed parametrically and updated in the model.

In the alternative case, where a brick location is close to a junction, the brick is repositioned so as to interlock with the adjacent wall segment (see Fig. 1, bottom), while the length of the neighboring brick within the same wall segment is adjusted accordingly. Simultaneously, the affected bricks in the connected wall segment are modified to accommodate the interlocking brick. As in the previous case, shape-related changes are handled through the associated type entities, whereas positional changes are reflected in the point list. Across layers, this interlocking is alternated between the involved wall segments. Within the connected wall segment, bricks that overlap with the interlocking brick are identified and trimmed as required. If a brick

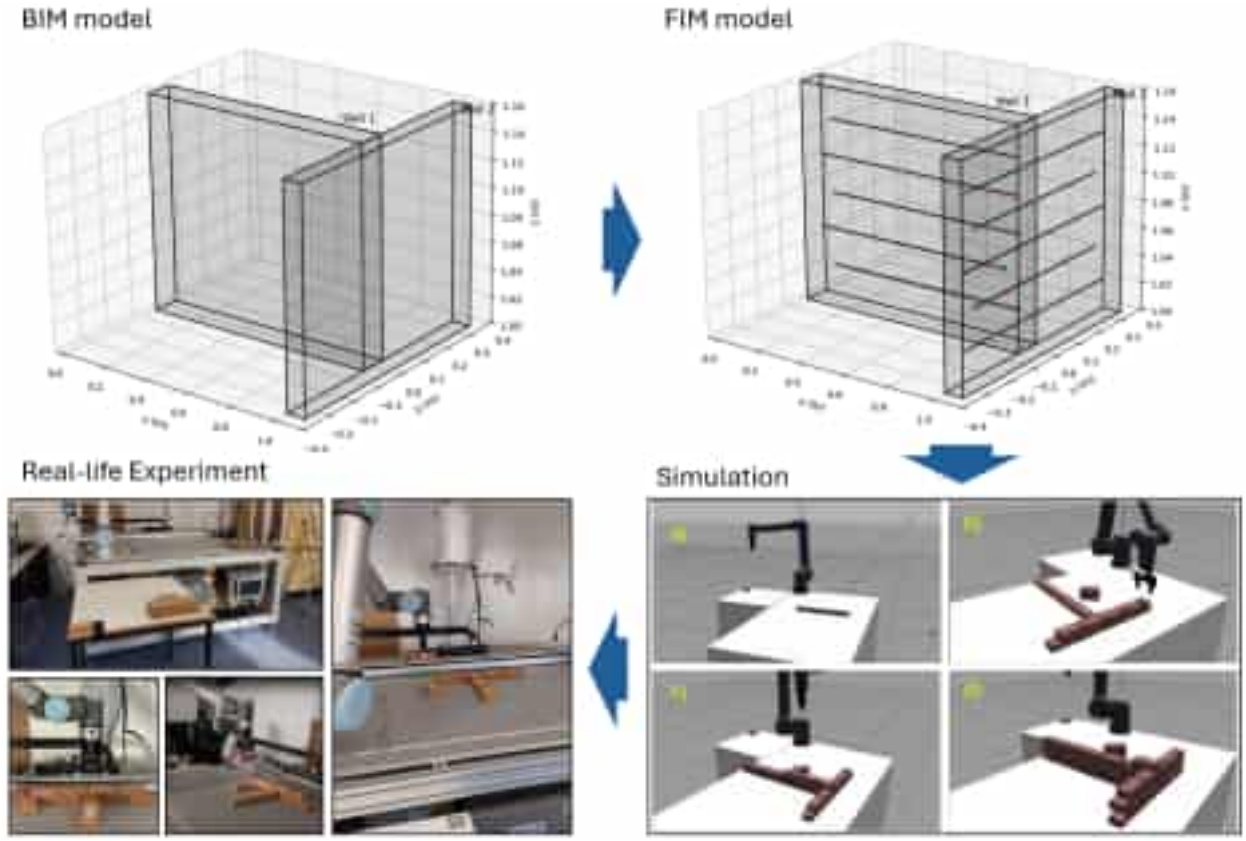


Figure 3. BIM model (top left); FIM model (top right); (a) Sequencing starts with priority wall, (b-c) Priority wall changes with layer change, (d) Successful execution in simulation (bottom right); and real-world execution using the UR10e robot (bottom left).

is split during this process, an additional discretization point and a corresponding `IfcBuildingElementProxy` are created. All resulting modifications are propagated back to the model by updating center point positions and brick dimensions accordingly.

Task planning. As an additional information layer built on top of the brick positions, the bricklaying sequence can be represented within FIM. To achieve this, each "IfcBrick" instance is associated with an `IfcTask` instance that abstractly describes the corresponding brick placement action (see Fig. 2). Consecutive placement actions are then linked using `IfcRelSequence`. In this study, sequencing is performed in a straightforward, layer-by-layer manner.

Each `IfcTask` instance represents the action of placing a single brick, includes attributes describing the task's execution state, and is linked to an `IfcTaskTime` instance that stores scheduling information. The task state can take the values *NOTSTARTED*, *STARTED*, or *COMPLETED*. The associated task time specifies the planned execution time and enables tracking of the actual execution duration.

FIM-based Robot Execution. Once the BIM-to-FIM pipeline is complete, the FIM model is ready for execution. Robot-interpretable commands are then generated sequentially for each modeled task. Each task is linked to an "IfcBrick" instance, from which the brick's position and orientation are retrieved via a simple Cypher query and transformed into a robot target pose. By additionally specifying a source position and orientation for the brick, motion paths can be generated that guide the robot to the pickup location (source pose) and from there to the drop location (target pose). Finally, gripper actions are defined to fully describe the pick-and-place action.

The streamed robotic commands consist of Tool Center Point (TCP) orientations, Cartesian movements, and gripper actions. The generated task sequences are recorded in JSON or YAML and transmitted via TCP/IP to ROS2 as granular brick action commands for execution. By streaming task data per brick instead of exporting a single static command file, the robot controller remains contin-

uously connected and synchronized with the component-level database during execution.

As an extra layer of safety, the whole process can be simulated first in a physics-based simulator, e.g., Gazebo, to detect potential issues, such as collisions or reachability problems. The simulation replicates the real robotic laboratory setup. After the simulation, the developed FIM-based fabrication can then be executed in real life using the robotic arm system.

In this study, the experiment was constrained to the reachable workspace of a single UR10e robot, with each wall discretized into a 6×6 brick grid. A T-junction case consisting of two connected walls was evaluated over 6 layers. T-junction handling resulted in 14 of 72 bricks being replanned as shown in Figure 3. After the simulation was executed without collisions, the process was successfully executed in real life using a UR10e robot arm equipped with a Robotiq 2-finger gripper 2F-85. The results of both simulation and real execution are illustrated in Fig. 3.

5 Results

The developed pipeline is responsible for the complete workflow, from geometric and semantic processing of BIM wall entities to producing the composite structure's FIM for robotic execution. Therefore, in this section, the end-to-end applicability of the developed FIM pipeline is demonstrated in the following steps: *i) Fabrication-relevant geometric and semantic information processing and mapping. ii) Deploying the adaptive planning loop. iii) Simulation execution. iv) Real-life robotic execution.*

6 Conclusion and Future Work

In this paper, a methodology for automating the robotic brick wall assembly task is presented as a BIM-to-FIM task and motion planning pipeline. The framework extends fabrication information modeling by enriching IFC-based wall geometry and semantic relationships to generate robot-interpretable process data, while leveraging this semantic information to enable rule-based adaptations to process-planning complexities, such as wall junctions. The previously introduced slicing approach is further extended to discrete Cartesian brick-placement tasks, and an adaptive planning method is proposed to maintain consistency. A streaming interface synchronizes the robot controller with FIM, ensuring that execution reflects the current fabrication state while preserving prior adaptations. The proposed approach was validated in the same simulation and real-life experiment with the UR10e setup, demonstrating that the pipeline bridges the gap between design and robotic execution in an integrated, automated workflow. However, the current evaluation should be interpreted as a proof of concept, as it was conducted in a constrained setup involving a single robot workspace, a 6*6 brick grid, and a six-layer T-junction configuration. Accordingly, the present results primarily demonstrate functional feasibility and workflow integration rather than quantitative performance in terms of execution time, planning efficiency, or geometric placement deviation.

The proposed approach demonstrates that brick-level process adaptation and junction handling can be automated directly by leveraging BIM semantics with an FIM-based planning process, while presenting the framework as a fabrication-centered, component-aware planning pipeline.

Overall, this work demonstrates that Fabrication Information Modeling can serve as a synchronized intermediate layer for robotic brick wall assembly, supporting automated sequencing and geometry-based task adaptation within a unified model. While highlighting FIM's task-agnostic and adaptable, the developed methodology opens the door to more comprehensive scenarios, such as multiple junctions and geometric complexities.

Future work will therefore extend the framework to larger and more diverse wall assemblies, including openings, corners, and multiple junction configurations, in order to assess its scalability under more realistic construction conditions. Furthermore, it will be extended by integrating process feedback mechanisms, such as sensor output and camera data, to enable continuous, process-aware, adaptive planning. In addition, future studies will incorporate quantitative evaluation metrics, including planning time, execution time, task success rate, and deviation between planned and executed brick placement. The current layer-by-layer sequencing strategy will also be extended toward optimization-based sequencing methods to improve process efficiency and reduce unnecessary robot motion.

Contribution statement

Martin Slepicka: Conceptualization, Writing, Reviewing, Supervision. **Canberk Yalçıklı**: Implementation, Writing. **Mohammad Reza Kolani**: Conceptualization, Writing, Reviewing, Supervision. **Stavros Nousias**: Reviewing, Supervision. **André Borrmann**: Writing, Reviewing.

Acknowledgements

The presented research was funded by the TUM Innovation Network "CoConstruct" and the German Research Foundation (DFG) in the frame of the Research Unit FOR 5672, grant number 517965147.

References

[1] Juan Manuel Davila Delgado, Lukumon Oyedele, Anuoluwapo Ajayi, Lukman Akanbi, Olugbenga

Akinade, Muhammad Bilal, and Hakeem Owolabi. Robotics and automated systems in construction: Understanding industry-specific challenges for adoption. *Journal of building engineering*, 26:100868, 2019.

[2] Z. Dakhli and Z. Lafhaj. Robotic mechanical design for brick-laying automation. *Cogent Engineering*, 4(1):1361600, 2017. doi:10.1080/23311916.2017.1361600.

[3] A. Fingrut, K. Crolla, and D. Lau. Automation complexity - brick by brick. *Proceedings of the International Conference on Computer-Aided Architectural Design Research in Asia*, January 2019. doi:10.52842/conf.caadria.2019.1.093.

[4] M. Slepicka, S. Vilgertshofer, and A. Borrmann. Fabrication information modeling: Closing the gap between building information modeling and digital fabrication. In *Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC)*, 2021. doi:10.22260/isarc2021/0004.

[5] M. Poljansek. Building information modelling (BIM) standardization (EUR 28977 EN). Technical report, Publications Office of the European Union, 2017.

[6] F. Craveiro, J. P. Duarte, H. Bartolo, and P. J. Bartolo. Additive manufacturing as an enabling technology for digital construction: A perspective on construction 4.0. *Automation in Construction*, 103:251–267, 2019. doi:10.1016/j.autcon.2019.03.011.

[7] M. Brinkhoff and M. R. Kolani. Deriving fabrication information from BIM for automated robotic task planning. In *Proc. of Forum Bauinformatik 2024*, Sep 2024. doi:https://doi.org/10.15480/882.13486.

[8] M. Slepicka, S. Vilgertshofer, and A. Borrmann. Fabrication information modeling: Interfacing building information modeling with digital fabrication. *Construction Robotics*, 6(2):87–99, 2022. doi:10.1007/s41693-022-00075-2.

[9] K. Dörfler, T. Sandy, M. Giftthaler, F. Gramazio, M. Kohler, and J. Buchli. *Mobile Robotic Brickwork*, pages 204–217. Springer International Publishing, Cham, 2016. ISBN 978-3-319-26378-6. doi:10.1007/978-3-319-26378-6_15.

[10] T. Bruckmann and R. Boumann. Simulation and optimization of automated masonry construction using cable robots. *Advanced Engineering Informatics*, 50:101388, August 2021. doi:10.1016/j.aei.2021.101388.

[11] Y. Zhou, B. Huang, B. Dong, Y. Wen, and M. Duan. Dynamic robotic bricklaying force-position control considering mortar dynamics for enhanced consistency. *Automation in Construction*, 174:106090, 2025. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2025.106090.

[12] Mohammad Reza Kolani, Stavros Nousias, and André Borrmann. Coordinating robotized construction using advanced robotic simulation: The case of collaborative brick wall assembly. *arXiv preprint arXiv:2405.17171*, 2024. doi:10.35869/proceedings_egice2024.

[13] KLB Klimaleichtblock GmbH. Walls and wall connections. `https://en.klb-klimaleichtblock.de/walls-and-wall-connections.html`, 2026.

[14] M. Slepicka, J. Helou, and A. Borrmann. Real-time data exchange (rtde) robot control integration for fabrication information modeling. In *Proceedings of the 40th ISARC*, Jul 2023. doi:https://doi.org/10.22260/ISARC2023/0016.

[15] M. Slepicka and A. Borrmann. Fabrication information modeling for closed-loop design and quality improvement in additive manufacturing for construction. *Automation in Construction*, 168:105792, 2024. doi:10.1016/j.autcon.2024.105792.

[16] S. Kim, M. Peavy, P.-C. Huang, and K. Kim. Development of bim-integrated construction robot task planning and simulation system. *Automation in Construction*, 127:103720, April 2021. doi:10.1016/j.autcon.2021.103720.

[17] P. Pauwels, R. De Koning, B. Hendrikx, and E. Torta. Live semantic data from building digital twins for robot navigation: Overview of data-transfer methods. *Advanced Engineering Informatics*, 56:101959, 2023. doi:10.1016/j.aei.2023.101959.

[18] A. Zhu, P. Pauwels, and B. De Vries. Component-based robot-prefabricated construction simulation using ifc-based building information models. *Automation in Construction*, 152:104899, 2023. doi:10.1016/j.autcon.2023.104899.

[19] X. Wang, S. Wang, C. C. Menassa, V. R. Kamat, and W. McGee. Automatic high-level motion sequencing methods for enabling multi-tasking construction robots. *Automation in Construction*, 155:105071, 2023. doi:10.1016/j.autcon.2023.105071.

[20] buildingSMART International. Ifc 4.3.2 documentation, 2024. URL `https://ifc43-docs.standards.buildingsmart.org`.

Scaffold Component Recognition and Joint-Level Understanding using 3D Gaussian Splatting and Vision-Language Models

Boyu Wang[1] and Borja García de Soto[1]

[1]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
bw3005@nyu.edu, garcia.de.soto@nyu.edu

Abstract

Scaffold inspection is critical for ensuring construction safety. However, scaffolds are typically dense, multilayered, and continuous structures, making them difficult to analyze automatically using traditional geometric methods. In this paper, a framework for automated scaffold component recognition and joint-level understanding is proposed by leveraging Vision–Language Models (VLMs) and 3D Gaussian Splatting (3DGS) to interpret complex scaffold assemblies. Domain knowledge is incorporated to decompose the structure using 3DGS, where photorealistic rendering is exploited to generate clear visualizations for semantic interpretation by pretrained VLMs. The semantic understanding of scaffold components lays a crucial foundation for future compliance-oriented analysis by enabling the precise identification and spatial reasoning of code-relevant and design-relevant scaffold components. This approach is expected to provide a reference for analyzing other continuous and complex structures in construction scenarios.

Keywords –

Scaffold; Vision-Language Model (VLM); 3D Gaussian Splatting (3DGS); Component Recognition

1 Introduction

Construction safety remains a critical concern worldwide, with falls from height consistently ranking among the leading causes of construction-related fatalities [1]. Temporary structures such as scaffolds are major contributors to these incidents, as their frequent erection and dismantling cycles significantly increase the risk of structural failure [2]. To mitigate these risks, safety regulations require that scaffolds be inspected regularly for compliance with both construction codes and design specifications throughout their lifecycle.

As shown in Figure 1, scaffold inspection tasks can be broadly categorized into two parts. The first is code compliance checking, which verifies whether key components meet regulatory requirements in terms of installation and dimensions. For example, whether base plates are properly placed, whether the vertical distance between adjacent guard rails is within the 600 mm limit, and whether toe board heights exceed the minimum of 90 mm, as mandated by standard scaffolding safety regulations [3]. The second is design compliance checking, which ensures that the scaffold has been assembled exactly as planned including checking for missing elements such as bracings, and verifying whether couplers match the design specifications.

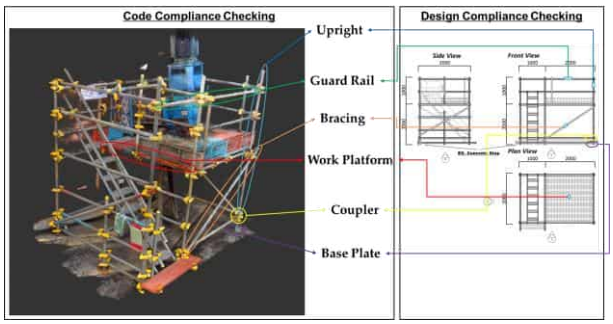


Figure 1. Conceptual workflow of scaffold component classification and compliance checking against code and design specifications. (Created by the authors based on the information provided by onsite engineers).

However, current inspection practices are still predominantly manual, relying heavily on the experience of safety managers and site supervisors. These methods are time-consuming, labor-intensive, and prone to human error, particularly in dense, multilayered scaffold systems. Although recent research has explored the use of computer vision for scaffold analysis, most existing approaches are limited to the geometric interpretation of textureless 3D point clouds [4-7]. These methods often

lack the photorealistic fidelity and semantic richness needed for detailed component recognition and are difficult to scale for complex real-world scenarios.

To address these limitations, this paper presents a framework for automated as-built scaffold understanding by integrating 3D Gaussian Splatting (3DGS) with Vision-Language Models (VLMs). Specifically, unlike sparse and textureless point clouds, 3DGS reconstructs 3D scenes with high photorealistic fidelity, preserving both accurate geometry and rich visual appearance (color, texture) to provide a high-quality foundation for downstream analysis [8-10]. Concurrently, VLMs solve the generalizability limitations of traditional deep learning methods by leveraging the photorealistic 2D renderings generated by 3DGS to perform flexible, prompt-based semantic segmentation without the need for extensive task-specific training [11-13]. This bypasses the need for extensive task-specific 3D training data and enables human-interpretable component recognition. Together, this integration serves as a robust first step toward enabling automated component recognition and compliance checking workflows.

To operationalize this integrated framework, the proposed pipeline follows a systematic workflow. Starting from site-captured point clouds and images, a photorealistic 3DGS model is constructed to represent the as-built scaffold with high visual and geometric fidelity. The structure is then decomposed into semantically meaningful subcomponents via 3DGS-based editing. Pretrained VLMs are applied to these substructures to extract and interpret semantic information, enabling fine-grained component segmentation and classification. Furthermore, critical joints can be examined in detail by adjusting viewpoints and zooming within the 3DGS model, supporting close-up inspection and providing the necessary information for compliance verification.

2 Literature Review

2.1 Scaffold Inspection

Previous research on scaffold inspection has primarily focused on two major directions. The first line of work aims to monitor scaffold deformation, which is crucial for early detection of structural instability. Two main approaches have been explored: sensor-based methods, which use strain gauges or other physical sensors to measure deformation directly [14,15], and vision-based methods, which infer deformation by analysing pixel-level changes across image sequences [16,17]. While the latter can generate heatmaps to visualize deformation trends, they often lack fine-grained, component-level information.

The second line of research focuses on semantic segmentation of scaffolds into discrete components such as uprights, guardrails, bracing, work platforms, stairs, and toe boards. These methods typically operate on 3D point cloud data, attempting to segment the scaffold into meaningful elements [4,7,18]. However, they often rely heavily on manually defined geometric rules, which are sensitive to noise and make the process unstable and difficult to generalize for large-scale scaffold inspection. While some deep learning–based methods have been proposed to improve robustness and generality, they usually require extensive annotated training data. Moreover, due to the black-box nature of deep models, it is often difficult to diagnose or fine-tune the system when performance is suboptimal in complex real-world conditions.

2.2 3D Gaussian Splatting

To remove the ambiguity of point cloud analysis–based component extraction, 3DGS provides a highly suitable solution. 3DGS is an emerging technique for real-time radiance field rendering, offering a compelling alternative to traditional mesh or point cloud based scene representations [9]. 3DGS models a scene as a collection of anisotropic 3D Gaussians, each parameterized by position, scale, orientation, color, and opacity. These Gaussians are efficiently splatted onto the image plane to generate photorealistic renderings from arbitrary viewpoints at interactive frame rates.

Recent studies have demonstrated the capability of 3DGS to capture fine surface details and complex lighting with high fidelity, which is especially valuable for detailed visual inspection in unstructured and cluttered environments. Compared to sparse and noisy point clouds, 3DGS produces a dense, view-dependent scene representation that preserves both geometric and appearance cues. This enables users to zoom into specific components, change viewpoints, and visually isolate substructures, which is essential for accurate and interpretable inspection workflows. Although 3DGS has been primarily applied in computer graphics and immersive rendering, its potential for component-level analysis in the Architecture, Engineering, and Construction (AEC) industry has not been fully explored. This paper leverages 3DGS to enhance semantic decomposition and inspection of as-built scaffold structures.

2.3 Vision-Language Model

While 3DGS enables photorealistic visualization of as-built scenes, VLMs provide potential solution to the semantic understanding of image content. VLMs have recently emerged as powerful tools for multimodal reasoning by jointly learning from both visual and textual data. These models, often pre-trained on large-scale

image text pairs, are capable of associating natural language descriptions with visual content, enabling tasks such as image captioning, visual question answering, and semantic segmentation through textual prompts. Recent advances, including CLIP [19], BLIP [20], and GPT-4V [21], have demonstrated strong generalization across domains, even with zero-shot or few-shot inputs. Such capabilities make VLMs especially promising for domains with limited labeled data or where human-like semantic understanding is needed [22-24].

In the context of construction automation, VLMs offer a new paradigm for interpreting complex scenes in a flexible and human-interpretable way. Unlike traditional vision models that require task-specific training datasets, VLMs can be prompted using natural language queries to identify or describe components, attributes, or spatial relationships. Despite their success in general vision tasks, the application of VLMs in 3D construction environments remains underexplored, and this paper aims to bridge that gap through an integration with 3DGS.

3 Methodology

This paper presents a method for as-built scaffold understanding by integrating 3DGS with VLMs, providing a foundation for downstream compliance inspection tasks. As illustrated in Figure 2, the proposed workflow consists of five main stages: (1) Mobile laser scanning is first employed to capture high-resolution point clouds and images of the as-built scaffold. (2) These multi-modal data are fused to generate a photorealistic 3DGS scene representation that preserves both geometric structure and visual appearance. (3) A 3DGS-based structure decomposition process is then applied to isolate scaffold substructures based on spatial layout. (4) Each substructure is further segmented through VLM-based semantic analysis, where pretrained VLMs interpret rendered 3DGS views to assign component-level labels (e.g., upright, bracing, guardrail). (5) Finally, a component-wise graph is constructed to encode the topological relationships among scaffold elements, enabling joint-level analysis that is relevant for code compliance and design conformance.

.

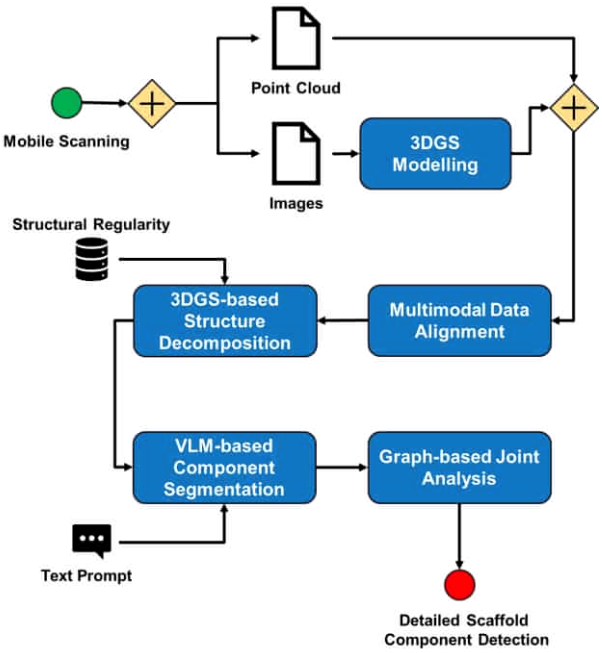


Figure 2. Pipeline for 3DGS-based, VLM-enabled understanding of as-built scaffold data.

3.1 Mobile Scanning

To capture as-built conditions of scaffold structures, a mobile scanning system was employed, as shown in Figure 3. The data acquisition platform integrates a LiDAR sensor and two high-resolution RGB cameras, enabling synchronous collection of both dense point clouds and photorealistic images. The scanning process is conducted manually by an operator walking around the scaffold with the handheld device, ensuring flexible access to complex and confined construction environments. The system adopts Fast-LIVO2 [25] as its Simultaneous Localization and Mapping (SLAM) backend, which provides robust real-time pose estimation and trajectory optimization. This allows for accurate spatial registration of the captured LiDAR frames and camera images, resulting in a globally consistent 3D representation of the scene. The acquired data serve as the foundation for 3DGS and subsequent semantic analysis in the pipeline.



Figure 3. Mobile scanning for data collection

3.2 3DGS Modelling

To obtain a high-fidelity and interpretable representation of the as-built scaffold structure, 3DGS is employed as the core modeling technique. Rather than relying on explicit mesh reconstruction or discrete point clouds, the scene was modeled as a radiance field parameterized by a set of volumetric 3D Gaussians. This representation enables photorealistic rendering from arbitrary viewpoints and provides strong support for visual inspection and downstream analysis.

The modeling process was initialized using a sparse Structure-from-Motion (SfM) reconstruction, from which camera parameters and initial 3D point distributions were estimated to guide the placement of Gaussians. An iterative optimization was then performed to refine each Gaussian's parameters including position, scale, opacity, and appearance (encoded via spherical harmonics) by minimizing a photometric loss between rendered and observed images. A visibility-aware renderer was used to simulate training views and guide updates. To ensure compactness and sufficient coverage, a density control mechanism was applied, dynamically adding or pruning Gaussians based on gradient feedback.

As shown in the left part of Figure 4, this process transforms a large set of unstructured images into a coherent 3D model that captures both photorealistic appearance and explicit spatial structure. Unlike traditional image-based methods, the resulting 3DGS model encodes geometric relationships in 3D space, allowing the scene to be interpreted as a spatially organized structure. This spatial coherence provides a robust foundation for structural decomposition, where scaffold components can be systematically separated based on their relative positions, orientations, and connectivity, enabling higher-level reasoning about component grouping and configuration in a data-driven yet interpretable manner.

3.3 Multimodal Data Alignment

Since the 3DGS reconstruction does not preserve an accurate global scale, SLAM-based point clouds were incorporated to leverage their precise geometric information. To integrate the SLAM point cloud with the 3DGS model, a registration step was performed. As the two representations exist in different coordinate frames, a two-stage alignment process was adopted. First, a coarse transformation was estimated using a three-point alignment based on corresponding anchor points identified in both datasets. The alignment was then refined using the Iterative Closest Point (ICP) algorithm to minimize geometric discrepancies. As a result, the photorealistic 3DGS model was accurately aligned with the globally referenced SLAM point cloud, as shown in Figure 4, enabling consistent spatial reasoning across subsequent processing stages.

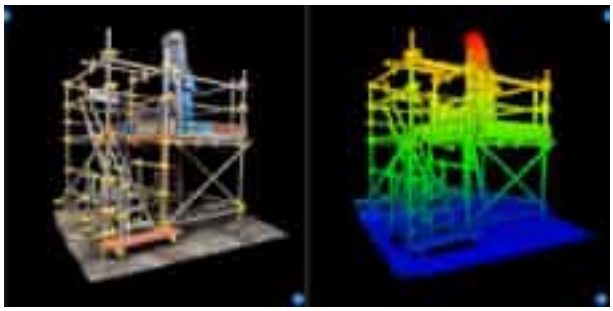


Figure 4. Geometrically aligned 3DGS (left) and point clouds (right).

3.4 3DGS-based Structure Decomposition

To simplify the analysis of complex scaffold assemblies, structure decomposition was performed directly on the 3DGS model. Based on the global coordinate system established during registration, the scaffold was segmented into meaningful substructures according to their geometric locations and spatial hierarchy. Specifically, spatial filters were applied along the X, Y, and Z axes to crop out regions corresponding to the front face, side frames, and intermediate bays. This axis-aligned cropping strategy leverages the inherent regularity of scaffold structures, which are typically aligned with global Cartesian axes. It enables clean separation of components while preserving substructure integrity, offering a simple and interpretable approach suitable for early-stage inspection workflows. By iteratively applying these geometric cropping operations, the continuous and highly layered scaffold structure was decomposed into a set of clear, modular components, each representing a distinct structural unit. As illustrated

in the left part of Figure 5, the front face has been isolated, enabling precise VLM-based analysis and component segmentation in subsequent stages.



Figure 5. 3DGS-based structure decomposition and VLM-enabled component recognition. The left image shows the cropped frontal view of the scaffold's 3DGS model; the right image illustrates the segmentation results using the text prompt "bar".

3.5 VLM-Based Component Segmentation

To achieve semantic understanding of scaffold components, a vision-language model (VLM) was employed to segment 3DGS-rendered views based on natural language prompts. Specifically, SAM3 [12] is adopted, a recent model designed for promptable concept segmentation, which supports segmentation of all instances matching a concept prompt across images and videos. SAM 3 accepts short noun phrases or image exemplars as prompts and returns corresponding segmentation masks along with object identities. Prompt-based segmentation was adopted to minimize manual labeling and avoid rigid class definitions, enabling flexible and interpretable semantic analysis. In the proposed pipeline, one rendering of the 3DGS model were generated for each substructure from calibrated camera poses, and each image was prompted with component-level concepts such as "bar", "stair", and "board". The SAM 3 model uses a DETR-based detector and a mask decoder, conditioned on both text and visual prompts, to produce high-quality masks. A presence token is used to decouple recognition and localization, improving segmentation accuracy in cluttered scenes. The segmentation results for the front side, obtained using the text prompt "bar," are presented in Figure 5.

The resulting 2D masks were projected back to the 3D space by leveraging the 2D–3D correspondences inherent in the 3DGS rendering process. Specifically, each 2D pixel in the segmentation mask was mapped to its nearest 3D point clouds by reprojecting the camera ray and using visibility-aware associations. In this way, the semantic labels predicted by the VLM were transferred to the corresponding subset of 3D point clouds, effectively labeling the scaffold model at the point level. By aggregating segmentation results across multiple views, a consistent 3D semantic segmentation was obtained as shown in the left part of Figure 6, forming the basis for subsequent component graph construction and compliance analysis.

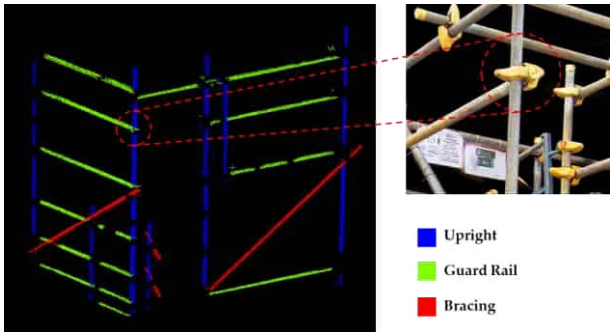


Figure 6. Scaffold component segmentation and corresponding as-built details. Left: segmented scaffold structure showing uprights (blue), guard rails (green), and bracings (red). Right: 3DGS of selected joints located at the intersection between a guard rail and an upright.

3.6 Graph-based Joint Analysis

Following instance segmentation, each scaffold component is represented as a labeled subset of 3D points corresponding to structural elements such as uprights, guard rails, and bracings. To capture the structural connectivity among these components, a spatial graph is constructed based on proximity-based interaction. Specifically, pairwise distances between segmented components are computed, and potential joints are inferred wherever two elements fall within a predefined spatial threshold. These candidate connection points serve as the edges in a component-level graph, while the scaffold members themselves define the nodes, forming a graph-based abstraction of the scaffold topology.

Once the graph is established, the precise 3D positions of joint candidates are localized using the high-fidelity 3DGS model. Since 3DGS preserves photorealistic appearance and scene geometry, rendered views of the detected joints allow for visual verification and detailed analysis of couplers and base plates. As illustrated in Figure 6, this enables the system to infer the presence of joints from geometry, providing detailed visual cues that assist in inspecting their physical condition and evaluating compliance with structural regulations. By combining geometric reasoning with appearance-based cues, this graph-based framework supports joint-level inspection in a scalable and interpretable manner. While a full-scale quantitative evaluation of the spatial graph and automated compliance checking are reserved for future work, this graph-based

abstraction serves as a crucial topological bridge. Currently, it is utilized to localize candidate joints, directly guiding the targeted detection and visual inspection of critical connection elements like couplers (as evaluated in Section 4). Thus, this joint-level understanding provides a necessary foundation for downstream analyses.

4 Experiments and Results

To evaluate the effectiveness of the proposed pipeline, experiments were conducted on real-world scaffold scenes captured via mobile laser scanning. The evaluation focused on the object detection accuracy of individual scaffold components, assessed by comparing the predicted labels to manual annotations. The tested scene serves as a representative real-world case intended to validate the feasibility of the proposed pipeline, rather than to provide a comprehensive performance benchmark. Therefore, the current findings should be considered preliminary. Broader validation across multiple sites and scaffold configurations is left for future work. Table 1 summarizes the detection performance across five key component categories (uprights, guard rails, braces, couplers, and base plates) using standard evaluation metrics: true positives (TP), false positives (FP), false negatives (FN), precision (Pre), recall (Rec), and F1-score (F1).

Table 1 Results of scaffold component detection

Component Type	TP	FP	FN	Rec.	Pre.	F1
upright	7	0	0	1.00	1.00	1.00
guard rail	11	0	0	1.00	1.00	1.00
bracing	4	0	0	1.00	1.00	1.00
coupler	28	0	2	1.00	0.93	0.96
base plate	4	0	0	1.00	1.00	1.00

A total of 56 components were annotated and used as ground truth. The results show that the proposed method achieves high precision across the evaluated categories in this specific scene, with nearly zero false negatives recorded among the 56 annotated components. Notably, the detection of couplers (achieving 28 true positives) and the joint-level visualization shown in Figure 6 (Right) serve as a validation of the graph-based joint analysis proposed in Section 3.6. By successfully localizing the intersections among linear structural elements through the spatial graph, the system was able to accurately isolate and identify the corresponding couplers. This demonstrates the practical utility of the graph-based abstraction in supporting detailed component inspection, even though a few false negatives occurred due to occlusions and limited visibility in the scanned data. While these initial results demonstrate the fundamental feasibility of the proposed 3DGS and VLM-based pipeline for component recognition, we acknowledge that this small-scale evaluation does not fully capture the complexities of diverse construction sites.

Although a direct quantitative comparison is challenging due to the absence of standardized public datasets for scaffold inspection, contrasting the proposed approach with existing literature underscores its novelty and practical feasibility. Traditional methods predominantly rely on fully supervised 3D point cloud segmentation networks (e.g., PointNet or RandLA-Net). In the broader computer vision domain, these general frameworks typically achieve mean Intersection over Union (mIoU) of around 80% on standard point cloud benchmarks [26]. Specifically for scaffolding, recent studies report mIoU scores ranging from 80% to 87% [4]. However, these approaches suffer from two major limitations: first, they are restricted to semantic segmentation and cannot provide the object-level (instance) representations essential for detailed compliance checking; second, they necessitate hundreds of manually annotated 3D frames for training. In contrast, our proposed 3DGS and VLM-based framework achieves highly competitive object-level recognition, yielding higher recall and precision on major components in our test case without requiring any task-specific 3D training data. The core novelty lies in synergizing the high-fidelity rendering of 3DGS with the zero-shot, open-vocabulary capabilities of VLMs. This paradigm fundamentally bypasses the data preparation bottleneck inherent in traditional point cloud methods, demonstrating strong potential for rapid, scalable deployment in dynamic construction environments.

Overall, the integration of 3DGS modeling and VLM-based segmentation enables high-quality detection for the tested component categories, demonstrating promising potential for automated inspection workflows.

5 Conclusion

This paper presents an automated pipeline for scaffold component recognition and joint-level understanding by integrating mobile laser scanning, 3DGS, and VLMs. Through precise 3D reconstruction and prompt-based segmentation, the system effectively identifies and labels structural elements such as uprights, guard rails, and couplers. A graph-based analysis further enables joint detection and spatial reasoning. Experimental results demonstrate high accuracy in component detection, with zero false positives and minimal false negatives.

While the current evaluation is based on a single representative scene with a limited number of components, the results confirm the initial feasibility of the approach as a proof-of-concept. However, several

critical limitations remain. The small dataset size does not sufficiently demonstrate the framework's generalizability across diverse scaffold configurations. Furthermore, the system's robustness in highly cluttered construction environments, where severe occlusions, varying lighting conditions, and dense overlapping structures are common, requires further investigation. These aspects, along with sensitivity to scan quality and prompt phrasing, will be addressed in future work through large-scale data collection and comprehensive benchmarking across multiple complex sites.

Overall, although a complete automated compliance checking system remains a subject for future work, the proposed approach successfully lays the essential foundation for such systems by providing reliable component recognition and joint-level topological understanding capabilities.

Acknowledgement

This research was partially supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013.

References

[1] S. Tan, Speech at SCAL Construction Safety, Health and Security Seminar, (25 March 2026) (2017)

[2] S.M. Whitaker, R.J. Graves, M. James, P. McCann, Safety with access scaffolds: Development of a prototype decision aid based on accident analysis, Journal of Safety Research, 34(3) (2003), pp. 249-261

[3] Workplace Safety and Health (Scaffolds) Regulations, Ministry of Manpower, Singapore, (2011)

[4] J. Kim, J. Kim, N. Koo, H. Kim, Automating scaffold safety inspections using semantic analysis of 3D point clouds, Automation in Construction, 166 (2024), pp. 105603

[5] J. Kim, J. Kim, Y. Kim, H. Kim, 3D reconstruction of large-scale scaffolds with synthetic data generation and an upsampling adversarial network, Automation in Construction, 156 (2023), pp. 105108

[6] J. Kim, D. Chung, Y. Kim, H. Kim, Deep learning-based 3D reconstruction of scaffolds using a robot dog, Automation in Construction, 134 (2022), pp. 104092

[7] Q. Wang, Automatic checks from 3D point cloud data for safety regulation compliance for scaffold work platforms, Automation in Construction, 104 (2019), pp. 38-51

[8] F. Luan, S. Sun, H. Zhang, Y. Yin, K. Wang, J. Yang, 3D Gaussian splatting technologies and extensions: A review, Neurocomputing, (2025), pp. 131629

[9] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, 3D Gaussian splatting for real-time radiance field rendering, ACM Transactions on Graphics, 42(4) (2023), pp. 139:131-139:114

[10] T. Lu, M. Yu, L. Xu, Y. Xiangli, L. Wang, D. Lin, B. Dai, Scaffold-gs: Structured 3d gaussians for view-adaptive rendering, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 20654-20664

[11] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu, Grounding DINO: Marrying dino with grounded pre-training for open-set object detection, Arxiv Preprint, (2023)

[12] N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris, C. Ryali, K.V. Alwala, H. Khedr, A. Huang, Sam 3: Segment anything with concepts, arXiv preprint arXiv:.16719, (2025)

[13] L.H. Li, P. Zhang, H. Zhang, J. Yang, C. Li, Y. Zhong, L. Wang, L. Yuan, L. Zhang, J.-N. Hwang, Grounded language-image pre-training, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10965-10975

[14] S. Sakhakarmi, J. Park, C. Cho, Enhanced machine learning classification accuracy for scaffolding safety using increased features, Journal of Construction Engineering Management, 145(2) (2019), pp. 04018133

[15] C. Cho, K. Kim, J. Park, Y.K. Cho, Data-driven monitoring system for preventing the collapse of scaffolding structures, Journal of Construction Engineering Management, 144(8) (2018), pp. 04018077

[16] Y. Jung, H. Oh, M.M. Jeong, An approach to automated detection of structural failure using chronological image analysis in temporary structures, International Journal of Construction Management, 19(2) (2019), pp. 178-185

[17] Y. Feng, F. Dai, H.-H. Zhu, Evaluation of feature-and pixel-based methods for deflection measurements in temporary structure monitoring, Journal of Civil Structural Health Monitoring, 5(5) (2015), pp. 615-628

[18] R. Luo, Z. Zhou, X. Chu, W. Ma, J. Meng, 3D deformation monitoring method for temporary structures based on multi-thread LiDAR point cloud, Measurement, 200 (2022), pp. 111545

[19] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, Learning transferable visual models from natural language supervision, International Conference on Machine Learning, 2021, pp. 8748-8763

[20] J. Li, D. Li, S. Savarese, S. Hoi, Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models, International Conference on Machine Learning, 2023, pp. 19730-19742

[21] OpenAI, GPT-4 technical report, (2023)

[22] B. Wang, Z. Chen, M. Li, Q. Wang, C. Yin, J.C. Cheng, Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes, Automation in Construction, 162 (2024), pp. 105384

[23] B. Wang, F. Lin, M. Li, Z. Liang, Z. Chen, M. Wang, J.C. Cheng, Informative As-Built Modeling as a Foundation for Digital Twins Based on Fine-Grained Object Recognition and Object-Aware Scan-vs-BIM for MEP Scenes, Advanced Engineering Informatics, 65 (2025), pp. 103382

[24] B. Wang, F. Lin, M. Li, Z. Liang, H. Yue, Q. Wang, J.C. Cheng, Aligning as-built and as-designed: Local point cloud to BIM registration via hybrid visibility map encoding for construction digital twins, Automation in Construction, 180 (2025), pp. 106551

[25] C. Zheng, W. Xu, Z. Zou, T. Hua, C. Yuan, D. He, B. Zhou, Z. Liu, J. Lin, F. Zhu, Fast-livo2: Fast, direct lidar-inertial-visual odometry, IEEE Transactions on Robotics, (2024),

[26] X. Wu, L. Jiang, P.-S. Wang, Z. Liu, X. Liu, Y. Qiao, W. Ouyang, T. He, H. Zhao, Point transformer v3: Simpler faster stronger, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 4840-4851

Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model for BIM Compliance Checking

Yicheng Zhao[1], Zhoupeng Wang[1], Xingbo Gong[2] and Xingyu Tao[1*]

[1] Dept. of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong 999077, China.
[2] Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong 999077, China.
yicheng0220.zhao@connect.polyu.hk, zhou-peng.wang@connect.polyu.hk, xgongai@connect.ust.hk, xingyu.tao@polyu.edu.hk

Abstract

Fire safety compliance checking is essential for BIM-based design review, yet it remains largely manual in practice, leading to substantial time costs. Although automated checking approaches (e.g., NLP-based methods) have been explored, two major challenges remain: (1) in practice, operationalization is largely limited to semantic requirements, whereas clauses involving geometric and topological structures are difficult to execute; (2) automation remains limited because NLP-based methods often require manual mapping tables due to terminology mismatch and unclear condition parsing, and the extracted rules are not sufficiently stable. To address these issues, this paper proposes the Framework of Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model (4FKE-LLM). First, this paper constructs a four-dimensional fire safety knowledge graph—Semantic, Geometric, Topological, and Scenario, which organizes and structurally represents regulations across the four dimensions to support subsequent LLM-based compliance checking. Second, this paper proposes an LLM-based two-layer compliance checking mechanism. The first layer pre-filters regulatory and model information to identify applicable regulations and relevant objects, and the second layer performs dimension-specific regulation matching and reasoning across the four categories.

Experiments on a dormitory building case study show that the 4FKE-LLM framework achieves F1 = 0.863 and Accuracy = 0.885, and outperforms the two selected baselines on this case in terms of classification performance and checkable coverage.

Keywords –

BIM Compliance Checking; Fire Safety Regulations; Knowledge Graph; Four-Dimensional Rule Classification; Large Language Model

1 Introduction

Fire safety inspection is of paramount importance during the BIM design process. In the event of a fire, the evacuation of personnel, smoke control, fire compartmentalization, and firefighting efforts are directly affected by the spatial and component layout of the building [1]. If issues are not detected and corrected during the design phase, the subsequent costs and risks of rework during construction and operations can be significant, impacting both project timelines and safety [2, 3]. Additionally, BIM designs must comply with fire safety regulations [4]. When a project fails to meet key regulatory clauses, designers are required to modify their plans, further delaying the construction schedule. Therefore, conducting timely and reliable fire safety compliance checks during the BIM design phase is essential [1].

Currently, there are three main approaches for fire safety compliance checking in BIM. The first method is manual inspection. Designers or reviewers examine the drawings and models, cross-referencing with checklists, applying judgment based on experience, and performing spot calculations to verify compliance [2]. While this method can address complex cases and incorporate engineering expertise, it is time-consuming, costly, and subjective, with varying results depending on the individual examiner [3]. To overcome this limitation, researchers begin using rule engines. This approach decomposes regulatory clauses into explicit conditions and thresholds, then uses geometric calculations, database queries, or scripts to automatically assess the model [4, 5]. This approach is faster and more stable, but is highly dependent on the complete accuracy of the translation of the rules into a machine-readable format and the ability of the model to contain the required parameters and relationships [6, 7]. When clauses are complex or models lack essential parameters, the rules can fail to execute, reducing the scope of checks. The

third method, based on natural language processing (NLP), attempts to automatically structure regulatory texts and align clauses with model fields for process-based checks [8, 9]. While this method reduces the need for manual clause breakdowns and organization, it still faces issues such as inconsistent terminology and unclear regulation decomposition [8]. These issues can lead to unstable clause-to-model matching, misclassification of requirements, and reduced checkable coverage in practical compliance workflows. More importantly, many fire safety clauses require spatial, component, and connectivity relationships, which are often insufficiently aligned with textual descriptions, limiting their reliability and verifiability [5-7].

In recent years, the capabilities of large language models (LLMs) have progressed rapidly. LLMs can better understand natural language clauses, normalize different expressions, and generate more coherent explanations [10]. This allows LLMs to address the shortcomings of traditional NLP in "semantic understanding" and "expressive rewriting". However, directly applying LLMs to compliance checking still leaves two key research gaps. First, LLMs can generate plausible but inaccurate conclusions and may produce inconsistent judgments across different reasoning rounds [3]. Second, fire safety compliance depends not only on semantics but also on geometric constraints, topological relationships, and attribute dependencies, such as “from where to where,” “which doors connect to which corridors,” “whether a compartment is independent,” and “whether a component’s fire resistance rating meets the requirement.” If these relations are missing from the BIM model or are not structured into usable data, the LLM may understand the text but still cannot produce stable or explainable clause-level judgments. As a result, it tends to work better on semantic clauses while struggling with geometric, topological, and scenario logic [3]. Therefore, LLMs function here not as standalone inspectors, but as supporting reasoning modules. Accordingly, this study aims to enable stable, clause-level fire safety checking by integrating these models with structured four-dimensional rules and enriched BIM data.

Thus, this paper proposes the Framework of Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model (4FKE-LLM) to solve these issues. (Figure 1). First, 4FKE-LLM constructs a four-dimensional fire safety regulation knowledge graph (Semantic, Geometric, Topological, and Scenario), which organizes and structurally represents regulations across the four dimensions to support subsequent LLM-based compliance checking. Second, 4FKE-LLM implements an LLM-based two-layer compliance checking mechanism, the first layer pre-filters regulation and model information to identify applicable regulations and relevant objects, while the second layer performs dimension-specific regulation matching and reasoning across the four dimensions. Rather than simply replacing traditional NLP modules, the key contribution of 4FKE-LLM lies in integrating a four-dimensional rule representation with a two-layer LLM checking workflow to improve classification performance and checkable coverage.

2 Methodology

Framework of Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model (4FKE-LLM)

The 4FKE-LLM Framework integrates BIM and fire safety regulation rules to provide an efficient fire safety compliance checking method (Figure 1). The overall framework comprises three main components: (1) the development of the 4D Fire Safety Knowledge Graph (steps 1-3), (2) BIM fire safety information extraction and data enrichment (steps 4-6), and (3) LLM-based reasoning using "Four-Dimensional Rule Knowledge" and "Model Data" to generate a compliance checking report (steps 7-11).

The first component is the development of the 4D Fire Safety Knowledge Graph. This process involves denoising raw regulatory texts and categorizing them into the four dimensions of semantic, geometric, topological, and scenario-based rules (step 1), enabling the LLM to better understand the structure and content of regulations. The categorized rules are then converted into a knowledge graph (step 2) and stored in a data exchange format (step 3) that is easily understood by the LLM, facilitating subsequent queries.

The second component involves BIM fire safety information enrichment and extraction. Custom parameters are added to the BIM model to enrich the key data necessary for fire safety compliance checks (step 4), which are typically missing in the original BIM model. The relevant parameters are then extracted and converted into the Industry Foundation Classes (IFC) standard (step 5), ensuring interoperability between different platforms. The enriched IFC model is then translated into a data format suitable for the LLM (step 6). This ensures that the model can read and process BIM fire safety information and provides data support for checking.

The third component is LLM-based Two-layer fire safety compliance checking. User can submit requests through a pre-configured prompt at the query interface (step 7). After ingesting the 4D Fire-Safety Knowledge Graph and the BIM data (step 8), the framework performs two-layer checking. In the first layer, the LLM applies predefined filtering rules to identify applicable clauses and select relevant BIM objects (step 9). In the second layer, it conducts rule matching and reasoning over the

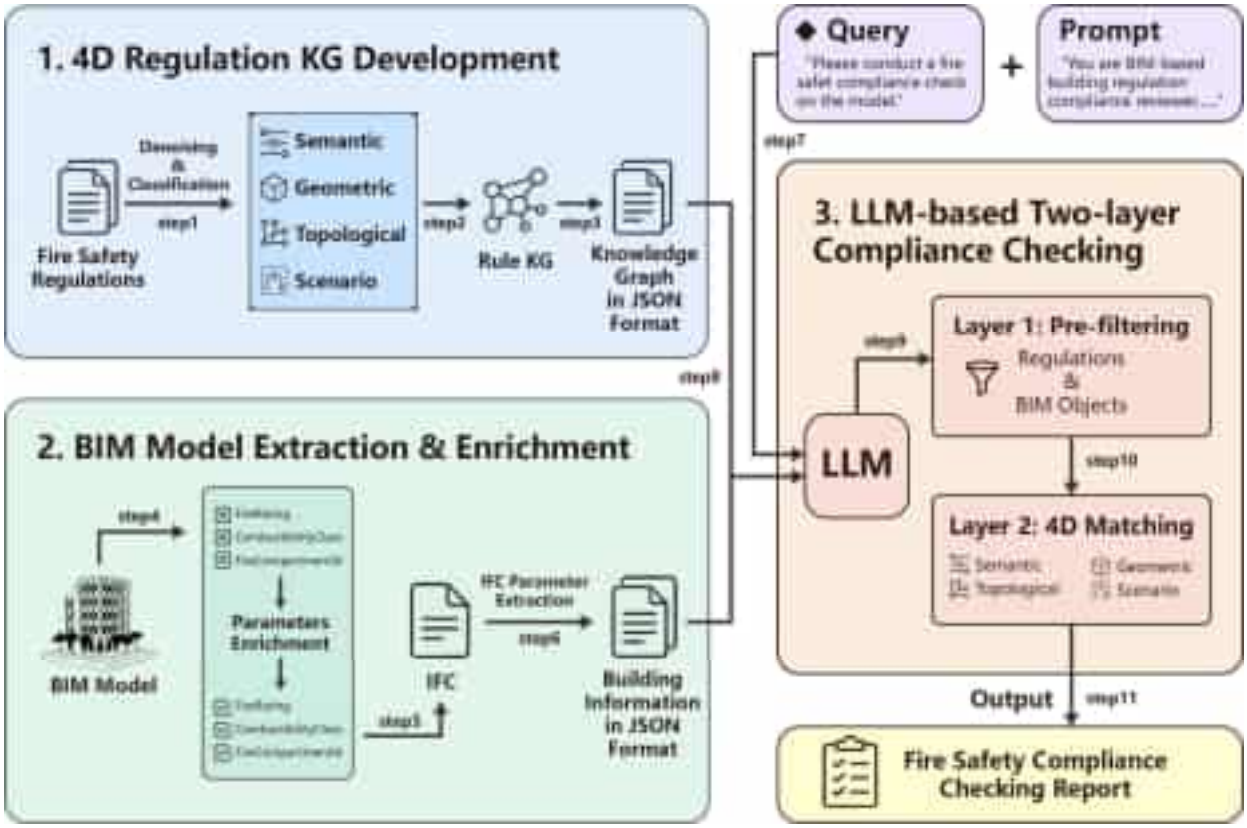


Figure 1: Framework of Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model

four-dimensional regulations using the combined knowledge graph and BIM model data to generate compliance reports (steps 10-11).

2.1 4D Rule Knowledge Graph Development

Figure 2 and Table 1 illustrate the classification of fire safety regulations into four types, each corresponding to different query designs. In this study, this four-type taxonomy is proposed as an operational representation for BIM-based fire safety compliance checking, and its broader applicability beyond the current fire regulations should be further tested in future work. The 4D Rule Knowledge Graph Development method aims to enhance the accuracy and automation of regulatory interpretation and compliance checking, reducing errors and omissions to achieve a more comprehensive and precise compliance evaluation.

Semantic Rules: Semantic regulations describe the relationships between regulations and the attributes that must be satisfied by specific components. For example, they specify the fire resistance rating required for fire doors.

$$(Regulation) - [HAS_ATTRIBUTE_REQ] -> (Component)$$

Geometric Rules: Geometric regulations focus on dimensional and distance constraints for components or spaces. These include minimum widths, maximum heights, and similar spatial metrics.

$$(Regulation) - [HAS_GEOMETRIC_REQ] -> (Component)$$

Topological Rules: Topological regulations define static relationships between components or spaces, such as connectivity, separation, or passage relations. These rules establish spatial connectivity between components using keywords such as "FROM" and "TO."

$$(Regulation) - [DEFINES] -> (TopoRule)$$

Scenario Rules: Scenario regulations define conditional logic in which a specific requirement becomes applicable only when a trigger condition is met. For instance, they specify that certain fire safety equipment must be installed when specific height requirements are met.

$$(Regulation) - [DEFINES_LOGIC] \rightarrow (Scenario)$$

Semantic rules focus on components, attributes, and constraints, mapping clauses to component nodes and abstracting fire resistance ratings and other requirements into attributes and constraint nodes. Geometric rules primarily address dimensional constraints, while topological rules define static relations among spaces or components, like connectivity and isolation. In contrast, scenario rules focus on trigger–requirement logic, following a two-phase structure with "trigger conditions" and "requirements" to manage actions based on specific scenarios, enabling automatic filtering and reasoning.

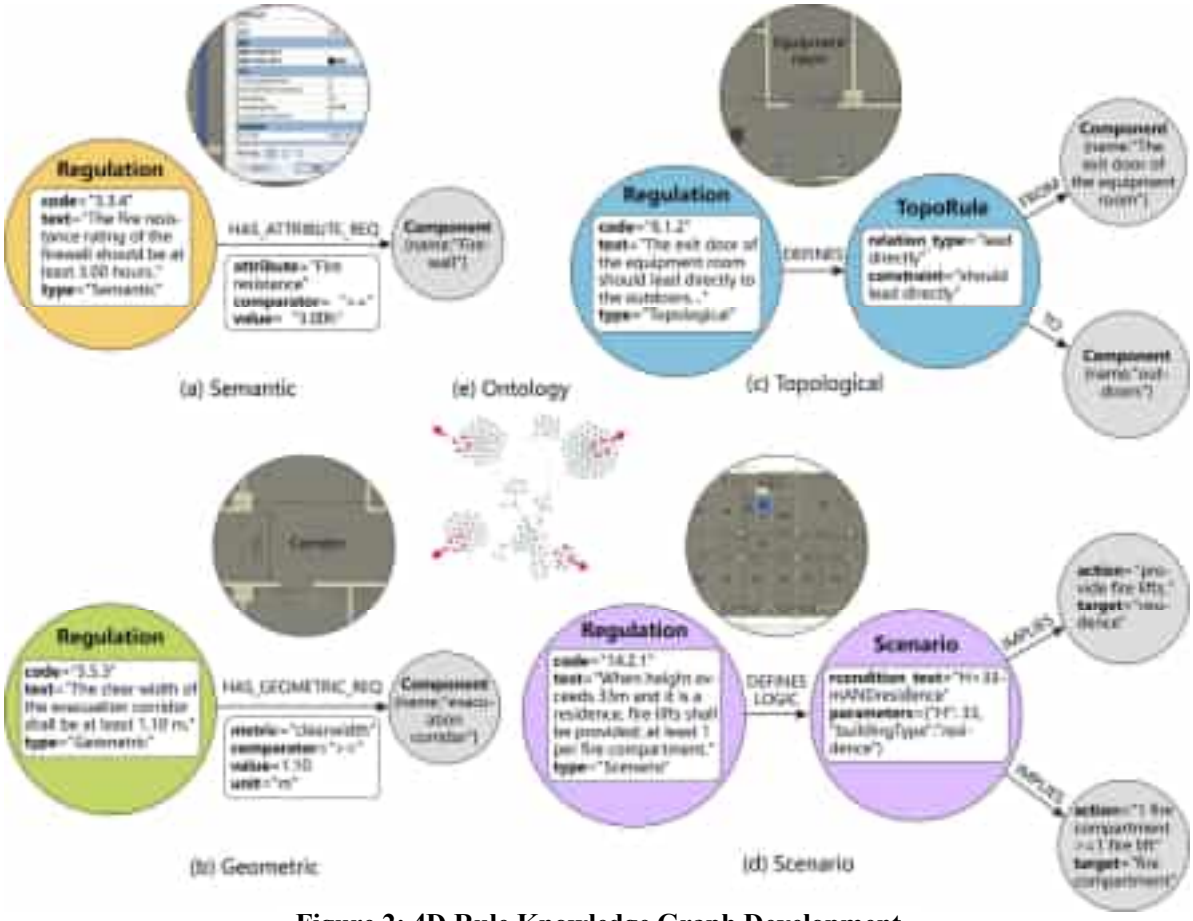


Figure 2: 4D Rule Knowledge Graph Development

Table 1
Node and Relationship Schema for the 4D Rule Knowledge Graph

Category	Node Design	Relationship Design
Semantic	Regulation: code, text, type = "Semantic". Component: name (the constrained component/object type).	**(Regulation)--[:HAS_ATTRIBUTE_REQ]→(Component)** • attribute: attribute name (e.g., fire resistance rating / combustibility) • value: required value (e.g., 2.00h / Class A / Grade A) • comparator: =, >=, <=, prohibition/permission, etc.
Geometric	Regulation: code, text, type = "Geometric". Component: name (component/space type).	**(Regulation)--[:HAS_GEOMETRIC_REQ]→(Component)** • metric: indicator (e.g., area / distance) • value: numeric value • unit: m, m², etc. • comparator: >=, <=, etc.
Topological	Regulation: code, text, type = "Topological". TopoRule: relation_type, constraint. Component: name (space/entity type).	**(Regulation)--[:DEFINES]→(TopoRule)** (the clause defines the rule) **(TopoRule)--[:FROM]→(Component_Source)** (source/endpoint) **(TopoRule)--[:TO]→(Component_Target)** (target/endpoint)
Scenario	Regulation: code, text, type = "Scenario". Scenario: condition_text, parameters. Requirement: action (required/prohibited action), target (affected object/scope).	**(Regulation)--[:DEFINES_LOGIC]→(Scenario)** (defines trigger conditions) **(Scenario)--[:IMPLIES]→(Requirement)** (requires compliance when the condition is met)

2.2 BIM Model Parameter Enrichment & Extraction

Figure 3 shows the BIM Model Parameter Enrichment and Extraction process. It adds key parameters for fire safety compliance checking, such as Fire Resistance Rating and Combustibility Class, to the BIM model so that most model elements contain information relevant to compliance checking. After parameter enrichment, the BIM model is exported in the IFC format. Fire safety attributes are bound to specific components using IFC standard property sets (Psets), enabling standardized cross-platform representation. Schedules and Mapping Tables are also exported to summarize parameters and standardize parameter mapping, naming, meaning, and units to ensure consistency across components and sources. These Mapping Tables are used for IFC/JSON field normalization and schema alignment, and they are not the manual regulation-to-model term mapping tables required by traditional rule-based compliance checking.

To ensure high-quality data input for compliance checks, this paper standardizes and denoises the data in two stages: "BIM to IFC" and "IFC to JSON." During the IFC export process, the Mapping Table is used to unify fire safety parameters' key names, meanings, and units, and essential relationships, such as those between components and rooms, are embedded directly into the data. During the conversion to JSON, the hierarchical and reference relationships in the IFC file are structured, and relevant attributes like GUIDs, component types, spatial contexts, and fire safety parameters are encapsulated in a directly accessible format, removing unrelated properties and complex geometric details. This process ensures that only the minimal necessary data required for the four types of rule checks remain in the final output.

The resulting model JSON is organized by "component—attribute—spatial context" and typically includes parameters such as element ID, IFC type, spatial context (e.g., storey, room), and fire safety-related attributes (e.g., Fire Rating). This JSON file serves as input for LLM-based fire safety compliance checks. It enables direct retrieval and reference, which improves the stability and efficiency of compliance checking.

2.3 LLM-based Two-layer Compliance Checking

The third component is LLM-based Two-layer fire safety compliance checking. Users submit requests through a pre-configured prompt via the query interface, specifying the checking scope (e.g., building type, floor range, or target spaces). After the 4D Fire Safety Knowledge Graph and enriched BIM model data are provided as inputs, the framework performs two-layer checking. In Layer 1, the LLM filters applicable clauses and retrieves the relevant BIM objects. In Layer 2, it evaluates each filtered clause against the corresponding BIM evidence under the semantic, geometric, topological, or scenario category and then generates the compliance report.

3 Validation of the Framework

Figure 4 illustrates the validation process for the framework, which was conducted on a dormitory building case study. The regulations selected for testing include the "General Fire Safety Code for Buildings" GB 55037-2022 [11], the "Fire Safety Code for Interior Building Decorations" GB 50222-2017 [12], and the "Building Fire Safety Code" GB 50016-2014 (2018 edition) [13]. First, the regulatory text was denoised, and

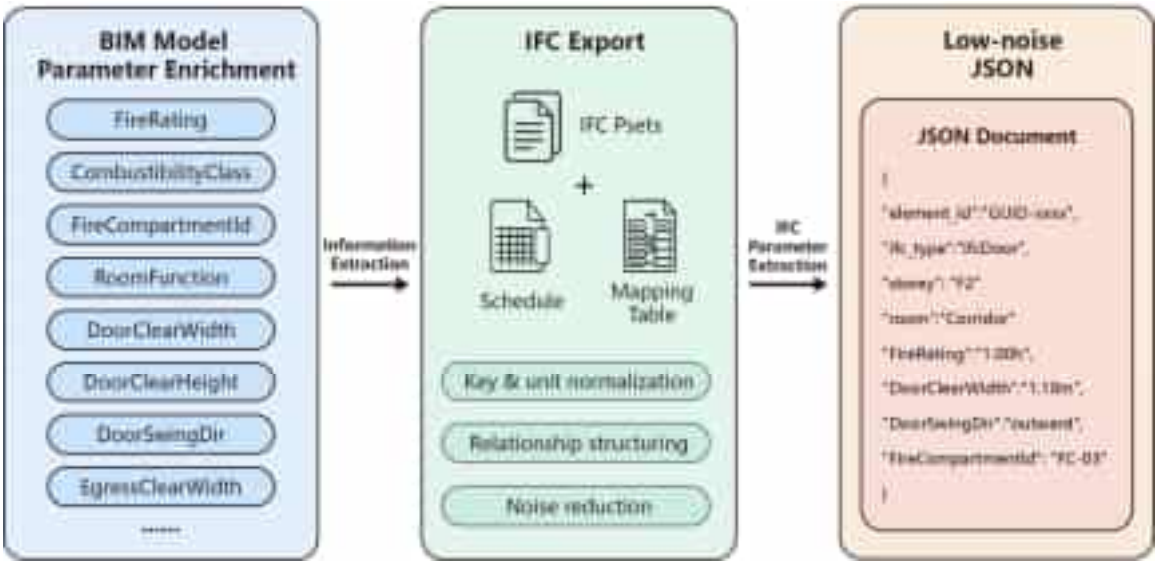
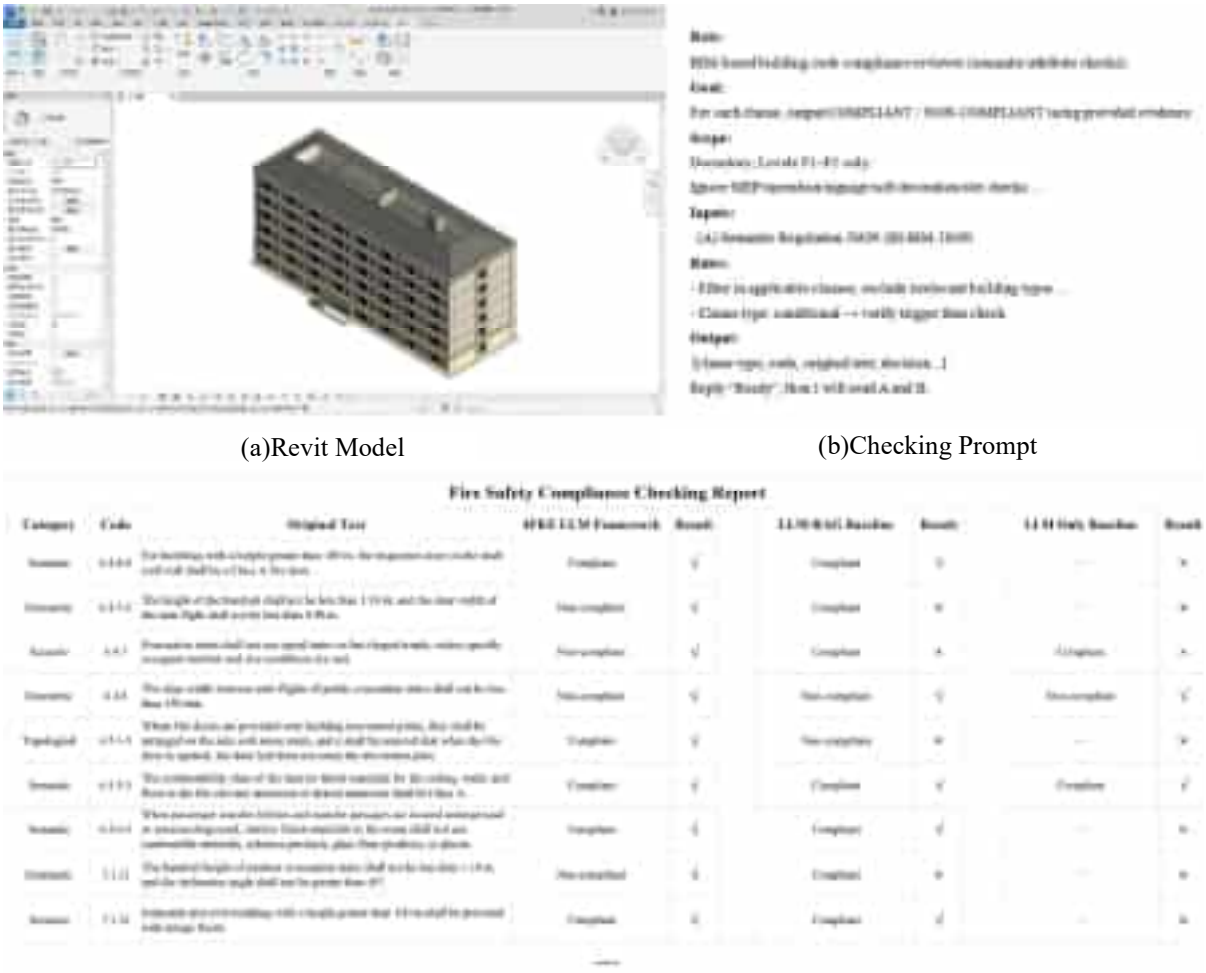


Figure 3: BIM Model Parameter Enrichment & Extraction



(a)Revit Model (b)Checking Prompt

(c)Checking Report

Figure 4: Validation of the Framework

tables were transcribed into natural language, yielding a total of 554 regulatory clauses. The rules were then classified into semantic, geometric, topological, and scenario categories and converted into Cypher queries, which were imported into Neo4j to construct the regulatory knowledge graph [14]. Structured rule data was exported as JSON for subsequent fire safety compliance checks using the LLM.

A BIM model of the dormitory building was created using Revit 2025 (Figure 4(a)), exported in IFC4 format to ensure that spatial boundary relationships and key fields were extractable. The model was then converted into structured JSON data, containing entity types, spatial hierarchies, connectivity relationships, and trigger conditions. During the compliance checking phase, the ChatGPT 5.2 Thinking model was used [15], with a pre-configured prompt (Figure 4(b)) that submitted the "4D Fire Safety Rule JSON + Model JSON" as input. This model was selected because the task requires long-context input handling and multi-step clause-level reasoning over rule JSON and model JSON. As the model is proprietary and may evolve over time, the reported results should be interpreted as a system-level evaluation under a fixed prompt setting rather than a strictly version-stable benchmark. First, the applicable clauses for the dormitory building were filtered, resulting in 122 relevant regulations. These clauses were then validated according to the four types of rules, and a compliance report was generated in Compliant/Non-compliant form (Figure 4(c)).

There are two baselines used for comparison in experiments. First, the LLM-RAG baseline takes the raw regulatory text together with the structured BIM model JSON as input. Second, the LLM-only baseline provides only the structured model JSON and asks the LLM to determine compliance without explicit rule inputs. The generated report is shown in Figure 4(c). For evaluation, each applicable clause-level decision is treated as a binary classification task, where the manually verified result is the ground truth label (Actual) and the model

output is the predicted label (Pred). We report the confusion matrix and compute Accuracy, Precision, Recall, and F1, where Precision measures the reliability of non-compliance predictions, Recall measures the coverage of true non-compliant clauses, and F1 summarizes the trade-off between them. The metrics are defined in Eqs. (1) – (4).

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

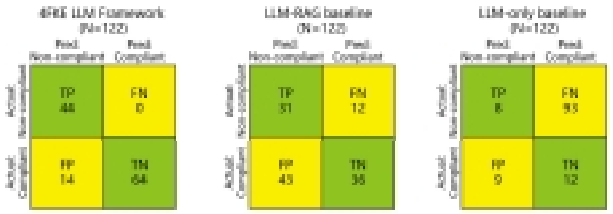
$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

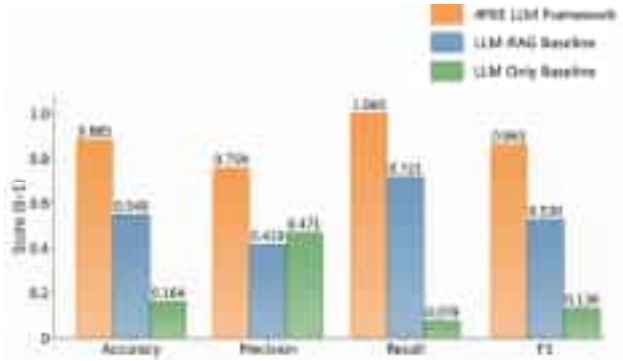
These four metrics are appropriate provided that the positive class is clearly defined. In this study, "Non-compliant" is treated as the positive class, so TP/FP/FN are counted with respect to non-compliance. Under this definition, Precision and Recall directly reflect the two key risks in fire-safety checking: false alarms (FP) and missed violations (FN). We choose these metrics because they capture complementary aspects of clause-level compliance checking: (1) Accuracy reports overall correctness, (2) Precision reflects the cost of false positives (unnecessary manual review or rework), (3) Recall reflects the safety-critical cost of false negatives (missed violations), and (4) F1 provides a balanced summary, avoiding misleading conclusions when class distributions are imbalanced.

Figure 5 summarizes the results (N = 122). In this context, N denotes the total number of clause-level compliance decisions evaluated, that is, the number of applicable clauses assessed across the case. The proposed Framework of Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model (4FKE-LLM) shows the best performance among the three evaluated settings, with TP = 44, TN = 64, FP = 14, FN = 0, yielding Accuracy = 0.885, Precision = 0.759, Recall = 1.000, and F1 = 0.863. This indicates zero missed violations (FN = 0) while maintaining a low false-positive rate. In contrast, the LLM-RAG baseline shows both higher false positives and false negatives (TP = 31, TN = 36, FP = 43, FN = 12), with reduced scores (Accuracy = 0.549, Precision = 0.419, Recall = 0.721, F1 = 0.530). The LLM-only baseline performs the worst, exhibiting severe missed violations (TP = 8, TN = 12, FP = 9, FN = 93; Accuracy = 0.164, Precision = 0.471, Recall = 0.079, F1 = 0.136), suggesting that providing model data alone without explicit regulatory rules does not yield stable clause-level compliance determinations.

In terms of practical expectations for safety-critical fire-safety compliance checking, this study adopts a recall-first evaluation criterion. It aims for Recall as close to 1 as possible (ideally FN = 0, or at least Recall ≥ 0.95) while keeping false positives within a manageable range (e.g., Precision ≥ 0.70 or a low FP rate), and it uses an overall target such as F1 ≥ 0.80 as a pragmatic usability threshold. This recall-first stance is consistent with prior automated fire code–compliance checking studies, which report very high precision/recall performance and, in some settings, argue that Recall is more critical than Precision [9, 16]. Under these criteria, 4FKE-LLM already falls within the expected range (Recall = 1.000, Precision ≈ 0.76, F1 ≈ 0.86), aligning with the risk preference of fire-safety review.



(a) Confusion Matrices



(b)Metrics Comparison

Figure 5: Comparison Result

4 Conclusion

This paper addresses the practical needs of BIM-based fire safety compliance checking by integrating a four-dimensional rule representation, BIM data enrichment, and a two-layer LLM checking workflow for clause-level reasoning. The Framework of Four-Dimensional Fire Safety Knowledge–Enhanced Large Language Model (4FKE-LLM) constructs regulatory knowledge graphs using a four-dimensional classification method and exports callable structured data. BIM model parameters are enriched and converted into structured data that includes attributes, spatial structures, and key relationships. In the dormitory building case study, the framework outperformed the two selected baselines, especially in recall and overall F1. These results suggest that the framework can improve clause-level fire safety compliance checking for the evaluated case. However, the current validation is limited to one

dormitory building and 122 applicable clauses, so broader testing across more building types and regulatory contexts is still needed to examine external validity. Furthermore, certain clauses still rely on finer-grained geometric calculations, and incomplete IFC exports or parameter enrichment may lead to situations that cannot be checked. Future work will focus on expanding the rule coverage and the clause template library. It will also improve IFC-based attribute mapping and the geometric/topological evidence computation modules to enhance cross-project compatibility and robustness.

Acknowledgement

This research is supported by the Hong Kong PolyU (UGC) Start-up Fund for New Recruits (P0056204) and Zhejiang Provincial "Pioneer" and "Leading Goose" Science and Technology Plan Project (2026C04040)

Reference

[1] S. Fischer, C. Schranz, H. Urban, and D. Pfeiffer, "Automation of escape route analysis for BIM-based building code checking," *Automation in construction,* vol. 156, p. 105092, 2023, doi: 10.1016/j.autcon.2023.105092.

[2] R. Amor and J. Dimyadi, "The promise of automated compliance checking," *Developments in the built environment,* vol. 5, p. 100039, 2021, doi: 10.1016/j.dibe.2020.100039.

[3] O. Iversen and L. Huang, "Leveraging large language models for BIM-based automated compliance checking," *Automation in construction,* vol. 182, p. 106707, 2026, doi: 10.1016/j.autcon.2025.106707.

[4] A. S. Ismail, K. N. Ali, N. A. Iahad, M. A. Kassem, and N. T. Al-Ashwal, "BIM-Based Automated Code Compliance Checking System in Malaysian Fire Safety Regulations: A User-Friendly Approach," *Buildings (Basel),* vol. 13, no. 6, p. 1404, 2023, doi: 10.3390/buildings13061404.

[5] J. Peng and X. Liu, "Automated code compliance checking research based on BIM and knowledge graph," *Scientific reports,* vol. 13, no. 1, pp. 7065-12, 2023, doi: 10.1038/s41598-023-34342-1.

[6] A. Li, P. K.-Y. Wong, X. Tao, J. Ma, and J. C. P. Cheng, "An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent," *Advanced engineering informatics,* vol. 66, p. 103375, 2025, doi: 10.1016/j.aei.2025.103375.

[7] X. Tao *et al.*, "Smarter smart contracts for automatic BIM metadata compliance checking in blockchain-enabled common data environment," *Advanced engineering informatics,* vol. 62, p. 102627, 2024, doi: 10.1016/j.aei.2024.102627.

[8] J. Zhang and N. M. El-Gohary, "Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking," *Journal of computing in civil engineering,* vol. 30, no. 2, 2016, doi: 10.1061/(ASCE)CP.1943-5487.0000346.

[9] Y. Wang, Y. Liu, H. Cai, J. Wang, and X. Zhou, "An Automated Fire Code Compliance Checking Jointly Using Building Information Models and Natural Language Processing," *Fire (Basel, Switzerland),* vol. 6, no. 9, p. 358, 2023, doi: 10.3390/fire6090358.

[10] S. Madireddy *et al.*, "Large Language Model-Driven Code Compliance Checking in Building Information Modeling," *Electronics (Basel),* vol. 14, no. 11, p. 2146, 2025, doi: 10.3390/electronics14112146.

[11] *General Fire Safety Code for Buildings (GB 55037-2022)*, Ministry of Housing and Urban-Rural Development of the PRC, Beijing, 2022.

[12] *Code for Fire Prevention in Design of Interior Decoration of Buildings (GB 50222-2017)*, Ministry of Housing and Urban-Rural Development of the PRC, Beijing, 2017.

[13] *Code for Fire Protection Design of Buildings (GB 50016-2014, 2018 Edition)*, Ministry of Housing and Urban-Rural Development of the PRC, Beijing, 2018.

[14] I. Neo4j. "Neo4j Graph Database (Web Version)." https://console.neo4j.io/ (accessed January 20, 2026).

[15] OpenAI. "ChatGPT (Version 5.2 Thinking model)." https://chat.openai.com (accessed January 20, 2026).

[16] J. Wu, X. Xue, and J. Zhang, "INVARIANT SIGNATURE, LOGIC REASONING, AND SEMANTIC NATURAL LANGUAGE PROCESSING (NLP)-BASED AUTOMATED BUILDING CODE COMPLIANCE CHECKING (I-SNACC) FRAMEWORK," *Journal of information technology in construction,* vol. 28, pp. 1-18, 2023, doi: 10.36680/j.itcon.2023.001.

Color-Enhanced Local Feature Fusion for FPFH-Based Point Cloud Registration

Zezhi Ding[1], Jiaqi Li[1], Xincong Yang[1,*], Yangfan Wu[1], Ruinan Tan[1]

[1]School of Intelligent Civil and Ocean Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong, 518055, PR China

Abstract –

To address the limitations of the traditional Fast Point Feature Histogram (FPFH) in heterogeneous building point cloud registration—over-reliance on geometric information, neglect of color cues, and susceptibility to lighting variations—this paper proposes a color-enhanced local feature fusion framework. First, the RGB color space is transformed to HSV, with histogram equalization applied to the Hue channel to improve illumination robustness and color discriminability. Then, normalized FPFH geometric features and PCFH color features are fused into a 44-dimensional joint vector. Experimental validation on real shopping mall indoor point clouds shows that the proposed method performs well in complex scenarios: RRE and RMSE are reduced by up to 74.9% and 74.6% under different initial poses; robust registration is achieved in asymmetric point-count scenarios where FPFH fails; and RMSE remains 27.73% lower than FPFH even with only 500 sampling points. An outdoor campus road experiment further verifies generalizability, with RRE, RTE, and RMSE reduced by 66.2%, 63.4%, and 59.8%, respectively. Overall, integrating compact color features introduces negligible computational overhead while improving registration accuracy and robustness.

Keywords –

Point cloud, Registration, Feature Fusion

1 Introduction

The application of point cloud technology in civil engineering has attracted widespread attention in recent years, serving as a key driver for the industry's digital transformation. Traditional surveying and mapping methods in this field are constrained by high labor costs, low efficiency, limited handling of complex scenarios, poor adaptability to dynamic environments, and time-consuming post-processing. In contrast, point cloud technology enables non-contact, high-precision capture of spatial information for engineering structures through high-density 3D data acquisition. Its advantages in efficiency, digitalization, 3D visualization, and adaptability to complex scenarios offer innovative solutions for full lifecycle management of engineering projects [1-4].

Point cloud registration, which estimates the rigid transformation (i.e., 3D rotation and translation) between source and target point clouds, is a fundamental challenge in applications such as 3D building reconstruction [5-7]. Among traditional methods, many rely on geometric feature descriptors, with the Fast Point Feature Histogram (FPFH) [8] being a common choice. By statistically analyzing angular distributions between query points and neighbors, FPFH demonstrates robustness to mild noise. However, when applied to heterogeneous building point clouds, FPFH has inherent limitations: its exclusive reliance on geometry overlooks rich color cues in RGB data. Furthermore, local geometric neighborhoods can suffer significant distortions when scanned from different viewpoints, reducing feature similarity.

As a complementary data type, color information is expected to enhance local feature discriminability. However, raw RGB values are sensitive to lighting variations, leading to suboptimal results in direct applications. For example, walls scanned on sunny days appear brighter than those captured on cloudy days, causing color discrepancies at the same physical locations [9]. Naively combining color features with FPFH fails to achieve reliable fusion; instead, targeted color enhancement is necessary to improve lighting invariance and feature distinctiveness.

To address these challenges, this paper presents a novel FPFH-based registration framework integrating color-enhanced local feature fusion. The core principle is to synergize lighting-invariant color features with geometric descriptors, leveraging their complementary strengths to improve matching accuracy and robustness for heterogeneous data. The contributions are summarized as follows:

(1) A tailored color enhancement strategy for building

point clouds: By converting RGB to HSV color space and applying histogram equalization to the Hue channel, lighting interference is mitigated, and color discriminability is enhanced.

(2) Fusion of local features by integrating color-enhanced attributes with FPFH descriptors. Through normalization and dimension balancing, both types of features effectively facilitate matching.

(3) Experimental validation on real-world point cloud datasets demonstrates superior registration accuracy and robustness compared to traditional FPFH methods.

2 Related work

Point cloud registration methods are generally classified into coarse and fine registration. Coarse registration estimates an initial transformation matrix between point clouds with significant pose differences. RANSAC [10] is a classic method that estimates the transformation by randomly sampling point pairs and evaluating their consistency. To improve robustness, 4PCS [11] selects coplanar point sets to reduce the influence of outliers, but it suffers from high computational cost on large-scale point clouds. Recently, optimal transport-based methods such as Sinkhorn algorithms [12-14] have been introduced to enhance correspondence confidence and reduce false matches. However, these methods still rely on discriminative initial features, highlighting the importance of high-quality descriptors.

For fine registration, the Iterative Closest Point (ICP) algorithm [15] and its variants [16] are widely used. These methods iteratively minimize the distance between corresponding points to achieve precise alignment, but they are highly sensitive to the initial pose. Poor initial alignment may lead to convergence to a local optimum rather than the global solution. Therefore, reliable coarse registration remains a critical prerequisite for successful point cloud registration.

Geometric feature descriptors form the foundation of point cloud registration by encoding local structural information. Early descriptors such as Shape Context [17] and Spin Images [18] laid the groundwork for 3D feature extraction but often involve high computational complexity. As an efficient variant of PFH, FPFH reduces point-pair calculations while maintaining strong geometric discriminability, making it a benchmark in geometry-based registration. Nevertheless, FPFH is sensitive to viewpoint changes and point-density variations in heterogeneous scenarios. More robust descriptors have therefore been developed. For example, SHOT [19] improves viewpoint invariance through structured neighborhood sectors but has high dimensionality, while ISS [20] detects key points based on geometric saliency but still requires auxiliary descriptors for matching.

As a complement to geometric information, color features have been increasingly used to enhance the discriminative capability of point cloud descriptors. Early studies directly incorporated RGB values to improve nearest-neighbor correspondence or pose estimation [21-24], but raw color data are highly sensitive to illumination changes, resulting in unstable matching. To improve illumination invariance, researchers have explored color space transformations—such as converting from RGB to HSV or Lab color spaces—which separate chromatic information from brightness, thereby mitigating the effects of variable lighting [25,26]. Inspired by these approaches, this work fully leverages the color information embedded in point cloud data and adopts the HSV color space, following established practices. This enables effective registration without compromising the integrity of spatial structure.

To summarize, existing registration methods face challenges in heterogeneous building environments: limited discriminability of geometric features alone, vulnerability of raw color features to illumination variations, and suboptimal performance of naive feature fusion. To address these issues, this paper proposes a color-enhanced local feature fusion framework. Compared with ColorPCR, which relies on multi-stage color–geometry fusion, and SHOT-color, which extends high-dimensional descriptors, the proposed method constructs a lightweight Hue-based color histogram and efficiently fuses it with FPFH while retaining low computational cost.

3 Methodology

3.1 Problem Statement

Given two color point clouds $\mathcal{P} = \{p_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}$ with $\mathcal{P}_c = \{p_{ci} \in \mathbb{R}^3 \mid i = 1, \dots, N\}$ and $\mathcal{Q} = \{q_i \in \mathbb{R}^3 \mid i = 1, \dots, M\}$ with $\mathcal{Q}_c = \{q_{ci} \in \mathbb{R}^3 \mid i = 1, \dots, M\}$, in which, $\mathcal{P}_c$ and $\mathcal{Q}_c$ denote the color of $\mathcal{P}$ and $\mathcal{Q}$ respectively, our goal is to estimate a rigid transformation $\boldsymbol{T} = \{\boldsymbol{R}, \boldsymbol{t}\}$ which aligns the two point clouds, comprising a 3D rotation $\boldsymbol{R} \in SO(3)$ and a 3D translation $\boldsymbol{t} \in \mathbb{R}^3$. With the estimated transformation, we can align the two color point clouds. To achieve the most precise alignment, we need to solve the following optimization problem:

$$\min_{\boldsymbol{R}, \boldsymbol{t}} \sum_{(p_i^*, q_i^*) \in C^*} \|R \cdot p_i^* + t - q_i^*\|_2^2 \tag{1}$$

3.2 Overall Algorithm Design

We improve coarse registration based on FPFH by incorporating color information into the FPFH

descriptors. The resulting combined features demonstrate robust performance even in environments with limited geometric features or surface details, and show strong resilience to noise and variations in point cloud density.

The main procedure is as follows (the pipeline is illustrated in **Figure 1**.):

(1) Downsample the input source and target point clouds;

(2) Estimate surface normals for both point clouds using normal estimation methods;

(3) Compute FPFH features for each point in both point clouds;

(4) Convert the RGB values to the HSV color space and extract color descriptors using only the Hue (H) channel;

(5) Apply L2 normalization separately to the geometric (FPFH) and color descriptors, then concatenate them into a unified feature vector;

(6) Perform registration using the RANSAC algorithm.

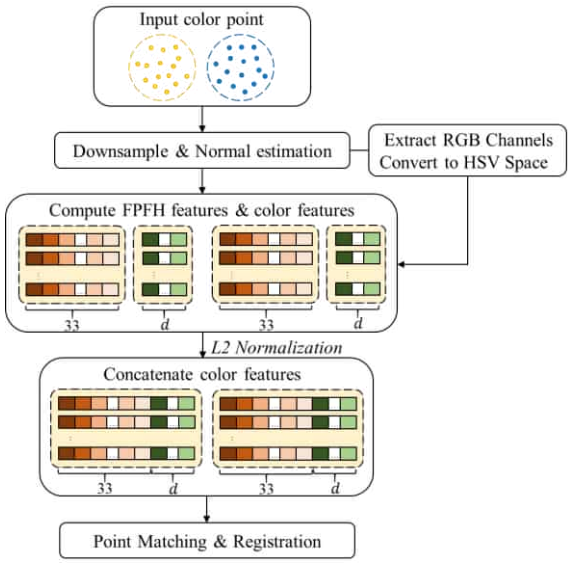
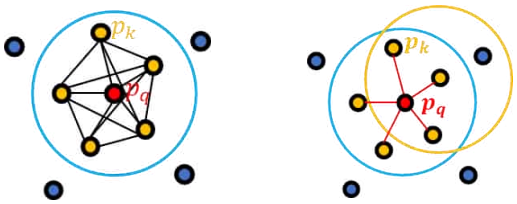


Figure 1. Pipeline of the proposed method

3.3 Computation of Fast Point Feature Histogram

The core idea of FPFH lies in first computing the Simplified Point Feature Histogram (SPFH) for each point in the k-neighborhood of the query point, and then weighting all the SPFH features via a formula to form the final FPFH feature. While FPFH reduces the computational complexity of the algorithm to O(nk), it still retains most of the discriminative properties of Point Feature Histograms (PFH). The computation process of FPFH is as follows:

(1) Only the three feature elements ($<\alpha, \phi, \theta>$) between each query point p_q and its neighborhood points are computed, which differs from PFH: while PFH calculates the feature elements for all combinations of neighborhood points, this step only computes those between the query point and its neighboring points. As shown in **Figure 2**, **Figure 2(a)** illustrates the feature computation process of PFH—i.e., the feature values for all combinations of neighborhood points (all lines in the figure, including but not limited to the lines between p_q and p_k, while **Figure 2(b)** depicts the computation scope of FPFH, which only requires calculating the feature elements between the query point p_q and its immediate neighboring points p_k (the red lines in **Figure 2(b)**). It can be seen that this design reduces the computational complexity, and this histogram is termed the Simplified Point Feature Histogram (SPFH).



(a) Spatial coverage of PFH (b) Spatial coverage of FPFH

Figure 2. The spatial coverage of FPFH and PFH

(2) A modified statistical method for k-neighborhood point pairs is proposed, comprising two components: point pairs between the query point and its k surrounding points; point pairs between each neighborhood point and its own k surrounding points. For the second component, the weight is usually set to the reciprocal of the distance, such that spatially closer neighborhood points make a greater contribution, as shown in **Eq. (2)**:

$$FPFH(p_q) = SPFH(p_q) + \frac{1}{k}\sum_{i=1}^{k}\frac{1}{w_k}\cdot SPFH(p_k) \quad (2)$$

3.4 Computation of Point Color Feature Histogram

We transform the RGB values of the point cloud data into the HSV color space and extract color features using only the Hue (H) channel. First, for each point, a d-dimensional Histogram of Hue Differences (HHD) is calculated between the query point p_q and its neighboring points p_k. Next, for each point again, the Point Color Feature Histogram (PCFH) is computed based on the HHD calculated in the first step, using **Eq. (3)-(4)**, where the weight is determined by the distance between the neighboring point and the query point.

$$HHD(p_q) = Histogram(|h_q - h_k|, d) \quad (3)$$

$$PCFH(p_q) = HHD(p_q) + \frac{1}{k}\sum_{i=1}^{k}\frac{1}{w_k}\cdot HHD(p_k) \quad (4)$$

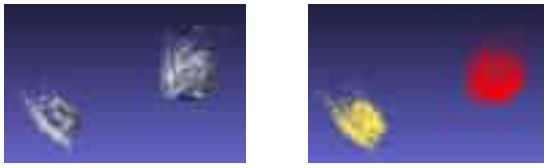
After obtaining the PCFH descriptor, we first apply L2 normalization to it, and then concatenate it with the FPFH descriptor to form a 44-dimensional feature vector (where the dimension of PCFH is fixed at 11 in subsequent experimental designs). The purpose of normalizing before concatenation is to facilitate better collaboration between the geometric and color descriptors.

4 Experiments and Results

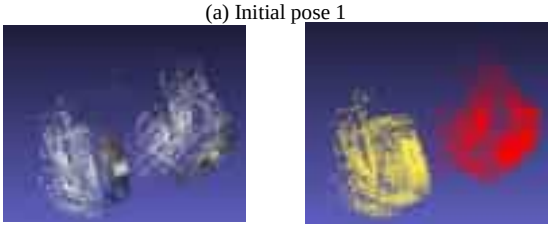
4.1 Experiment Design

The point clouds used in this study were collected from a large shopping mall using the Leica BLK360 imaging scanner. The source and target point clouds were obtained from two scans at the same location. Since the scanner does not provide initial pose information, the ground truth transformation matrix in Eq. (5)–(7) was generated by applying an artificial transformation to the target point cloud, ensuring an exact reference. Five experiments were designed: (1) registering two 10,000-point clouds under three initial poses; (2) registering asymmetric point counts of 12,000 and 4,000 points; (3) testing reduced densities of 4,000, 2,000, 1,000, and 500 points; (4) comparing the computation time of FPFH and FPFH + PCFH; and (5) testing outdoor generalization on a campus road segment with two 10,000-point clouds.

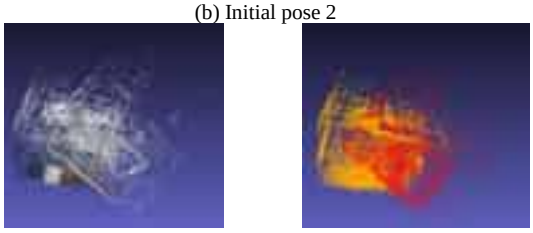
Experiment 1: The three initial poses are detailed in Table 1. This experiment evaluated the registration accuracy of the Color-enhanced FPFH algorithm under diverse initial pose conditions. Figure 3 illustrates the downsampled point clouds and their initial relative poses, with the source cloud colored red and the target orange. The RANSAC algorithm was set to 600,000 iterations, with an inlier threshold of 0.2 m, and the results were compared with FPFH in Section 4.2.



(a) Initial pose 1



(b) Initial pose 2



(c) Initial pose 3

Figure 3. Three different initial poses in Experiment 1

Table 1 Three different initial poses in Experiment 1

	Initial pose 1	Initial pose 2	Initial pose 3
Roll/°	55.08	60.63	33.68
Pitch/°	-9.61	-84.00	41.03
Yaw/°	-55.52	97.76	17.52
X-axis translation distance/m	-20.097	8.376	-3.523
Y-axis translation distance/m	20.474	-7.014	-0.243
Z-axis translation distance/m	24.060	8.380	1.931

Experiment 2: The source and target point clouds were downsampled to 12,000 and 4,000 points, respectively, to amplify point-count discrepancy and reduce the proportion of feature-similar points. The initial pose is shown in Figure 4 and Table 2. The RANSAC settings were the same as in Experiment 1, and the results were compared with FPFH in Section 4.2.



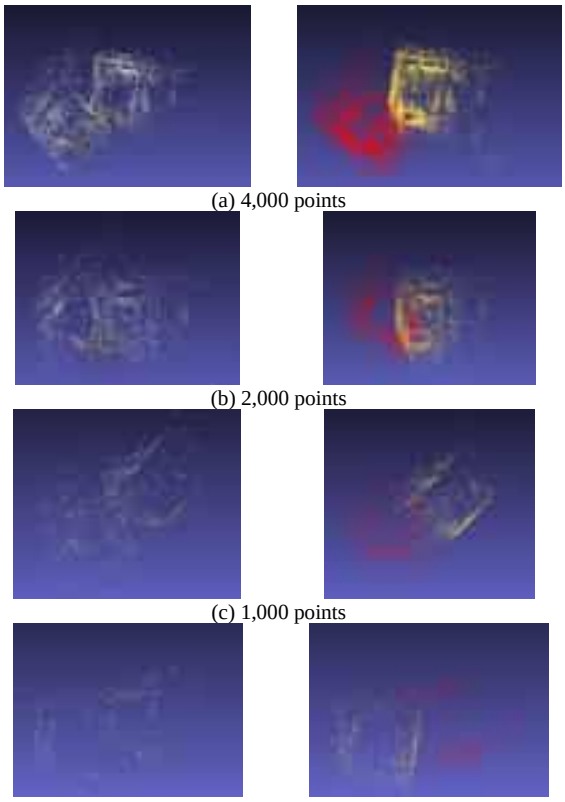
Figure 4. The initial pose in Experiment 2

Table 2 The initial pose in Experiment 2

	Initial pose
Roll/°	-2.38
Pitch/°	-68.74
Yaw/°	30.71
X-axis translation distance/m	3.362
Y-axis translation distance/m	3.933
Z-axis translation distance/m	-3.000

Experiment 3: Four downsampling point counts were tested: 4,000, 2,000, 1,000, and 500. The objective was to evaluate the accuracy and robustness of the Color-enhanced FPFH algorithm under continuously decreasing sampling densities. The point clouds and initial poses are shown in Figure 5 and

Table 3.



(a) 4,000 points

(b) 2,000 points

(c) 1,000 points

(d) 500 points

Figure 5. The four downsampling settings in Experiment 3

Table 3 The initial poses in Experiment 3

	Initial pose (4k)	Initial pose (2k)	Initial pose (1k)	Initial pose (0.5k)
Roll/°	55.97	-39.97	-11.23	-28.28
Pitch/°	14.39	35.47	-58.75	-61.12
Yaw/°	-34.77	-18.12	-22.35	-19.60
X-axis translation distance/m	4.479	0.666	4.966	-3.183
Y-axis translation distance/m	2.380	0.087	4.855	4.865
Z-axis translation distance/m	1.579	-2.754	-4.406	1.291

Experiment 4: To quantify the computation time of the fused descriptor versus the FPFH, we designed a comparative experiment. To ensure a fair comparison, both methods utilized the same parameters.

Experiment 5: The initial pose is detailed in Table 4. This experiment evaluated outdoor generalization using a campus road segment point cloud. Figure 6 shows the source and target point clouds and their initial relative poses, with the source colored red and the target orange.

Table 4 The initial pose in Experiment 5

	Initial pose
Roll/°	52.47
Pitch/°	23.51
Yaw/°	−56.22
X-axis translation distance/m	−97.50
Y-axis translation distance/m	81.86
Z-axis translation distance/m	81.90



Figure 6. The initial pose in Experiment 5

In this paper, normal estimation employs a hybrid neighborhood search strategy, with a search radius of 2 meters and a maximum of 30 nearest neighbors. Similarly, the FPFH feature extraction algorithm utilizes a hybrid search strategy, setting the FPFH search radius to 5 meters and the maximum number of neighbors to 100.

The custom PCFH descriptor (dimension = 11) is constructed based on the Hue (H) channel. Its neighborhood search adopts a spherical neighborhood with a radius of 2.0 m (skipping points with fewer than 3 neighbors). The Hue difference calculation accounts for the 360° cyclic nature of the H channel by taking the minimum angular difference (0–180°) and normalizing it to the [0,1] range. Histogram binning uses 11 equal-width bins covering the full [0,1] range.

During feature matching, the nearest neighbor in the target point cloud is first identified for each source point in the concatenated 44-dimensional feature space. Mutual nearest neighbor filtering retains bidirectional correspondences, and incorrect matches are further filtered using edge length verification (threshold 0.9) and a correspondence distance threshold of 0.2 m. The valid correspondences are then used for RANSAC, configured with 600,000 iterations, confidence of 0.999, three sample points, and the same 0.2 m distance threshold.

4.2 Discussion

To further quantify the registration results, the Relative Rotation Error (RRE), Relative Translation Error (RTE), and Root Mean Square Error (RMSE) for different algorithms were calculated. The formulas for RRE, RTE and RMSE are as follows:

$$RRE = \arccos\left(\frac{tr(R_{gt}^{T}R_{est}) - 1}{2}\right) \times \frac{180}{\pi} \quad (5)$$

$$RTE = \left\| t_{gt} - t_{est} \right\|_2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum \left\| (T_{gt} \cdot P) - (T_{est} \cdot P) \right\|_2^2} \qquad (7)$$

where R_{gt} and t_{gt} denote the ground truth rotation and translation, respectively; R_{est} and t_{est} denote the estimated rotation and translation; T_{gt} denote the ground truth transformation matrix; T_{est} denotes the estimated transformation matrix, and P denotes the source point cloud.

Under controlled conditions where all other experimental parameters were fixed, Tables 5-9 summarize the registration results. As shown in Table 5, under different initial poses, the Color-enhanced FPFH-based algorithm consistently outperformed standard FPFH. The most significant improvement occurred under Initial Pose 2, where RRE, RTE, and RMSE improved by 74.9%, 67.4%, and 74.6%, respectively. Although the RTE under Initial Pose 1 slightly increased from 0.0638 m to 0.104 m, this value remains within the 0.2 m RANSAC correspondence threshold and does not compromise overall stability. In Experiment 2 (Table 6), standard FPFH essentially failed under significantly different point counts, while the proposed method achieved low errors of 2.100° (RRE), 0.048 m (RTE), and 0.2 m (RMSE), indicating better matching robustness under asymmetric density. In Experiment 3 (Table 7), the proposed method maintained high accuracy at 4,000, 2,000, and 1,000 points, and even with only 500 points still improved RRE, RTE, and RMSE by 26.2%, 30.82%, and 27.73%, respectively. Table 8 shows that adding the 11-dimensional color feature causes almost no increase in computation time. Table 9 further demonstrates outdoor generalization, with RRE, RTE, and RMSE reduced by 66.2%, 63.4%, and 59.8%, respectively.

Table 5 Comparison of registration results (Experiment 1)

	FPFH-based algorithm			Color-enhanced FPFH-based algorithm		
Initial pose	1	2	3	1	2	3
RRE/°	2.904	4.084	1.313	0.824	1.025	0.635
RTE/m	0.0638	0.279	0.099	0.104	0.091	0.070
RMSE/m	0.234	0.379	0.169	0.127	0.096	0.108

Table 6 Comparison of registration results (Experiment 2)

	FPFH-based algorithm	Color-enhanced FPFH-based algorithm
RRE/°	172.510	2.100
RTE/m	2.947	0.048
RMSE/m	8.550	0.200

Table 7 Comparison of registration results (Experiment 3)

	FPFH-based algorithm			
Point Count	4k	2k	1k	0.5k
RRE/°	1.440	39.795	10.109	29.795
RTE/m	0.066	1.063	0.201	2.164
RMSE/m	0.142	3.012	0.818	2.517
	Color-enhanced FPFH-based algorithm			
Point Count	4k	2k	1k	0.5k
RRE/°	1.633	2.108	2.987	21.990
RTE/m	0.065	0.045	0.261	1.497
RMSE/m	0.147	0.185	0.329	1.819

Table 8 The processing times in Experiment 4

Point Count	4k	2k	1k	0.5k
FPFH	0.0034s	0.0025s	0.0010s	0.0006s
FPFH + PCFH	0.00036s	0.0027s	0.0011s	0.0009s

Table 9 Comparison of registration results (Experiment 5)

	FPFH-based algorithm	Color-enhanced FPFH-based algorithm
RRE/°	0.281	0.095
RTE/m	0.254	0.093
RMSE/m	0.241	0.097

5 Conclusions

This paper proposes a color-enhanced local feature fusion framework to address the limitations of FPFH-based registration in heterogeneous building point clouds, including over-reliance on geometry, neglect of color cues, and sensitivity to lighting interference. The method transforms RGB to HSV, applies histogram equalization to the Hue channel, and fuses normalized FPFH and PCFH into a 44-dimensional joint descriptor. Experiments on real shopping mall point clouds demonstrate strong performance under different initial poses, asymmetric point counts, and sparse sampling: RRE and RMSE are reduced by up to 74.9% and 74.6%; robust registration is achieved when FPFH fails under asymmetric point counts; and at 500 points, RMSE remains 27.73% lower than FPFH. The computation-time comparison shows negligible additional overhead, and the outdoor experiment confirms promising generalization ability.

Despite the promising results, this study has limitations. The impact of PCFH dimensionality on registration performance has not been fully explored; the value of 11 was chosen following the common bin setting of PFH and FPFH. Future work will optimize the color feature dimensionality and weighting strategy through sensitivity analysis or adaptive learning. Moreover, this study only uses the Hue channel, while the influence of Saturation and Value channels should be investigated in more complex scenarios.

6 Acknowledgement

This research is supported by Key Technologies R&D Program (Grant No. 2024YFC3808602), Shenzhen Science and Technology Programs (Grant No. KJZD20230923114310021, Grant No. JCYJ20241202123722029).

References

[1] H. Yue, Q. Wang, H. Zhao, N. Zeng, and Y. Tan. Deep learning applications for point clouds in the construction industry. Automation in Construction, 168:105769, 2024.

[2] Q. Wang and M. K. Kim. Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. Advanced engineering informatics, 39:306–319, 2019.

[3] L. J. Sanchez-Aparicio, F. L. del Blanco-Garcia, D. Mencias-Carrizosa, P. Villanueva-Llaurado, J. R. Aira-Zunzunegui, D. Sanz-Arauz, and J. Garcia-Gago. Detection of damage in heritage constructions based on 3D point clouds. A systematic review. Journal of Building Engineering, 77:107440, 2023.

[4] P. Stałowska, C. Suchocki, and M. Rutkowska. Crack detection in building walls based on geometric and radiometric point cloud information. Automation in Construction, 134:104065, 2022.

[5] M. R. Sanzana, T. Maul, J. Y. Wong, M. O. M. Abdulrazic, and C. C. Yip. Application of deep learning in facility management and maintenance for heating, ventilation, and air conditioning. Automation in Construction, 141:104445, 2022.

[6] C. Zhang and D. Arditi. Advanced progress control of infrastructure construction projects using terrestrial laser scanning technology. Infrastructures, 5(10):83, 2020.

[7] R. Maalek, D. D. Lichti, and J. Y. Ruwanpura. Automatic recognition of common structural elements from point clouds for automated progress monitoring and dimensional quality control in reinforced concrete construction. Remote Sensing, 11(9):1102, 2019.

[8] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In 2009 IEEE International Conference on Robotics and Automation, pages 3212–3217. IEEE, 2009.

[9] J. Mu, L. Bie, S. Du, and Y. Gao. ColorPCR: Color point cloud registration with multi-stage geometric-color fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21061–21070, 2024.

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

[11] D. Aiger, N. J. Mitra, and D. Cohen-Or. 4-points congruent sets for robust pairwise surface registration. In ACM SIGGRAPH 2008 papers, pages 1–10, 2008.

[12] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. The Annals of Mathematical Statistics, 35(2):876–879, 1964.

[13] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. Pacific Journal of Mathematics, 21(2):343–348, 1967.

[14] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. Advances in Neural Information Processing Systems, 26, 2013.

[15] P. J. Besl and N. D. McKay. Method for registration of 3-D shapes. In Sensor Fusion IV: Control Paradigms and Data Structures, volume 1611, pages 586–606. SPIE, 1992.

[16] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling, pages 145–152. IEEE, 2001.

[17] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4):509–522, 2002.

[18] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5):433–449, 2002.

[19] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In European Conference on Computer Vision, pages 356–369. Springer Berlin Heidelberg, 2010.

[20] Y. Zhong. Intrinsic shape signatures: A shape descriptor for 3D object recognition. In 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pages 689–696. IEEE, 2009.

[21] J. Park, Q. Y. Zhou, and V. Koltun. Colored point cloud registration revisited. In Proceedings of the IEEE International Conference on Computer Vision, pages 143–152, 2017.

[22] M. Korn, M. Holzkothen, and J. Pauli. Color supported generalized-ICP. In 2014 International Conference on Computer Vision Theory and Applications, VISAPP, volume 3, pages 592–599. IEEE, 2014.

[23] J. Servos and S. L. Waslander. Multi-Channel Generalized-ICP: A robust framework for multi-

channel scan registration. Robotics and Autonomous Systems, 87:247–257, 2017.
[24] J. H. Joung, K. H. An, J. W. Kang, M. J. Chung, and W. Yu. 3D environment reconstruction using modified color ICP algorithm by fusion of a camera and a 3D laser range finder. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3082–3088. IEEE, 2009.
[25] J. Mu, L. Bie, S. Du, and Y. Gao. ColorPCR: Color point cloud registration with multi-stage geometric-color fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21061–21070, 2024.
[26] H. Men, B. Gebre, and K. Pochiraju. Color point cloud registration with 4D ICP algorithm. In 2011 IEEE International Conference on Robotics and Automation, pages 1511–1516. IEEE, 2011.

Domain-specific Large Language Model Pretraining for Construction Specification Review

Shuyi Wang[1] and Jinwoo Kim[2]

[1]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
[2]Department of Civil and Environmental Engineering, Hanyang University, Seoul, Republic of Korea
shuyi003@e.ntu.edu.sg, jinwookim@hanyang.ac.kr

Abstract –
With advancements in large language models (LLMs), there has been growing attention to automatic construction specification reviews, enabling faster and more standardized assessments. However, general-purpose LLMs' effectiveness has been bounded when applied to construction contexts, because they lack necessary domain-specific vocabulary and knowledge. Therefore, we propose a domain-specialization LLM pretraining method to leverage construction terms and knowledge and quantitatively evaluate its efficacy in two specification review tasks. To this end, we generate two construction-specific corpora, i.e., close-domain and in-domain, and investigate their hybrid effects on LLM performance at varying scales. Results reveal that construction-specific contextualization led to a significant improvement, with a 6.9% increase in accuracy for extractive question answering and an 8.0% enhancement in information retrieval accuracy. We also observe that augmenting in-domain corpus with close-domain data, can achieve performance comparable to models trained with extremely large in-domain data. These will facilitate the adoption of LLMs in construction and support specification reviews.

Keywords –
Large language model; Domain-specific; Pretraining; Construction specification

1 Introduction

Construction specification review is a critical process in the construction industry, focused on ensuring that project specifications align with required standards, project goals, and regulatory compliance [1]. Specifications are detailed documents that outline project requirements, including material quality, design standards, construction methods, and safety protocols. However, manual review of such extensive documents is time-consuming and resource-intensive, leading researchers to explore automated methods using natural language processing (NLP) technologies [2,3]. Previous studies have primarily focused on two types of NLP tasks in specification reviews: (i) Contextual understanding of provisions: Its application involves interpreting and classifying regulatory provisions based on their semantic contents, such as identifying risky contractual terms [4] or categorizing provisions into classes such as technical requirements, financial regulations, and material standards [5]. This is also foundational for tasks like extractive question answering (QA) from provisions, which requires precise extraction of textual span according to user intent; and (ii) Information retrieval for specific details: Its application focuses on recognizing and extracting critical entities including subjects, requirements, actions, and standards through rule-based NLP algorithms and deep learning techniques [2,6].

Recently, large language models (LLMs), such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT), have become prominent in NLP applications due to their ability to capture rich contextual information [7]. For instance, Ahmadi et al. [8] leveraged existing LLMs like Gemini and ChatGPT to classify construction accident reports into categories including injury cause, root cause, affected body part, severity, and accident type, demonstrating the promising potential of general-purpose LLMs. Although LLMs perform well in everyday language tasks, adapting them to a specific domain, particularly construction documents with many technical terminologies, presents significant challenges. In particular, LLMs pretrained on generic corpora often struggle with domain-specific vocabulary, contextual nuances, and technical language characteristic [9]. Consequently, these models lack the precise understanding required for accurately interpreting construction specifications, demonstrating the need for target domain adaptation.

Some researchers have recognized this challenge and addressed it by employing domain-adapted LLMs, such as BERT variants pretrained on domain-contextualized corpora. In this regard, two types of corpora have

typically been used: (i) in-domain corpus, which is narrowly focused on a specific target domain (i.e., a particular type of documents) and incorporate specialized language and contextual terminology; and (ii) close-domain corpus, which encompasses a wide spectrum of construction documents, capturing diverse linguistic patterns and terminology relevant to the field. For instance, Zheng et al. utilized both close-domain and in-domain Chinese corpora and validated their individual benefits as well as the combined effects of mixing two corpora on NLP-based construction document analysis. Zhong and Goodfellow [11] also introduced pretrained language models specifically designed for construction management systems, by leveraging in-domain corpora with and without sentence concentration. While these studies indicate that domain-specific pretraining enhances LLM performance in NLP tasks, several knowledge gaps remain to be addressed. First, there is limited quantitative analysis on the effects of domain-specific knowledge on LLM performance in construction specification reviews. Although untrainable models like ChatGPT demonstrate strong capabilities, a comprehensive evaluation of domain-specific adaptation on small-scale trainable models can inform their effective utilizations in construction. This trial will also help address the scarcity of labeled domain-specific data that limits supervised methods, as pretraining on unlabeled domain data offers a promising approach to reduce dependence on extensive labeled datasets [12]. Second, although previous studies theoretically identify the advantages of incorporating domain knowledge including technical vocabulary and language patterns into LLMs, few studies have investigated how varying quantities of close-domain and in-domain corpora impact model's performance. Specifically, it remains unclear how much data can motivate the model's capability in contextual understanding. Moreover, emphasis has largely been placed on in-domain corpus due to its direct relevance to target documents. However, when the available in-domain corpus is limited, its impact on model adaptation becomes marginal, making it necessary to consider supplementary close-domain data. However, the hybrid effects of combining close-domain and in-domain corpora are rarely examined, leaving unanswered questions about the optimal quantity of these corpora for efficient domain adaptation in LLM pretraining.

Therefore, we propose a LLM pretraining method to combine close-domain and in-domain corpora and specialize LLMs to the construction field. We then fine-tune and validate LLMs in two construction specification review tasks, extractive QA and information retrieval. Particularly, we evaluate the effectiveness of domain corpora on LLM performance and investigate their impacts at varying sizes. Our findings will facilitate the domain adaptation of LLMs and offer insights on optimizing training efficacy with varying domain corpora sizes.

2 Research Method

This study followed a three-stage method comprising corpus generation, LLM pretraining, and task-specific finetuning (Figure 1). First, we collected various types of publicly accessible construction documents and generated close-domain and in-domain corpora. Subsequently, we utilized masked language modeling (MLM) for LLM pretraining. Lastly, the pretrained LLMs were fine-tuned for two construction specification review tasks: extractive QA for provisions and information retrieval from provisions. More details are provided in the following subsections.

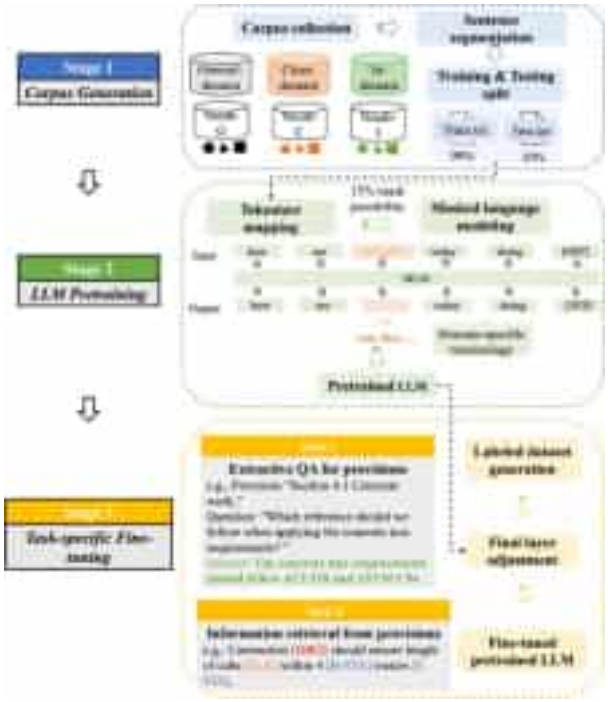


Figure 1. Overall pipeline for pretraining a construction-specific large language model

2.1 Stage 1: Corpus Generation

In Stage 1, we generated comprehensive corpora encompassing various types of documents in the construction domain, categorized into two distinct types: (i) Close-domain documents, which include a range of construction-related documents such as contracts, regulatory laws, and management guidelines; and (ii) In-domain documents, which are specific to a single document type that closely aligns with a particularly type of document (i.e., construction specifications in this study). To construct these corpora, we initially collected construction documents publicized in two different

formats. The first format consisted of Portable Document Format files, which were converted into machine-readable texts using Optical Character Recognition tools. The second format included web-based documents, for which we employed web scraping techniques to extract and consolidate textual information. Ultimately, we gathered 752 close-domain documents comprising contracts, law regulations, and management guidelines. Additionally, we obtained 56 in-domain construction specification documents from four countries, United States, United Kingdom, Canada, and Australia, covering a total of 56 distinct states and regions. Following conversions of texts for these documents, we adopted data cleaning techniques including removing extraneous characters and irrelevant formatting. We then applied text preprocessing techniques to both corpora, including sentence segmentation and training-testing splitting. Sentence segmentation enables the model to process structured linguistic units, which is critical for enhancing model understanding for complex contents. Detailed statistical information of two corpora is illustrated in Table 1. Finally, we allocated 90% of each corpus for training and 10% for testing [13]. This split ensures that the model is evaluated on a representative portion of the data while maintaining sufficient training data to capture domain-specific patterns and vocabulary.

Table 1. Statistical information of domain corpora

	Numbers of sentences	Numbers of tokens
Close-domain	1,028,086	23.6 million
In-domain	1,097,925	22.4 million

2.2 Stage 2: LLM Pretraining

After generating close-domain and in-domain corpora, we proceeded to the pretraining phase. Each sentence was initially tokenized to its numeric values that can be processed by computer programming language [10]. We then utilized MLM by randomly masking a certain percentage of input tokens within sentences, requiring the model to predict these obscured tokens. In this process, the final hidden vectors corresponding to the masked tokens were passed through an output softmax layer over the vocabulary. According to BERT implementation [13], 15% of all WordPiece tokens were randomly selected and replaced with the [MASK] token in each sequence, prompting the model to predict these tokens by learning contextual representations from neighboring words. This random selection process was managed by our data collator, which ensures 15% of token positions are designated for prediction. The cross-entropy loss function was finally computed by the average prediction error over all masked tokens in a sequence L following Eq.1 [14]:

$$Loss = -\frac{1}{|M|}\sum_{i \in M} logP(x_i|h_i^L) \quad (1).$$

where $|M|$ represents the total number of masked tokens. For each masked position i, x_i represents the original token and h_i^L is the corresponding output representation from the final layer of the model.

Notably, MLM serves as a self-supervised pretraining process, eliminating demands for additional manual labeling. Upon completion of pretraining on construction-specific corpora, this pretrained LLM acquires contextual domain knowledge, enhancing its performance in following downstream NLP tasks.

2.3 Stage 3: Task-specific Fine-tuning

Building on the pretrained LLM developed in Stage 2, we fine-tuned the final output layer by applying task-appropriate loss functions tailored for construction specification review tasks. We particularly focused on two primary tasks: extractive QA from provisions and information retrieval. With respect to the first task, we concentrated on the model's ability to identify and extract the precise span of text within a provision that answers a given question. To achieve this, we adapted the final output layer of the pretrained encoder by adding two classification heads $[Answer_s, Answer_e]$ to predict the start and end positions of the answer span within the input token sequence. During pretraining, the input consisted of a concatenated $[Provision, Question]$ pair, and the model was supervised using a span-based objective function—specifically, the cross-entropy losses for the start and end positions. This formulation enabled the model to learn fine-grained semantic alignment between the question and the contextual tokens of the provision, facilitating accurate answer localization. Meanwhile, information retrieval task refers to extract core elements from extensive documents based on their important semantic contents, which is a classical classification problem [6]. For this purpose, we appended a new classification layer $[CLS]$ to assign class labels (e.g., organization, action, element, standard, and reference) to each token based on a labeled information retrieval dataset, enabling the model to accurately categorize information of construction specifications.

3 Experiments and Results

As illustrated in Figure 2, we conducted two experiments to investigate the effects of domain-specific corpora on LLM performances (Experiment #1); and to analyze the hybrid effects of close-domain and in-domain corpora at varying sizes (Experiment #2). In both experiments, we adopted BERT as our foundational LLM due to its well-established capability in contextual understanding, which is particularly suitable for

specification review. Furthermore, it supports computationally efficient pretraining compared with larger-scale open-source LLMs, assisting practical experimentation [9]. It was then fine-tuned for two downstream tasks, extractive QA for provisions and retrieving information from provisions.



Figure 2. Experimental setups

Prior to task-specific fine-tuning, we conducted domain-adaptive pretraining to better align the LM with the semantic and syntactic characteristics of construction documents. The pretraining was carried out with the following parameters informed by [13]: a batch size of 32, a learning rate of 5×10^{-5}, a maximum sequence length of 512, and a total of 5,000 training steps. All pretraining was performed on NVIDIA A100 GPU. The evolution of both training and testing losses is demonstrated in Figure 3. It was observed that both losses converged and stabilized at around 5,000 training steps, indicating that the model reached convergence. To further validate the success and robustness of pretraining, we evaluated the model on a masked word prediction task using sentences from construction specifications. In this setting, 15% of tokens were randomly masked, and the model was tasked with recovering the original tokens. The masked token prediction accuracy reached 93.2%, demonstrating the model's strong capacity to learn and generalize over the domain terminology.

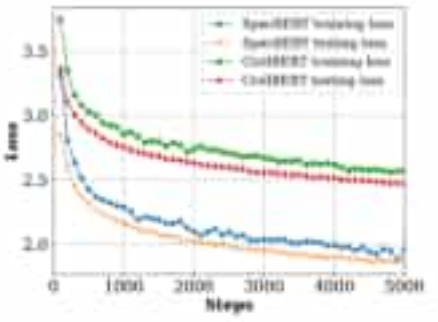


Figure 3. Losses for domain-specific pretraining

With respect to fine-tuning for experiment, we initially built fine-tuning task-specific training and test datasets as follows. To prepare the extractive QA dataset, we initially employed GPT-4o with prompt engineering to produce initial QA pairs from 201 construction provisions, as introduced in [15]. The generation process simulated real-world domain experts engaged in reviewing construction specifications, covering five question types: responsibility identification, procedure specification, numerical constraint, material standard, and regulatory references [6]. These automatically generated pairs then underwent a manual structural and semantic verification. Structurally, we ensured each pair complied with the Stanford Question Answering Dataset format—answers had to be explicitly found in the provision, with correct span indices [16]. Semantically, we first removed redundant questions that queried the same information. Second, inaccurate or contextually irrelevant questions were corrected or excluded. Third, to increase reasoning complexity, a subset of questions was reformulated. For instance, a basic question like "What curing compound is required for the concrete pouring?" was rewritten as "Which curing compound specified for concrete pouring should also satisfy water retention standard?". After this manual refinement, a total of 855 high-quality extractive QA pairs were obtained and subsequently reviewed by domain experts to ensure its quality. This dataset was divided into a training dataset (684 pairs) and testing dataset (171 pairs) with an 8:2 ratio. Using the dataset, we trained and tested models with two quantitative metrics: F1-score and Sentence-BERT (S-BERT). Specifically, F1-score provides a strict measure of token-level correspondence by balancing precision and recall, thus assessing how closely the predicted answer matches the ground truth. However, such span-based metrics penalize semantically correct answers but differ in wording. To this end, S-BERT is employed to evaluate semantic overlap by computing the cosine similarity between sentence-level embeddings of the predicted and reference answers. Collectively, these two metrics reflect both the surface-level accuracy and semantic relevance.

For the other fine-tuning task, retrieving information from provisions, a total of 1,250 provisions were extracted from the identical specifications used in the former task. These provisions were then categorized into five classes according to [6] to extract specification review information: Organization, Action, Element, Standard, and Reference. Each token was manually labeled as one entity using the open-source annotation tool Label Studio, with contiguous tokens assigned consistent labels and explicit boundaries defined to prevent overlap between entity types. In total, 15,969 tokens were annotated with corresponding semantic entity labels. This dataset was then divided into a training

dataset (1,000 provisions) and testing dataset (250 provisions) with an 8:2 ratio. To assess the model's performance, two evaluation metrics, F1-score and accuracy, were utilized. Specifically, F1-score integrates both precision and recall and requires a predicted entity match the gold standard in both its semantic label and boundary to be counted as correct. In contrast, accuracy measures performance at the token level by calculating the proportion of tokens assigned the correct label, including non-entity ("O") tokens. More detailed settings and results of each experiment are provided in the following subsections.

3.1 Experiment #1: Effects of Domain-specific Corpora on LLM

In Experiment #1, we compared the performance of four types of LLMs in construction specification review tasks: (i) an untrainable LLM like GPT-4o; (ii) BERT, which is a general-purpose LLM pretrained on a broad corpus; (iii) CivilBERT, a model pretrained on a close-domain corpus of 20 million tokens; and (iv) SpecBERT, a model pretrained on an in-domain corpus of 20 million tokens. Baseline models were determined as untrainable and pretrained LLMs from general corpus (i and ii).

3.1.1 Downstream task #1 results: Extractive QA for provisions

As illustrated in Table 2, general-purpose LLMs, such as GPT-4o, exhibited relatively lower performance. Specifically, GPT-4o achieved an F1-score of 74.0% and an S-BERT of 79.1% on the extractive QA. The BERT model, pretrained on a broad general corpus and fine-tuned on the extractive QA dataset, showed a modest improvement of 6.4% in F1-score and 4.0% in S-BERT, reaching final scores of 78.8% and 82.3% respectively. It was noteworthy that LLMs pretrained with domain-specific data, however, displayed significantly enhanced performance. CivilBERT attained an F1-score of 80.1% and an S-BERT of 86.4%, surpassing BERT by 1.4% in F1-score and 4.1% in S-BERT. Moreover, SpecBERT delivered the highest performance, with an F1-score of 82.4% and S-BERT of 89.2%, representing increases of 3.7% and 6.9% over BERT in two metrics. Overall, these results indicated that with domain-specific pretraining, smaller-scale LMs could surpass out-of-the-box commercial LLMs on specialized extractive QA tasks.

Table 2. Comparisons of LLM performances on extractive QA dataset

	F1-score	S-BERT
GPT-4o	74.0%	79.1%
BERT	78.7%	82.3%
CivilBERT	80.1%	86.4%
SpecBERT	**82.4%**	**89.2%**

Figure 4 presents a representative example of qualitative behaviors of baselines and domain-specific LLMs. Specifically, GPT-4o tended to prioritize completeness in answering the question by extracting all information related to the question and adding additional descriptions inferred from the contents. While this resulted in comprehensive responses, it often introduced unnecessary elaboration beyond the exact answer span. In contrast, BERT fine-tuned on the dataset adhered strictly to the question and focus on retrieving spans that appear directly relevant. However, due to the lack of pre-existing semantic understanding of the entire provision, BERT often extracted multiple loosely related segments. CivilBERT and SpecBERT, on the other hand, demonstrated stronger contextual comprehension. Both models exhibited a better grasp of how to localize and extract answer spans based on the semantic alignment between the question and the provision. This reflected their enhanced capacity for domain-aware understanding and more accurate extractive reasoning.



Figure 4. Qualitative behaviors of baselines and domain-specific LLMs for extractive QA pair

3.1.2 Downstream task #2 results: Information retrieval from provisions

Table 3 presented comparative results across three corpus types for the information retrieval dataset. General-purpose LLMs demonstrated relatively lower performance, with GPT-4o achieved F1-score at 53.9% and accuracy at 54.6%. BERT displayed substantial

improvement, with an F1-score of 71.7% and accuracy of 84.1%. Domain-specific pretraining further enhanced performance. CivilBERT outperformed BERT with an F1-score of 76.3% and accuracy of 87.7%, representing increases of 6.4% and 4.3% in two metrics, respectively. The best performance was observed in SpecBERT, which achieved an F1-score of 77.4% and accuracy of 88.9%, surpassing BERT by 8.0% in F1-score and 5.7% in accuracy compared to BERT.

Table 3. Comparisons of LLM performances on information retrieval dataset

	F1-score	Accuracy
GPT-4o	53.9%	54.6%
BERT	71.7%	84.1%
CivilBERT	76.3%	87.7%
SpecBERT	**77.4%**	**88.9%**

As illustrated in Figure 5, GPT-4o demonstrated limited capability to accurately classify domain-specific terms due to the lack of prior construction-specific knowledge in its general-purpose pretraining. To be specific, it failed to recognize specialized terminology such as "geotechnical fill materials," instead generalizing to broader terms like "materials." Additionally, its interpretation of standards and references tended to be ambiguous, frequently resulting in overlapping or misclassified outputs. In contrast, the fine-tuned BERT model showed improved performance, demonstrating the capability to distinguish between standards and references, while also preserving the integrity of domain-specific terminology. CivilBERT further enhanced this performance, though it struggled with certain terms, for instance, simplifying "resident engineer" to "engineer," and missed critical references, especially those pertaining to project-specific or organization-defined specifications beyond national standards. Notably, SpecBERT achieved the best overall performance. Benefiting from domain-specific pretraining, it was capable of extracting a more comprehensive set of relevant terms and correctly categorizing them semantically.

Overall, these results highlighted that domain-specific pretraining, particularly with in-domain corpora (e.g., SpecBERT), substantially enhanced performance of LLMs in both contextual comprehension and information retrieval. By integrating construction-specific technical terminologies, the model's ability to capture contextual contents and extract semantic elements was markedly strengthened, addressing unique linguistic complexities of the construction domain. This domain-aware capability proves especially valuable across a range of practical construction management applications [8,11]. Specifically, it facilitates information retrieval for decision support, allowing site engineers and project managers to quickly locate technical requirements—such as materials, construction methods, and tolerances—during on-site execution. Through aligning the model's linguistic capacity with the operational realities of construction workflows, domain-specific pretraining contributes to a more intelligent, efficient, and trustworthy document processing pipeline**,** supporting both high-level regulatory reviews and real-time field-level implementation.



Figure 5. Qualitative behaviors of baselines and domain-specific LLMs for information retrieval

3.2 Experiment #2: Hybrid Effects of Close-domain and In-domain Corpora on LLM

Experiment #2 examined the effects of varying corpus quantities by evaluating different proportions $x\%$ ($x \in \{10, 20, 30, 40, 50, 100\}$) of corpus data from single domains (close-domain or in-domain) and mixed domains in finetuning tasks. A maximum of 20 million tokens from each corpus was defined as the full-scale setting. Accordingly, 2 million tokens of in-domain data produced a 10%-SpecBERT, while mixture of 2 million tokens from each domain (total 4 million tokens) yielded a 10%+10%-MixBERT. Baseline models were established as LLMs pretrained on single-domain data with the same corpus size as the mixed-domain LLMs.

3.2.1 Downstream task #1 results: Extractive QA for provisions

Figure 6 demonstrated model performance varying with corpora quantity in extractive QA for provisions. It was found that incorporating 10% close-domain data (2 million) or 10% in-domain data (2 million) led to a slight increase in S-BERT to 82.6% and 83.1%, respectively, surpassing the baseline (82.3%). As the corpus size ranged from 2 million to 20 million tokens, models pretrained exclusively on in-domain data improved from 83.1% to 89.2% while those pretrained on a mixture of close-domain and in-domain data increased from 83.8% to 88.1%. In contrast, models pretrained totally in close-domain corpus exhibited the lowest performance. These findings indicated the substantial advantages of maximizing in-domain data for contextual comprehension. Meanwhile, when in-domain data was limited, augmenting the in-domain corpus with close-domain data during pretraining really enhanced model performance. For instance, a model pretrained on mixture of 4 million tokens of in-domain data and 4 million tokens of close-domain data attained a S-BERT of 85.9%, outperforming the 84.7% S-BERT of a model pretrained exclusively on 4 million tokens of in-domain data.

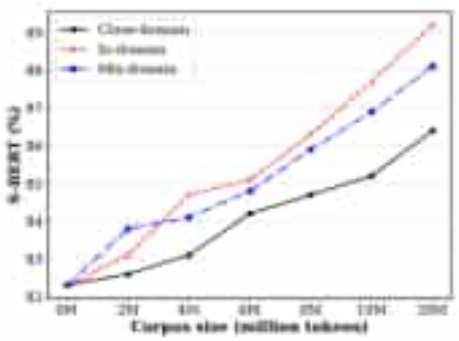


Figure 6. Model performance varying with corpora quantity in extractive QA

3.2.2 Downstream task #2 results: Information retrieval from provisions

Figure 7 illustrated how model performance varies with corpus size across three domains in information retrieval. Initial results showed that with a limited corpus of 2 million tokens (10%), models pretrained on close-domain and in-domain data achieved F1-scores of 71.1% and 71.4%, respectively, both slightly below the F1-score of 71.7% for the model without domain corpus. However, a hybrid approach combining 5% close-domain and 5% in-domain data yielded a notable improvement with an F1-score of 73.3%, a 1.9% increase over model pretrained from 10% in-domain data. This improvement suggested that the hybrid strategy effectively mitigated the overfitting risk associated with small amounts of either close-domain or in-domain data alone. As corpus size increased from 4 million to 20 million tokens, models pretrained on hybrid corpora exhibited a consistent upward trend in F1-scores, ranging from 73.5% to 76.8%. These performances were comparable to those of models pretrained exclusively on in-domain data from 73.9% to 77.4%, much higher than models pretrained using only the close-domain corpus.

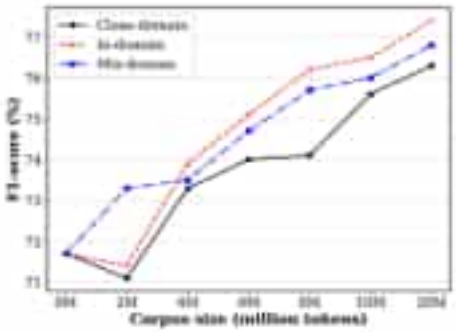


Figure 7. Model performance varying with corpora quantity in information retrieval

4 Conclusion

In this study, we proposed a three-stage LLM pretraining method to integrate domain-specific terminologies, enhancing its generalization and performances in construction specification review tasks. First, we constructed both close-domain and in-domain corpora to capture relevant construction-specific vocabulary and contexts. We then applied MLM for pretraining on these different types of corpora. Finally, we fine-tuned these pretrained models for two specification review tasks: extractive QA for provisions and information retrieval from provisions. To validate the effectiveness, we conducted two experiments. In Experiment 1, SpecBERT, our in-domain pretrained LLM, achieved the highest and at 82.4% and 89.2% in F1-score and S-BERT, surpassing the fine-tuned BERT model by 3.7% and 6.9% respectively, Additionally, SpecBERT demonstrated superior results in information retrieval from provisions, achieving a 77.4% F1-score and 88.9% accuracy, outperforming the state-of-the-art BERT model by 7.9% and 5.7%, respectively. Experiment #2 evaluated varying quantities and hybrid effects of close-domain and in-domain corpora. Results demonstrated that increasing quantity of domain corpora benefits the performance of LLMs in downstream tasks. When the size of in-domain corpus is limited, it is advisable to incorporate additional close-domain corpus data. Notably, a hybrid approach, mixture of close-domain and in-domain corpora, can yield performance levels comparable to models trained on sufficient quantities of in-domain data alone. In overall, this study

offered a practical approach to develop domain-aware pretrained LLMs adaptable to various construction specification review tasks under varying quantity.

Despite the promising potential of the work, there still exist two limitations. First, the size of the fine-tuning dataset was relatively limited. Increasing the volume and diversity of the dataset could further enhance model performance and robustness. Second, the pretraining corpus was English-centric, which may limit the model's generalizability to specifications in other languages. Future work will extend the approach to multilingual settings through cross-lingual pretraining and adaptation.

Acknowledgement

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2026-25524415).

References

[1] S. Moon, G. Lee, S. Chi, Automated system for construction specification review using natural language processing, Advanced Engineering Informatics 51 (2022) 101495. https://doi.org/10.1016/j.aei.2021.101495.

[2] J. Zhang, N.M. El-Gohary, Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking, J. Comput. Civ. Eng. 30 (2016) 04015014. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346.

[3] R. Zhang, N. El-Gohary, Natural language generation and deep learning for intelligent building codes, Adv. Eng. Inf. 52 (Apr) (2022) 101557. https://doi.org/10.1016/j.aei.2022.101557.

[4] S. Moon, S. Chi, S.-B. Im, Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT), Autom. Constr. 142 (Oct) (2022) 104465. https://doi.org/10.1016/j.autcon.2022.104465.

[5] J. Zhou, Z. Ma, Named Entity Recognition for Construction Documents Based on Fine-Tuning of Large Language Models with Low-Quality Datasets, (2024). https://doi.org/10.2139/ssrn.4914418.

[6] S. Moon, G. Lee, S. Chi, H. Oh, Automated Construction Specification Review with Named Entity Recognition Using Natural Language Processing, Journal of Construction Engineering and Management 147 (2021) 04020147. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001953.

[7] Y. Zhu, H. Yuan, S. Wang, J. Liu, W. Liu, C. Deng, H. Chen, Z. Dou, J.-R. Wen, Large Language Models for Information Retrieval: A Survey, (2024). http://arxiv.org/abs/2308.07107 (accessed July 9, 2024).

[8] E. Ahmadi, S. Muley, C. Wang, Automatic Construction Accident Report Analysis Using Large Language Models (LLMs), J. Intell. Constru 3 (2024) 9180039. https://doi.org/10.26599/JIC.2024.9180039.

[9] Y. Kim, J.-H. Kim, Y.-M. Kim, S. Song, H.J. Joo, Predicting medical specialty from text based on a domain-specific pre-trained BERT, International Journal of Medical Informatics 170 (2023) 104956. https://doi.org/10.1016/j.ijmedinf.2022.104956.

[10] Z. Zheng, X.-Z. Lu, K.-Y. Chen, Y.-C. Zhou, J.-R. Lin, Pretrained domain-specific language model for natural language processing tasks in the AEC domain, Computers in Industry 142 (2022) 103733. https://doi.org/10.1016/j.compind.2022.103733.

[11] Y. Zhong, S.D. Goodfellow, Domain-specific language models pre-trained on construction management systems corpora, Automation in Construction 160 (2024) 105316. https://doi.org/10.1016/j.autcon.2024.105316.

[12] L. Zheng, N. Guha, B.R. Anderson, P. Henderson, D.E. Ho, When does pretraining help?: assessing self-supervised learning for law and the CaseHOLD dataset of 53,000+ legal holdings, in: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law, ACM, São Paulo Brazil, 2021: pp. 159–168. https://doi.org/10.1145/3462757.3466088.

[13] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, (2019). http://arxiv.org/abs/1810.04805 (accessed April 17, 2024).

[14] J. Salazar, D. Liang, T.Q. Nguyen, K. Kirchhoff, Masked Language Model Scoring, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: pp. 2699–2712. https://doi.org/10.18653/v1/2020.acl-main.240.

[15] S. Wang, Y. Fu, J. Kim, Toward construction-specialized, small language models: The interplay of domain adaptation, model scale and data volume, Adv. Eng. Inf. 69 (Jan) (2026) 104035. https://doi.org/10.1016/j.aei.2025.104035.

[16] P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, SQuAD: 100,000+ Questions for Machine Comprehension of Text, (2016). https://doi.org/10.48550/arXiv.1606.05250.

Multi-Dimension Simulation Framework for Operating Room Construction Using Omniverse

Chia-Ying Lin[1], Lung-Hsiang Su[2] and I-Chen Wu[1]

[1] Department of Civil Engineering, National Kaohsiung University of Science and Technology, Taiwan
[2] JE MON Engineering Co., Ltd., Taiwan

cylin813@nkust.edu.tw, kwu@nkust.edu.tw

Abstract –

The construction of operating rooms involves multiple coordinated procedures. In practice, time and cost estimation are often uncertain. Additionally, project control still relies heavily on the manual interpretation of 2D drawings, resulting in inconsistencies between the design intent and on-site execution. There is one other thing that is important for the operating room. Operating rooms impose stringent requirements on illuminance and uniformity of illuminance to support long-duration surgical tasks. Inadequate lighting distribution can lead to increased visual fatigue and introduce shadows that negatively impact task performance. Therefore, this study integrates illuminance and uniformity simulation into the design workflow, allowing for the immediate evaluation of lighting performance after each layout configuration and the adjustment of lighting performance at an early stage. To address the above issues, this paper proposes a Multi-Dimension Simulation Framework based on the Omniverse platform. The proposed workflow can automatically generate the operating room interior framework (e.g., metal studs and antibacterial panels) from CAD floor plans, thereby reducing modeling time and supporting on-site installation processes. The system then placed the ceiling grid using a modular, center-out strategy, and the related attributes were recorded for subsequent quantity takeoff. At the same time, the user can place modular cabinet components, and then the system detects openings and boundaries, generating the bracing and support studs for enhancing stability. Finally, the component of the model will bind the schedule and cost data for 4D/5D simulation. Therefore, the results of this research can raise the effectiveness and efficiency of the operating room construction and project management.

Keywords –

Omniverse; Design Automation; Operating Room Construction; Illuminance Simulation; 4D/5D

1 Introduction

As a specialized interior environment defined by high sterility and process-intensive operations, the operating room demands strict adherence to specific protocols in design and construction. Guided by cleanroom design manuals, walls and ceilings must employ a modular, seamless construction integrated with a metal stud system; corners require rounded profiles to prevent dust accumulation, while openings necessitate reinforcement and edge sealing to ensure compliance with ISO Class 5–7 cleanliness standards [1]. However, current practices predominantly rely on manual interpretation of 2D schematics for modeling. This often results in inconsistent stud modularity, edge detailing, and opening-reinforcement execution, making it difficult to translate design requirements into computable parameters—such as module size and spacing, minimum clear distances, and corner radii—that serve as inputs for subsequent geometry generation and verification.

In the domain of automated configuration and digital workflows, academic research has made significant progress in recent years. Lee et al. [2] utilized Building Information Modeling (BIM) technologies to establish a template matching protocol, proposing a methodology wherein model geometry is deconstructed to allocate standard components based on rules and priorities, while defining non-configurable zones. This approach effectively illustrates the implementation path of transforming "regulation to rule to geometry." Addressing integration and collaboration across diverse data sources, Abbasianfar [3] posited that distributed construction simulation—encompassing IoT, imagery, and numerical models—relies critically on interoperability and consistency. Visualization platforms anchored by Universal Scene Description (USD) and NVIDIA Omniverse were identified as effective tools for preserving data integrity and supporting visual synchronization. Regarding the practical integration of 4D/5D dimensions, Inzerillo et al. [4] used a railway station project to demonstrate the interfacing of Revit

(3D/Quantity Take-off) with Navisworks (4D scheduling) and supplementary plugins (ArchVISION/PriMus) for cost linkage. This integration achieved real-time updates of quantities and costs in response to model changes, thereby establishing a unified management system for both schedule and expenditure.

In the domain of lighting simulation, the research focus has shifted from solely managing energy consumption to prioritizing human-centric visual comfort and optimizing illuminance distribution. Since Kennedy and Eberhart [5] introduced Particle Swarm Optimization (PSO), which employs swarm intelligence to determine global optima, this methodology has become a cornerstone of lighting control systems. Subsequently, Madias et al. [6] integrated multi-objective evolutionary algorithms into interior lighting control, demonstrating that compliance with the EN 12464-1 standard could align with the dual objectives of average illuminance and uniformity, effectively enhancing energy efficiency. Wagiman et al. [7] advanced this research by employing Multi-Objective Particle Swarm Optimization (MOPSO) alongside the Illuminance Uniformity Deviation Index (IUDI); by simultaneously minimizing power consumption and non-uniformity, their simulations achieved compliance with standards requiring a uniformity of ≥ 0.6 and an average illuminance of 500 lx. From an engineering practice perspective, Wati [8] highlighted that the Lumen Method (also known as the Zonal Cavity Method) retains a distinct efficiency advantage, enabling the rapid estimation of average illuminance and luminaire requirements, which serve as a reference for electrical configuration. In the context of operating rooms, where zoned lighting and scenario transitions (e.g., preparation, surgery, sanitation) are operational necessities, illuminance simulation must occur concurrently with geometric configuration. Only through this synchronization can the stringent requirements for shadowless illumination, uniformity, and maintenance accessibility be fully met.

In light of the foregoing analysis, three critical challenges persist within contemporary practice that demand immediate resolution. First, operating room renovation regulations remain unsystematically parameterized, rendering them unable to directly drive automated generation and verification mechanisms. Second, the integration of geometric modeling with 4D/5D data relies predominantly on manual linkage, resulting in Information not synchronized and inadequate version consistency. Third, illuminance simulation is typically introduced only in the latter stages of design, which prevents the establishment of a closed feedback loop with stud and panel configurations. To address these impediments, this study establishes an automated interior renovation workflow anchored by Universal Scene Description (USD). By implementing a rule engine using Python and Omniverse, the system translates construction principles into geometric constraints and processes. The system will generate the operating room model based on these rules. The scene utilizes the USD format to preserve readability and modification advantages, reserving metadata fields for 4D/5D integration to link scheduling and cost data simultaneously. Concurrently, the system synthesizes generated geometry for real-time assessment of illuminance and uniformity, thereby realizing a holistic workflow that unifies design, construction, and performance simulation.

2 System Framework

This study uses an operating room as the case study and targets renovation projects characterized by complex interfaces and frequent design changes. It establishes a Multi-Dimension Simulation Framework that integrates Python automation scripts with Universal Scene Description (USD). As shown in Fig. 1, the framework follows a data-driven approach and organizes the workflow into three sequential stages:

1. Drawing parsing input layer: reads 2D vector drawings, parses entities, and extracts geometric feature data.
2. Rule-based automation core: parameterizes the extracted geometry and converts it into USD assets according to predefined construction logic (e.g., opening reinforcement and light-gauge partition framing layout).
3. Simulation application layer: imports the generated model into NVIDIA Omniverse to support integrated simulations of construction sequencing and environmental performance.



Figure 1. System framework for this research.

2.1 System design

The system design is shown in Fig. 2. The workflow is implemented by two main classes, Interactor and USDScene, and three supporting components: OmniKit_Context, JSON_Data_Source, and pxr.Usd.Stage. Interactor is responsible for UI interaction and state management. It queries OmniKit_Context (e.g., get_context() and get_selection()) to access the Omniverse editor context and user selections. It also reads the initial layout inputs from JSON_Data_Source (parsed_data.json) and stores them as baseline data

(baseline_data). During editing, Interactor records user actions in a command list (cabinet_tasks) and derives the current logical state (tubes) by overlaying these commands on the baseline data, without directly editing USD geometry.

When the logical state is updated, Interactor instantiates USDScene and triggers a rebuild (Fig. 2, "Instantiates & Triggers Rebuild"). USDScene performs full-scene regeneration and writes geometry to pxr.Usd.Stage (Fig. 2, "Generates/Writes Geometry"), including floor/grid construction from wall segments (wall_segments), column placement from structural-chain data (cols_by_chain), panel and opening generation, and detailing operations such as corner correction, C-channel reinforcement, and ceiling-keel placement (create_ceiling_keels). The stage is then saved as USD-compliant outputs. In addition, Interactor reads/monitors pxr.Usd.Stage changes via a Tf.Notice listener (Fig. 2, "Reads/Monitors") to synchronize stage updates with the internal command flow.

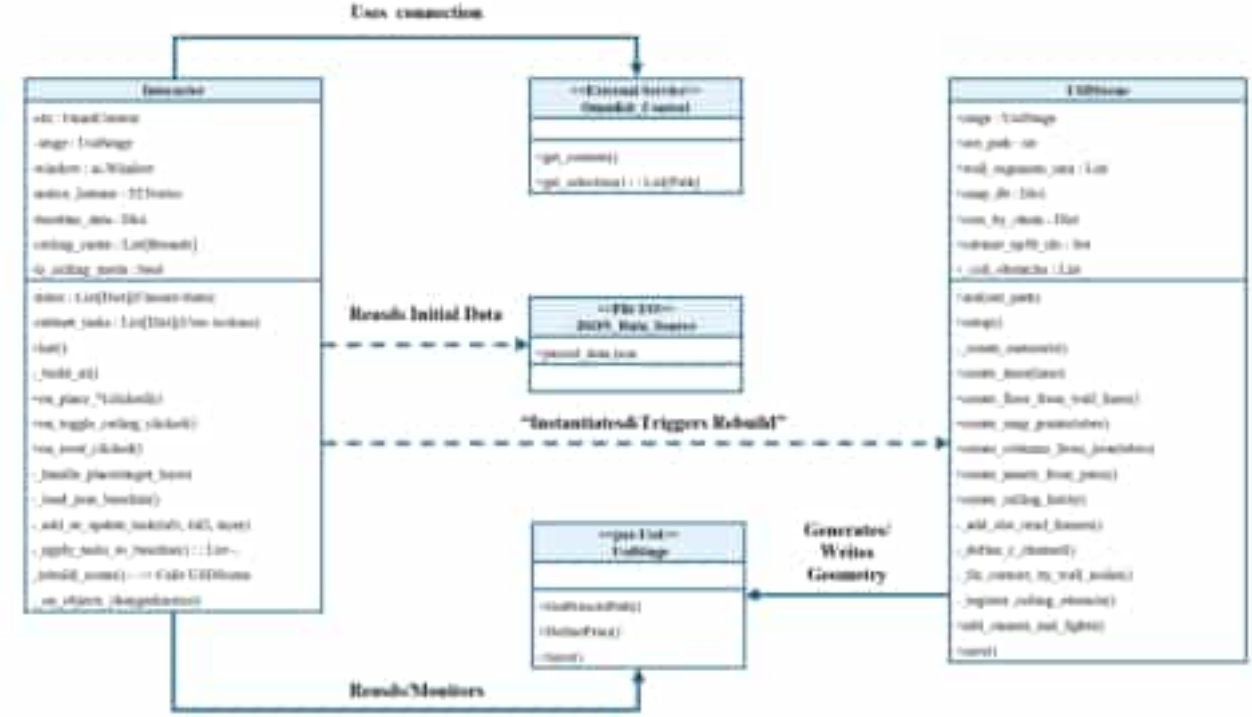


Figure 2. Schematic of the System Design.

3 Operating Room 3D Model Generation

This section outlines the comprehensive workflow for automatically generating a 3D operating room model from 2D drawings. The system divides the process into four stages. First, it parses the raw CAD data to obtain spatial coordinates. Next, it generates building components according to procedural construction rules. And then, it outputs a USD model that includes engineering information. Finally, the user can place modular cabinet components for time and cost simulation.

3.1 Layout Data Extraction

This study establishes a standardized preprocessing workflow that converts AutoCAD drawings into computable data. The program will filter out non-geometry data, such as dimensions, text, and hatches, to reduce noise, and then export only the geometry of walls and columns. The extracted coordinates of lines and points are written into the basic data as the geometric basis for subsequent automated generation and validation. The CAD drawing of the operating room layout is shown in Fig. 3.

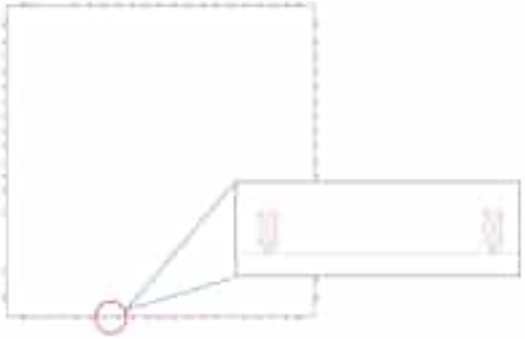


Figure 3. The CAD drawing of the operating room layout.

3.2 3D Model of Operating Room

The automated workflow in this study outputs not only 3D geometry but also a set of semantic OpenUSD assets. The resulting 3D model of the operating room is shown in Fig. 4. The assets are stored in USDA format, which features an open data structure that allows for direct access and attribute comparison, thereby improving modification and version control during the development stage. During component generation, the system writes the required metadata so that the geometry can be used as "smart objects" for engineering management. Examples include:

- Classification and model ID: bim:model (e.g., CeilKeel-3x6, Panel-<len>).
- Dimensional information: bim:length_mm, grid width, and related fields for 5D quantity takeoff.
- Construction sequence: sim:order for 4D animation and progress mapping.



Figure 4. The results of the operating room 3D model.

To ensure consistency across tools, the scene utilizes a unified upaxis and unit definition. After importing the assets into NVIDIA Omniverse, 4D/5D binding and illuminance analysis can be performed directly without additional coordinate or unit adjustments. Overall, the USDA maintains a clear linkage among "geometry–attributes–process." This supports rule verification and change detection, and provides a stable data structure for schedule- and cost-related computations.

As shown in Table 1, this research uses the general surgery operating room as a comparison with manual modeling. The proposed system not only generates the operating room geometry but also writes the required attributes and processes the data simultaneously. This allows the generated model to be directly used for later 4D scheduling, 5D cost calculation, and illuminance simulation. In contrast, manual modeling usually focuses on geometry reconstruction, while the data needed for later simulations must be added, organized, or transferred to other platforms afterward.

Table 1. Comparison between manual modeling and system-generated modeling

Item	Manual modeling	System-generated modeling
Modeling time	1 hr 6 min	1 min
Geometry	Antibacterial panels, floor, ceiling, and equipment	Antibacterial panels, floor, ceiling, equipment, peripheral framing, and equipment reinforcement structures
Data	Additional organization or manual input required	Written automatically during generation
4D / 5D / illuminance simulation	Cross-platform integration required	Can be directly continued within the proposed system

3.3 Implementation for Rule-based Engine

This study implements a rule-based engine for modeling of operating room construction, covering four functions: geometry generation, interactive reinforcement, 4D/5D simulation, and illuminance verification. To support review and implementation, Table 2 summarizes each rule set by its trigger conditions, processing steps, and output attributes.

3.4 Modular Cabinet Components Placement

The user interface and interactive mechanism for placing and reinforcing modular cabinet components are illustrated in Fig. 5. The cabinet instances used for placement are created in Revit and exported as Revit family files (RFA), and are also converted into USD assets for interaction in the scene. When the user places a cabinet asset, the system evaluates its geometric relationship to openings and boundaries and generates the corresponding reinforcement components with associated attributes for sequence simulation and cost/schedule linkage.

Table 2. Mapping of system generation rules.

Rule Category	Trigger Condition	Algorithmic Action	Generated Components	Key Attribute
Framing Generation	Closed boundary provided	Place studs and panels along the path using modular spacing; generate edge trims at boundaries.	C-Studs, Antibacterial Panels, Edge Trims	bim:model、bim:length_mm
Ceiling Grid Generation	Geometric center defined	Expand ceiling grids outward from the center, splitting at intersections and trimming at boundaries.	Ceiling Grids, Ceiling Panels	bim:model、bim:length_mm
Interactive Reinforcement	Equipment Placement	Analyze topological relations, generate opening frames, horizontal supports, and corner stiffeners, and densify studs as required.	Opening Frames, Horizontal Supports, Corner Stiffeners, Densified Studs	bim:model、bim:length_mm
4D Simulation	Visibility Activation	Toggle visibility sequentially to reconstruct the assembly order.	Step-by-step Visualization	sim:order
5D Cost Estimation	User Initiation	Convert length to modules and cost in real time; remove duplicates by selecting the longest overlapping path segment.	Itemized / Total Cost	bim:length_mm
Illuminance & Uniformity	Manual Simulation Trigger	Run illuminance mapping and uniformity checks based on the generated scene.	Illuminance Map, Uniformity Verification Map	-



Figure 5. User interface and interactive mechanism for placing and reinforcing modular cabinet components.

4 4D/5D Data Binding and Visualization

This study proposes an integrated strategy that packages geometric models, construction schedules, and cost data into a single scene description file, thereby reducing information fragmentation in conventional BIM workflows. The system uses the USD file as the computational core, where required attributes are stored in an embedded data structure within the scene. Accordingly, 4D and 5D computations are driven by reading the schema attributes attached to each component, rather than querying a separate data store.

In terms of data structure, the system uses the

component model identifier (bim:model) and geometric dimensions (e.g., bim:length_mm) as the basis for quantity measurement, and uses the sequence attribute (sim:order) to label the construction order. After parsing these attributes, both 4D reconstruction and 5D cost estimation can be performed simultaneously, as shown in Fig. 6. Each 4D event—defined as a component switching from hidden to visible and treated as "completed"—also serves as the trigger point for 5D costing. Because progress milestones and cost calculations share the same component ID and attributes, the system can accumulate costs along the timeline with synchronized data updates.

First, the 4D schedule reconstruction applies a "sequential reveal" strategy. At the start of the simulation, the entire scene is hidden, and component visibility is then switched on in batches according to sim:order to reproduce the assembly sequence and serve as the basis for computing the progress timeline. Second, 5D cost estimation is performed concurrently during the 4D animation. For components that have been revealed, the system performs real-time quantity takeoff and applies unit prices, enabling a "cost-on-completion" workflow：

1. Linear components (e.g., ceiling grids and C-channels): Quantities are accumulated using length attributes and then converted into the number of standard modules as the budgeting basis.
2. Wall panel deduplication: For the same wall-path segment, only the largest Panel-<len> is used as the calculation value to avoid duplicate counting caused by overlaps.
3. Unit-based modules (e.g., ceiling panels): The number of pieces is derived from width attributes and predefined modular rules.

These rules ensure consistency in geometry, scheduling, and cost computations within a single USD scene, facilitating tracking and comparison.

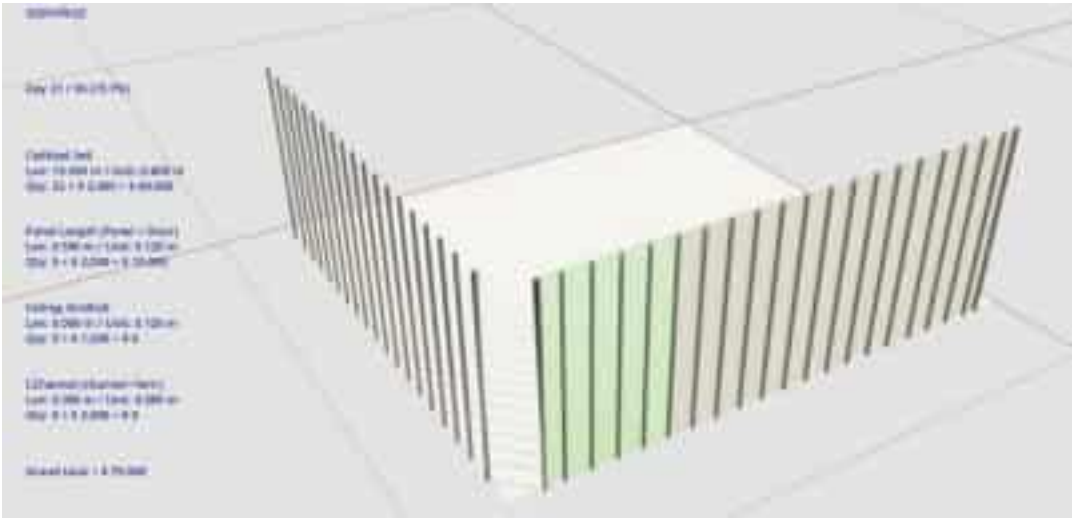


Figure 6. The Interface for 5D Process Simulation and Real-Time Cost Analysis.

5 Illuminance and Uniformity Simulation

This study undertakes a comprehensive evaluation of operating room lighting in strict compliance with the EN 12464-1 standard. Addressing the specific illuminance requirements for medical environments, the benchmark for recovery rooms is set at 500 lx, and general operating room illumination at 1000 lx. Conversely, the surgical cavity mandates an intensity ranging from 10,000 to 100,000 lx, as outlined in Table 3. Regarding the assessment of uniformity, computations are executed utilizing Equation (1):

$$U_0 = \frac{E_{min}}{E_{avg}}$$

Where U_0 denotes the uniformity of illuminance; E_{min} represents the minimum illuminance within the evaluation zone; and E_{avg} signifies the mean illuminance of the respective region. In accordance with established standards, the scope of evaluation is delineated into the task area and the immediate surrounding area:

- Task Area: Defined as the primary locus where visual tasks are executed (exemplified by the operative surface of the surgical table).
- Immediate Surrounding Area: Defined as a band-shaped zone that surrounds the task area and lies within the field of view. According to the relevant requirements, its width shall be at least 0.5 m. As illustrated in Fig. 7, the red box indicates the primary surgical task area, and the blue box represents the area expanded outward by 0.5 m. The system uses this defined range for grid discretization and uniformity verification.

In instances where the requisite illuminance upon the working plane equals or exceeds 750 lx—with this study stipulating a general operating room illuminance of 1000 lx—the objective uniformity is mandated as follows: ≥ 0.7 for the task area and ≥ 0.5 for the immediate surrounding area, as delineated in Table 4.

In the implementation, the system integrates geometric configuration with optical simulation. After the user places objects about surgical equipment, luminaires, and the surgical table in the scene, the system enables NVIDIA Omniverse Illuminance Arbitrary Output Variables (AOVs) to perform ray tracing and outputs a high-dynamic-range EXR image. At the moment the surgical table is placed, the system simultaneously detects its bounding box and defines it as the task area. The immediate surrounding area is then automatically generated by extending the task area outward by 0.5 m in accordance with EN 12464-1. The EXR data are subsequently read by a Python program developed in this study, which performs grid discretization and numerical sampling within these predefined regions for illuminance and uniformity verification. As shown in Fig. 8, a false-color map is first used to present the illuminance distribution, where red/orange indicates high illuminance, yellow/green indicates medium levels, and blue indicates low illuminance; the corresponding lx values are mapped to the color bar below, allowing designers to quickly identify hotspots and dark zones. As shown in Fig. 9, the system further performs a quantitative grid-based verification for the task area by computing illuminance values at each sampling point and checking whether the uniformity index U_0 meets the target criteria.

Table 3. Design Illuminance Specifications for Medical Environments [10].

Type of interior, task or activity	$\bar{E}_m$	Remarks
Pre-op and recovery rooms	500	
Operating theatre	1000	
Operating cavity	-	$\bar{E}_m$:10,000 to 100,000 lx

Table 4. The Relationship of Illuminance and Uniformity between the Task Area and the Immediate Surrounding Area [10].

Task illuminance (lx)	Illuminance of immediate surrounding areas (lx)	Uniformity: Task area (U_0)	Uniformity: Immediate surrounding area (U_0)
≥ 750	500	≥ 0.7	≥ 0.5



Figure 7. Schematic definition of the uniformity evaluation regions.

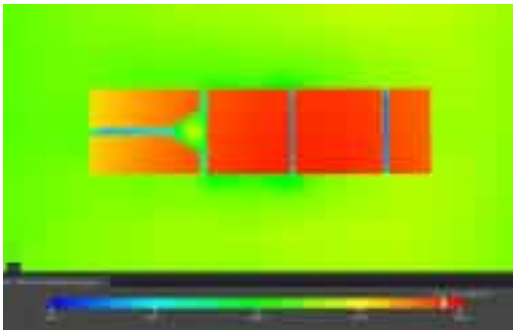


Figure 8. Simulated overall illuminance distribution for the operating table.

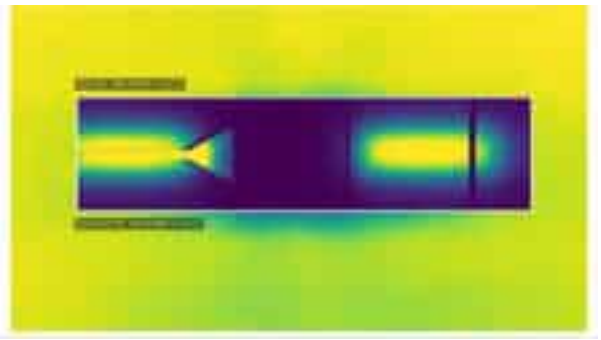


Figure 9. Analytical Diagram Verifying Illuminance Uniformity within the Task and Immediate Surrounding Areas.

6 Conclusions and Future Work

This study uses operating-room renovation as the validation scenario and develops a Multi-Dimension Simulation Framework. The framework separates computational logic from geometric data, providing a workflow from drawing parsing to multidimensional simulation. The main contributions are summarized as follows:

1. 3D Modeling: The system generates structure framing that includes metal studs and ceiling grids automatically. When the user places the cabinet components, the system updates the opening–boundary relationships and generates the required bracing and support studs. This can reduce the time and cost of 3D modeling.

2. 4D Simulation: The installation sequence is replayed by reading sequence attributes embedded in each component. This enables pre-construction checks of task precedence and supports early detection of process conflicts.
3. 5D Cost: 3D Model components are linked to a unit-price table. When wall or ceiling configurations change, quantities and total costs for panels and framing are recalculated and updated to reflect the current design state, allowing budget impact assessment without manual takeoff.
4. Illuminance and uniformity verification: Illuminance images (EXR) are generated via ray tracing. The system will calculate the average illuminance and uniformity on the operating table and verify compliance against EN 12464-1, enabling early verification during the design phase.

In summary, the results indicate that using USD as a unified data storage can integrate geometry, time, cost, and physical performance data within a single workflow, thereby reducing data discontinuities across heterogeneous tools. Future work will extend the framework to computational fluid dynamics (CFD) and human-factor evaluations, including airflow simulation for operating room ventilation, to support quantitative decision-making in healthcare space planning.

References

[1] White E. Cleanroom design, construction, and qualification. *Journal of Validation Technology*, 15(4):30, 2009.

[2] Lee B., Choi H., Min B., Ryu J. and Lee D. E. Development of formwork automation design software for improving construction productivity. *Automation in Construction*, 126:103680, 2021.

[3] Abbasianfar V., Karim M. R., Hague S. and Mohamed Y. Evaluating the interoperability of 3D visualization platforms in federated construction simulation settings. In *Proceedings of the International Conference on Computing in Civil and Building Engineering*, pages 258–271, Cham, Switzerland, 2024.

[4] Inzerillo L., Acuto F., Pisciotta A., Mantalovas K. and Di Mino G. Exploring 4D and 5D analysis in BIM environment for infrastructures: A case study. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:233–240, 2024.

[5] Kennedy J. and Eberhart R. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, pages 1942–1948, Perth, Australia, 1995.

[6] Madias E. N. D., Kontaxis P. A. and Topalis F. V. Application of multi-objective genetic algorithms to interior lighting optimization. *Energy and Buildings*, 125:66–74, 2016.

[7] Wagiman K. R., Abdullah M. N., Hassan M. Y. and Radzi N. H. M. A new metric for optimal visual comfort and energy efficiency of building lighting system considering daylight using multi-objective particle swarm optimization. *Journal of Building Engineering*, 43:102525, 2021.

[8] Wati G., Suardi S., Mubarak A. A., Fabila N., Simanjuntak L. G., Adila I. R. and Myistro E. Analysis of generator power requirements for lighting distribution using LED lights on a 500 DWT Sabuk Nusantara. *Indonesian Journal of Maritime Technology*, 1(2), 2023.

[9] NVIDIA Corporation. Importing USD— Omniverse Unity Connector documentation. Online:https://docs.omniverse.nvidia.com/connect/latest/unity/manual/import.html,Accessed:09/12/2025.

[10] CEN. *Lighting of work places - Part 1: Indoor work places*, EN 12464-1. European Committee for Standardization, Brussels, 2002.

A Computational Pipeline for Design and Fabrication of Building Slab Units with 3D Printing and Robotic Production Cells

Yuchen Tang[1], Emad Shakour[1], Oded Amir[1] and Rafael Sacks[1]

[1]Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Israel

yuchen.tang@campus.technion.ac.il, semad@technion.ac.il, odedamir@technion.ac.il, cvsacks@technion.ac.il

Abstract -

Full-scale 3D concrete printing for horizontal structural components is challenging due to large unsupported spans, early-age material instability, and geometric limitations of extrusion-based processes. A practical solution is to fabricate modular slab units by printing only the structural ribs of the slab while casting the top and bottom layers conventionally and achieving composite action through post-tensioning. This paper presents a computational pipeline for the integrated design and feasibility evaluation of such hybrid slab units. A design pipeline generates topology-optimized rib layouts, segments the slab into prefabricated units, refines local structural features, and performs constructability checks before producing a detailed BIM model. A complementary simulation pipeline evaluates fabrication and feasibility of erection within the same workflow. Motion-level simulations in Gazebo assess reachability, collision avoidance, and lifting trajectories for production cells, while rule-based risk evaluators identify constraints beyond rigid-body simulation.

Keywords -

3D concrete printing; Constructability evaluation; Hybrid construction; Ribbed slab units; Topology optimization; Robotic simulation.

1 Introduction

Recent advances in 3D concrete printing have enabled efficient automated fabrication of vertical components, such as walls and columns, where material is deposited layer by layer in a predominantly vertical direction and is continuously supported[12, 19]. In contrast, full-scale slab printing remains largely infeasible. Slabs require depositing fresh material horizontally over large spans, which requires traditional formwork and cannot be printed. In addition, many current extrusion-based systems are designed for prismatic, single-direction geometries that are not well suited to multi-directional structural behaviour.

An alternative is to prefabricate units in production cells where printing is restricted to the structural ribs of the slab. In such a hybrid system, ribs can be generated using topology optimization to align with principal stress trajectories. The top and bottom layers can be cast conventionally, and post-tensioning can be used to activate composite action and enable large-span performance. The slab can be segmented into prefabricated units to satisfy transportation and lifting constraints. However, this hybrid approach produces a highly coupled workflow involving rib printing, concrete casting, unit lifting, on-site assembly, and post-tensioning. Assessing feasibility across this workflow is challenging due to geometric, structural, and robotic limitations.

In this work, we formulate a computational pipeline for the integrated design and feasibility evaluation of such slab units. The pipeline is structured into two tightly coupled components:

1. **Design**: design of hybrid ribbed slab units, from initial input conditions to parametric BIM geometry, with a focus on topology optimization. This workflow not only serves to provide input for simulation of the system as a whole, it also forms the basis for a fully parametric and automated design system for eventual production of such slab units.
2. **Simulation**: simulation of the fabrication process that integrates robot motion and 3D printing equipment with rule-based scheduling and risk evaluation, and simulation of the erection and post-tensioning process on site.

The paper provides a background review of the state-of-the art for fabrication of horizontal elements with 3D printing in building construction, software for structural design and topology optimization, software for simulation of robotic and 3D printing production cells, and software for 4D planning and analysis of construction operations. It then details feasible computational workflows for each of the two steps listed above.

2 Related work

2.1 3D printing of structural components

3D printing of cementitious materials, particularly 3D concrete printing (3DCP), is gaining interest for enabling

complex geometries and material-efficient structural design. This is increasingly relevant given the difficulty of realizing optimized concrete structures using conventional formwork-based construction.

3DCP allows formwork-free fabrication of complex geometries, but applications have primarily focused on vertical elements. Horizontal members remain challenging due to bending-dominated response and limited reinforcement options, which motivates compression-dominated designs. For beams and slabs, prestressing enables such behavior and improves structural performance [11], while also supporting segmental post-tensioned and modular construction.

Hybrid construction, combining printed stay-in-place formwork with cast concrete, has emerged as an alternative approach. Initially applied to vertical elements, it has recently been extended to horizontal members [9]. These developments highlight the potential of 3DCP and the need for integrated computational workflows that link structural design, optimization, fabrication, and construction.

2.2 Structural optimization

Structural optimization, particularly topology optimization, enables substantial material savings by aligning geometry with internal force flow. Owing to its design freedom, topology optimization can be tailored to address multiple design stages, from conceptual layout to fabrication-aware realization.

A simultaneous shape and topology optimization framework for prestressed concrete systems was introduced in [1, 2] and later integrated with 3DCP through CAD-based interpretation and post-tensioned fabrication [16]. This demonstrated its potential for segmented, prestressed printed structures, while also highlighting the need for tighter design–fabrication integration. The framework has since been extended to fully three-dimensional free-form optimization of prestressed beams and slabs [14], which is adopted in the present study.

An additional key challenge in realizing optimized geometries with 3DCP is early-age material behavior during printing. To address this, optimization formulations incorporating buildability constraints have been proposed to ensure printability [7, 13].

Overall, these developments demonstrate the potential of structural optimization for efficient 3DCP, particularly when structural, material, and fabrication considerations are integrated within a unified design framework.

2.3 Robotics simulation in construction

Robotics simulation environments such as Gazebo and V-REP have been increasingly used in construction research to visualize and validate robotic workflows [5]. However, most studies treat simulation as a visualization or debugging tool, focusing on kinematics and collision avoidance, rather than using it as part of a risk-evaluated construction logic. The proposed methodology embeds motion-level simulation into a validation pipeline, combining Gazebo-based action simulation with external rule-based evaluators to assess constructability, printability, and erection risks of full slab systems—bridging the gap between motion simulation and feasibility-driven design.

2.4 4D CAD simulation for construction

4D CAD is a well-established method for evaluating construction methods and processes [10]. It is commonly used for complex construction projects and numerous software tools offer 4D functionality. However, they are restricted to predetermined motion paths and fixed element geometries - they lack the operational detail and the physics engines needed for planning and validating the motion of robotic and 3D concrete printing machinery needed for design of automated production cells contemplated in this work. For this reason, we use Gazebo Simulation[8] with Robotic Operating System (ROS) integration to plan and validate the constructability of slab units.

2.5 Software interoperability

The increasing use of BIM, parametric modeling, and robotics in construction demands robust software interoperability. However, workflows often remain fragmented, with tools such as Revit, Rhino/Grasshopper, FEM solvers, and ROS/Gazebo used in isolation. Prior efforts have explored parametric and BIM-integrated workflows, but few address interoperability from early design to robotic fabrication [6, 15, 18]. The proposed pipeline demonstrates a connected chain: BIM authoring $\rightarrow$ parametric modeling $\rightarrow$ structural analysis $\rightarrow$ robotic simulation, with data exchanged through mesh conversion, scripting, and task encoding—supporting a full feedback loop across design and construction domains.

3 Solution overview

The proposed method integrates design and simulation into a two-stage computational pipeline.

First, a Design Pipeline processes an input architectural slab layout into a manufacturable set of hybrid precast units. This design stage encompasses geometry extraction, global topology optimization, segmentation into units, local structural design, constructability checks, and parametric BIM geometry generation.

Second, a Simulation Pipeline takes the BIM model as input and evaluates feasibility and safety. It converts geometry into formats suitable for robotic simulation, defines tasks, generates toolpaths and trajectories, sets up Gazebo

scenes, runs simulations for both fabrication and erection, and complements Gazebo with rule-based risk evaluators.

The two pipelines are linearly connected by default: the output of the Design Pipeline feeds into the Simulation Pipeline sequentially. Critically, the framework also supports an iterative feedback loop (stage S7) whereby detected risks in simulation prompt design modifications in earlier stages, enabling design-for-constructability through repeated cycles if needed. This ensures that the slab system not only meets structural and architectural requirements but can also be fabricated with robotic 3D-printing and assembled on-site.

4 Design pipeline

4.1 D1 – Input and boundary conditions

The Design pipeline (Figure 1) begins with a formal definition of the slab design problem, consolidating all relevant architectural, structural, and constructability parameters. In this stage, the system gathers and encodes the boundary conditions, design objectives, and fabrication constraints that govern the remainder of the pipeline.

Architectural input includes the slab layout, openings, and support configuration, typically extracted from BIM models in Revit or IFC format. Structural input includes material properties , design load cases, and permissible deflection criteria. Constructability-related constraints are also explicitly defined. These include limits on maximum unit weight based on crane capacity, unit dimensions based on transportation and printer workspace, and minimum rib thickness or curvature radius based on extrusion nozzle geometry.

All inputs are consolidated into a structured configuration file, which may be implemented as a .json or .yaml file and is readable by downstream modules. Each parameter is associated with its source and its status. This configuration serves as the canonical reference for the remainder of the pipeline, guiding geometry extraction, optimization, segmentation, constructability checks, and simulation. This stage enhances clarity, reusability, and cross-project consistency.

4.2 D2 – Geometry extraction and pre-processing

This stage converts architectural slab definitions from BIM into a clean and analysis-ready geometric domain. The process begins by extracting the slab's physical outline and internal voids using Dynamo for Revit or Python scripts with IfcOpenShell for IFC-based models. Structural supports—such as columns, load-bearing walls, or beams—are identified by intersecting BIM elements with the slab geometry. These intersections are translated into support boundary conditions for structural analysis. Unlike conventional approaches that assume regular or rectangular shapes, this method accepts arbitrary slab boundaries, including non-convex, curved, and multiple-voided geometries.

The extracted geometry is then pre-processed to remove inconsistencies and modeling noise. Small fillets, disconnected edges, and planar irregularities are cleaned. The cleaned geometry is exported as a three-dimensional solid mesh and used as input for 3D topology optimization in stage D3.

4.3 D3 – Global topology optimization

This stage applies three-dimensional topology optimization to generate the global rib-and-void geometry of the slab. Using the meshed slab domain from Stage D2 and the material and boundary conditions from Stage D1, the slab is modeled as a three-dimensional solid domain. The optimization follows the prestressed concrete topology optimization framework developed in [14], incorporating common prestressed concrete design practices in the objective formulation and volume constraint, including the structural effects of post-tensioning on load transfer and stiffness. We note that the framework incorporates simultaneous optimization of the concrete distribution and of the tendons' geometry, such that the effect of prestressing is captured from the early design stage by means of equivalent forces.

An illustrative example is presented for a simply supported prestressed concrete slab with dimensions $12.0 \times 12.0 \times 0.36$ m, subjected to uniformly distributed dead and live loads of 10 kN/m^2 each. The top and bottom two element layers are prescribed as solid material, corresponding to a thickness of 0.06 m for the top and bottom slab parts, while the optimization is restricted to the intermediate region to identify an efficient rib layout connecting the lower and upper concrete layers. The overall allowable concrete volume fraction, including the top and bottom parts, is limited to 50%. A mid-height cross-section illustrating the top and bottom parts of the optimized design is shown in Figure 2.

The resulting rib geometry is freeform, shaped by the optimized layout of post-tensioning cables and additional rib stiffeners. The output is a three-dimensional solid geometry capturing the full layout of ribs and internal voids, which is exported as an STL mesh or converted into editable BREP surfaces representing the printable and structurally informed volume of the slab core.

4.4 D4 – Unitization (segmentation into prefabricated units)

After the generation of the global rib geometry, the slab must be divided into constructible units suitable for robotic fabrication, transport, and on-site lifting. This unitization

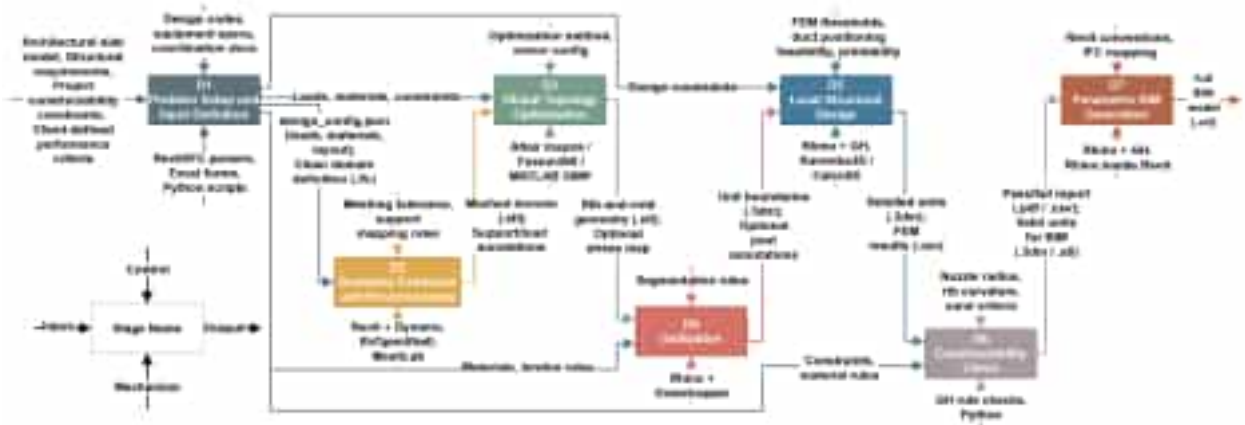


Figure 1. Design pipeline using IDEF0 notation



(a) Top layer.



(b) Bottom layer.

Figure 2. Optimized slab geometry.

stage defines the boundaries between slab modules and directly affects their structural continuity, printing feasibility, and assembly workflow. This modular decomposition aligns with DFMA-oriented prefabrication approaches [4].

Segmentation begins by identifying regions of the rib network where natural breaks can occur. Cutting lines are guided by the rib topology: they are deliberately routed through void zones or rib junctions, and avoid slicing directly through the middle of major load-bearing ribs. The goal is to minimize the disruption of continuous load paths and reduce the total rib cross-sectional area that must be interrupted. Curved or angled segmentation lines are allowed if they better follow the rib architecture, rather than imposing an artificial rectilinear grid.

Each proposed unit is then evaluated against fabrication and logistics constraints defined earlier in stage D1. These include limits on unit footprint, estimated weight for lifting, and clearance requirements for handling equipment. At this stage, the evaluation is rule-based and geometric—no finite element analysis is performed. Units that violate basic constraints are iteratively re-partitioned. Where unavoidable, segmentation boundaries that intersect rib regions are marked for later reinforcement or anchorage detailing in stage D5.

The unitization workflow is implemented in Rhino using Grasshopper, with rib geometry imported as STL or mesh files from D3. Segmentation curves are generated and refined parametrically, then exported as BREP sub-volumes. These definitions are passed to downstream detailing (D5) and BIM integration (D7), ensuring traceability and continuity across design stages.

4.5 D5 – Local structural design and detailing

Once the slab has been divided into prefabricated units, each segment undergoes local structural design to ensure printability, post-tensioning feasibility, lifting safety, and full structural performance. This stage converts the global optimization mesh into detailed, buildable geometry.

Since the ribs generated in stage D3 are non-parametric and typically represented as STL meshes, a geometric reconstruction process is performed. The discrete optimization mesh is converted into parametric BREP geometry through cross-section sampling and lofting to enable downstream detailing.

Tendon ducts are embedded within the rib volume using solid Boolean operations. Each duct follows a defined path from one anchorage point to another, and is constrained by minimum curvature, wall clearance, and interface geometry. At rib intersections, ducts are either limited to principal structural directions or offset vertically to avoid conflict. This ensures smooth threading and successful anchorage during post-tensioning. Void regions between ribs form unoccupied spaces that are temporarily filled with sand to support the top concrete layer during casting. Rib spacing and topology are adjusted to avoid excessive overhangs and ensure that voids remain accessible for sand placement and removal.

The slab unit is completed with a bottom and top concrete layer, each containing its own steel mesh reinforce-

ment. The three components form a sandwich structure. For enhanced interlayer shear resistance, starter bars may be embedded into or alongside the ribs. In high-curvature or support-bearing regions, ribs may be thickened, or vertical reinforcement zones introduced. Each unit undergoes structural validation using lightweight finite element models, such as Karamba3D or CalculiX, to evaluate lifting deformation, internal stress distribution, and tendon alignment, reflecting the influence of prestressing on structural response. Units that fail to meet structural or fabrication constraints are flagged for revision of rib geometry, duct routing, or anchorage region shaping.

The output of stage D5 is a fully detailed, fabrication-ready slab unit model—including solid ribs, tendon ducts, sand-supportable voids, and cast concrete layers—passed to stages D6 for constructability verification and D7 for BIM integration.

4.6 D6 – Constructability check

Before the slab unit models proceed to simulation and BIM integration, each unit undergoes a constructability check to ensure that the geometry complies with fabrication, transportation, and assembly constraints, verifying whether the geometry can be realistically fabricated, lifted, and handled. This stage checks a series of rule-based geometric and physical constraints, such as: minimum rib thickness and printable curvature radius to avoid nozzle failure or material sagging; maximum unit dimensions and weight based on gantry size, crane capacity, and transport limits; duct minimum bending radius to ensure tendon threading is possible.

These checks are implemented in Grasshopper using parametric scripts. Thresholds and constraints are imported from stage D1 and updated as needed. Units that fail any criteria are flagged for revision in rib geometry, segmentation boundary, or detailing logic.

The constructability check also produces embedded metadata such as center of gravity, expected lifting points, and estimated assembly tolerances. This information is essential for planning site erection (S5) and exporting to BIM (D7).

The output of stage D6 is a verified set of slab units with constructible geometry and constraint-satisfying properties. These models are then passed to the simulation pipeline.

4.7 D7 – Parametric BIM generation

After structural and constructability validation, each slab unit is converted into a parametric BIM representation for coordination, documentation, and integration with broader project workflows. This ensures the design outputs from previous stages are compatible with industry-standard BIM platforms and ready for simulation, fabrication, and site operations.

Each slab unit is composed of multiple components—including cast-in-place top and bottom concrete layers, 3D-printed rib structures, embedded tendon ducts, temporary voids, and reinforcement meshes. These are modeled in Rhino and Grasshopper, then transferred into Revit using the Rhino.Inside.Revit framework. Geometries are classified into appropriate Revit categories (e.g., floors, structural framing, and generic models) with embedded metadata as Revit instance parameters to support scheduling and clash detection.

The final output of stage D7 is a complete, structured BIM model comprising all prefabricated slab units, their internal components, and associated metadata. It serves as the authoritative digital representation of the system and links design, fabrication, and erection phases into a unified digital workflow.

5 Simulation pipeline

5.1 S1 – Model conversion

The Simulation Pipeline (Figure 3) begins by converting the detailed BIM output from D7 into simulation-ready assets for robotic and construction environment modeling. Each slab unit is exported from BIM (Revit or Rhino3D) as a triangulated mesh in STL or OBJ format. Supporting elements such as columns, walls, and floors are also extracted to define the physical context for virtual erection. Prior to export, a mesh simplification step is applied using tools like MeshLab or open3D to reduce polygon count, improving simulation efficiency without altering the slab's global shape. Geometry export is automated via scripting, while the assignment of lifting points and tendon tags currently requires manual annotation in Grasshopper or Revit. These assets are converted into Gazebo-compatible formats: static objects are encoded in Simulation Description Format (SDF), while movable objects (e.g., slab units subject to lifting) are converted into Unified Robot Description Format (URDF) with optional joint definitions and inertial tags.

Embedded metadata (e.g., element ID, volume, mass) are extracted automatically from the BIM model. Additional metadata, such as tendon duct alignment or grouting zone tags, can also be embedded using Gazebo plugins for later validation. The result is a library of simulation-ready models reflecting the BIM geometry, but optimized for robot control and physical simulation in Gazebo. This library is used in later stages (S4–S6) for kinematics, task execution, and collision-based validation.

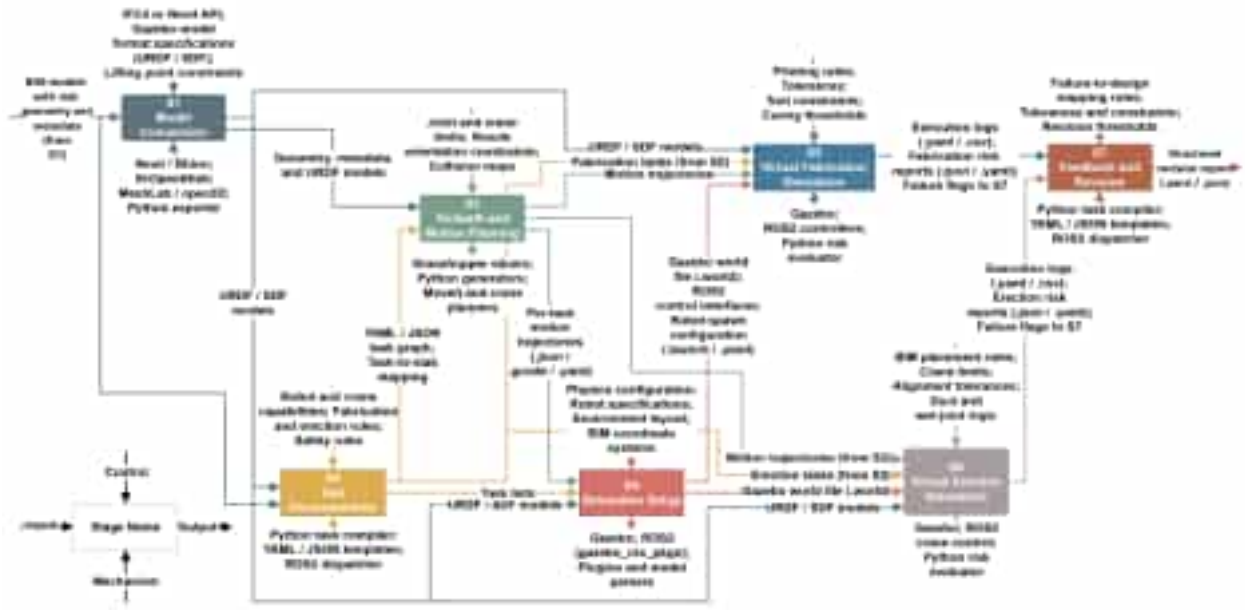


Figure 3. Simulation pipeline using IDEF0 notation

5.2 S2 – Task decomposition and sequencing

This stage defines a structured execution plan for fabrication and erection, decomposing the overall slab construction process into discrete, simulation-ready tasks. Based on the BIM-derived models and metadata from S1, each slab unit is associated with a specific sequence of robotic or crane-based operations. There are eight fabrication steps, three storage and transportation steps, and six erection steps. Most geometry-to-task mappings—such as identifying target slab units and assigning agents based on lifting points or print zones—are automatically inferred from BIM metadata or spatial relationships, while certain parameters (e.g., pour speed or human involvement) still require manual annotation at this stage.

Each task is encoded in a machine-readable YAML or JSON format that specifies: 1) Task type (e.g., print_ribs, place_ducts) 2) Target slab unit or geometry 3) Assigned agent (gantry, robotic arm, crane) 4) Required inputs (e.g., toolpath file, lifting anchor) 5) Temporal dependencies (e.g., curing complete before lifting).

This results in a Directed Acyclic Graph (DAG) of tasks that enables logical sequencing and parallelization. The task graph can be parsed and executed by ROS2 nodes, providing a control interface between planning and simulation. Task definitions can also be linked to BIM elements for 4D scheduling or progress monitoring. Where available, task hints embedded in BIM models—such as lifting zones or build order tags—are incorporated. The generated graph serves as the input for motion planning (S3) and simulation execution (S5–S6).

5.3 S3 – Toolpath and motion planning

This stage translates each task from the construction graph (S2) into executable robot motion. For fabrication, geometric inputs from S1 are automatically sliced into layer-based toolpaths, exported as .json or .gcode files. Each path includes 3D positions, print speed, nozzle orientation, and pause points. Ducts are inserted using a "pause-and-insert-from-above" method. After printing several layers of the rib, ducts are lowered vertically into grooves and sealed by continuing the print (conceptually consistent with in-process insertion and pause-and-insert strategies reported in [17, 3]). This method avoids lateral interference and fits well with robot-assisted insertion.

For each path, ROS2 MoveIt performs inverse kinematics and collision checks. Gantries and arms are automatically evaluated for reachability, joint limits, and obstacle avoidance. Unreachable or unsafe segments are flagged for revision. Tasks like sand filling and mesh placement are handled as partially scripted, parametric sweeps or pick-and-place motions, with manual definition of start/end poses.

Erection planning uses simplified crane kinematics. Given start/end poses from S2, safe lifting paths are automatically generated considering swing limits and clearance from obstacles. The planner ensures loads stay within the crane's envelope and avoids sharp accelerations.

All results are saved in ROS2-compatible formats (.json, .yaml), indexed by unit ID and task name. These trajectories are passed to the simulation system in S5–S6, forming the robotic execution layer.

5.4 S4 – Simulation environment setup in Gazebo

This stage builds a simulation-ready digital twin in Gazebo that replicates both the fabrication and site environments for the slab system. The environment integrates geometric models (from S1), robotic tasks (from S2), and planned motion trajectories (from S3), enabling time-sequenced motion playback, real-time collision checking, and dynamic constructability validation. The simulation environment comprises two zones: 1) a factory layout containing the printing gantry, formwork, and manipulators, and 2) a site layout including BIM-derived structural supports, the crane system, and positioning frames.

Each object is instantiated from its SDF or URDF model with its physics properties. ROS2 control plugins are attached to each robot to enable real-time command execution. For instance, the gantry receives Cartesian print trajectories; the crane interprets lifting paths and winch cable lengths; and all actions are synchronized via ROS2 messages. Slab units spawn with embedded metadata to support accurate tracking and placement.

Environmental dynamics such as gravity, surface contact friction, and optional wind disturbances are enabled. The goal is to achieve realistic behaviour of printed structures, suspended loads, and placement operations. Contact modeling between slab units and supports allows validation of alignment and load transfer.

By the end of S4, a complete virtual environment is assembled, with all elements positioned in BIM-coherent coordinates. This environment forms the runtime context for executing and analyzing fabrication and erection tasks in S5 and S6.

5.5 S5 – Fabrication simulation

This stage executes each prefabrication task in the virtual environment built in S4, validating robotic execution under physical constraints. Using ROS 2 controllers, each robot agent performs the motion trajectories generated in S3. Gazebo simulates kinematics, contact, and collisions based on physical parameters, while task scheduling and execution follow the construction graph from S2.

The fabrication sequence includes: bottom mesh placement, bottom-layer concrete casting, rib printing with duct insertion pauses, sand filling, top mesh placement, and top-layer casting. The rib printing process follows layered paths; at designated heights, the print head pauses and a duct insertion motion is executed.

A custom Risk Evaluator monitors execution in real-time, flagging issues such as nozzle collisions, reachability violations, or curing time constraints based on task timestamps and predefined rule-based thresholds (e.g., minimum clearance distance, reach limits, and time intervals). All fabrication actions are logged with timestamps, unit IDs, task names, and pass/fail tags. Warnings from the Risk Evaluator are saved in structured YAML/JSON reports and linked to slab unit metadata. If a failure is detected the system records which unit and stage requires revision. These reports are forwarded to stage S7 for design iteration.

5.6 S6 – Erection simulation

This stage simulates the robotic and crane-assisted assembly of prefabricated slab units in a virtual construction site, evaluating geometric feasibility, placement accuracy, and interface alignment. Using the digital twin environment established in S4, each slab unit is lifted, positioned, and logically integrated into the structural frame.

A crane model, a URDF with boom joints and winch mechanism, is commanded via ROS2 controllers to follow the lifting trajectories defined in S3. Each unit is lifted from the ground or factory bed, transported across the site, and lowered onto structural supports derived from the BIM geometry. Upon placement, the system verifies whether the unit rests stably on pre-positioned temporary supports. Contact is evaluated using bounding box intersections and pose thresholds (e.g., allowable tilt angle and support coverage ratio). If a corner lacks adequate support or the unit exceeds tilt limits, the Risk Evaluator flags it for temporary shoring.

Adjacent slabs undergo geometric validation for duct alignment and wet joint feasibility. Co-oriented duct outlets within tolerance (e.g., positional and angular deviation limits) enable logical tendon threading, while misalignments trigger a redesign flag. Post-tensioning is abstracted as a logic state rather than a physical simulation, while its structural effects are incorporated in the design stage through topology optimization and structural analysis, set only when valid alignment and closeable joint gaps are confirmed.

All outcomes—placement success, duct alignment, propping needs, and tendon connection feasibility—are logged per slab unit in structured YAML/JSON records. These records are passed to the feedback module (S7), enabling targeted revision in design.

5.7 S7 – Feedback loop and design revision

This stage closes the simulation-design loop by interpreting the results from virtual fabrication (S5) and erection (S6), and routing issues back to the appropriate stages of the design pipeline (D3–D6) for revision. The goal is to enable targeted design refinement without restarting the entire process.

6 Discussion and conclusion

This paper proposed a dual-pipeline method for the design and simulation of 3D-printed slab systems with

embedded post-tensioning, integrating a D-stage design pipeline (D1–D7) with a S-stage simulation pipeline (S1–S7). Unlike prior work focusing on small-scale printing or isolated robotic tasks, our pipeline supports full-scale prefabricated units with embedded structural logic, and accounts for site-level erection and post-tensioning.

A key contribution is the integration of rule-based feedback between simulation and design, enabling early-stage decisions to be evaluated against real-world constraints. The structured task graph format and use of ROS2+Gazebo ensure that both fabrication and erection steps can be virtually tested before deployment.

Limitations include the simplification of material behaviors and the lack of full automation in geometry-to-task mapping in the simulation stage . Future work will extend risk evaluation to cover time-dependent rheology and incorporate real-world feedback from physical trials. The proposed pipeline serves as a foundation for digital twin–enabled robotic prefabrication in building-scale applications.

Acknowledgements

This work was funded by the European Union under Grant Agreement No. 101162318 (AM2PM).

References

[1] O. Amir and E. Shakour. Simultaneous shape and topology optimization of prestressed concrete beams. *Structural and Multidisciplinary Optimization*, 57(5):1831–1843, 2018. Publisher: Springer.

[2] O. Amir and E. Shakour. Topology optimization of post-tensioned concrete beams. In *Proceedings of IASS Annual Symposia*, volume 2018, pages 1–8. International Association for Shell and Spatial Structures (IASS), 2018. Issue: 19.

[3] G. Bai, L. Wang, Z. Li, Y. Qu, and G. Ma. Integrating prestress into 3d printed ultra-high performance concrete composite beams for superior flexural performance. *Engineering Structures*, 343:121255, 2025.

[4] D. Dong, Y. Zou, H. Pan, G. Zhou, Y. Feng, and Y. Tang. Dfma-oriented modular and parametric design and secondary splitting of vertical pc components. *Scientific Reports*, 13(1):3457, 2023.

[5] S. M. Kargar, B. Yordanov, C. Harvey, and A. Asadipour. Emerging trends in realistic robotic simulations: A comprehensive systematic literature review. *IEEE Access*, 12:191264–191287, 2024. ISSN 2169-3536.

[6] S. Kim, M. Peavy, P.-C. Huang, and K. Kim. Development of bim-integrated construction robot task planning and simulation system. *Automation in Construction*, 127:103720, July 2021.

[7] M. Mogra. Design optimization of building elements for manufacturing using 3D concrete printing. Master's thesis, Technion - Israel Institute of Technology, 2023.

[8] Open Source Robotics Foundation. Gazebo simulator, 2024. URL `https://gazebosim.org/`.

[9] S. Raza, M. Sakha, Z. Hassan, B. Manshadi, X. Wang, H. Fan, B. Dillenburger, and M. Shahverdi. Flexural behavior of stay-in-place load-bearing 3d-printed concrete formwork for ribbed slabs. *Engineering Structures*, 338:120531, 2025.

[10] R. Sacks, G. Lee, L. Burdi, and M. Bolpagni. *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors and Facility Managers*. John Wiley and Sons, Hoboken, NJ, 4th edition, 2025. ISBN 978-1-394-22222-3.

[11] T. A. M. Salet, Z. Y. Ahmed, F. P. Bos, and H. L. M. Laagland. Design of a 3D printed concrete bridge by testing. *Virtual and Physical Prototyping*, 13(3): 222–236, 2018.

[12] R. Sedghi, K. Rashidi, and M. Hojati. Large-scale 3d wall printing: From concept to reality. *Automation in Construction*, 159:105255, 2024.

[13] E. Shakour and O. Amir. An integrated computational framework for structural optimization of prestressed concrete structures with 3d-printed formwork. *engrXiv*, 2026.

[14] E. Shakour, A. Shaked, and O. Amir. Topology and shape optimization of 3D prestressed concrete structures. *Engineering Structures*, 321:118936, 2024.

[15] M. Slepicka, S. Vilgertshofer, and A. Borrmann. Fabrication information modeling: Interfacing building information modeling with digital fabrication. *Construction Robotics*, 6(2):87–99, June 2022. ISSN 2509-8780.

[16] G. Vantyghem, W. D. Corte, E. Shakour, and O. Amir. 3D printing of a post-tensioned concrete girder designed by topology optimization. *Automation in Construction*, 112:103084, 2020. ISSN 0926-5805.

[17] Z. Wu, A. M. Memari, and J. P. Duarte. State of the art review of reinforcement strategies and technologies for 3d printing of concrete. *Energies*, 15(1):360, 2022.

[18] S. Zhang, Y. Tang, Y. Zou, H. Yang, Y. Chen, and J. Liang. Optimization of architectural design and construction with integrated bim and plm methodologies. *Scientific Reports*, 14(1):26153, 2024.

[19] Y. Zhi, H. Chai, T. Teng, and M. Akbarzadeh. Automated toolpath design of 3d concrete printing structural components. *Additive Manufacturing*, 100: 104662, 2025.

A Hybrid Retrieval-Augmented Generation Driven Multi-Agent System for Quality Control in Off-Site Construction

Fanfan Meng[1], Mi Pan[1] and Wei Pan[2]

[1]Department of Civil and Environmental Engineering, University of Macau, Macao SAR, China
[2]Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
yc47936@um.edu.mo, mipan@um.edu.mo, wpan@hku.hk

Abstract –

Quality control is critical for off-site construction, but it still depends heavily on individual expertise and manual inspection. The management and coordination of heterogeneous and fragmented knowledge among the stakeholders involved in quality control of off-site construction remains challenging. To address this knowledge gap, this paper proposes QuosRA, a novel framework for **qu**ality control in **o**ff-**s**ite construction empowered by a hybrid **r**etrieval-augmented generation (RAG)-driven multi-**a**gent system. In particular, QuosRA breaks down multi-disciplinary quality control tasks into role-based subtasks handled by dedicated large language model agents with specific expertise. Each agent is assigned a particular role, expertise, and abilities for automatic quality compliance checking and quality control task guidance. The performance of QuosRA is evaluated using 30 quality inspection cases and 10 task guidance queries, compared against single-agent hybrid RAG and unstructured multi-agent system baselines. The results indicate that structured task coordination in QuosRA achieves better performance in both quality compliance checking (retrieval precision of 82.4% and determination accuracy of 93.3%) and quality control task guidance (retrieval precision of 92.9% and high-quality response score of 80.0%). Overall, QuosRA improves the efficiency and consistency of quality control workflows in off-site construction projects, offering an effective pathway toward scalable and knowledge-driven quality management in the construction industry.

Keywords –

Off-Site Construction; Quality Control; Large Language Model; Multi-Agent System; Retrieval Augmented Generation

1 Introduction

Off-site construction (OSC) is an innovative construction method that refers to the manufacturing and pre-assembly of building elements, components, or modules before their installation at their final locations [1]. It has attracted growing attention with its potential benefits in enhanced productivity, shortened construction periods, improved safety, and minimization of construction waste [2]. Quality control (QC) refers to a set of measures and procedures implemented to ensure that OSC complies with predefined quality standards and satisfies relevant regulatory requirements [3], involving quality compliance checking, quality data collection, and quality issue remediation. In particular, the widespread production of prefabricated components imposes stringent QC requirements to reduce potential quality defects before delivery to the construction site. This highlights the crucial importance of precisely monitoring and controlling the quality of prefabricated components during manufacturing.

Despite its importance, existing research on implementing QC in OSC, spanning design to construction activities, remains inadequate [3], and the current practice is still challenging. QC is highly specialized and complex, wherein quality knowledge is diverse and scattered among different stakeholders and markets [4]. OSC quality inspections highly depend on cross-disciplinary participants with varying knowledge, experiences, and responsibilities. The vast and intricate body of regulations and standards involved across regions further exhibits multi-role practitioners to track, identify, and enforce quality requirements for OSC projects [5]. Besides, the shortage and frequent turnover of skilled QC personnel greatly hinder knowledge sharing and circulation within teams, resulting in limited collective understanding and overreliance on individual capabilities in quality diagnoses [6]. Ultimately, existing quality information systems remain constrained by static rule-based mechanisms and offer limited capacity to interpret diverse QC contexts, integrate fragmented knowledge, or support the interactive decision processes required in complex OSC environments [7].

Recent advancements in LLMs, retrieval-augmented generation (RAG), and multi-agent systems (MAS) provide a promising solution to address these complexities. LLMs offer strong natural language

interpretation and context-aware reasoning capabilities for knowledge-intensive tasks in construction, while MAS coordinates multiple intelligent agents to manage complex and distributed workflows [8]. Meanwhile, RAG synergistically merges intrinsic knowledge of LLMs with dynamic and domain-specific information, enhancing the accuracy and credibility of response generation [9]. The integration of LLM-based semantic understanding, RAG-driven knowledge retrieval, and MAS-supported collaborative orchestration could automate highly specialized tasks that demand nuanced comprehension and flexible coordination, such as QC in OSC. This integrated paradigm moves beyond the limitations of traditional agent systems in construction [10], enabling richer interpretation, trustworthy knowledge reuse, and adaptive multi-role coordination.

However, how the complementary strengths of LLMs, RAG, and MAS can be integrated into a coherent and domain-oriented framework for QC in OSC is still underexplored. Such integration is essential for managing fragmented knowledge, reducing inspection inconsistencies, and coordinating multiple roles across distributed off-site settings. To fill this knowledge gap, this paper aims to propose a novel framework for **qu**ality control in **o**ff-**s**ite construction empowered by a hybrid **R**AG-driven multi-**a**gent system, named QuosRA, that supports dynamic interaction, collaborative analysis, and systematic decision-making of QC-related tasks in OSC. The feasibility and efficiency of QuosRA are also demonstrated by comparing with the single-agent hybrid RAG and native MAS baselines.

2 Literature Review

2.1 Quality Control in Off-site Construction

Quality issues have been identified as one of the most critical factors affecting the implementation and development of OSC. Yin et al. [11] emphasized that OSC projects require increased emphasis on quality control measures to address prevalent quality issues. However, QC in OSC projects predominantly relies on traditional construction management guidelines, with relatively slow progress in knowledge and technological advancements [3].

Studies have explored the applications of advanced informatics technologies for quality management in OSC projects, such as building information modeling [12] and ontology [13]. However, current approaches still face limitations that result in fragmented and labor-intensive workflows, impeding effective domain knowledge capture, retrieval, and reuse. In particular, research rarely provides a systematic workflow for QC in OSC to support efficient decision-making, cross-stakeholder collaboration, and consistent quality assurance.

2.2 LLM-based Multi-agent Applications

LLMs demonstrate great potential for knowledge-intensive tasks without labor-intensive and time-consuming efforts in knowledge representation, which have been increasingly applied in construction, such as construction compliance checking and automated reporting [14]. RAG has been used to further strengthen LLM performance by mitigating hallucinations in domain-specific tasks. In the construction domain, RAG contributes to complex information retrieval and question answering in domain document parsing and regulation inquiry tasks. He et al. [7] designed an improved RAG framework for regulation information extraction in construction quality checks. Wan et al. [15] explored a hybrid RAG strategy by incorporating knowledge graph metadata into vector retrieval, supporting design for additive manufacturing tasks.

However, these RAG methods remain reactive and primarily respond to the given query, offering limited capability for complex task breakdown and multi-step task handling. Recent attention has concentrated on LLM-based MAS, since it allows agents with different roles to integrate diverse inputs, initiate internal reasoning, and coordinate planned, goal-oriented actions rather than passively reacting to user queries [16]. Each agent in the multi-agent system has been assigned a specific responsibility and specializes in a particular task. This highlights a transition from RAG-based reactive retrieval to MAS-enabled proactive orchestration, offering a more autonomous and proactive paradigm for complex and dynamic construction scenarios.

Currently, LLM-based multi-agent systems have gained increasing attention in construction, mainly on automated building design [17], building energy analysis [18], and human-machine collaboration [19]. However, few studies have explored how to employ RAG-driven reliable retrieval and deeper reasoning to enhance LLM-based MAS in handling domain-specific and knowledge-intensive tasks, such as QC in OSC projects. Meanwhile, there is still a lack of a well-designed agentic workflow for QC in OSC that enables multiple stakeholders to achieve more accurate knowledge retrieval, trustworthy compliance checking, coherent guidance generation, and coordinated decision-making.

3 Methodology

This study proposes a novel LLM-based multi-agent system for QC in OSC, named QuosRA, that integrates hybrid RAG with a centralized orchestrator architecture, as shown in Figure 1. Within QuosRA, QC-related tasks are explicitly broken down into specialized subtasks, such as query analysis, knowledge retrieval, and output evaluation, with responsibilities assigned to dedicated agents driven by LLMs tailored to distinct subtasks.

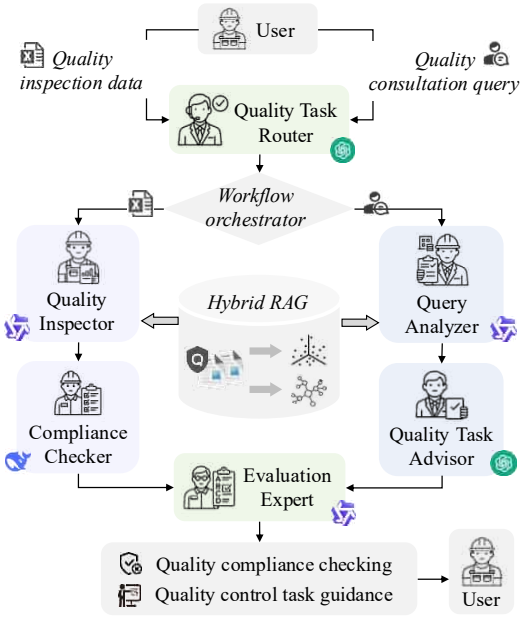


Figure 1. Conceptual framework of QuosRA

QuosRA is organized around a central workflow orchestrator that governs the overall interaction workflow. When a user query is input, Quality Task Router determines and dispatches it to the corresponding processing workflow. Quality inspection data recorded in Excel files are directed to the left-side workflow, where Quality Inspector processes data and searches requirements for Compliance Checker to check compliance. Task guidance queries are routed to the right-side workflow, where Query Analyzer parses user intent and retrieves interpretations for the downstream Quality Task Advisor to generate context-aware guidance. Quality Inspector and Query Analyzer are supported by a shared domain-specific knowledge base empowered by a hybrid RAG [20]. Evaluation Expert further reviews the synthesized results for both workflows to provide coherent and reliable QC assistants.

This centralized-orchestrated architecture of QuosRA adopts the hybridization of hierarchical and decentralized mechanisms, which has been identified as a promising approach for achieving scalability and robustness in MAS [21]. The central orchestrator ensures stable workflow sequencing, whereas the domain-specific agents perform specialized subtasks with local autonomy. It enables a balance between centralized oversight and decentralized adaptability for complex QC-related tasks.

3.1 Hybrid RAG for Knowledge Retrieval and Reasoning

Domain-specific knowledge of QC tasks in QuosRA is provided by the hybrid RAG [20], which integrates vector-based search for semantic matching with graph-based retrieval for explicit relational reasoning, aiming to enhance retrieval accuracy, trustworthiness, and reliability. The hybrid RAG pipeline is illustrated in Figure 2 and briefly elaborated below. More details can be found in previous work [20].

Quality-related documents are preprocessed and segmented into text chunks. Entities and relationships are extracted from text chunks and stored as node and edge chunks. Based on this representation, subgraphs are constructed and subsequently merged into a unified knowledge graph. Simultaneously, both node and edge chunks are encoded into vector embeddings to capture semantic features. The knowledge graph is further consolidated by assigning importance weights to nodes and edges, where edge weights are determined using frequency-based scores ($Weight_{edge}$) and node weights are computed using PageRank scores (PR_{node}).

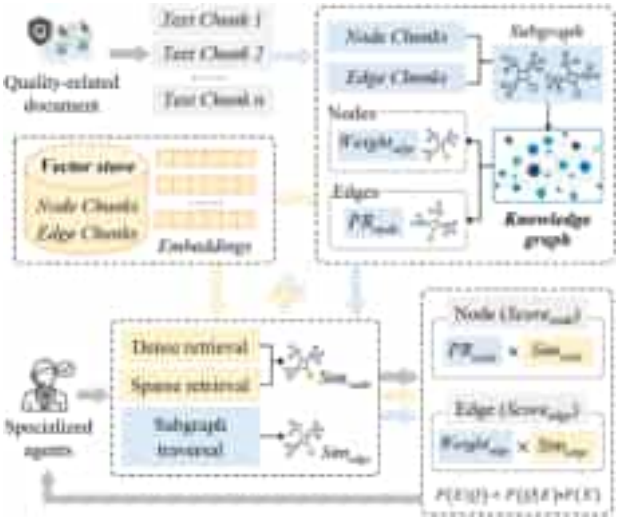


Figure 2. Procedural pipeline of the hybrid RAG

Following the role constraints encoded in the system message, Quality Inspector and Query Analyzer perform external knowledge retrieval and subsequent reasoning by invoking the hybrid RAG. The retrieval process integrates dense and sparse retrieval, as well as subgraph traversal. Specifically, dense and sparse retrieval are employed to support embedding-level and keyword-level search for entity matching. Dense retrieval measures semantic similarity using cosine similarity ($Socre_D$), while sparse retrieval adopts a Robertson-Sparck Jones Inverse Document Frequency (RSJ-IDF)-based weighting scheme without term frequency, with named entity recognition (NER) and part-of-speech (POS) tagging information to prioritize related entities ($Socre_S$). The similarity score of nodes is integrated based on the relative score fusion (RSF) (see Equation 1). In addition, subgraph traversal is employed to enable multi-hop reasoning over the knowledge graph (see Equation 2), where h denotes the hop distance.

$$Sim_{node} = 0.7Score_D \times 0.3Score_S \quad (1)$$

$$Sim_{edge} = \frac{Sim_{node}}{h+2} \quad (2)$$

To align semantic similarity with graph structural importance, a fusion scoring strategy is designed inspired by Bayes' theorem (see Equations 3-4). Driven by hybrid RAG, agents retrieve and prioritize QC domain knowledge under task-specific contexts to support the execution of their assigned tasks.

$$Score_{node} = Sim_{node} \times PR_{node} \quad (3)$$

$$Score_{edge} = Sim_{edge} \times Weight_{edge} \quad (4)$$

3.2 Agent Responsibilities Analysis

QuosRA is a distributed intelligence mechanism where specialized agents handle distinct QC subtasks in OSC, coordinated through a central router. QuosRA consists of six key agents, each with one particular role, expertise, and specialized abilities:

- ***Quality Task Router*** serves as the initial interface with users and processes both Excel and textual inputs. This agent analyzes queries, classifies them into compliance checking or guidance tasks, and routes them to the appropriate workflow handled by specialized agents.
- ***Quality Inspector*** processes structured quality inspection data recorded in Excel files and extracts inspection items. This agent is allowed to retrieve the corresponding quality requirements associated with each inspection item through the hybrid RAG. The responsibility is limited to focusing on mapping inspection items to relevant regulatory requirements.
- ***Compliance Checker*** performs quality compliance determinations based on processed inspection data and quality requirements provided by the upstream Quality Inspector. To reduce the spread of potential uncertainties, this agent only conducts compliance checking reasoning without extra retrieval behaviors.
- ***Query Analyzer*** rewrites user queries of task guidance and extracts domain-specific entities as keywords for retrieval. Based on query analyses, the agent invokes the hybrid RAG to retrieve supporting quality knowledge for query intent interpretation, which is passed downstream to guidance generation.
- ***Quality Task Advisor*** receives retrieved knowledge from Quality Analyzer to further generate coherent task guidance, which can include the clarification of quality issues, preliminary cause analysis, and recommended remediation measures, synthesized with explicit consideration of applicable standards, engineering practices, and project-specific contexts.
- ***Evaluation Expert*** consolidates and reviews the outputs produced by Quality Task Advisor and compliance checker agent to ensure accuracy and consistency of the response. Based on the final verification, this agent sends validated compliance results and QC task guidance to the user.

Through coordinated interactions among specialized agents, QuosRA adopts clear responsibility divisions and constrained interactions, which help mitigate the spread of potential uncertainties across subtasks, enabling more reliable and consistent QC in OSC.

3.3 Model Selection

In this study, agents within QuosRA are powered by three open-source medium-sized LLMs, selected based on their complementary strengths. Each model is chosen to align with subtask requirements of specialized agents, such as knowledge retrieval, reasoning, or structure output generation. Table 1 summarizes the mapping between selected open-source LLMs and agents, highlighting capability alignment.

Table 1 Model selection for agents in QuosRA

Open-source LLM	Agent	Capability	References
GPT-OSS-20b	Quality Task Router Quality Task Advisor	Task abstraction capability; Instruction-following robustness High-quality synthesis; Logically long-form coherence; Well-structured expression	[20, 22]
Qwen3-14b	Quality Inspector Quality Analyzer Evaluation Expert	Information retrieval stability Conservative yet rigorous style; Low hallucination tendency	[20, 23]
Deepseekr1-14b	Compliance Checker	Strong logical reasoning capability	[20, 24]

GPT-OSS-20B is chosen to drive Quality Task Router and Quality Task Advisor. Previous studies have shown that GPT-OSS exhibits strong logical coherence, concise expression, and clear structural organization in output generation [20, 22], which is suitable for task orchestration and guidance synthesis grounded in reliably retrieved evidence. Qwen3-14B powers Quality Inspector, Quality Analyzer, and Evaluation Expert, leveraging its reliable and stable knowledge retrieval capabilities. Qwen3-14B has been proven to show a

cautious yet rigorous and robust style in knowledge retrieval and reasoning performance [20], making it particularly effective for retrieval-oriented tasks [23]. Moreover, the low hallucination tendency and balanced performance of this model ensure reliability and consistency, which are critical for comprehensive evaluation tasks. Compliance Checker is powered by DeepSeek-R1-14B, chosen for its robust logical reasoning capability [24]. Although DeepSeek-R1 excels in structured reasoning tasks, it exhibits a higher risk of hallucination compared to more conservative models [20]. To reduce the risk of uncertainty spread and improve overall trustworthiness, Compliance Checker is strictly restricted from any knowledge retrieval behaviors, relying exclusively on inspection data and pre-retrieved requirements provided by Quality Inspector.

4 Experiments

The proposed QuosRA is validated considering quality compliance checking and quality control task guidance. QuosRA builds upon the capabilities of the hybrid RAG that have been shown in prior work to provide reliable retrieval and reasoning performance [20]. Accordingly, the validation focuses on assessing the efficiency of the structured workflow design, as well as the feasibility of explicit responsibility division and constrained agent interactions in QC of OSC, rather than pursuing further improvements in domain knowledge retrieval or reasoning performance.

4.1 Experimental Design

The evaluation is conducted based on an OSC quality reference integrating both regulatory documents and practical field knowledge, including twelve regulations and standards issued by mainland China, Hong Kong SAR, and Macao SAR, and five industry guidelines and confidential project documents from industry practitioners to capture procedural knowledge and field practices in precast concrete production.

Baseline configurations are designed to isolate the impact of structured subtask coordination in QuosRA. Specifically, two baselines are applied to validate the overall performance of QuosRA. The first is a single-agent hybrid RAG powered by Qwen3-14B, which reactively responds to the given query for all QC tasks without agent decomposition. The second baseline is a native MAS with an unstructured workflow, in which both the hybrid RAG and all agents are powered by Qwen3-14B. The agent roles set in this baseline are identical to those defined in QuosRA. However, role-specific responsibilities are not enforced through explicit task decomposition or coordination mechanisms. All agents are allowed to directly access the hybrid RAG and independently process the complete QC-related task in parallel, without interaction constraints or subtask assignments. This baseline is designed to examine the impact of structured subtask decomposition and constrained coordination in QuosRA under equivalent agent and retrieval configurations. Furthermore, the LLM configurations of specialized agents in QuosRA are illustrated in Table 1, where the hybrid RAG setting is identical to baseline methods. This experimental design ensures consistent retrieval behavior across all settings, enabling a focused examination of the impact of role specialization and structured coordination.

4.2 Evaluation Task and Metrics

To fully capture real-world tasks faced by quality inspectors and decision makers in OSC projects, two types of QC-related tasks are designed.

Quality compliance checking involves determining compliance with regulatory requirements. For instance, a measured deviation of 7 mm indicates that the longitudinal reinforcement spacing in precast concrete column satisfies the ±10 mm allowable limit defined in clause 4.3.8 of DBJT 15 171-2019 and is therefore determined as compliant. The outputs are required to both extract relevant normative clauses and reasonably judge correctly. This evaluates the accuracy of compliance determinations and the relevance of the supporting references to the ground-truth, which were manually annotated to include specific normative clauses and guideline provisions. An Excel file recording 30 inspection items was provided as input, covering inspection objects, inspection points, measured data, inspectors, and inspection time. For each inspection item, QuosRA and tested baselines are asked to provide the corresponding regulatory requirements and explicit compliance determination. Two metrics are adopted to reflect the reliability of supporting evidence and the correctness of compliance judgment. ***Retrieval precision (RP)*** measures the proportion of given references that are actually correct and relevant among N inspection items, which is defined in Equation (5).

$$Retrieval\ precision = \frac{1}{N}\sum_{i=1}^{N}\frac{|S_i \cap R_i|}{S_i} \tag{5}$$

where S_i denotes the set of references retrieved by the tested method for inspection item i, and R_i refers to the ground-truth set. ***Determination accuracy (DA)*** measures the proportion of inspection items for which the correct compliance determination is given.

Quality control task guidance, which represents complex queries requiring procedural or planning advice, such as "*what is the definition of cellular defects, and how to deal with them*?" Such task involves analyzing query intention and rewriting the retrieved knowledge according to the indications made by the user query. This task emphasizes professional relevance, response

coherence, and content completeness. A total of 10 representative QC task guidance queries are designed based on common quality concerns in OSC projects [20], such as defect clarification, regulatory interpretation, and measure suggestions. Two metrics are used. ***Retrieval precision (RP)*** is adopted as the same metric used in quality compliance checking to explore the reliability of supporting references. ***High-quality response score (HQS)*** evaluates the overall quality of the generated guidance, emphasizing how effectively retrieved knowledge is synthesized into actionable and coherent guidance. Specifically, this metric jointly assesses the response coherence and content completeness of outputs based on five criteria: (1) whether the quality issue is explicitly identified; (2) whether at least one relevant regulatory clause is clearly referenced; (3) whether clear and actionable measures are provided; (4) whether the guidance is logically consistent without obvious contradictions; and (5) whether the response sufficiently addresses the key aspects of the query without major omissions. A response is considered high-quality if it satisfies at least four criteria. Over a set of R responses, HQS is computed as:

$$HQS = \frac{1}{R}\sum_{r=1}^{R} 1\left(\sum_{j=1}^{5} c_{r,j} \geq 4\right) \quad c_{r,j}\epsilon\{0,1\} \qquad (6)$$

where r indexes the responses and $c_{r,j}$ indicates whether the j-th criterion is satisfied for response r (1 if satisfied, 0 otherwise).

5 Results and Analyses

For quality compliance checking, Table 2 illustrates the results of the retrieval precision and determination accuracy across three tested methods. Experimental results demonstrate that QuosRA exhibits a modest yet systematic improvement over two baseline methods, underscoring the effectiveness of its role-specialized architecture and constrained agent interactions. Specifically, compared with the hybrid RAG baseline (RP 80.6%, DA 90.0%), QuosRA achieves a moderate increase in RP (82.4%) and a more noticeable improvement in DA (93.3%). The limited enhancement in RP suggests that the primary contributions of QuosRA are not centered on greatly improving retrieval precision, but on stabilizing the downstream determination process once relevant regulatory knowledge has been retrieved. This observation aligns with the design intent of QuosRA and highlights its ability to promote greater stability and reliability in decision-making, particularly in quality compliance scenarios that require precise interpretation and integration of regulatory knowledge. For example, when retrieving the allowable deviation of the centerline position of reserved holes in concrete prefabricated slabs, Quality Inspector passes validated outputs (no more than 5 mm) to Compliance Checker, which performs consistent and reliable determinations under the structured workflow. The explicit separation between retrieval and compliance checking stages facilitates a direct and dependable translation of accurate retrieval results into compliance decisions, highlighting the scaling principle that coordination architecture can stabilize model capabilities in complex logical tasks [25]. In addition, the comparatively lower performance of the native MAS (RP 76.32%, DA 86.67%) provides empirical evidence of the negative returns from unstructured coordination [25], which indicates that multi-agent interaction without explicit role separation and coordination constraints often introduces logical interference and inconsistent interpretation of retrieved references, thereby undermining overall compliance accuracy. In contrast, QuosRA exhibits a closer alignment between RP and DA, mitigating the influence of irrelevant or conflicting references and supporting more consistent determinations of retrieved requirements.

Table 2 Experimental results of compliance checking

Method	Description	RP	DA
Hybrid RAG	Single-agent hybrid RAG	80.6%	90.0%
Native MAS	Multi-agent hybrid RAG without role separation	76.3%	86.7%
QuosRA	Role-specialized Multi-agent hybrid RAG	82.4%	93.3%

For quality control task guidance, responses produced by QuosRA show stronger reliability and higher quality than the other two baselines. The results are illustrated in Table 3. For RP, QuosRA attains the highest score (92.9%), outperforming both the hybrid RAG (88.2%) and the native MAS (83.3%). This improvement indicates that the role-specialized workflow in QuosRA contributes to more accurate alignment between query intent and retrieved regulatory references. By comparison, the native MAS shows a modest degradation in RP compared to the hybrid RAG, suggesting that solely introducing multiple agents without structured workflow design and explicit role constraints may lead to coordination noise and overlapping knowledge retrieval. Regarding HQS, both the hybrid RAG and the native MAS achieve the same HQS of 60.0%. This indicates that without structured design in agentic workflows, baseline methods may struggle with fragmented and incoherent guidance outputs. QuosRA improves the HQS to 80.0%, reflecting a higher proportion of responses that provide clear problem identification, explicit regulatory grounding, actionable measures, and logically consistent guidance. Besides, a cross-metric comparison further reveals that although the hybrid RAG achieves relatively strong retrieval precision, its overall quality of generated guidance remains limited, highlighting a clear gap

between accurate retrieval and effective task-level synthesis. This observation further corroborates our previous findings in [20], which showed that Qwen3-14b exhibit more reliable and conservative retrieval behavior, yet tend to underperform in generating well-structured and coherent task guidance. By decoupling knowledge retrieval and guidance generation and assigning specialized agents powered by different LLMs, QuosRA demonstrates that a well-architected MAS can transform raw model capacity into superior task-oriented performance through strategic labor division [25]

Table 3 Experimental results of QC task guidance

Method	Description	RP	HQS
Hybrid RAG	Single-agent hybrid RAG	88.2%	60.0%
Native MAS	Multi-agent hybrid RAG without role separation	83.3%	60.0%
QuosRA	Role-specialized Multi-agent hybrid RAG	92.9%	80.0%

Regarding time efficiency, experimental results under both standalone (all-on-server) and distributed (laptop-server coordination) deployment modes indicate that QuosRA introduces a 2.7- to 3.2-fold increase in latency compared with the hybrid RAG baseline. Compared with the native MAS baseline, QuosRA achieves a 41%-49% reduction in execution time through explicit role separation and structured task decomposition. Considering that QC in OSC does not usually require strict real-time responses, the increased latency could be an acceptable trade-off for the improved reasoning reliability and structured decision support.

6 Discussion

This study proposes a centralized-orchestrated architecture multi-agent system tailored for QC in OSC. The evaluation results across QC-related tasks, involving quality compliance checking and quality control task guidance, demonstrate the effectiveness of QuosRA in supporting complex QC scenarios in OSC. Specifically, the clear responsibility divisions and constrained interactions in QuosRA mitigate interference and inconsistency in unconstrained multi-agent systems, leading to more accurate compliance checking and coherent, actionable QC guidance outputs. This design echoes the necessity of structured knowledge sharing among stakeholders in OSC [26] and offers a strategic solution to the pervasive challenges of labor shortages and expert scarcity. By mimicking a professional labor division where expert-level agents supervise technique-specific agents, QuosRA also provides a resilient management template for hierarchical workforce division, expert-level oversight, and task decomposition for QC in OSC. The cross-metric analysis further suggests that baseline methods are often limited in simultaneously achieving retrieval rigor, reasoning reliability, and generating structured responses. This finding further verifies that the lack of communication and cooperation may struggles in vital information sharing, leading to inefficiencies, conflicts, and knowledge misuse [21]. QuosRA exhibits promising potential to facilitate the transformation of retrieved knowledge into decision-relevant outputs, which demonstrates the strength of LLM-based multi-agent applications in cross-coordination and evaluation for more accurate, reliable, and robust solutions [17].

Practically, QuosRA enables non-experts to conduct QC-related tasks through natural language interactions without requiring expertise in regulatory coding or rule formalization, allowing QC to be implemented in an automated and consistent manner. Natural language can be used to obtain domain knowledge, check compliance, and access expert-level guidance in OSC. Although these interactions can also be implemented through RAG pipelines [6, 15, 20], such approaches remain reactive and limited in capability for proactive task breakdown and subtask coordination. In this regard, this paper also contributes to a scalable paradigm to expand the application of RAG and LLM-based MAS into QC coordination in OSC, and can be generalized to other complex engineering scenarios.

7 Conclusions

This study proposes QuosRA, a novel framework for QC in OSC driven by a hybrid RAG-driven multi-agent system, and validates its feasibility and efficiency through quality compliance checking and quality control task guidance. Evaluation results demonstrate that QuosRA improves the reliability and logical consistency of outputs for relevant queries. This research contributes an agentic workflow that explicitly decomposes QC-related tasks into coordinated subtasks, supported by a centralized orchestration mechanism. QuosRA provides an effective solution to address the challenges in QC of OSC projects, containing fragmented knowledge utilization, high manual effort, and limited decision support. Overall, this study establishes a solid foundation for applying LLM-based MAS for knowledge-driven management tasks in construction. Future work will extend QuosRA toward a multimodal assistant by integrating vision-language models for QC scene-aware perception. Furthermore, comprehensive ablation studies and resource efficiency analysis will be conducted to evaluate the contributions of different agent roles and further optimize system performance across diverse deployment settings.

8 Acknowledgements

This work was funded by The Science and Technology Development Fund, Macau SAR (File no. 0101/2024/RIB2) and University of Macau (File no. SRG2023-00006-FST).

References

[1] C. Goodier, and A. Gibb. Future opportunities for off-site in the UK. *Construction Management and Economics*, 25(6):585-595, 2007.

[2] M. Pan, Y. Yang, Z. Zheng, and W. Pan. Artificial intelligence and robotics for prefabricated and modular construction: A systematic literature review. *Journal of Construction Engineering and Management*, 148(9):03122004, 2022.

[3] Y. Liu, F. Yao, Y. Ji, et al. Quality control for offsite construction: Review and future directions. *Journal of Construction Engineering and Management*, 148(8), 2022.

[4] Z. Ma, S. Cai, N. Mao, et al. Construction quality management based on a collaborative system using BIM and indoor positioning. *Automation in Construction*, 92:35-45, 2018.

[5] Y. Zhang, M. Pan, W. Pan, Y. Yang, and J. Wu. Enhancing modular construction supply chain: Drivers, opportunities, constraints, concerns, strategies, and measures. *Developments in the Built Environment*, 18:100408, 2024.

[6] R. Rachel, and K. Han. Automated model-based 3D scan planning for prefabricated building components. *Journal of Computing in Civil Engineering*, 37(2):04022058, 2023.

[7] C. He, W. He, M. Liu, et al. Enriched construction regulation inquiry responses: A hybrid search approach for large language models. *Journal of Management in Engineering*, 41:04025001, 2025.

[8] T. Guo, X. Chen, Y. Wang, R. et al. Large language model based multi-agents: A survey of progress and challenges. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 1-10, Jeju, Korea, 2024.

[9] Y. Gao, Y. Xiong, X. Gao, et al. Retrieval-augmented generation for large language models: A survey. *arXiv preprint*, 2(1):2312.10997, 2023.

[10] L. Xiang, Y. Tan, G. Shen, and X. Jin. Applications of multi-agent systems from the perspective of construction management: A literature review. *Engineering, Construction and Architectural Management*, 29(9): 3288-3310, 2022.

[11] X. Yin, H. Liu, Y. Chen, and M. Al-Hussein. Building information modelling for off-site construction: Review and future directions. *Automation in Construction*, 101:72-91, 2019.

[12] Y. Zhou, Z. Zheng, J. Lin, X. Lu. Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Computers in Industry*,142:103746, 2022.

[13] F. Meng, and X. Cao. An intelligent framework for rework risk identification in prefabricated construction processes based on compliance checking. *Engineering, Construction and Architectural Management*, 32(8):5066-5098, 2025.

[14] H. Pu, X. Yang, J. Li, and R. Guo. AutoRepo: A general framework for multimodal LLM-based automated construction reporting. *Expert Systems with Applications*, 255:124601, 2024.

[15] Y. Wan, Z. Chen, Y. Liu, et al. Empowering LLMs by hybrid retrieval-augmented generation for domain-centric Q&A in smart manufacturing. *Advanced Engineering Informatics*, 65:103212, 2025.

[16] S. Chen, Y. Liu, W. Han, W. Zhang, and T. Liu. A survey on LLM-based multi-agent system: Recent advances and new frontiers in application. *arXiv preprint*, 2412.17481, 2024.

[17] Y. Dong, Z. Zhan, Y. Hu, et al. AI BIM coordinator for non-expert interaction in building design using LLM-driven multi-agent systems. *Automation in Construction*, 180:106563, 2025.

[18] L. Zhang, V. Ford, Z. Chen, and J. Chen. Automatic building energy model development and debugging using large language models agentic workflow. *Energy and Buildings*, 327:115116, 2025.

[19] J. Liu, H. Li, C. Chai, K. Chen, and D. Wang. A LLM-informed multi-agent AI system for drone-based visual inspection for infrastructure. *Advanced Engineering Informatics*, 68:103643, 2025.

[20] F. Meng, and M. Pan. Advancing quality control in off-site construction with large language models enhanced by hybrid retrieval-augmented generation. *Advanced Engineering Informatics*, 71:104381, 2026.

[21] L. Sun, Y. Yang, Q. Duan, Shi, Y., et al. Multi-agent coordination across diverse applications: A survey. *arXiv preprint*, 2502.14743, 2025.

[22] S. Agarwal, L. Ahmad, J. Ai, et al. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint*, 2508. 10925, 2025.

[23] A. Yang, A. Li, B. Yang, B. et al. Qwen3 technical report. *arXiv preprint*, 2505.09388, 2025.

[24] J. Chua and O. Evans. Are DeepSeek R1 And Other Reasoning Models More Faithful? *arXiv preprint*, 2501.08156, 2025.

[25] Y. Kim, K. Gu, et al. Towards a Science of Scaling Agent Systems. *arXiv preprint*, 2411.11504, 2024.

[26] S. Ma, Z. Li, L. Li, et al. Exploring the effect of stakeholder relationship quality on technological innovation in off-site construction: the mediating role of the knowledge sharing. *Journal of Civil Engineering and Management*, 29(1):77-92, 2023.

Multi-granularity infrastructure-BIM object classification and decomposition for highway project delivery

Mengtian Yin[1], Junxiang Zhu[1], Xingbo Gong[2], Fengqiao Zhang[3], Yixiong Jing[1], and Ioannis Brilakis[1]

[1]Department of Engineering, University of Cambridge, United Kingdom
[2]Department of Civil and Environmental Engineering, The Hong Kong University Science and Technology, Hong Kong, China
[3]Department of Engineering Structures, Delft University of Technology, Netherlands

my424@cam.ac.uk, jz652@cam.ac.uk, xgongai@connect.ust.hk, F.Zhang-5@tudelft.nl, yj401@cam.ac.uk, ib340@cam.ac.uk

Abstract -

Highway project information delivery using building information modelling (BIM) often suffers from inefficiencies when converting 3D Industry Foundation Classes (IFC) models into network representations that conform to agency-specific asset data standards. These inefficiencies primarily arise from inconsistent object classification and mismatches in modelling granularity. To address this challenge, this paper proposes a novel method to automatically classify multi-granularity IFC BIM models and decompose coarse elements into finer components that satisfy asset information exchange requirements. A context-aware mesh neural network (CA-MNN) is developed to incorporate contextual geometric information, enabling more accurate prediction of face-level mesh semantics for IFC elements. Triangulated faces are then grouped based on connectivity, and predictions are smoothed using a majority voting strategy, enabling the decomposition of coarse elements into finer-grained asset models. The proposed method is evaluated using 6,058 IFC elements from real-world highway BIM models in the UK. The results demonstrate that the CA-MNN with smoothing achieves the highest classification accuracy of 86.38%. This study contributes a new BIM object classification and decomposition approach that supports the identification and segmentation of coarse IFC elements for infrastructure project delivery.

Keywords -

Building information modelling (BIM); project information delivery; highway infrastructure management; geometric deep learning

1 Introduction

Highway infrastructure systems comprise roads, bridges, tunnels, traffic signs, drainage, utilities, lighting, and road restraint systems [1]. Effective information delivery is essential for highway construction and reconstruction projects, as reliable asset data strongly influence decisions in subsequent operation and maintenance (O&M) phases. Building information modelling (BIM) technology [2] is widely adopted for information exchange and project data delivery in construction projects. BIM provides semantically rich 3D models representing the geometry and functionality of buildings and civil infrastructure, which can be exchanged across stakeholders and software systems using vendor-neutral Industry Foundation Classes (IFC) specifications [3]. In recent years, BIM-based information delivery has been adopted for infrastructure projects in several countries. For example, Norway has implemented nationwide road project delivery using a common data environment (CDE) platform (Trimble Quadri), where heterogeneous virtual models are uploaded and transformed into IFC-compliant software data models [4]. BIM-based delivery provides more accurate and structured as-built data for operation and maintenance (O&M). Furthermore, when combined with sensing technologies, artificial intelligence, and robotics, infrastructure-BIM (I-BIM) models can evolve into digital twins [5], enabling dynamic asset monitoring and intelligent maintenance decision-making.

Despite this potential, several challenges hinder BIM-based information delivery in highway infrastructure projects. First, there is a mismatch between IFC BIM standards and the asset data standards used by different road agencies [6]. These local asset standards are typically developed around existing asset databases and legacy systems; without clear mappings between data models, contractors have limited incentive to produce well-classified IFC models. Second, the coverage and usability of IFC standards remain limited in the infrastructure domain. For example, a Federal Highway Administration (FHWA) report highlights that road-related IFC entities, such as earthworks, drainage, and ancillary facilities, require further extension to support BIM-based infrastructure management [7]. Third, although recent IFC versions (e.g., IFC 4.3 [3]) introduce broader civil infrastructure concepts, few BIM authoring tools support seamless exchange based on the latest standards; for instance, the highest officially supported IFC export option in Autodesk Revit 2026 remains IFC4 [8]. Consequently, IFC models for highway project delivery are often misclassified or exported using generic types such as *IfcBuildingElementProxy* [9]. In addition, granularity mismatches are common: a single BIM object (e.g., a "deck") may need to be decomposed into multi-

ple asset components (e.g., slabs, pavement, girders) to support downstream asset management, whereas multiple fine-grained objects (e.g., railings) may need to be aggregated into a single asset such as a "pedestrian restraint system."

Previous studies have proposed various technical solutions for automatic BIM object classification using geometric features [10, 9], materials [11], attributes [12], and topological relationships [13, 14]. These approaches generally assume that BIM object instances are already modelled at an appropriate level of granularity and therefore focus on reclassifying objects to assign correct IFC types. However, real-world I-BIM models often exhibit not only typing errors but also inappropriate granularity, where object instances must be either aggregated or decomposed to meet the requirements of hierarchical asset data standards. This shifts BIM object classification from the object level to the part level, requiring finer-grained semantic understanding of 3D models.

To address these challenges, this study proposes an automated method for multi-granularity infrastructure-BIM (I-BIM) object classification and decomposition. The method performs per-face classification of BIM object meshes and decomposes coarse object instances into multiple parts that can be mapped to asset components defined by asset data standards. A context-aware mesh neural network (CA-MNN), built on the MeshNet backbone [15], is developed to incorporate geometric context by selecting K-nearest neighbouring objects and learning both object-level and context-level features to predict semantic labels for each triangulated face. A smoothing and clustering algorithm is then applied to both improve classification and automatically split coarse objects into meaningful parts when multiple asset types are present. The method is evaluated using three types of real-world IFC models, i.e., bridge, culvert, and road, aligned with asset classes defined in the UK Asset Data Management Manual (ADMM) [16]. ADMM provides a tree-structure asset class hierarchy, covering 10 parent asset types (e.g., ancillary, drainage, lighting, pavement), subtypes, and asset components. The best performing model achieves a per-face accuracy of 86.38% on a dataset of 6058 object instances with varying levels of granularity.

The remainder of the paper is structured as follows. Section 2 reviews related work. Section 3 describes the proposed method. Section 4 presents the evaluation results. Section 5 concludes the paper.

2 Literature review

Existing approaches for automatic BIM object classification can be broadly categorized into rule-based methods, machine learning-based methods, and hybrid methods.

Early studies primarily relied on manually defined domain-specific rules. Belsky et al. [17] developed a semantic enrichment engine for BIM (SeeBIM) that enables the composition of inference rule sets based on properties, geometry, topology, and functional relationships. Sacks et al. [13] further extended the SeeBIM engine with additional operators to classify BIM models of highway bridges. Ma et al. [14] developed matching algorithms for object identification by incorporating domain knowledge of shape features (e.g., orientation, volume) and pairwise relationships. Wu and Zhang [10] proposed an algorithm for classifying structural objects using inherent geometric features, such as the number of subcomponents, number of faces, and cross-sectional profiles.

Rule-based methods tend to be rigid and difficult to scale across diverse object types and features. Consequently, machine learning- and deep learning-based approaches have been increasingly explored. Wu et al. [18] proposed invariant signatures based on geometric and material features to describe IFC entities and evaluated classification performance using five machine learning models (e.g., random forest). Kim et al. [19] proposed an approach for automatically classifying building object instances in BIM models using a 2D convolutional neural network (CNN). Koo et al. [9] applied two geometric deep learning models, i.e., multi-view CNN and PointNet [20], to classify I-BIM objects such as culverts, retaining walls, and wing walls, and further extended this work to classify subtypes of BIM elements for walls and doors. Collins et al. [21] developed a graph convolutional network-based approach that operates on native triangle meshes and automatically generated local features for BIM object classification. Emunds et al. [22] proposed SpaRSE-BIM, a neural network based on sparse convolutions for classifying IFC-based geometry. Wang et al. [23] applied graph neural networks for room type classification and semantic enrichment of BIM models. More recently, Teclaw et al. [24] proposed Context-NET, which integrates contextual information from surrounding components by incorporating spatial relationships to improve object classification performance.

Recent studies have also explored hybrid approaches that integrate multiple types of features to improve classification accuracy. Utkucu et al. [25] applied an ensemble framework using a stacking strategy that combines four BIM classification tools: keyword search, rule-based inference, machine learning based on geometric features, and deep learning based on visual shape features. Luo et al. [26] introduced a two-branch geometric–relational deep learning framework that fuses geometric and relational features for IFC element classification. Yang et al. [27] proposed IFCGeoNet, which integrates explicit shape features and implicit representation features through graph learning. Similarly, Han et al. [28] and Wang and Ying

[29] developed generic graph neural network-based frameworks that embed semantic, spatial, and topological features of BIM models, where BIM element classification is treated as one of the downstream tasks.

In summary, existing context-aware methods (e.g., Context-NET) perform object-wise classification, where each IFC element in a BIM model is treated as an independent instance. However, they do not address granularity mismatch issues commonly observed in BIM-enabled infrastructure project delivery, where coarse IFC elements should be decomposed and overly fine-grained elements should be aggregated. Addressing this challenge requires hierarchy-aware classification and part-level semantic segmentation of infrastructure-BIM (I-BIM) objects.

3 Proposed solution

3.1 Overview

In this study, we propose a CA-MNN-based approach to automatically classify I-BIM object instances while enabling the decomposition of coarse elements. Figure 1 presents an overview of the proposed framework. The method begins with preprocessing I-BIM models exported from IFC-authoring software. The geometry is tessellated into triangulated meshes composed of vertices and faces. Quadric edge collapse is then applied to obtain decimated meshes that preserve the core geometry. The decimated mesh of each object instance is stored together with its GUID, which serves as a reference for generating contextual geometry.

For each object instance in an IFC model, the goal of CA-MNN is to predict the class types of all triangulated faces. The class types are derived from asset data standards used by highway agencies. In this study, three types of I-BIM models (i.e., bridge, culvert, and road) are used, covering a total of 30 asset classes defined in ADMM (see Table 2). The predicted face semantics are used to determine (a) the type of the IFC element and (b) whether its modelling granularity is too coarse and therefore requires decomposition. If an IFC element contains faces belonging to different asset classes, it is segmented into multiple parts by clustering faces of the same class.

The CA-MNN pipeline consists of three steps (see Figure 1). First, the K-nearest IFC elements surrounding the target element are retrieved. The decimated meshes of these neighbouring elements are combined as contextual faces until a maximum size is reached to satisfy memory constraints. Using decimated meshes instead of raw meshes allows more contextual elements to be incorporated. Second, a mesh neural network architecture is designed to embed both target geometry and contextual geometry into dense feature descriptors. The contextual feature descriptor is concatenated with the feature vectors of each triangulated face of the target element. A multilayer perceptron (MLP) then outputs class scores for each asset type. Third, connected faces are grouped and overlapping groups are merged. A final class is assigned to each group using majority voting, where the most frequent predicted class within a cluster is selected.

Based on these three steps, IFC object instances are classified and coarse elements are decomposed into parts with appropriate granularity according to asset data standards. The processed IFC models can then be seamlessly exchanged with infrastructure asset management systems (AMS). The following subsection details the architecture of CA-MNN.

3.2 Context-aware mesh neural network

The architecture of CA-MNN is illustrated in Figure 2. The network extends the original MeshNet [15] backbone with (a) a lightweight context branch and (b) a face-wise classifier. The inputs consist of meshes for both the target element and its surrounding context: the target geometry contains F_t faces, while the context mesh provides up to F_c faces. Each face includes a center point, three corner points, one normal vector, and neighbourhood indices.

The target mesh and context mesh are processed independently through the same descriptor modules and mesh convolution pipeline. First, center vectors ($n \times 3$), where n denotes the number of faces, are processed by a spatial descriptor, implemented as a multilayer perceptron (MLP) that embeds spatial features into a 64-dimensional vector. In parallel, corner points and normal vectors are processed by a structural descriptor that captures shape features. In the MeshNet [15] backbone, two types of structural descriptors are used. The first performs face rotation convolution over corner vectors to capture intra-face structural features. The second applies face kernel correlation, which uses normal vectors of each face and its neighbours to encode local geometric affinities.

In both branches, spatial and structural features are passed to two mesh convolution layers to expand the receptive fields of faces by aggregating information from neighbouring faces. The mesh convolution layer includes two parts. The first part combines spatial and structural features and applies a MLP to generate 512-dimensional global features. The second part aggregates the structural features of a face with those of connected faces through vector concatenation.

After two mesh convolution layers, spatial and structural features are concatenated in each branch, respectively. Regarding target branch, we apply an MLP to generate feature vectors for all faces ($F_t \times 1024$). In terms of context branch, we employ an adaptive average pooling layer that collapses the variable-length face dimension to size of 1, meaning that per-face features are transformed into a

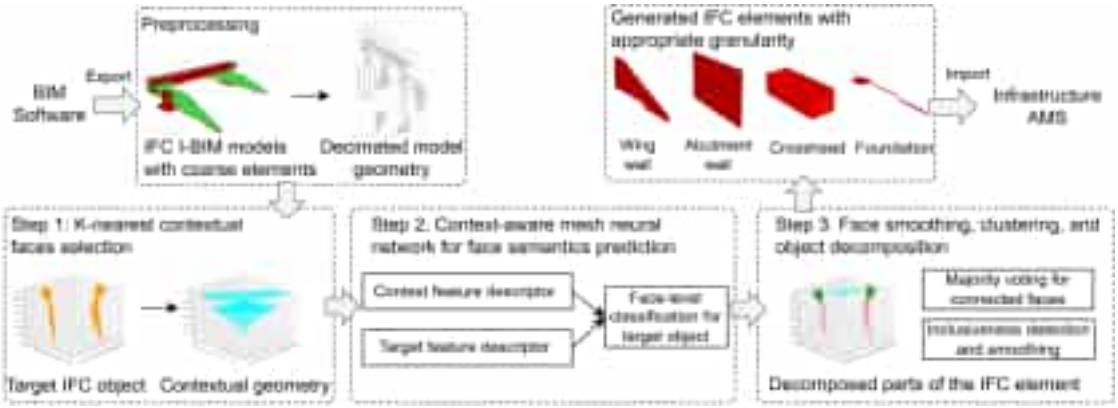


Figure 1. Overview of the proposed CA-MNN-based approach for granularity-aware I-BIM object classification and decomposition.

global feature of context geometry. Further, we use one MLP to generate a 256-dimensional vector, which is concatenated with per-face feature of target geometry (F_t × (1024+256)). Finally, we exploit a 1d convolutional layer to output final scores(F_t ×num_classes) for face-level classification, where num_classes represents the total number of asset class types.

4 Performance evaluation

4.1 Experiment design

We evaluate the proposed method using real-world BIM models from UK highway projects. The dataset includes five bridge models, five culvert models, and five road models, covering 30 asset classes defined in ADMM. All IFC models were imported into Blender for manual annotation. Specifically, ground-truth labels were assigned to each IFC element in the I-BIM models. For coarse IFC elements containing multiple asset units, Blender tools were used to split the geometry and generate finer IFC elements with corresponding ADMM labels. The dataset was further expanded by synthesizing coarse IFC elements through recombination of real elements from the original models. The generation process follows the statistical distribution of element types and group sizes observed in the original dataset, while applying controlled sampling to balance the occurrence of different object classes. Semantic consistency is preserved via asset hierarchy constraints, and duplicate combinations are eliminated based on source element IDs to ensure the uniqueness of all samples.

The final evaluation dataset contains 6,058 IFC elements, including 5,376 coarse elements requiring decomposition. The proportion between synthetic and original samples is 51:1. The class distribution is illustrated in Figure 3. The dataset was split into training and test sets using an 8:2 ratio, resulting in 4,845 training samples and 1,213 test samples.

The CA-MNN model was trained using the training set. The maximum numbers of target faces and context faces were set to 30,000 and 10,000, respectively. The batch size was set to 2, and the model was trained for a fixed number of 100 epochs to ensure sufficient convergence during training. Convergence is evidenced by the stabilisation of the training loss, where the relative decrease in loss is less than 1% over 10 consecutive epochs, occurring around the 90th epoch. The final model obtained after 100 epochs was used for evaluation on the test set.

4.2 Test results

We evaluate the proposed method using face-level classification accuracy. As shown in Table 1, CA-MNN correctly classifies 8,919,146 faces out of 10,325,920, achieving an accuracy of 80.80%. After applying the face smoothing operation, the accuracy increases significantly to 86.38%, demonstrating the effectiveness of smoothing and clustering connected faces to eliminate local misclassifications caused by ambiguous geometric features. However, although connectivity-based smoothing improves overall performance, it may introduce additional errors for tightly connected components such as abutment walls and wing walls.

In comparison, the mesh neural network without the context branch achieves an accuracy of 74.68%, which increases to 80.0% after smoothing. The performance gap indicates that the context branch in CA-MNN effectively incorporates surrounding information for better understanding of IFC geometry semantics and granularity.

Per-class classification performance is reported in Table 2. The macro-average accuracy across classes is 77.65%, indicating that performance is influenced by the imbalanced distribution of face counts across classes, which stems from disparate mesh complexity. Major

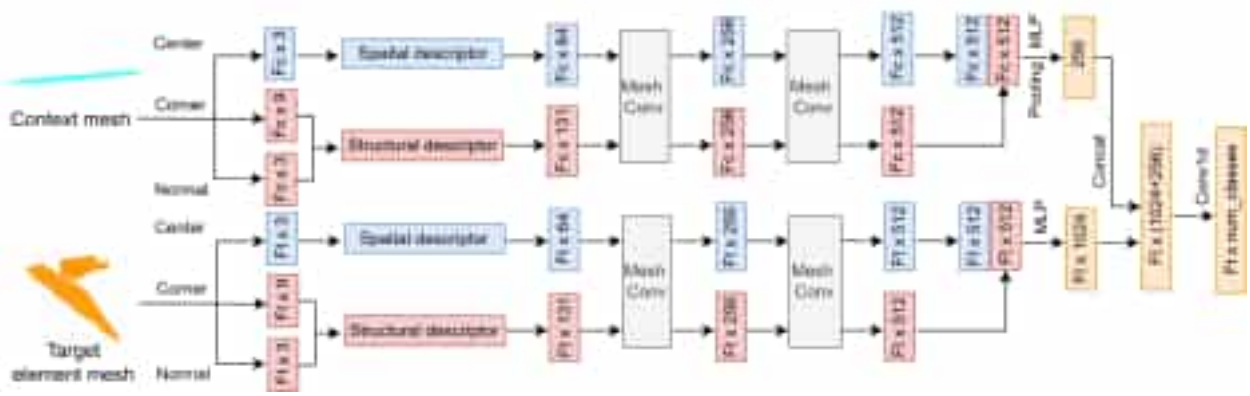


Figure 2. The architecture framework of the context-aware mesh neural network for face-level IFC geometry classification.

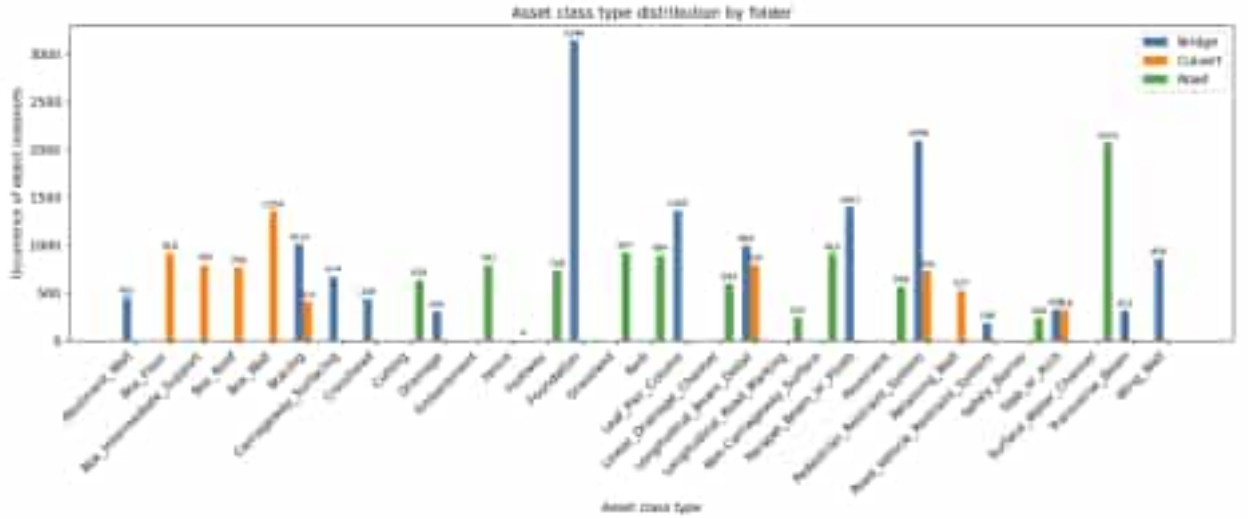


Figure 3. The distribution of different asset class types in the dataset.

classes with more faces generally achieve accuracies above 90%, whereas minority classes perform worse. For example, the accuracy for abutment walls (428 faces) is only 44.9%. Accuracy rates for bridge, culvert, and road models are 82.05%, 89.18%, and 78.84%, respectively, indicating reasonably balanced performance across infrastructure types.

Classification errors arise from several factors. First, CA-MNN struggles with geometrically featureless object types such as carriageways, footways, and safety barriers, which are often represented as simple cuboids. Table 2 shows consistently lower accuracy for these classes, and they are frequently misclassified as one another. For example, safety barriers are most often misclassified as linear drainage systems (16.6%). These errors typically occur across the entire element, indicating that CA-MNN fails to correctly interpret the object as a whole. Incorporating additional features, such as colour for recognising road markings, could be beneficial. However, as noted by Sacks et al. [13], semantic enrichment methods should ideally function under the worst-case scenario where only geometry is available, and the integration of additional features depends on the quality of the input BIM models. Second, the imbalanced distribution of mesh quantities across classes reduces performance on minority classes. Third, the current context extraction strategy may retrieve surrounding objects with large mesh sizes, which can dilute the representation of global contextual information. In addition, K-nearest neighbour retrieval is based on distance to the centroid of the target object, which is suboptimal for objects consisting of spatially separated components, such as groups of bridge foundations.

In terms of part decomposition performance, the mIoU achieves 0.701 across the entire test set and 0.667 for

Table 1. Model performance comparison

Model	Correct faces	Accuracy (%)
CA-MNN + smoothing	8919146	86.38
CA-MNN	8343047	80.80
MNN + smoothing	7692810	79.20
MNN	8178128	74.50

coarse IFC elements that contain multiple classes. In comparison, the overall mIoU drops to 0.603 without smoothing operation.

5 Conclusion

BIM-based highway information delivery faces challenges in aligning IFC models with agency-specific asset data standards, particularly due to granularity mismatches between construction and operation requirements. This often leads to labour-intensive manual processing in BIM or CDE platforms. To address this, this paper proposes a mesh neural network-based method for automatic classification of multi-granular infrastructure-BIM (I-BIM) objects, where coarse elements are identified and decomposed into finer components that better satisfy asset data requirements.

The proposed geometric deep learning (GDL)-based approach begins by tessellating IFC models into triangulated meshes, followed by geometry decimation using quadric edge collapse. The inference process of CA-MNN consists of three steps. First, the K-nearest surrounding objects of a target element are retrieved to form the contextual mesh. Second, the CA-MNN processes both target and contextual meshes through dual network branches to capture spatial and structural features, respectively. The global context feature is then concatenated with per-face features of the target object to perform face-level classification. Third, connectivity-based smoothing and clustering are applied to group faces with the same class labels, enabling both classification of IFC elements and decomposition of coarse elements into finer parts for downstream integration with infrastructure AMS.

The proposed approach is evaluated using 6,058 IFC elements from real-world BIM models, covering 30 asset classes defined in the Asset Data Management Manual (ADMM) used in the UK highway industry. The results show that CA-MNN with smoothing correctly classifies 8919146 faces out of 10325920, achieving an accuracy of 86.38%. Without smoothing, the accuracy decreases to 80.80%. We further evaluate a mesh neural network without the context branch, where accuracy drops to 79.20% with smoothing and 74.50% without smoothing. These results demonstrate the effectiveness of incorporating contextual geometry into GDL models for capturing semantic and topological relationships among infrastructure elements.

The contributions of this study are twofold. First, we propose a novel pipeline for automatic classification of multi-granular I-BIM objects using face-level semantics, which enables coarse IFC elements to be decomposed into parts with suitable granularity. Second, by combining geometry decimation, K-nearest context selection, and a dual-branch CA-MNN architecture, the method effectively learns contextual geometry and improves classification accuracy, as confirmed by the experimental results.

This study also has some limitations. First, CA-MNN struggles to reliably recognise elements with featureless geometry, such as roadways. Second, model training is negatively affected by class imbalance, leading to poorer performance on minority classes. Third, the current context selection strategy may retrieve irrelevant surrounding geometry. Future work will focus on improving context selection strategies and mesh neural network architectures to further enhance performance. In addition, a comprehensive AI-assisted workflow will be developed to support fully automated BIM-based information delivery for highway projects.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sk lodowska-Curie grant agreement No 101034337. We wish to thank Kieran Mulvey, Dr. Pieter Pauwels, Dr. Brian Sheil, Dr. George Economides, Dr. Stefano Cavazzi, Yogesh Patel, Matt Peck, and Simon Hayton for their contributions to this paper.

References

[1] Mengtian Yin, Varun Kumar Reja, Ran Wei, Ioannis Brilakis, Brian Sheil, Federico Perrotta, Alix Marie d'Avigneau, and Linjun Lu. Exploring the value of digital twins for information management in highway asset maintenance. *Developments in the Built Environment*, 21:100614, 2025.

[2] Chuck Charles M Chuck Eastman, Chuck Charles M Chuck Eastman, Paul Teicholz, Rafael Sacks, and Kathleen Liston. *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley Sons, Hoboken, 2nd edition, 2011. ISBN 0470541377.

[3] buildingSMART International Ltd. IFC4.3 RC2 - Release Candidate 2 [Draft], 2020. URL `https://standards.buildingsmart.org/IFC/DEV/IFC4_3/RC2/HTML/`.

[4] Norwegian Public Roads Administration Selects Trimble for Asset Life-

Table 2. Per-class face accuracy

Asset class	Face numbers (Correct/Total)	Accuracy (%)
Leaf_Pier_Column	94879 / 96692	98.10
Foundation	8629 / 8862	97.40
Crosshead	1330 / 1763	75.40
Pedestrian_Restraint_System	995093 / 1112537	89.40
Parapet_Beam_or_Plinth	108739 / 119893	90.70
Slab_or_Arch	21421 / 23113	92.70
Longitudinal_Beam_Detail	133677 / 164546	81.20
Carriageway_Surfacing	117051 / 201155	58.20
Bracing	31278 / 31604	99.00
Drainage	60519 / 107115	56.50
Transverse_Beam	8744 / 10741	81.40
Road_Vehicle_Restraint_System	127039 / 128017	99.20
Abutment_Wall	192 / 428	44.90
Wing_Wall	2236 / 2865	78.00
Box_Intermediate_Support	10151 / 10290	98.60
Box_Roof	2453 / 2552	96.10
Box_Wall	24433 / 24630	99.20
Box_Floor	9641 / 9748	98.90
Retaining_Wall	87891 / 101671	86.40
Embankment	169556 / 262089	64.70
Cutting	167176 / 230291	72.60
Surface_Water_Channel	1181163 / 1210525	97.60
Linear_Drainage_Channel	24343 / 31277	77.80
Footway	126153 / 136696	92.30
Non-Carriageway_Surface	163343 / 231190	70.70
Kerb	104775 / 106488	98.40
Longitudinal_Road_Marking	113190 / 242742	46.60
Safety_Barrier	41743 / 128411	32.50
Pavement	156934 / 285642	54.90

cycle Management. URL https://www.prnewswire.com/news-releases/norwegian-public-roads-administration-selects-trimble-for-asset-lifecycle-management-301717319.html.

[5] Yan Gao, Shuyue Qian, Zihan Li, Ping Wang, Feiyue Wang, and Qing He. Digital twin and its application in transportation infrastructure. *Proceedings 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence, DTPI 2021*, pages 298–301, 2021. doi:10.1109/DTPI52967.2021.9540108.

[6] G. S. Floros, G. Boyes, D. Owens, and C. Ellul. Developing IFC for infrastructure: A case study of three highway entities. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(4/W8):59–66, 2019. ISSN 21949050. doi:10.5194/isprs-annals-IV-4-W8-59-2019.

[7] Federal Highway Administration. Bridge Information Model Standardization VOLUME II: SCHEMA ANALYSIS, 2016. URL https://www.fhwa.dot.gov/bridge/pubs/hif16011/hif16011c.pdf.

[8] Autodesk. Autodesk Revit 2026. 2026. URL https://help.autodesk.com/view/RVT/2026/ENU/.

[9] Bonsang Koo, Raekyu Jung, and Youngsu Yu. Automatic classification of wall and door BIM element subtypes using 3D graph convolutional neural networks. *Advanced Engineering Informatics*, 47: 101200, 2021. ISSN 1474-0346.

[10] Jin Wu and Jiansong Zhang. New Automated BIM Object Classification Method to Support BIM Interoperability. *Journal of Computing in Civil Engineering*, 33(5), 2019. ISSN 08873801. doi:10.1061/(ASCE)CP.1943-5487.0000858.

[11] Wu Jin, Akanbi Temitope, and Zhang Jiansong. Constructing Invariant Signatures for AEC Objects to Support BIM-Based Analysis Automation through Object Classification. *Journal of Computing in Civil Engineering*, 36(4): 4022008, jul 2022. doi:10.1061/(ASCE)CP.1943-5487.0001012. URL https://doi.org/10.1061/(ASCE)CP.1943-5487.0001012.

[12] Duygu Utkucu, Huaquan Ying, Zijian Wang, and Rafael Sacks. Classification of architectural and mep bim objects for building performance evaluation. *Advanced Engineering Informatics*, 61:102503, 2024.

[13] Rafael Sacks, Ling Ma, Raz Yosef, Andre Borrmann, Simon Daum, and Uri Kattel. Semantic enrichment for building information modeling: Procedure for

compiling inference rules and operators for complex geometry. *Journal of Computing in Civil Engineering*, 31(6):4017062, 2017. ISSN 0887-3801.

[14] Ling Ma, Rafael Sacks, Uri Kattel, and Tanya Bloch. 3D Object Classification Using Geometric Features and Pairwise Relationships. *Computer-Aided Civil and Infrastructure Engineering*, 33(2):152–164, 2018. ISSN 14678667. doi:10.1111/mice.12336.

[15] Yutong Feng, Yifan Feng, Haoxuan You, Xibin Zhao, and Yue Gao. Meshnet: Mesh neural network for 3d shape representation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8279–8286, 2019.

[16] National Highways. Asset Data Management Manual (ADMM) and other management and maintenance guides, 2023. URL https://nationalhighways.co.uk/suppliers/design-standards-and-specifications/admm-and-other-management-and-maintenance-guides/.

[17] Michael Belsky, Rafael Sacks, and Ioannis Brilakis. A Semantic Enrichment Engine for Building Information Modelling. *Computer-Aided Civil and Infrastructure Engineering*, pages 261–274.

[18] Jin Wu, Hezha Lutfalla Sadraddin, Ran Ren, Jiansong Zhang, and Xiaoyun Shao. Invariant signatures of architecture, engineering, and construction objects to support BIM interoperability between architectural design and structural analysis. *Journal of construction engineering and management*, 147(1): 4020148, 2021. ISSN 0733-9364.

[19] Jinsung Kim, Jaeyeol Song, and Jin-Kook Lee. Recognizing and classifying unknown object in BIM using 2D CNN. In *International Conference on Computer-Aided Architectural Design Futures*, pages 47–57. Springer, 2019.

[20] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[21] Fiona C Collins, Alexander Braun, Martin Ringsquandl, Daniel M Hall, and André Borrmann. Assessing ifc classes with means of geometric deep learning on different graph encodings. In *Proc. of the 2021 European Conference on Computing in Construction*, 2021.

[22] Christoph Emunds, Nicolas Pauen, Veronika Richter, Jérôme Frisch, and Christoph van Treeck. SpaRSE-BIM: Classification of IFC-based geometry via sparse convolutional neural networks. *Advanced Engineering Informatics*, 53:101641, 2022. ISSN 1474-0346.

[23] Zijian Wang, Rafael Sacks, and Timson Yeung. Exploring graph neural networks for semantic enrichment: Room type classification. *Automation in Construction*, (November):104039, 2021. ISSN 09265805. doi:10.1016/j.autcon.2021.104039. URL https://doi.org/10.1016/j.autcon.2021.104039.

[24] Wojciech Teclaw, Artur Tomczak, Mateusz Kasznia, Marcin Luczkowski, Nathalie Labonnote, and Eilif Hjelseth. ContextNET-Contextual classification of building components using deep neural networks. *Advanced Engineering Informatics*, 67: 103484, 2025. ISSN 1474-0346.

[25] Duygu Utkucu, Huaquan Ying, Zijian Wang, and Rafael Sacks. Classification of architectural and MEP BIM objects for building performance evaluation. *Advanced Engineering Informatics*, 61:102503, 2024. ISSN 1474-0346.

[26] Hairong Luo, Ge Gao, Han Huang, Ziyi Ke, Cheng Peng, and Ming Gu. A geometric-relational deep learning framework for BIM object classification. In *European Conference on Computer Vision*, pages 349–365. Springer, 2022.

[27] Mingsong Yang, Xinhong Hei, Haining Meng, Kehai Chen, Xinyu Tong, YuChao Li, and Qin Zhao. Lightweight framework for IFC element classification using multi-view and heterogeneous graph with decision-level fusion. *Automation in Construction*, 181:106660, 2026. ISSN 0926-5805.

[28] Jin Han, Xin-Zheng Lu, and Jia-Rui Lin. Pretrained graph neural network for embedding semantic, spatial, and topological data in building information models. *Computer-Aided Civil and Infrastructure Engineering*, 40(26):4607–4631, 2025. ISSN 1093-9687.

[29] Zijian Wang and Huaquan Ying. Graph Representation Learning: Embedding Multimodality BIM Models into Graphs. In *EC3 Conference 2025*, volume 6, page 0. European Council on Computing in Construction, 2025. ISBN 9083451313.

Probabilistic Flood Damage Cost Estimation for Residential Buildings Using Bayesian Information Updating

Yitong Li[1] and Jie Gong[2]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
[2]Department of Civil and Environmental Engineering, Rutgers University, USA
yitong2@ualberta.ca, jg931@soe.rutgers.edu

Abstract – Flood damage estimation for residential buildings is a critical component of post-disaster recovery planning, budgeting, and construction management. Conventional depth–damage curves, which remain the standard tool in flood risk assessment, rely on deterministic and monotonic assumptions that fail to capture the considerable variability observed in real-world flood losses. As a result, these methods often provide limited support for uncertainty-aware cost estimation and decision-making. This study proposes a probabilistic framework for residential flood damage cost estimation based on Bayesian inference and Markov Chain Monte Carlo simulation. Building damage is modeled as a percentage of replacement value and represented using beta distributions. The framework sequentially updates damage distributions as new post-flood observations become available and propagates uncertainty into cost estimates through Monte Carlo simulation. The proposed approach is demonstrated through case studies in Manville and Cranford, New Jersey, using residential flood damage data from Hurricane Ida. Results show that posterior predictive cost estimates consistently encompass observed damage costs within credible intervals, while traditional depth–damage curve-based estimates exhibit limited agreement with observations. The findings highlight the value of probabilistic, uncertainty-aware modelling through improved representation of cost uncertainty, thereby providing useful insights for residential construction cost estimation and post-disaster recovery management.

Keywords –
Flood damage cost estimation; Probabilistic modelling; Markov Chain Monte Carlo simulation; Uncertainty quantification

1 Introduction

Flooding is among the most impactful natural hazard worldwide, affecting the built environment and residential buildings that underpin community lifelines. As damage to residential buildings directly drives recovery needs and resource demands, accurate damage estimation is essential for construction-related decision-making, including defining repair scopes, allocating contingencies, and evaluating mitigation alternatives such as home elevation or buyout programs [1][2]. In practice, however, even small errors in translating physical flood characteristics into monetary losses can distort cost estimates. Such uncertainty complicates budget approval, delays reconstruction efforts, and weakens the economic justification for long-term residential flood mitigation investments [3].

The most commonly used approach for estimating flood damage to residential buildings is the depth–damage curve (DDC), which relates floodwater depth to an expected percentage of structural or contents loss. In the United States, standardized residential DDCs developed by the U.S. Army Corps of Engineers (USACE) and the Federal Emergency Management Agency (FEMA) are widely applied in housing damage assessments, insurance analyses, and construction cost estimation workflows [4][5]. Their simplicity makes them particularly attractive during early post-disaster stages, when rapid estimates are required and detailed building condition data are limited.

Despite their widespread adoption, existing studies have identified several critical limitations of DDC-based methods when applied to cost estimation. Specifically, standard DDCs are typically derived from aggregated historical data or expert judgment and represent an "average" response for broad residential building categories [6]. This approach doesn't consider the variability in residential construction practices, such as foundation type, framing systems, finish materials, and construction quality [3]. As a result, DDCs are typically used for high-level cost estimation, with limited ability to account for cost variability across individual homes.

Moreover, most conventional DDCs are built on the assumption that flood damage is a monotonic function of water depth, meaning that expected losses increase continuously as water levels rise [6]. However, recent

study reveals that observed flood losses are often not monotonic and instead follow a bimodal distribution where losses are concentrated at the minimal and maximal extremes [6]. As a result, residential flood losses at a given depth exhibit variation, with similar inundation levels leading to different damage levels. From a construction management perspective, such systematic bias can influence residential recovery budget allocation and hinder effective sequencing and resourcing of repair activities, which highlights the importance of uncertainty-aware estimation approaches [7][3][8].

Although multivariate damage models that incorporate additional variables such as flood duration or building condition have been proposed [9][10][11], their extensive data requirements and limited transferability restrict their use in residential contexts. As a result, there is a need for probabilistic damage estimation frameworks tailored to residential buildings that can represent uncertainty while remaining practical for construction-oriented cost estimation. Probabilistic approaches, with particular emphasis on Bayesian inference–based approaches, are capable of addressing these limitations by producing distributions of cost estimates rather than single-point values. By combining prior knowledge of building information with post-flood damage observations, these methods enable uncertainty quantification and reliable cost estimation.

This study proposes a Bayesian framework that integrates residential building attributes with Markov Chain Monte Carlo (MCMC) simulation for building cost estimation following flood impacts. Rather than assuming a monotonic depth–damage relationship, building damage is modelled independently at each inundation depth. For each depth and residential building type (e.g., two-story buildings with basements), the proportion of building damage, bounded between zero and one, is modelled using a beta distribution. MCMC is then employed to update these distributions as new damage observations become available. The resulting damage distributions are subsequently converted into cost distributions by incorporating building replacement values. Through this updating process, cost estimates are continuously refined while uncertainty is explicitly quantified, thereby supporting more robust budgeting, contingency planning, and construction management decision-making in flood-prone residential areas.

2 Literature Review

One of the most widely used approaches for estimating flood damage to residential buildings is the DDC, which relates inundation depth directly to expected economic loss. Although there is no universally accepted framework for assessing flood damage in urban build environments, most existing damage models rely on DDCs because of their conceptual simplicity and ease of implementation [12]. In practice, these curves are frequently used to support cost estimation, post-disaster budgeting, and recovery planning, particularly during early-stage damage assessments when detailed building inspections are not yet available.

DDCs are typically constructed under the assumption that flood damage increases monotonically with water depth. In practice, however, this assumption is frequently violated. Recent analyses of more than two million National Flood Insurance Program (NFIP) claims show that observed losses are often non-monotonic and instead exhibit a bimodal distribution, with damage clustered near both minimal and maximal loss levels [6]. From a construction management perspective, this behaviour complicates repair cost estimation, as similar flood depths can lead to vastly different repair needs, ranging from minor finishes replacement to full system and structural rehabilitation [3]. Studies further indicate that conventional DDCs tend to overestimate damage from shallow flooding while underestimating losses associated with deep inundation. In many cases, their predictive performance is sufficiently limited that a simple mean of observed damages outperforms the curve-based estimates [6]. Another limitation of traditional DDCs is their reliance on a “one-to-one” relationship between depth and damage, where each inundation depth is associated with a single representative damage value, typically expressed as the median or modal loss. Empirical evidence suggests that flood damage at a given depth is highly variable and poorly described by central tendency measures. For example, at one foot of inundation, the middle 50% of observed claims may span from approximately 4% to 30% of building value [6].

The inability to explicitly model uncertainty is widely recognized as a major bottleneck in flood risk assessment, as deterministic approaches typically produce single-point estimates that obscure the inherent variability and stochastic nature of flood damage processes. To address this limitation, numerous studies have developed multivariable flood damage models that extend beyond depth–damage formulations by incorporating additional factors, such as flood duration, the timing and quality of early warnings, building quality, and residents’ socioeconomic characteristics [9][10][11]. A range of statistical techniques has been applied, including random forest models [13], tree-structured models [14], and empirically derived equations [15]. Despite being useful, these approaches require large, high-quality datasets, and limitations in data availability and consistency can introduce bias [16]. Moreover, these multivariable models are often highly site-specific, limiting their transferability across regions.

A growing body of research has shifted toward

probabilistic flood vulnerability modeling to improve the reliability and transparency of flood damage cost estimates [8][3]. Probabilistic frameworks, particularly those grounded in Bayesian inference, allow uncertainty in depth–damage relationships to be explicitly represented and propagated into cost estimates. By combining prior knowledge with limited empirical data, Bayesian methods yield robust, observation-informed cost estimates while explicitly quantifying uncertainty [17][18]. When empirical damage and cost data are scarce, prior information, such as established damage functions or historical repair cost data, can guide early-stage cost estimation and budget planning. As additional building-specific damage observations become available, estimates can be progressively refined to reflect observed cost estimates. Rather than producing single-point damage or cost estimates, Bayesian models generate probability distributions of expected losses, with credible intervals that reflect data availability. For residential building categories with small sample sizes, wider credible intervals reflect increased uncertainty, whereas narrower intervals correspond to greater confidence under improved data availability.

3 Methodology

This study proposes a Bayesian framework that integrates residential building attributes with MCMC simulation for building cost estimation following flood impacts. The detailed steps are illustrated in Figure 1.

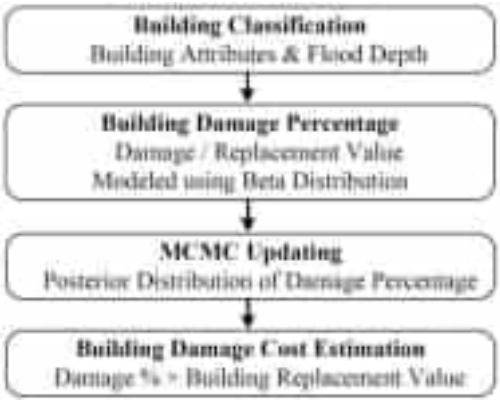


Figure 1. Methodology framework.

First, buildings are classified according to the following attributes: occupancy type, which describes the building use (e.g., single-family residence or two- to four-unit residential building); number of stories (e.g., one-story or two-story structures); basement type (e.g., no basement, finished basement, or crawlspace); and flood depth, defined as the difference between the flood elevation and the building's first-floor elevation.

Next, beta distributions are used to model the distribution of building damage percentage. This percentage is used as the damage metric because depth–damage curves, which remain the standard tool in flood damage assessment, express damage as the percentage of the estimated building damage value to the total building replacement value. Modelling damage in percentage terms provides a normalized representation of loss that is independent of absolute building value and allows direct compatibility with existing depth–damage–based cost estimation workflows. The resulting damage ratios are subsequently converted into monetary losses by incorporating building replacement values.

The beta distribution is selected to model building damage percentage because this metric is defined as a proportion bounded between zero and one. Unlike unbounded distributions, the beta distribution ensures physically meaningful damage ratios and can flexibly represent a wide range of empirical damage patterns, including skewed and highly variable loss behaviour observed in flood events. In addition, the beta distribution is well suited for Bayesian inference and MCMC updating, enabling progressive refinement of damage estimates as new observations become available.

For a given inundation depth and residential building class, observed damage is expressed as a percentage,

$$y_i = \frac{estimated\ building\ damage\ value_i}{building\ replacement\ value_i} \quad (1)$$

where $y_i \in (0,1)$, with 0 indicating no damage and 1 indicating total loss. These observations are modelled using a beta likelihood, which inherently enforces the physical bounds of building damage percentages. The parameters of the beta distribution, $\theta = (\alpha, \beta)$, are updated using Bayes' rule as new building damage percentage observations D_{new} become available:

$$p(\alpha, \beta | D_{old}, D_{new}) \propto p(D_{new}|\alpha, \beta) \times p(\alpha|D_{old}) \times p(\beta|D_{old}) \quad (2)$$

with the likelihood given by:

$$p(D_{new}|\alpha, \beta) = \prod_{i=1}^{n} Beta\,(y_i|\alpha, \beta). \quad (3)$$

The parameters α and β are assigned as independent gamma prior distributions,

$$p(\alpha|D_{old}) = Gamma(\alpha|a_1, b_1), \quad (4)$$
$$p(\beta|D_{old}) = Gamma(\beta|a_2, b_2), \quad (5)$$

which can also be specified as informative or weakly informative depending on data availability. When substantial prior knowledge of residential building damage exists, informative priors can be selected to reflect established damage behaviour. In data-scarce contexts, weakly informative priors (e.g.,

$\mathrm{Gamma}(0.001, 0.001)$) can be used minimize prior influence.

MCMC sampling is used to approximate the posterior distribution of building damage percentage. Each MCMC draw represents a plausible realization of damage percentage for the given depth and building class. These sampled damage percentages are subsequently converted into building damage cost samples by multiplying them by the corresponding building replacement value. This transformation propagates uncertainty from damage percentage directly into cost estimation, resulting in a probabilistic representation of residential flood damage cost estimation.

4 Results

To demonstrate the applicability of the proposed framework, a case study was conducted using residential flood damage data from Hurricane Ida. Two townships in New Jersey—Manville and Cranford—were selected due to their repeated exposure to riverine flooding and their history of residential flood damage. Both communities experienced inundation during Hurricane Ida, resulting in a large number of reported residential losses. The analysis focuses on a representative residential building class: single-family dwellings with two stories and basements, evaluated at an inundation depth of 1 ft above the first-floor elevation. Building damage percentage data for this building class were obtained from the OpenFEMA National Flood Insurance Program (NFIP) dataset [19], which provides publicly available, claim-level flood damage information derived from post-event assessments.

Figures 2 and 3 illustrate three posterior predictive distributions (mean, lower 95%, upper 95%) of building damage percentage for Manville and Cranford, respectively, obtained through Bayesian updating using MCMC simulation. Weakly informative gamma priors were assigned to the beta shape parameters α and β to represent minimal prior knowledge of building damage behaviour. In the figures, the histogram shows the empirical distribution of observed damage percentages. The three curves represent beta distributions corresponding to the posterior mean and the lower and upper 95% credible bounds of the shape parameters, derived from the updated MCMC samples. While only three representative distributions are shown for clarity, they reflect two sources of uncertainty implied by the posterior: variability within each beta distribution, which represents the range of possible damage percentages for a given set of shape parameters, and variability across beta distributions arising from uncertainty in the shape parameters themselves. Together, these distributions capture the range of plausible damage behaviours, including cases with density concentrated near both low and high damage extremes, demonstrating the flexibility of the beta model to represent outcomes from minimal damage to near-total loss.

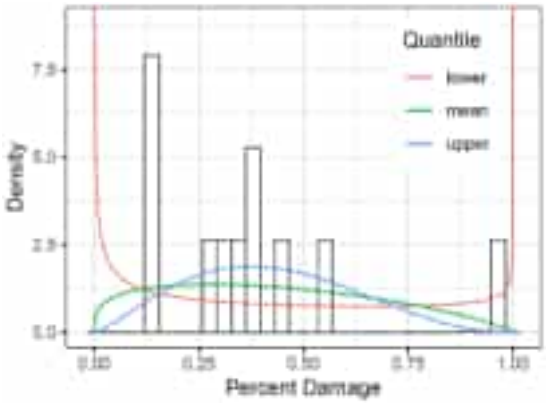


Figure 2. Posterior predictive distributions (mean, lower 95%, upper 95%) of residential building damage percentage for Manville, NJ.

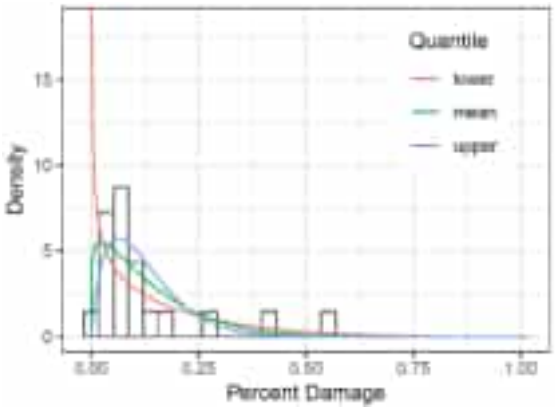


Figure 3. Posterior predictive distributions (mean, lower 95%, upper 95%) of residential building damage percentage for Cranford, NJ.

For practical cost estimation, these damage percentages can be directly translated into monetary losses by multiplying them by the building replacement value. As a result, each beta distribution corresponds to a distribution of potential damage costs rather than a single estimate. This enables probabilistic cost estimation, allowing practitioners to quantify not only expected losses (e.g., using the mean) but also the associated uncertainty (e.g., using credible intervals), which is critical for risk-informed decision-making, budgeting, and resilience planning.

Figure 4 compares estimated residential flood damage costs against observed damage costs for Manville, including results from both the proposed probabilistic approach and traditional depth–damage curves. For the probabilistic approach, the vertical lines

represent the 95% credible intervals of damage cost derived from posterior predictive distributions of building damage percentage, with blue circles indicating the corresponding mean estimates. These cost estimates are obtained by transforming the predicted damage percentages into monetary values using building replacement costs. The diagonal line represents perfect agreement between estimated and observed costs. Intersections between the credible intervals and the diagonal line indicate that the observed damage cost is captured within the uncertainty bounds of the proposed method. The results show that a substantial portion of the observed costs fall within the estimated credible intervals, demonstrating the ability of the probabilistic framework to represent uncertainty and capture observed outcomes.

In contrast, the red triangles represent deterministic cost estimates obtained from traditional depth–damage curves. These point estimates are frequently farther from the diagonal line and, unlike the probabilistic intervals, do not account for uncertainty. Comparing the mean probabilistic estimates (blue circles) with the deterministic estimates (red triangles), the probabilistic approach generally yields estimates closer to the diagonal, indicating improved agreement with observed costs. While a few cases show the deterministic estimates performing comparably or slightly better, this is likely due to limited data informing the posterior in those instances. Overall, the results highlight the advantage of the proposed method in providing uncertainty-aware estimates while maintaining improved accuracy relative to traditional approaches.

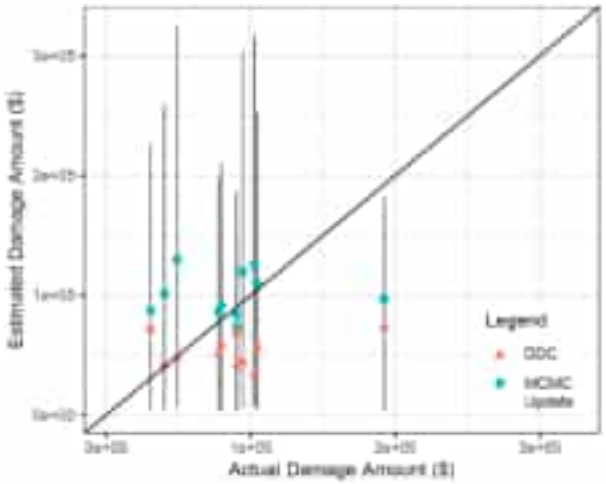


Figure 4. Observed versus estimated residential flood damage costs for Manville, NJ: probabilistic (95% credible intervals) and deterministic comparisons.

Figure 5 compares estimated and observed residential flood damage costs for Cranford, NJ, including both the proposed probabilistic approach and traditional depth–damage curve estimates. Similar to the Manville results, the posterior predictive cost estimates are represented by 95% credible intervals, with mean estimates shown as blue circles. A large proportion of these intervals intersect the diagonal line, indicating that the observed costs are well captured within the uncertainty bounds of the proposed method. In contrast, the deterministic depth–damage curve estimates (shown as red triangles) are more frequently distant from the diagonal.

With a larger dataset available for Cranford, the posterior distributions of the model parameters are better constrained, resulting in more accurate and consistent cost estimates. This is reflected in the mean probabilistic estimates, which are generally closer to the diagonal line than the deterministic estimates. The results demonstrate that, with sufficient data, the proposed probabilistic framework not only captures uncertainty but also provides improved accuracy compared to traditional depth–damage approaches.

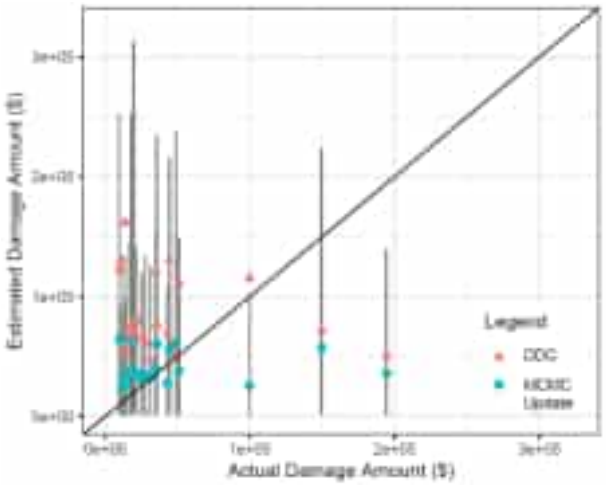


Figure 5. Observed versus estimated residential flood damage costs for Cranford, NJ: probabilistic (95% credible intervals) and deterministic comparisons.

Overall, the Manville and Cranford case studies demonstrate the practical applicability of the proposed Bayesian framework for residential flood damage cost estimation. Across both communities, the probabilistic approach consistently captures observed damage costs within estimated uncertainty ranges, while traditional depth–damage curve–based estimates show more limited agreement with observed outcomes. Additionally, with increased data availability in Cranford, the probabilistic approach provides not only reliable uncertainty quantification but also improved accuracy in mean cost estimates compared to the deterministic method. These results highlight the value of uncertainty-aware modelling for residential flood damage cost estimation.

5 Conclusion

This study presents a probabilistic framework for estimating residential flood damage costs that explicitly accounts for uncertainty in building damage behaviour. By modelling building damage percentage using a Bayesian inference through MCMC simulation, the proposed method moves beyond conventional deterministic depth–damage curves that rely on single-point estimates and monotonic assumptions. Modelling damage independently at each inundation depth allows the framework to represent multiple plausible damage states for the same flood level, better reflecting observed residential flood loss patterns.

The case studies conducted in Manville and Cranford, New Jersey, using post-event data from Hurricane Ida, demonstrate the practical applicability and advantages of the proposed approach. Across both communities, posterior predictive cost estimates consistently encompassed observed residential damage costs within their credible intervals, whereas traditional depth–damage curve–based estimates showed limited agreement with observations. These results highlight the limitations of deterministic damage functions for construction cost assessment and illustrate the value of uncertainty-aware modelling for capturing the variability inherent in residential flood damage.

From a construction management perspective, the proposed framework provides decision-support outputs that are directly applicable to budgeting, contingency planning, and post-disaster recovery management. Rather than producing a single expected cost, the method generates distributions cost estimations, enabling stakeholders to assess financial risk and allocate resources more effectively under uncertainty. The ability to update damage estimates as new data become available further enhances the framework's suitability for post-disaster contexts, where information is often incomplete and evolves over time. While this study focuses on a specific residential building class and flood depth, the framework is flexible and can be extended to additional building types, inundation levels, and geographic regions as data become available. Future research will incorporate additional building attributes to enable more detailed cost estimation. Overall, this research demonstrates that probabilistic, Bayesian-based damage modelling offers a robust and practical alternative to traditional depth–damage curves for residential flood damage cost estimation, thereby informing reliable planning and budget allocation for residential building construction and recovery.

References

[1] Y. Li and W. Ji. Robustness of stakeholder response networks for infrastructure system protection. *Journal of Management in Engineering*, 37(6):04021063, 2021.

[2] Y. Li, F. Zhang, and W. Ji. Integrated data-driven and equity-centered framework for highway restoration following flood inundation. *Journal of Management in Engineering*, 39(3):04023012, 2023.

[3] H. McGrath, A. Abo El Ezz, and M. Nastev. Probabilistic depth–damage curves for assessment of flood-induced building losses. *Natural Hazards*, 97(1):1–14, 2019.

[4] Federal Emergency Management Agency. Hazus Flood Model Technical Manual. On-line: https://www.fema.gov/sites/default/files/documents/fema_rsl_hazus-7-fltm_06272025_0.pdf, Accessed: 24/04/2026.

[5] Federal Emergency Management Agency. Hazus. On-line: https://www.fema.gov/flood-maps/products-tools/hazus, Accessed: 24/04/2026.

[6] O. E. Wing, et al. New insights into US flood vulnerability revealed from flood insurance big data. *Nature Communications*, 11(1):1444, 2020.

[7] W. Mobley, K. O. Atoba, and W. E. Highfield. Uncertainty in flood mitigation practices: Assessing the economic benefits of property acquisition and elevation in flood-prone communities. *Sustainability*, 12(5):2098, 2020.

[8] K. Angeles and T. Kijewski-Correa. Bayesian data integration framework for the development of component-level fragilities derived from multiple post-disaster datasets. *Structural Safety*, 99:102260, 2022.

[9] B. Merz, H. Kreibich, and U. Lall. Multi-variate flood damage assessment: A tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, 13(1):53–64, 2013.

[10] K. Schröter, et al. How useful are complex flood damage models? *Water Resources Research*, 50(4):3378–3395, 2014.

[11] A. H. Thieken, et al. Estimating changes in flood risks and benefits of non-structural adaptation strategies—a case study from Tyrol, Austria. *Mitigation and Adaptation Strategies for Global Change*, 21(3):343–376, 2016.

[12] R. Hasanzadeh Nafari and T. Ngo. Predictive applications of Australian flood loss models after a temporal and spatial transfer. *Geomatics, Natural Hazards and Risk*, 9(1):416–430, 2018.

[13] M. B. Malgwi, M. Schlögl, and M. Keiler. Expert-based versus data-driven flood damage models: A comparative evaluation for data-scarce regions. *International Journal of Disaster Risk Reduction*, 57:102148, 2021.

[14] T. Sieg, et al. Tree-based flood damage modeling of companies: Damage processes and model performance. *Water Resources Research*, 53(7):6050–6068, 2017.
[15] A. Paleo-Torres, et al. Vulnerability of Florida residential structures to hurricane induced coastal flood. *Engineering Structures*, 220:111004, 2020.
[16] T. Endendijk, et al. Flood vulnerability models and household flood damage mitigation measures: An econometric analysis of survey data. *Water Resources Research*, 59(8):e2022WR034192, 2023.
[17] Y. Li and W. Ji. Bayesian-based forecasting of infrastructure restoration progress following extreme events. *International Journal of Disaster Risk Reduction*, 103519, 2022.
[18] Y. Li, C. Yang, and W. Ji. Enhanced prediction of highway flood inundation through Bayesian generalized linear geostatistical models. *Advanced Engineering Informatics*, 61:102451, 2024.
[19] Federal Emergency Management Agency. OpenFEMA Data Sets. On-line: https://www.fema.gov/about/openfema/data-sets, Accessed: 24/04/2026.

A hybrid RAG approach for generating ontology-based information containers from building permit application submissions using text-based LLMs

Philipp Hagedorn[1], Jayesh Sudhakar Adlinge[2], Judith Fauth[2] and André Borrmann[2]

[1]Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany
[2]Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

philipp.hagedorn-n6v@rub.de, jayesh.adlinge@tum.de, judith.fauth@tum.de andre.borrmann@tum.de

Abstract -

Building permit applications usually consist of multiple heterogeneous, unstructured, and non-machine-readable data. To efficiently check and process applications, relevant information must be extracted, and documents must be classified and linked to the application and process steps in the review. To address this, this study presents an end-to-end method for generating ontology-structured information containers from PDF-based building permit applications by combining Large Language Models with a hybrid Retrieval Augmented Generation (RAG) layer. The proof-of-concept demonstrates that the proposed pipeline can populate OntoBPR entities while producing standard-compliant container structures in accordance with ISO 21597. Remaining challenges relate to scalability, handling of ambiguous regulatory language, and quantitative evaluation of the pipeline. Overall, the study demonstrates the potential of retrieval-augmented LLM pipelines for advancing automation and semantic integration in digital building permit processes.

Keywords -

Architecture, Engineering and Construction (AEC); Digital Building Permitting (DBP); Knowledge Graphs (KG); Retrieval Augmented Generation (RAG); Neuro-symbolic AI

1 Introduction

Building permit authorities typically receive application submissions as heterogeneous collections of documents: (scanned) PDF application forms, text-based reports, drawings, exports from specialist software, and occasionally models from Building Information Modeling (BIM) serialized in the Industry Foundation Classes (IFC) [1]. Even when these materials are available in digital form, they are usually not structured to support automated reasoning, traceability, or even code-compliance checking [2]. In particular, the formal review is identified as an essential step in building permit review, because missing documents and information delay the permit procedure, as it is depicted in the study of Fauth et al. [3] reporting that 65 % of building applications were delayed because of missing or erroneous application documents. Moreover, previous research has identified the relevant taxonomy for the building permit process [4], which has been incorporated into an ontology representing processes, stakeholders, and documents in the building permit review, denominated as the OntoBPR framework. The OntoBPR framework initially included a set of ontologies as the formal semantic representation of permit-related terms [2]. In recent research, this approach was integrated with the Information Container for linked Document Delivery (ICDD) to model not only the documents but also the interconnections between them [5], and here, in particular, between the documents submitted in an application package. While ICDD containers in the OntoBPR framework are generated manually, the objective of this article is to automatically generate information containers and ontology-based structures from the submitted documents and to classify the submitted documents with metadata to facilitate formal checks of the application's completeness.

Recent advances in Large Language Models (LLMs) and multimodal AI systems enable automatic extraction, structuring, and semantic classification of information from heterogeneous, unstructured documents [6]. These capabilities allow for the generation of metadata and knowledge representations that support automated reasoning, traceability, and completeness checks [7, 8]. The capabilities to generate and summarize text, as well as to adapt to additional given context, also known as few-shot learning, are the significant benefits of LLMs that are identified in the review of Ploennings et al. [9], who analyzed the potentials, challenges, and research directions for applying LLMs and Large Vision Models (LVM) in the Architecture, Engineering, and Construction (AEC) domain. Moreover, document generation and compliance checks were identified as one of the main use cases for Large Language Models (LLMs) in the construction domain in a literature review by Kampelopoulos et al. [10]. Wu et al. [11] argue that LLMs are trained on general knowledge rather than AEC-specific knowledge. While

training an LLM on AEC knowledge is an option, it is impractical due to the enormous costs and training time [11]. A solution to this issue is Retrieval Augmented Generation (RAG), in which information or documents are segmented into smaller chunks, vectorized, and injected into LLMs [11]. Based on the literature, this research aims to employ LLMs and customize them for the use case of analyzing building permit documents by a state-of-the-art RAG implementation. Following up on this aim, the research question formulated in this article is the following: *How can unstructured building permit applications be converted into standard-compliant ICDD containers and OntoBPR-aligned RDF metadata?*

To answer this research question, the remainder of the paper is structured as follows: First, related work on customizing LLMs for specific document processing is provided. Second, the methodology and concept of this approach are described. Third, results from implementing the conceptualized pipeline are presented, and, lastly, the results are discussed and conclusions drawn.

2 Related Work

2.1 OntoBPR and Information Containers

Structuring building permit applications in information containers was introduced with OntoBPR in [2] and offers advantages, including facilitating formal checks on the application, increasing transparency in the permit review by sharing progress updates, and supporting decision-making through rule-based checks of machine-readable reports. The OntoBPR schema is implemented as an ontology, which is available at *https://w3id.org/ontobpr#*. For structuring the submission of a building application, the ontology is instantiated in an ICDD container, which itself is compliant with ISO 21597-1 [12]. Fig. 1 depicts an exemplary OntoBPR ICDD container of a submitted building application in the file system

These ICDD containers have a predefined structure defined through ontologies, thus semantically describing documents and their interconnections. OntoBPR is a semantic layer on top of the generic description of documents, adding context to the meaning of documents and to the purpose of the information container as a building application container. Main classes that are relevant to this study are the `ct:Document` class and its subclasses (c.f. [12]), as well as the classes `obpr:BuildingApplication` and `obpr:Building` from OntoBPR. The OntoBPR framework and the ICDD use a shared technology stack comprising the Resource Description Framework (RDF) and the Web Ontology Language (OWL) for the formal description of the knowledge graph, and Shapes and Constraint Language (SHACL) for validation [2].

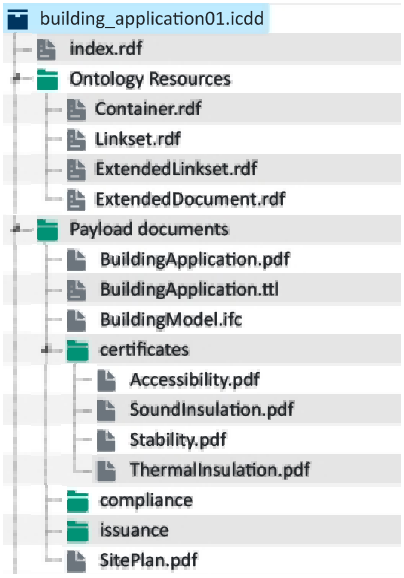


Figure 1. OntoBPR ICDD Container example

2.2 Extraction from Documents

Recent methodological developments have shifted information extraction in the AEC domain from rule-based systems to approaches driven by LLMs. This transition is primarily motivated by the need to handle heterogeneous, complex, and weakly structured data sources that are common in construction projects. The extraction of metadata from engineering drawings remains a longstanding challenge due to limited standardization across documents. Recent research introduces multimodal processing pipelines that combine lightweight Convolutional Neural Networks with visual language models to detect and interpret title blocks and embedded annotations. Empirical evaluations demonstrate clear performance gains over classical optical character recognition techniques, particularly in terms of robustness to layout variability [13].

A further emerging direction is the use of LLMs for knowledge graph construction. This line of work focuses on the automated extraction of entities and relations from unstructured text to populate graph-based representations. Recent studies demonstrate the direct generation of RDF triples from natural language inputs and the population of domain-specific ontologies through zero-shot and few-shot prompting strategies. Applications include knowledge models, e.g., for building materials, indicating the growing maturity of generative approaches for semantic structuring in the AEC domain [14, 15, 16].

2.3 Retrieval Augmented Generation (RAG)

Conventional RAG systems often exhibit limited performance when operating on flat data representations (i.e., lacking explicit structural, relational, or semantic mod-

eling). Therefore, recent research focuses on hybrid retrieval-augmented generation to address the structural complexity of heterogeneous documents that combine text, tables, and code. Structure-aware RAG extends classical pipelines by explicitly modeling structural relations. Approaches such as *GraphRAG*, *StructRAG*, and *LightRAG* integrate knowledge graphs into the retrieval stage. Context retrieval is no longer based on isolated text segments but on linked entities and relations, such as document hierarchies or citation networks. This strategy yields a coherent, semantically connected context. The explicit coupling of neural language models with symbolic knowledge graph structures also aligns these approaches with contemporary interpretations of neuro-symbolic artificial intelligence. Empirical studies report improved answer diversity and superior performance on multi-hop reasoning tasks compared to standard RAG [17, 18].

Ontology-guided retrieval augmented generation addresses domain-specific requirements by embedding formal ontologies into the retrieval and generation pipeline. Systems such as *GrOWL RAG* and *OntologyRAG* use structured domain knowledge to constrain and guide language model outputs, which is particularly relevant for tasks that require strict schema compliance, including document-level relation extraction and code-to-concept mapping. Reported results indicate reduced hallucination rates and increased precision in specialized application scenarios [19, 20]. Hybrid and collaborative RAG aims to overcome the limitations of single retrieval strategies. State-of-the-art systems combine sparse retrieval methods, such as BM25 for lexical matching, with dense retrieval based on vector embeddings for semantic similarity. This combination improves recall and robustness across query types [21, 22]. Recent work introduces a concept called *retrieval augmented retrieval*, in which LLMs generate candidate answers or query expansions prior to retrieval. This mechanism expands the effective query space and improves zero-shot performance in low-resource settings [23].

3 Methodology and Concept

This article presents a proof-of-concept pipeline that uses a hybrid RAG approach to ingest building application documents and generate a compliant ICDD container populated with the OntoBPR knowledge graph. Therefore, a pipeline is implemented and evaluated using a use case as a proof-of-concept. The hybrid RAG approach was chosen after studying several methods of RAG that are referred to in the literature in the previous section.

3.1 Ingestion and Preprocessing

The implemented pipeline comprises four main steps, as shown in Fig. 2, which are described in detail. In the first step, the ingestion component stages and cleans a ZIP file containing a building application submission. The included files are copied into an ICDD-like layout under a case-specific output directory `Payload documents/<CASE_ID>/`. During this process, the system constructs the ICDD container description (index.rdf) using *rdflib*[1]. It creates a `ct:ContainerDescription` individual for the container. Each payload file is registered as a `ct:InternalDocument` with a relative path and MIME type derived from the file extension and heuristics.

The text extraction focuses on PDF documents identified in the first step of the MIME type analysis. The system first attempts to use *pdfminer.six*[2] to read any text layer directly. If this yields insufficient output, an OCR fallback is triggered: pages are rasterized using *pdf2image*[3], and *pytesseract*[4] is applied to recognize text in the resulting images. The extracted text is segmented into overlapping chunks of configurable size. Each chunk is represented by a *TextChunk* record that carries the case identifier, a document identifier, page information, a sequence index, and a generalized role specification inferred from the source filename using a small set of patterns. These chunks, stored in a JSON Lines[5] file, form the basis for retrieval.

3.2 Hybrid Indexing and Retrieval

The retrieval layer combines vector-embedded semantic similarity search with a simple notion of document structure in a temporary knowledge graph. On the vector side, each text chunk is encoded using a SentenceTransformers model (*all-MiniLM-L6-v2*)[6]. The resulting embeddings are stored and indexed for fast cosine-similarity search. Given a textual query describing an information need (for instance, ”application reference” or ”site address”), this index returns chunks whose content is semantically similar, even when the exact query phrase does not appear. On the graph side, a Neo4j[7] database temporarily stores relationships between cases, documents, and chunks. This graph records which chunks belong to which document and preserves their sequence. When answering a query, the system can follow these edges to retrieve contextually neighbouring chunks within the same document or section, thereby complementing vector search with structural context. A dual retrieval function draws from both the vector index and the graph, merges the results, removes duplicates, and orders them into a single context window. This combined context is provided to the LLM for extraction in the next step.

[1] rdflib, v.7.1.4, https://github.com/RDFLib/rdflib
[2] pdfminer.six, v.20250506, https://github.com/pdfminer/pdfminer.six
[3] pdf2image, v.1.17.0, https://github.com/Belval/pdf2image
[4] pytesseract, v.0.3.13, https://github.com/madmaze/pytesseract
[5] see https://jsonlines.org/
[6] https://github.com/huggingface/sentence-transformers
[7] Neo4j, v.5.25.0, https://github.com/neo4j/neo4j

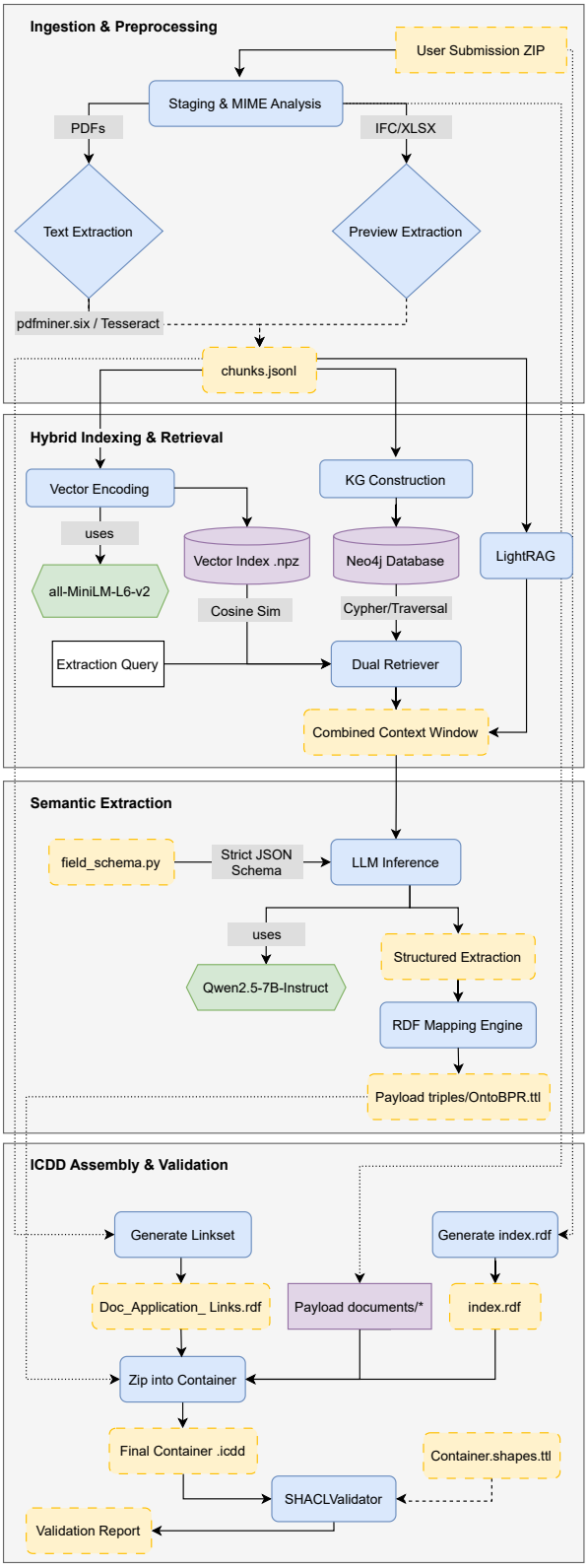


Figure 2. Implementation overview

3.3 Semantic Extraction

In Step 3, the utilized *Qwen 2.5-7B-Instruct*[8] LLM is not required to generate novel RDF structures or arbitrary JSON formats (c.f. also in Fig. 2). Instead, the pipeline relies on a predefined schema, which delineates all relevant data fields. At the current development stage, this schema covers application-level information, such as an internal identifier, application type, authority name, and site address, as well as building-level information, including the building's intended use and, where available, simple quantitative descriptors. For each case, the prompt provided to the LLM combines both the dual retrieval context and the JSON schema. The accompanying instructions emphasize that the model must rely exclusively on the supplied context, leave fields blank when no information is available, and produce syntactically valid JSON that is schema-compliant. The result is a structured JSON document containing the predicted values for all fields, organized under entities such as `building_application` and `building`. These JSON files are stored alongside the other case-specific outputs and subsequently serve as inputs to the OntoBPR mapping and evaluation script.

The JSON extraction in Step 5 is then transformed into an RDF graph aligned with the OntoBPR ontology. The mapping logic implemented using *rdflib* creates individuals representing the building application and, where applicable, the building itself. JSON fields are converted into datatype properties, e.g., an extracted application reference is mapped to `ontobpr:applicationIdentifier` or a site address to `ontobpr:siteAddress`, while the building use type is mapped to `ontobpr:buildingUseType`.

Moreover, for employing also the capability of the LLM to classify the documents based on the extracted terms semantically, OntoBPR does include the class `ontobpr:DocumentClassification` with a set of named individuals with German and English labels, such as `ontobpr:ApplicationForm`, `ontobpr:EngineeringProofs`, `ontobpr:PlotRelatedPlans`, or `ontobpr:Unclassified`.

The mapping uses a fixed set of OntoBPR classes and properties defined in code. The LLM therefore contributes the values, but the ontology vocabulary and URI patterns remain under deterministic control. This choice avoids the common problem of LLMs inventing nonexistent ontology terms (hallucinating) and keeps the RDF graph compliant with OntoBPR for the implemented subset. The resulting triples are written to a Turtle file in the output directory and later included as part of the container's semantic payload.

[8] https://huggingface.co/Qwen/Qwen2.5-7B-Instruct

3.4 ICDD Assembly and Validation

To maintain traceability between semantic data and source documents, a simple linkset is generated using the ICDD linkset vocabulary. This file expresses relationships between documents and, where appropriate, between documents and semantic entities. Before packaging, the combined RDF graph, including the container description, internal documents, linksets, and OntoBPR triples, is validated using SHACL. The prototype employs *pySHACL*[9] together with SHACL shapes `Container.shapes.ttl` derived from the ICDD standard validation resources. Finally, the container description, payload documents, ontology resources, linksets, and generated triples are assembled into a directory and compressed as a *.icdd* file. The overall pipeline and the assembly of generated artifacts are displayed in detail in Figure 2.

4 Proof of Concept and Results

The current prototype has been tested on a representative real-world permit application (`Apl_25-04543-FULL`). The corresponding submission archive is approximately 25 MB in size and consists of twelve distinct files, including text-based PDF forms, scanned or image-based PDFs, and several additional attachments such as images and spreadsheets (see Tab. 1). The documents exhibit typical characteristics of administrative practice: overlapping content across versions, varying terminology for similar concepts, and partial mixing of languages. Some information is distributed across different sources, for example, addresses that appear only in headers or footers, or quantitative data that appears only in drawings.

The system was executed in an environment with a single NVIDIA A100 GPU. *Qwen-2.5-7B-Instruct* was used as the extraction model, and *all-MiniLM-L6-v2* served as the sentence embedding model for the vector index. The hybrid retrieval layer combines dense vector search with a small *Neo4j*-based document graph. The precise configuration details (such as quantization settings or batch sizes) are less critical for the analysis presented here than the overall architectural pattern: retrieval-supported extraction followed by deterministic mapping to RDF.

For the examined case, the pipeline successfully produced an ISO 21597-1 container. The resulting `index.rdf` file lists all payload documents as internal documents, assigns the expected MIME types based on file extensions, and attaches a container description node with the required core metadata. SHACL validation against the available container shapes confirmed that the basic structural constraints of Part 1 are satisfied for this example, after the usual relaxation of strictly closed shapes to allow additional metadata. A simple linkset was generated to demonstrate the integration of Part 2. The linkset associates documents with each other and, where applicable, connects them to semantic entities. In the tested container, these links were successfully resolved to internal paths, and the linkset passed the structural checks applied in this prototype.

[9] pySHACL, v.0.30.1,https://github.com/RDFLib/pySHACL/

Table 1. Composition of the submission archive

Type	Count	Description
Administrative documents	3	Application form, receipt of application, and cover letter.
Planning documents	1	Planning Statement providing policy-based and contextual justification.
Technical drawings and plans	3	Architectural and technical drawings submitted as part of the application.
Supporting documents	6	Consultation, notification, and objection documents, including schedules, photographs, and spreadsheets.
Total	13	**6.53 MB zipped archive**

The OntoBPR payload produced by the mapping step is intentionally modest in scope. For the examined case, the pipeline created a `ontobpr:BuildingApplication` individual and associated it with a small number of datatype properties, such as an application identifier, an authority name, and a site address (see Fig. 3, lines 6–16). From a purely syntactic and ontological point of view, the graph is consistent in that it uses declared OntoBPR properties and datatypes and does not introduce ad hoc vocabulary (hallucination). However, this does not imply that all values are correct or complete; it only confirms that the mapping logic respects the chosen subset of the ontology. The behavior of the hybrid retrieval layer is best understood through concrete examples. For fields whose mentions are relatively explicit in the text, such as the application identifier or the name of the authority, vector-based retrieval generally retrieves appropriate chunks from application forms or covering letters. In such cases, the LLM can fill the JSON schema with plausible values, and the subsequent mapping to OntoBPR is straightforward.

For more distributed or abstract fields, such as building use type (see Fig. 3, line 20), the benefit of combining vector and graph retrieval becomes more apparent. When the relevant information is spread across a design statement and other supporting documents, vector search alone may return passages that are topically related but do not contain a clear, machine-readable statement of the use type. In these situations, extending the context with structurally neighboring chunks from the same document, identified via the graph, can provide the model with enough continuity to extract a meaningful value. This observation is



Figure 3. Excerpt of the generated OntoBPR-KG

qualitative but suggests that structural information helps guide the model toward relevant parts of long documents.

The classification of documents (see Fig. 3, lines 30ff.) according to OntoBPR classified ~39 % of the submitted documents, while eight documents remain unclassified after the procedure. While this classification is not the main achievement of this research, it requires a dedicated evaluation on a large-scale dataset and with variations in the classification descriptions in future work.

The complete pipeline during the case study execution required approximately 170 seconds on the described hardware, with retrieval preparation accounting for the majority of the overall runtime (see Tab. 2). Preprocessing of PDF documents, particularly OCR for scanned pages, accounts for a substantial portion of the overall runtime. Building the vector index and populating the *Neo4j* graph adds additional overhead but remains manageable for a single case. LLM inference for the relatively small schema used here contributes comparatively little to total runtime.

Table 2. Reported runtime for the test case

Pipeline Component	**Runtime (s)**
Ingestion and staging	0.12
Text extraction and chunking	28.00
Retrieval preparation	135.50
Retrieval query execution	3.84
LLM extraction and OntoBPR mapping	1.85
Container generation and validation	0.42
Total Runtime	**169.73**

In terms of container generation, the prototype was able to create basic link relations from the semantic layer back to the primary application document, and SHACL validation showed no structural violations for these links. This demonstrates that the principle of traceability from semantic statements to source documents can be implemented within the ICDD framework.

5 Discussion and Conclusions

OntoBPR was introduced in prior work as an ontology-based framework to support digital building permit review processes. This study extends that work by operationalising OntoBPR through standardized semantic information delivery using the Information Container for linked Document Delivery (ICDD) and by integrating a retrieval-augmented generation (RAG) approach to support information access during the building permit phase. Rather than proposing a new ontology, the study demonstrates how OntoBPR can be embedded into an interoperable, AI-supported submission and review mechanism that addresses fragmentation in digital building permitting.

Many digital permitting approaches remain limited to isolated functions, particularly automated code checking, and rely on document-centric information exchange. Integrating OntoBPR with ICDD bundles heterogeneous documents, datasets, and models into a semantically enriched container, where relationships between regulations, submitted evidence, and procedural steps are explicitly represented using RDF. This shifts the building application from a static document collection to a machine-interpretable information package that supports validation, traceability, and reuse. The integration of a RAG approach further extends this capability. By using the structured semantic content of ICDD containers as a trusted knowledge base, RAG enables context-aware retrieval and synthesis of relevant regulatory provisions and decision-relevant information, supporting building officials while maintaining transparency through semantic grounding.

The pipeline can generate an ISO 21597-1 container with an embedded OntoBPR application graph for a realistic building-permit submission, passing all SHACL checks. The results of this study must be interpreted in light of several intentional design constraints and limitations of the prototype implementation. OntoBPR is employed in a deliberately restricted manner. Only a subset of classes and properties related to the building application and core building attributes is instantiated. More expressive components of the ontology, such as workflow stages, involved actors, decision-making processes, and explicit representations of building-code knowledge graphs, are intentionally excluded in the first step of this research. Con-

sequently, the findings should be understood as a proof-of-concept for semantic extraction and representation using LLMs with the presented hybrid RAG approach, rather than a comprehensive OntoBPR instantiation.

The evaluation dataset is small, and the quantitative metrics produced by the evaluation script should therefore be considered indicative rather than statistically conclusive. The reported values primarily demonstrate overall feasibility and expose typical error patterns in document restructuring, attribute extraction, and semantic mapping. Although the ingestion pipeline is conceptually capable of handling heterogeneous document types, the current experimental focus is predominantly on PDF-based submissions. Other common artifacts in building permit applications include IFC models, which are central to the future development of digital building permitting. These artifacts are not yet analyzed with comparable semantic depth. Extending the pipeline towards multimodal extraction and tighter integration of model-based information remains an important avenue for future research, particularly given IFC's role in digital building permitting.

While this study is a first approach towards the automatic processing of building applications in practice, larger and more heterogeneous datasets in a real-world scale will be required to enable statistically meaningful evaluations and comparative benchmarking, which will be provided in future work. Finally, although the system can operate in a fully automated manner once configured, the generated ICDD containers and RDF graphs still require expert review for critical or legally relevant use. Integrating systematic human-in-the-loop refinement mechanisms for assessing the quality of the generated container and annotations would not only increase trust and adoption but also enable iterative quality improvement through expert feedback loops. Future research must therefore focus on empirical validation through pilot deployments with building authorities, assessing the scalability, usability, and acceptance of OntoBPR-enabled ICDD submissions and RAG-based decision support.

Acknowledgments

This research uses publicly available and research-oriented data sources. Annotated building permit documents were obtained from the ACCORD dataset, supplemented by real-world permits from public application portals (e.g., the City of Westminster). All documents were used solely for research, processed locally, and any identifying information was anonymized or excluded. JF received funding through the R&D project "Development of a process model and identification of innovative approaches to the building permit procedure in the building life cycle" funded by brain-SCC GmbH.

References

[1] Jeroen Werbrouck, Philipp Hagedorn, Donkers, Alex, and Judith Fauth. Semantic Web Technologies for Building Permitting. In Fauth, Guler et al., editor, *Digital Transformation of Building Permitting*. Taylor & Francis, 2026. URL `https://www.routledge.com/9781041091752`.

[2] Philipp Hagedorn, Judith Fauth, Sven Zentgraf, Sebastian Seiß, Markus König, and Ioannis Brilakis. OntoBPR: An ontology-based framework for performing building permit reviews using standardized information containers. *Advanced Engineering Informatics*, July 2025(66), 2025. ISSN 14740346. doi:10.1016/j.aei.2025.103369.

[3] Judith Fauth, Aurica Pötz, Michael Brenner, and Hannes Heitzhausen. Das Baugenehmigungsverfahren und dessen Digitalisierung: eine Bestandserhebung. *Bautechnik*, 102(10):571–580, 2025. ISSN 0932-8351. doi:10.1002/bate.202500011.

[4] Judith Fauth, Tanya Bloch, Francesca Noardo, Nicholas Nisbet, Stefanie-Brigitte Kaiser, Peter Nørkjær Gade, and Jernej Tekavec. Taxonomy for building permit system - organizing knowledge for building permit digitalization. *Advanced Engineering Informatics*, 59:102312, 2024. ISSN 14740346. doi:10.1016/j.aei.2023.102312.

[5] Anne Göbels and Jakob Beetz. Using ICDD Containers for documentation data archives: Semi-automated creation of linked documentation data archives using an Information Container for linked Document Delivery. *Proceedings of the 27th International Conference on Cultural Heritage and New Technologies (CHNT), Vienna, Austria*, 2022.

[6] Chen Han, Yuanyuan Li, and Xijin Tang. DocPolicyKG: A Lightweight LLM-Based Framework for Knowledge Graph Construction from Chinese Policy Documents. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pages 4753–4757, New York, NY, USA, 2025. ACM. doi:10.1145/3746252.3760904.

[7] Kiara M. A. Arevalo, Shruti Ambre, and Rene Dorsch. AutOnto: Towards A Semi-Automated Ontology Engineering Methodology. In *Knowledge Graphs and Semantic Web*, pages 225–241, Cham, 2025. Springer Nature. doi:10.1007/978-3-031-81221-7_16.

[8] Lea Höltgen, Sven Zentgraf, Philipp Hagedorn, and Markus König. Utilizing Large Language Models for Semantic Enrichment of Infrastructure Con-

dition Data: A Comparative Study of GPT and Llama Models. *AI in Civil Engineering*, 4(1), 2025. doi:10.1007/s43503-025-00055-9.

[9] Joern Ploennigs, Markus Berger, Thomas Wortmann, Jakob Kirchner, Jakob Beetz, Alina Roitberg, Karsten Menzel, and Björn Ommer. Building Foundation Models - Potentials, Challenges and Research Directions for Using LLM and LVM in AEC. In *Proceedings of the 2025 European Conference on Computing in Construction*, 2025. doi:10.35490/EC3.2025.268.

[10] Dimitrios Kampelopoulos, Athina Tsanousa, Stefanos Vrochidis, and Ioannis Kompatsiaris. A review of LLMs and their applications in the architecture, engineering and construction industry. *Artificial Intelligence Review*, 58(8), 2025. ISSN 1573-7462. doi:10.1007/s10462-025-11241-7.

[11] Chengke Wu, Wenjun Ding, Qisen Jin, Junjie Jiang, Rui Jiang, Qinge Xiao, Longhui Liao, and Xiao Li. Retrieval augmented generation-driven information retrieval and question answering in construction management. *Advanced Engineering Informatics*, 65:103158, 2025. ISSN 14740346. doi:10.1016/j.aei.2025.103158.

[12] ISO 21597-1. *Information container for linked document delivery: Exchange specification - Part 1: Container*. International Organization for Standardization, Geneva, Switzerland, 1 edition, 2020. URL `https://www.iso.org/standard/74389.html`.

[13] Alessio Lombardi, Li Duan, Ahmed Elnagar, Ahmed Zaalouk, Khalid Ismail, and Edlira Vakaj. Title Block Detection and Information Extraction for Enhanced Building Drawings Search. In *Proceedings of the 2025 European Conference on Computing in Construction*, 2025. doi:10.35490/EC3.2025.344.

[14] Soichi Onozuka and Takaaki Ohnishi. Analysis of LLMs for RDF Triple Generation: Semantic and Syntactic Evaluation Using WebNLG. In *2025 19th International Conference on Semantic Computing (ICSC)*, pages 136–143, 2025. doi:10.1109/ICSC64641.2025.00025.

[15] Teeradaj Racharak, Tongyu Wang, and Watanee Jearanaiwongkul. An automated medical rdf knowledge graph construction from text using in-context learning. In *2024 16th International Conference on Knowledge and System Engineering (KSE)*, pages 465–471, 2024. doi:10.1109/KSE63888.2024.11063495.

[16] Sarah Ayad and Lydia Khelifa Chibout. Ontology Population With Large Language Models (LLMs): A Case Study on Asbestos Ontology. *Applied Ontology*, 20(1):89–106, 2025. doi:10.1177/15705838251334528.

[17] Runsong Jia, Bowen Zhang, Sergio José Rodríguez Méndez, and Pouya G. Omran. StructRAG: Structure-Aware RAG Framework with Scholarly Knowledge Graph for Diverse Question Answering. In *Companion Proceedings of the ACM on Web Conference 2025*, page 2567–2573, New York, NY, USA, 2025. ACM. doi:10.1145/3701716.3717819.

[18] Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. Lightrag: Simple and fast retrieval-augmented generation, 2025. URL `https://arxiv.org/abs/2410.05779`.

[19] Wilma Johanna Schmidt, Diego Rincon-Yanez, Irlan Grangel-González, Adrian Paschke, and Evgeny Kharlamov. Document-Level Relation Extraction with Ontology-Guided RAG. In *Linking Meaning: Semantic Technologies Shaping the Future of AI*, Studies on the Semantic Web. IOS Press, 2025. doi:10.3233/SSW250006.

[20] Hui Feng, Yuntzu Yin, Emiliano Reynares, and Jay Nanavati. OntologyRAG: Better and Faster Biomedical Code Mapping with Retrieval-Augmented Generation (RAG) Leveraging Ontology Knowledge Graphs and Large Language Models, 2025. URL `https://arxiv.org/abs/2502.18992`.

[21] Ye Yuan, Chengwu Liu, Jingyang Yuan, Gongbo Sun, Siqi Li, and Ming Zhang. A Hybrid RAG System with Comprehensive Enhancement on Complex Reasoning, 2024. URL `https://arxiv.org/abs/2408.05141`.

[22] Jaedong Lee, Hyosoung Cha, Yul Hwangbo, and Wonjoong Cheon. Enhancing Large Language Model Reliability: Minimizing Hallucinations with Dual Retrieval-Augmented Generation Based on the Latest Diabetes Guidelines. *Journal of Personalized Medicine*, 14(12), 2024. ISSN 2075-4426. doi:10.3390/jpm14121131.

[23] Tao Shen, Guodong Long, Xiubo Geng, Chongyang Tao, Yibin Lei, Tianyi Zhou, Michael Blumenstein, and Daxin Jiang. Retrieval-augmented retrieval: Large language models are strong zero-shot retriever. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 15933–15946, Bangkok, Thailand, 2024. doi:10.18653/v1/2024.findings-acl.943.

Research on the Structure and Transmission of Enhanced IFC for LLM-Driven Intelligent Design

Yuwen Chen[1], Ziyang Jiang[1],Kaifeng Zhou[1], Zhao Xu[1],

[1]Dept. of Civil Engineering, Southeast University, China
chenyw_seu@seu.edu.cn, jzy916@seu.edu.cn, Leozhou2003@seu.edu.cn, xuzhao@seu.edu.cn

Abstract

With the growing use of artificial intelligence in architectural design, BIM models now need to record not only geometric shapes but also process information and external knowledge for intelligent design and optimization. This paper introduces an enhanced semantic IFC (Industry Foundation Classes) model to transmit process and semantic data during design. By aligning large-scale model data and leveraging RAG for intelligent design, we extend the IFC standard to support multi-stage design exchanges. Additionally, blockchain and AES encryption are integrated into the IFC model to address copyright and security concerns, enabling traceable verification and encrypted storage. This approach improves the intelligence, security, and integrity of architectural design data, offering new directions for the future of information modeling technologies.

Keywords –

Intelligent Design; LLM; Blockchain; Enhanced Semantic; IFC

1 Introduction

Building Information Modeling (BIM) is the cornerstone of digital transformation in architecture, enhancing management accuracy and cross-stage interoperability [1, 5]. Recently, Generative AI and Large Language Models (LLMs) have shifted the design paradigm from manual exploration to human-machine collaboration, offering exceptional capabilities in conceptual generation and parameter recommendation [2, 3]. However, while these intelligent workflows significantly boost efficiency, they introduce complex semantic contexts, reasoning chains, and implicit knowledge that are vital for downstream visualization and review [4].

Despite its maturity, the Industry Foundation Classes (IFC) standard exhibits significant deficiencies in capturing these AI-driven outputs. Existing IFC schemas excel at encoding geometry but struggle to represent design intent, knowledge associations, and the generative reasoning produced by LLMs [6]. This results in semantic fragmentation within BIM-integrated workflows. Current academic efforts—primarily focused on RDF/OWL-based lightweight annotations—lack a comprehensive mechanism to map the multi-dimensional semantic structures of generative models directly to the IFC framework [7].

To bridge this gap, this study integrates Knowledge Graph (KG) technology into the semantic enhancement of IFC [8]. KGs excel at modeling complex entity-relationship structures and support multi-hop reasoning, making them inherently compatible with the graph-based outputs of large models [9]. By leveraging the structured nature of KGs, this approach not only improves semantic retrieval and interoperability but also facilitates the construction of controlled, privacy-protected knowledge bases within industrial BIM practices [10].

Based on the above background and needs, this paper proposes an IFC semantic enhancement and encryption method driven by AI large-model generative design. The main contributions of this study are as follows: First, a systematic analysis of the semantic features of large-model generative design is conducted, identifying key elements in terms of semantic types, structured requirements, and visualization integration. Second, the IFC data schema is extended in a targeted manner, mapping knowledge graph representation mechanisms into IFC or its associated external files, thereby increasing semantic coverage while maintaining compatibility with BIM visualization software. Third, a corresponding encryption verification process is established based on smart contracts and encryption algorithms, ensuring the security of IFC design data during transmission and the adaptability of each design version.

Existing BIM semantic enrichment (e.g., ASK-BIM [11]) fully converts the BIM physical model into a Large-Scale Knowledge Graph (KG) for LLM-based query. However, this approach faces significant challenges in full-scale IFC conversion, token costs during reasoning, and computational redundancy. In contrast, our proposed SE-IFC architecture achieves

critical breakthroughs, as summarized in Table 1.

Table 1. Comparison: ASK-BIM vs. SE-IFC

Item	ASK-BIM	Proposed SE-IFC	Advancement
Graph Strategy	Full-Scale conversion of all physical BIM data	Partial Mounting on-demand semantic enrichment	Token Efficiency & Scalability for LLM reasoning
Structure	Difficult to align BIM with external datasets	On-demand Mounting: Seamlessly links multi-layer external VLM data	High Scalability & Ease of implementation
Reasoning	Monolithic graph lacking versioning capabilities	Records reasoning chains and knowledge-base rules	Full Traceability of the design process

2 Intelligent Design Based on LLM

This study introduces an intelligent design paradigm leveraging the contextual reasoning of LLMs, rather than data-intensive deep learning. By utilizing few-shot prompting with exemplar-based learning, the system generates layouts via linguistic reasoning without manual annotation. To ensure controllability and interpretability, three mechanisms are implemented:

- semantic knowledge-base: Explicitly structures design codes and spatial logics as hard constraints.
- self-verification: Automatically inspects and corrects LLM outputs based on knowledge-base feedback.
- Cross-platform parametric generation: Translates structured instructions into 3D models via BIM APIs

This integrated pipeline—from linguistic reasoning to autonomous verification and parametric modeling—balances model flexibility with practical engineering deployability.

2.1 Preprocessing of Housing Layout Data

To support the LLM-based intelligent design, this study utilizes the 3D-FRONT dataset [12] as a source of high-quality spatial priors. Rather than traditional energy-intensive training, a subset of 980 bedroom and living room layouts is employed to construct a structured example library. Each layout, comprising components like tables, beds, and wardrobes (Figure 1), is preprocessed into a standardized format.



(a)Front view (b)Top view (c)Side view (d)Oblique view

Figure 1. Original indoor layout data

To support the LLM-based design process, 3D-FRONT layout data are preprocessed and standardized into a consistent format. Key geometric features—including room dimensions and object-level attributes (type, size, coordinates, and orientation)—are extracted and serialized into structured JSON. This machine-readable representation ensures that the LLM can efficiently parse both spatial and semantic information through few-shot prompting, facilitating high-fidelity layout generation without manual intervention. Table 2 illustrates the extracted parameters and their structured representation.

Table 2. JSON Schema for Indoor Layouts

Parameter Group	Description	Example
Spatial Envelope	Room type, length, width (m)	LivingRoom, 6.0, 5.0
Object Geometry	Type, L, W, H, Depth (m)	wardrobe, 2.2, 2.4, 1.1
Position	Left/top offsets (m), Orientation (°)	1.8, 1.4, 90°

2.2 Adaptive Knowledge Base and Agent Updates in RAG Framework

In intelligent design systems, especially for complex tasks such as residential layout planning, large language models (LLMs) can efficiently generate design solutions based on preloaded datasets. However, they often rely on fixed rules and historical data, which limits flexibility for iterative and interactive design. Traditional digitized floor-plan methods lack dynamic rule-update capabilities, constraining adaptive design processes. To overcome the rigidity and "black-box" nature of traditional LLMs, this study integrates an external Retrieval-Augmented Generation (RAG) knowledge base. This repository explicitly stores furniture specifications (e.g., 3D-FRONT dimensions), design codes (e.g., clearances), and accessibility constraints. By integrating these knowledge categories into the RAG framework, the system can adaptively update design rules, providing context-aware guidance

that enhances flexibility, safety, and practicality in AI-generated layouts.

Taking residential layout design as an example, the overall process is illustrated in Figure 2. The system employs three functionally distinct agents to collaboratively complete the task. First, the Review agent performs knowledge-base matching and constraint validation on the initial design generated by the LLM, outputting any non-compliant elements in JSON format. Next, the Revision agent leverages these flagged items and invokes built-in operational functions (including move, rotate, and parameter adjustments) to revise and regenerate the layout. Finally, the Modeling agent converts the revised layout into parametric modeling instructions, achieving a complete pipeline from conceptual design generation to practical implementation.

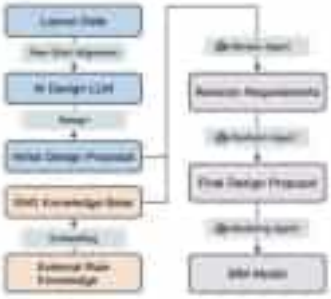


Figure 2. Full process of LLM-Driven intelligent design

2.3 Cross-Platform BIM Visualization

BIM Integration & MCP Execution. The transition from textual layout parameters to physical building components relies on a parametric component library built via BIM software APIs. These components function as a programmable "toolset," ensuring consistent data formats between LLM outputs and internal agents.

As shown in Figure 3, the workflow proceeds from parsing LLM drafts to invoking specific toolset functions. These functions are registered at the Model Context Protocol (MCP) layer and transmitted via TCP communication to a 3D modeling server for execution [13]. Each server-side function maintains a one-to-one mapping with its MCP counterpart, ensuring traceable execution and automated visual reconstruction within the BIM environment.



Figure 3. MCP-Based parametric component visualization in BIM

3 Semantic Enhancement of IFC for LLM-Based Design

In traditional BIM workflows, data interoperability across different software platforms typically relies on the IFC format, which separates geometric information from semantic attributes, ensuring a stable and extensible information structure.However, for the semantic-driven, large-model‐based design paradigm proposed in this study, the native IFC schema is no longer sufficient to capture the full spectrum of design intent, reasoning logic, semantic constraints, and generative context produced by large language models. To address this limitation, we introduce a semantic extension to the traditional IFC format, upgrading it from a geometry-centric exchange standard to a Semantic-Enhanced IFC capable of representing design knowledge, generative context, and the underlying reasoning chain produced during intelligent layout generation.

This study expands the semantic hierarchy of IFC geometry files using the open-source IFCOpenShell library. The enhanced IFC retains the standard spatial organization structure IFCPROJECT → IFCSITE → IFCBUILDING → IFCBUILDINGSTOREY—thereby ensuring full compatibility with mainstream BIM authoring tools.

On this foundation, we introduce two additional semantic carriers: a Display Modification Layer and a Hidden Semantic Layer. These extensions enrich the IFC with design-generation semantics while preserving its original geometric and attribute structure. The overall framework is illustrated in Figure 4, where the blue region represents native IFC content and the pink region denotes the newly added semantic enhancements.



Figure 4. Structural schema of the semantic-enhanced IFC

Display Semantic Layer. Integrated at the IFCBUILDINGSTOREY level, this layer functions as a global annotation container for cross-component design logic. It archives three core elements:

1. **Agent Execution Logs**：Records intelligent agent

types and tool functions (e.g., AgentFunctionCall, ToolName, Timestamp) to ensure full traceability of the generation-validation-optimization chain.

2. **Retrieved RAG Knowledge and Associated Constraint Rules：** Embeds external knowledge retrieved during iterations — such as furniture specifications, safety clearances, and circulation rules — preserving a transparent record of the normative logic guiding design adjustments.
3. **Modification Localization and Interaction Context:** Tracks semantic localization of modifications by logging natural-language prompts and agent rationales. This multimodal documentation transforms the IFC from a static geometry container into an interaction-aware semantic entity.

Through this enriched recording mechanism, the IFC model evolves from a static geometric container into an interpretable, interaction-aware semantic entity.

At the IFCFURNITURE level, an additional Hidden Semantic Layer is introduced to capture deep semantic features that extend beyond the explicit information stored in the display layer. This layer records latent visual and structural attributes that emerge during the later stages of intelligent layout generation, including furniture color composition, material textures, fine-grained structural details, and other non-essential micro-parameters.these types of features are typically difficult to extract directly through conventional parametric design workflows, as their complete representation often depends on the APIs provided by third-party modeling software. However, most commercial BIM platforms expose only limited geometry or attribute parameters, resulting in missing information related to visual details and semantic completeness.

To address this limitation, this study incorporates a Vision-Language Model (VLM) – based semantic enhancement mechanism. Leveraging the visual encoding capabilities of VLMs, the system extracts implicit feature representations of furniture objects directly from rendered images. This enables multimodal alignment between images and textual descriptors, allowing implicit semantic modeling of appearance attributes, material characteristics, and spatial composition relationships—without requiring access to low-level geometric APIs.

Ultimately, the Hidden Semantic Layer serves as a complementary semantic container within the IFC hierarchy. While the display semantic layer focuses on explicit interaction logs and design revisions, the hidden layer concentrates on reconstructing and enriching implicit visual semantics. Together, these two layers establish a dual-semantic structure for the enhanced IFC model, enabling IFC files to describe not only parametric geometric properties but also deep semantic information derived from large-model reasoning. This results in an interpretable and evolvable information representation framework tailored for intelligent design systems.The specific augmented IFC schema is illustrated in Figure 5.

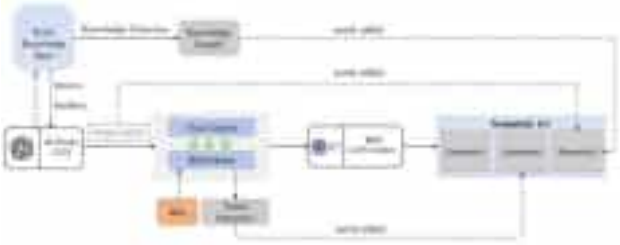


Figure 5. Flowchart of the semantic-enhanced IFC implementation process

Overall, this framework provides a robust foundation for semantic BIM modeling and intelligent reasoning, enabling secure, scalable, and interoperable knowledge integration within intelligent design systems.

4 Blockchain- and AES-Based Secure Management of IFC Files

This paper proposes an innovative IFC information security method that combines blockchain verification with AES encryption to address data integrity and privacy concerns during multi-party collaboration. Rather than conventional full-file encryption or on-chain storage—which introduces prohibitive overhead for large BIM models — our approach employs a selective, layered security architecture that decouples geometric accessibility from semantic confidentiality.

The security framework operates through three distinct layers with strict separation of concerns:

Layer 1: Geometric Data (Off-Chain, Unencrypted)

IFC geometric entities remain in standard cleartext format,Stored via conventional BIM repositories.

Layer 2: Semantic Knowledge (Off-Chain, Encrypted)

Knowledge Graph triples (design intent, reasoning chains, constraint rules) serialized to JSON,AES-256-GCM encryption applied to JSON payload, producing compact ciphertext,Result: Semantic data remains confidential even if IFC file is exfiltrated.

Layer 3: Integrity Verification (On-Chain)

SHA-256 hash of (JSON KG + encryption) anchored to private Ethereum testnet,Only the hash resides on-chain; no geometric data, no semantic content, no IFC file fragments,the complete transmission process follows this Figure 6:



Figure 6. Semantic Encryption Workflow

4.1 The introduction of blockchain technology

In intelligent design applications, ensuring the uniqueness, immutability, and transparency of knowledge graphs is one of the key challenges. Blockchain technology, with its decentralized, tamper-proof, and highly transparent characteristics, provides a feasible technical solution to address these issues. By submitting the JSON sequence of the knowledge graph to the blockchain for smart contract management, each knowledge graph can be effectively recorded, ensuring the reliability and security of its data. The advantages of this approach over traditional IFC-based storage are illustrated in Table 3.

Table 3. Blockchain vs. Traditional IFC Storage

Feature	Traditional IFC Storage	Blockchain Technology
Data Integrity	Lacks immutability	Unique Hash
Traceability	Incomplete/Manual	Interaction-based Blocks
Governance	Centralized (Siloed)	Decentralized (Distributed)

4.2 Construction and Serialization of Blockchain-Based Knowledge Graphs

A primary constraint of blockchain technology is its limited capacity for high-volume data storage, which renders the direct uploading of extensive design documentation computationally expensive and inefficient. To circumvent this, we propose a semantic abstraction and serialization framework. By distilling complex design narratives into structured Knowledge Graphs (Figure 7), we capture essential engineering entities and their interdependencies while stripping away linguistic redundancy. These graphs are subsequently serialized into compact JSON-based triples (Entity – Relationship – Attribute), significantly reducing the storage footprint compared to raw text. This "dehydrated" representation allows for the efficient anchoring of design logic onto the blockchain, ensuring data integrity and traceability without compromising system performance. This approach balances the robust security of decentralized ledgers with the practical requirements of large-scale BIM data management.

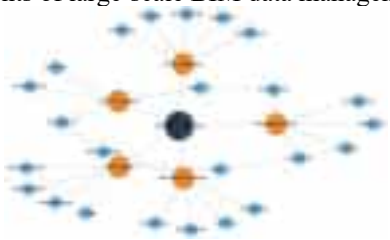


Figure 7. Normalized Knowledge Graph

4.3 Integration of Blockchain-Based Cryptographic Verification in Web-BIM Environment

To circumvent the data redundancy typical of traditional BIM workflows, this study adopts a selective encryption strategy rather than full-scale on-chain storage.he practical implementation within a Web-BIM interface is depicted in Figure 8. As shown in the component property panel, the fields "Knowledge Graph Hash" and "Semantic Ciphertext" confirm that the underlying knowledge graph has been securely processed. Crucially, the geometric data remains decoupled and fully accessible for real-time visualization. This architecture successfully achieves a security-functional decoupling between public geometric representations and private semantic metadata. This ensures that while the model ' s collaborative utility is preserved, sensitive design intelligence is restricted to authorized stakeholders via blockchain-anchored verification.



Figure 8. Layered Encryption Display of BIM

5 Experimental Evaluation

This chapter validates the feasibility of the LLM-driven intelligent design framework and the performance of the SE-IFC data structure. The evaluation focuses on three dimensions: layout reliability, computational efficiency, and blockchain-based cryptographic robustness.

The experimental configuration utilized a workstation equipped with an Intel Core i9-13900H CPU and an NVIDIA GeForce RTX 4070 GPU. On the software and model layer, Qwen3.5-4B was employed as the core design engine, while DeepSeek-VL-7B

served as the vision-language backbone. The programming environment was developed in Python 3.10, utilizing IfcOpenShell (v0.7.0) for the manipulation of IFC 2×3 standardized files. Blockchain integration and smart contract interactions were implemented using the Web3.py library.

5.1 Reliability Assessment of Intelligent Layout Generation

The intelligent design paradigm proposed in this research fully leverages the generative capabilities of Large Language Models (LLMs), enabling the automated transformation from conceptual descriptions to structured spatial layouts through natural language interaction. However, existing studies predominantly employ traditional Natural Language Processing (NLP) metrics—such as ROUGE-1, ROUGE-2, and ROUGE-L to evaluate model performance. These metrics essentially measure n-gram overlap within text sequences, reflecting the surface-level similarity between generated and reference texts rather than the functional rationality or spatial quality of the final design solutions.

This study shifts the evaluation paradigm from "model performance" to a utility-oriented "design outcome assessment," establishing a multi-variate framework that prioritizes spatial physical performance over linguistic fluency. The proposed framework employs three key metrics to quantify the rationality of LLM-generated layouts: Space Utilization Rate, the Coefficient of Variation in Furniture Spacing, and the Minimum Corridor Width. Specifically, the Space Utilization Rate (SUR) measures the ratio of the total furniture area to the floor area, characterizing the "physical saturation" of the space and reflecting the layout's ability to maximize available area while preventing over-congestion. The calculation is defined in Equation (1):

$$R = \frac{\sum_{i=1}^{n} A_i}{L * W} \tag{1}$$

Where **R** represents the Space Utilization Rate , **A** denotes the total area occupied by furniture, and **L** and **W** represent the width and height of the room, respectively.

The Coefficient of Variation in Furniture Spacing (CV) quantifies the structural order and distribution regularity of the layout by calculating the Euclidean distances between the centroids of all furniture pairs. A lower CV indicates higher spatial uniformity, whereas higher values suggest potential clustering or irregular spacing. The calculation is defined in Equation (2):

$$CV = \frac{\sigma_{dist}}{\mu_{dist}} \tag{2}$$

In this， **CV** denotes the Coefficient of Variation of the layout, μ epresents the mean value of the distances between furniture units, and σ signifies the standard deviation.

The Minimum Corridor Width metric simulates the physical clearance between furniture-to-furniture edges and furniture-to-wall boundaries. To ensure the reliability of the spatial analysis, a rationality filter ranging from 0.3m to 3.0m was implemented. This metric primarily evaluates whether the generated layout preserves sufficient circulation space for human movement and accessibility. The calculation process is defined as follows Equation (3):

$$W_{\min} = \min\{\text{gap}(i, j) \mid \delta < \text{gap} < \text{threshold}\} \tag{3}$$

Admittedly, indoor layout design is a relatively open-ended task with high degrees of freedom; therefore, quantitative metrics alone cannot fully capture the nuances of design quality. To address this, this study randomly sampled 100 representative expert designs from the 3D-FRONT dataset to establish a reference benchmark for reasonable spatial distribution. The LLM was then prompted to generate 10 independent layout iterations. If the metrics of the LLM-generated designs fall within the statistical range of the expert samples, the generative results can be considered to possess a high degree of reliability. The statistical distributions of the three key dimensions are illustrated in Figure 9:

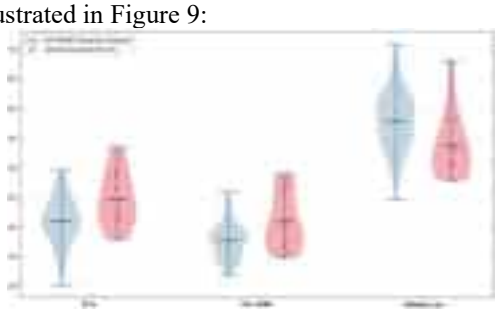


Figure 9. Comparative Distribution of Interior Layout Metrics

Statistical results (Figure 9) indicate that LLM-generated layouts exhibit a high degree of macro-level statistical alignment with expert designs from the 3D-FRONT dataset. Regarding local performance, LLMs enhanced with Retrieval-Augmented Generation (RAG) knowledge bases demonstrate a proactive tendency to increase functional density (e.g., via additional furniture units) to precisely map the multi-scenario requirements defined in user dialogues.

Notably, it was observed that the Minimum Corridor

Width for both expert samples and generated solutions remains within a relatively compact range. This phenomenon is attributed to the inherent spatial constraints of the East Asian residential footprints prevalent in the source dataset. In summary, the current generative model demonstrates remarkable reliability in both semantic synthesis and spatial organization, effectively reaching the statistical thresholds of professional architectural design.

5.2 Storage Overhead Analysis of SE-IFC

To evaluate the SE-IFC framework, we conducted stress tests by scaling metadata by 100x using IfcOpenShell. Storage analysis shows that while a baseline IFC is 3,307 KB, the JSON-serialized semantic layer adds only 79 KB. Due to IFC's internal attribute compression, the integrated package totals 3,349 KB—significantly less than the sum of separate components—confirming minimal storage overhead.

Full-lifecycle loading tests on an i9-13900H workstation (Figure 10) reveal that computational cost is dominated by geometric triangulation (84.3%). The proposed 'Semantic Stream' accounts for only 2.4% of total latency. These results demonstrate that the dual-layer semantic integration achieves seamless model binding with negligible impact on real-time system responsiveness.



Figure 10. Computational Latency Distribution of the BIM Model Loading Pipeline.

5.3 Evaluation and Performance Analysis of BIM Data Anchoring

In this study, a private Ethereum testnet was established using Ganache, with smart contracts deployed via Remix IDE to simulate a real-world BIM data anchoring environment. The experiment involved batch-encapsulating encrypted JSON data packets containing between 10 and 1,000 RDF triples. We recorded the end-to-end latency, encompassing the entire process of "encryption – signing – consensus-based anchoring – retrieval feedback." As illustrated in Figure 11, the total processing time for up to 1,000 triples remained consistently between 0.03s and 0.06s. These results demonstrate exceptional real-time responsiveness and linear scalability, ensuring that the proposed architecture does not impede high-frequency interactions within BIM models.

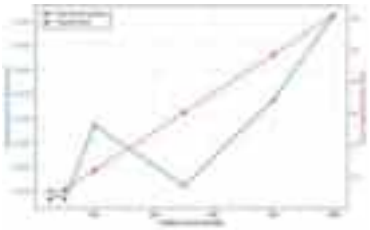


Figure 11. Performance Evaluation of BIM Semantic Data Anchoring

Experimental data further reveals that in a stress-test scenario involving 1,000 BIM semantic triples, the resource consumption for a single anchoring transaction is 1,023,456 Gas. While storage costs on the Ethereum mainnet can be substantial, the smart contracts in this research were authored in Solidity, providing inherent cross-chain interoperability and protocol consistency. This allows the system to dynamically bind to various underlying settlement tokens or blockchain environments (e.g., EVM-compatible Layer 2 solutions or sidechains) based on project budgets and security requirements. In these alternative EVM-compatible environments, the Gas unit price is typically 1% or less than that of the Ethereum mainnet. Consequently, the operational cost for the same scale (1.02M Gas) of BIM data anchoring can be significantly reduced to between 0.07 and 0.28 USD. Given the scale of large-scale construction projects requiring tens of thousands of attribute updates, this cost magnitude offers high economic viability and sustainability.

6 Conclusions and Discussion

In this paper, we propose an IFC semantic enhancement and encryption method driven by large-model generative design. Through a systematic analysis of the semantic features of large-model generative design, we identify key semantic types, structured requirements, and visualization integration elements. The IFC data schema is then extended in a targeted manner, integrating knowledge graph mechanisms to increase semantic coverage while maintaining compatibility with BIM software. Additionally, we design an encryption verification process using smart contracts and encryption algorithms to ensure the security of IFC design data during transmission.

The innovation of this study lies in integrating AI technology with the traditional BIM storage architecture, effectively addressing the limitations of traditional IFC in handling complex design semantics. This approach offers a new perspective for the intelligent transformation of architectural design.

Regarding the inherent limitations of Large Language Models (LLMs) in executing complex architectural design tasks, both academia and industry

have bifurcated into two primary evolutionary trajectories: the "Harness-oriented" approach and the "Model-oriented" approach. The former advocates for the construction of multi-level, progressive skill sets and knowledge base systems to constrain model outputs through external logic. Conversely, the latter focuses on enhancing the domain-specific reasoning capabilities of the models themselves via domain pre-training and fine-tuning.

This study posits that, compared to the standardized environments of code generation, the challenges of "regulatory abstraction" and "data sparsity" within the architectural design domain are significantly more profound. As noted by industry leaders such as Anthropic (Claude), the acquisition and structural complexity of high-quality architectural drawing data far exceed those of general-purpose code datasets. Consequently, developing a robust "External Architectural Framework"—rather than relying solely on increasing model scale—is of paramount importance for enhancing the practical utility of LLMs in the AEC (Architecture, Engineering, and Construction) sector.

The research focus of this study consistently aligns with the "Harness-oriented" (Architectural Augmentation) pathway, dedicated to constructing a comprehensive automated upgrade system. The specific strategies are as follows:

Multi-stage Task Decomposition: By constructing hierarchical systems and serialized command streams, this approach effectively mitigates "attention dispersion" and "logical drift" caused by excessive one-time context injection. This ensures the stability of the generative process across complex design sequences.

Multi-Agent Cooperation Architecture: The framework introduces functionally diverse Agents for parallel processing and cross-verification, establishing a robust error-correction mechanism.

Acknowledgments

This research is funded by National Key R&D Program of China (2021YFF0500900)

References

[1] M. Xie, Y. Qiu, Y. Liang, Y. Zhou, Z. Liu, and G. Zhang. Policies, applications, barriers and future trends of building information modeling technology for building sustainability and informatization in China. *Energy Reports*, 8:7107–7126, 2022.

[2] W. L. Weigang, J. A. Maia, E. Stenzel, and L. R. Siefert. Eixão-UAM: LLM-assisted iterative design of a low-altitude urban air mobility corridor in Brasilia. *Frontiers of Information Technology & Electronic Engineering*, 2025.

[3] C. Li, T. Zhang, X. Du, Y. Zhang, and H. Xie. Generative AI models for different steps in architectural design: A literature review. *Frontiers of Architectural Research*, 14(3):759–783, 2025.

[4] C. Mirarchi, M. Gholamzadehmir, B. Daniotti, and A. Pavan. Semantic Enrichment of BIM: The Role of Machine Learning-Based Image Recognition. *Buildings*, 14(4):1122, 2024.

[5] S. Jiang, X. Feng, B. Zhang, and J. Shi, "Semantic enrichment for BIM: Enabling technologies and applications," *Advanced Engineering Informatics*, vol. 56, p. 101961, 2023.

[6] S. Tang, D. R. Shelden, X. Gao, C. Zhang, J. Hao, F. Guo, and H. Zhang. A framework for integrating BIM, building automation systems, and IoT: An application in design and operation using open standards. *Journal of Building Engineering*, 113:114092, 2025.

[7] Y. Wei, and X. Li. Knowledge-enhanced ontology-to-vector for automated ontology concept enrichment in BIM. *Journal of Industrial Information Integration*, 45:100836, 2025.

[8] G. L. Wang, M. Katsigarakis, D. Mavrokapnidis, I. Korolija, and D. Rovas, "BIM-based semantic enrichment and knowledge graph generation via geometric relation checking," A*utomation in Construction*, vol. 173, p. 106081, 2025.

[9] L. Bahr, C. Wehner, J. Wewerka, J. Bittencourt, U. Schmid, and R. Daub. Knowledge graph enhanced retrieval-augmented generation for failure mode and effects analysis. *Journal of Industrial Information Integration*, 45:100807, 2025.

[10] Z. Wang, H. Ying, R. Sacks, and A. Borrmann. CBIM: A graph-based approach to enhance interoperability using semantic enrichment. *In Proceedings of the 30th International Conference on Intelligent Computing in Engineering*, 2023.

[11] IBBA A, ALONSO R, REFORGIATO RECUPERO D. ASK-BIM: A knowledge graph-powered AI system for natural language querying of BIM models. *Data & Knowledge Engineering*, 2026.

[12] H. Fu, B. Cai, L. X. Gao, L. Zhang, J. Wang, C. Li, Q. Zeng, C. Sun, R. Jia, B. Zhao, and H. Zhang. 3D-FRONT: 3D Furnished Rooms with layOuts and semaNTics. *In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10933–10942, 2021.

[13] B. K. Nithyanantham, T. Sesterhenn, A. Nedungadi, S. Peral Garijo, J. Zenkner, C. Bartelt, and S. Lüdtke. MCP4IFC: IFC-based building design using large language models. *arXiv preprint arXiv:2511.05533*, 2025.

Exploiting Point Cloud Semantic Segmentation for Automatic Geometric Measurement in Modular Construction

Hongxu Chen[1], Tingtian Li[1], Jie Hong[1], Xiao Li[1]

[1]CI[3] Lab, Department of Civil Engineering, The University of Hong Kong, Hong Kong SAR, China

hongxu.chen@connect.hku.hk, tingtian@hku.hk, jiehong@hku.hk, shell.x.li@hku.hk,

Abstract -

Geometric measurement is critical for Modular Construction (MC) after fabrication because it helps prevent dimensional mismatches and safety risks that might otherwise be discovered only during on-site assembly. Terrestrial laser scanning (TLS) provides dense 3D measurements, but automated measurement in cluttered MC scenes remains challenging without robust semantic understanding and geometric reasoning. This paper presents an automated TLS point-cloud-based measurement framework for MC, consisting of data collection, deep-learning semantic segmentation, and geometry-driven post-processing. We evaluate PointNet, PointNet++, and Point Transformer V3 under transfer-learning and training-from-scratch settings. Using the segmented point clouds, we estimate room dimensions, assess wall verticality and orthogonality, estimate window dimensions, and evaluate wall flatness using RMS and point-to-plane distance metrics. Experiments on representative MC rooms demonstrate that the proposed pipeline achieves practical measurement accuracy and produces full-coverage deviation maps to locate nonconforming regions.

Keywords -

Modular Construction; Geometric Measurement; Semantic Segmentation in Point Clouds

1 Introduction

Modular Construction (MC) is an innovative and efficient building paradigm in which fully finished or semi-finished functional modules, including structural elements, interior fittings and utility systems, are prefabricated in factories, transported, and assembled on site [1]. However, integrating substandard MC modules can introduce quality and safety risks, potentially leading to demolition, repair, or replacement, which in turn delays construction schedules and increases costs. Therefore, geometric measurement is essential in MC projects.

Typical geometric tasks include room-dimension measurements, verification of the dimensions and placement of doors and windows, and assessment of wall surface flatness.

Conventional geometric measurement in the construction industry largely relies on manual measurements using tools such as measuring tapes and levels. These procedures are time-consuming and highly dependent on workers' experience and skill levels [2]. To mitigate these limitations, emerging sensing technologies, including computer vision [3, 4], laser scanning [5, 6], and infrared measurement technologies [7, 8] have been increasingly explored for construction geometric measurement.

Image-based approaches are easy and cost-effective to deploy and enable rapid data collection. However, due to inherent limitations in imaging and depth recovery, they often struggle to reach the precision required for fine-grained geometric inspection [9]. In contrast, terrestrial laser scanning (TLS) provides accurate 3D spatial information and has been widely used for measuring building elements such as doors and windows [10].

Early studies demonstrated the feasibility of TLS for automated dimensional inspection of prefabricated components. Kim et al. [11, 12] developed TLS-based pipelines to measure the dimensions, positions, and squareness of precast concrete elements by combining geometric feature extraction, error compensation, and BIM-based deviation analysis, achieving millimeter-level accuracy under controlled conditions. To extend TLS-based inspection to indoor environments, Tan et al. [13] proposed a geometry-driven approach to evaluate wall flatness, verticality, and door/window opening dimensions. Despite their accuracy, such methods typically rely on strong geometric assumptions and relatively clean scenes with limited occlusion, which restricts their applicability to complex and cluttered MC environments.

To improve robustness in unstructured scenes, recent studies have explored learning-based semantic understanding and component detection in point clouds. Tang et al. [14] integrated 3D deep learning with morphological processing to reconstruct semantically enriched BIM models from LiDAR and RGB-D data. Boudarbala et al. [15] proposed an automated workflow for door/window detection by jointly exploiting geometric and radiometric cues. Zhang [16] further addressed limited labeled data by training RandLA-Net on BIM-generated synthetic point clouds with data augmentation; real scans were aligned to the

BIM model via manual selection of feature points, after which the trained network was used to segment on-site point clouds and quantify dimensional deviations from the BIM design using cloud-to-cloud distance metrics. Despite improved automation and semantic understanding, these methods often rely on manual tuning or human intervention (e.g., achieving high-accuracy alignment between BIM-derived and as-built point clouds), and may provide limited accuracy for high-precision dimensional measurement.

Beyond dimensional measurement, TLS has also been applied to surface-related indicators. Wei et al. [17] quantified surface flatness of tunnel initial supports using deviation-based metrics on TLS data. Li et al. [18] developed a full-coverage TLS-based flatness inspection method for large-scale industrial floors by virtualizing the traditional 2-m ruler and feeler-gauge procedure on dense point clouds. However, existing TLS-based surface flatness methods are often tailored to relatively structured and simplified scenarios, making them difficult to extend directly to component-level surface inspection in cluttered, diverse, and heterogeneous environments.

In summary, existing workflows for building inspection can be broadly categorized into purely geometry-driven methods and learning-based point-cloud segmentation methods. Geometry-driven approaches can be accurate under controlled conditions but rely on strong assumptions and are best suited to isolated components in structured scenes. Learning-based methods improve automation, yet in complex MC environments they may suffer from degraded segmentation and limited accuracy for fine-grained geometric measurement. Moreover, many pipelines still require substantial manual tuning and support only a limited set of component types.

To address these limitations, this paper proposes an automated geometric measurement framework based on TLS point clouds for MC applications. We adopt a more efficient point-cloud semantic segmentation model to improve component-level recognition accuracy in MC scenes, and develop more geometry-aware post-processing algorithms to enable reliable room-scale dimensional measurement. The main contributions of this paper are summarized as follows:

- A real-world TLS dataset of MC modules with diverse structural configurations is collected, providing practical data support for geometric measurement research in complex MC environments.
- An automated geometric measurement framework for MC is proposed based on TLS point clouds that integrates semantic segmentation with geometry-aware post-processing to enable robust component-level geometric measurement and analysis.

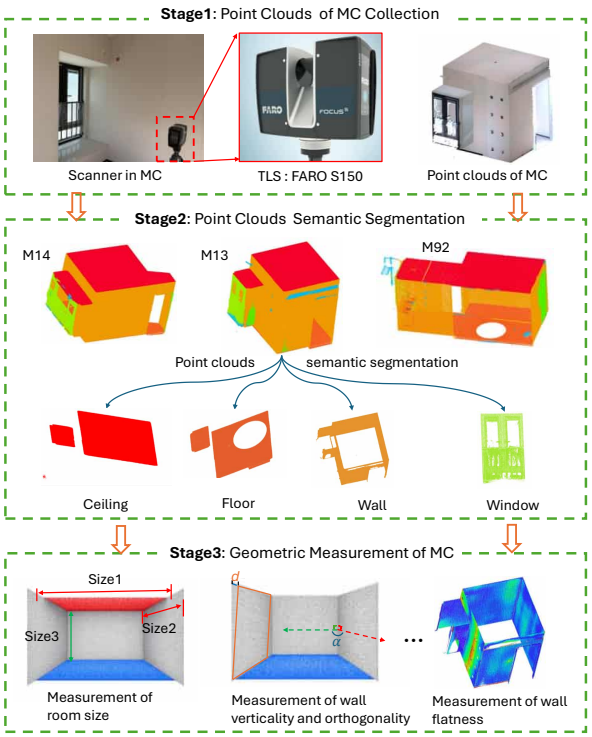


Figure 1. Proposed MC geometric measurement pipeline. TLS-acquired point clouds are first segmented by a semantic segmentation model, and geometric measurements are then performed for each predicted class.

2 Methodology

This paper proposes an automated geometric measurement workflow tailored to MC scenarios, as illustrated in Fig. 1. The framework consists of three stages. First, TLS point clouds of MC modules are acquired and organized. Second, a neural-network-based semantic segmentation model is used to identify key building components in the point clouds. Third, geometry-aware analysis is applied to the segmented components to compute dimensional and surface-flatness metrics, enabling fine-grained inspection in complex indoor environments.

2.1 Semantic Segmentation

To automatically identify MC components in raw scanner point clouds, we employ neural network-based semantic segmentation. Given the variety of MC elements and the complexity of indoor scenes, learning-based segmentation offers a practical alternative to purely geometry-driven pipelines. We evaluate three representative point cloud segmentation models on the proposed MC dataset: PointNet [19], PointNet++ [20], and the recently proposed

Point Transformer V3 [21]. These models cover different design paradigms, including point-wise global feature learning (PointNet), hierarchical local feature aggregation (PointNet++), and transformer-based representation learning for large-scale point clouds (Point Transformer V3).

Each point is represented using scanner-provided attributes, including 3D coordinates (xyz), RGB color, and surface normals (nx, ny, nz). These features are concatenated as the network input to capture both geometric cues and appearance-related information of MC components.

To evaluate learning-based semantic segmentation on MC point clouds, we consider two training strategies: (1) transfer learning from a large-scale indoor dataset followed by staged fine-tuning on MC data, and (2) training from scratch using only the collected MC dataset. All models are evaluated using the same train–test split.

Pretraining on S3DIS and two-stage fine-tuning. Given the limited amount of annotated MC point cloud data and the similarity between MC scenes and indoor environments, we initialize the networks with S3DIS-pretrained weights. The segmentation head is re-initialized to predict the MC classes. Fine-tuning on the MC dataset is performed in two stages to stabilize optimization:

- **Stage I (head-only training):** the backbone is frozen and only the segmentation head is trained on the MC dataset to adapt to the new class taxonomy.
- **Stage II (full fine-tuning):** the backbone is unfrozen, and the entire network is further fine-tuned on the MC dataset to better fit MC-specific geometry and component distributions.

Training from scratch on MC data. In addition to transfer learning, we train each model from scratch using only the MC dataset as an alternative training strategy. In this setting, all network parameters are randomly initialized and optimized end-to-end on MC data under the same train–test split and evaluation protocol.

Segmentation performance is evaluated using Intersection over Union (IoU), defined as the ratio of the intersection between the predicted segmentation and the ground-truth region to their union. A higher IoU indicates more accurate and complete segmentation of the MC components, thereby reducing subsequent measurement errors.

2.2 Geometric Measurement

This subsection presents geometric measurement modules operating on segmented MC point clouds, including (i) room size measurement, (ii) ceiling/floor elevation range evaluation, (iii) wall flatness evaluation, (iv) verticality/orthogonality assessment, and (v) window size measurement.

2.2.1 Room Size

Given a semantically segmented point cloud $\mathcal{P} = \{(\mathbf{x}_i, \ell_i)\}_{i=1}^{N}$, we estimate the room dimensions (length, width, and height) using the ceiling, floor, and wall regions, which provide stable planar references. The corresponding point subsets are $\mathcal{P}^{\text{ceil}} = \{\mathbf{x}_i \mid \ell_i = \ell_{\text{ceil}}\}$, $\mathcal{P}^{\text{floor}} = \{\mathbf{x}_i \mid \ell_i = \ell_{\text{floor}}\}$, and $\mathcal{P}^{\text{wall}} = \{\mathbf{x}_i \mid \ell_i = \ell_{\text{wall}}\}$.

For each structural subset, isolated points are first removed using a radius-based filter. Plane candidates are then extracted by RANSAC, and DBSCAN is subsequently applied to separate spatially disconnected but geometrically coplanar fragments. For each candidate plane, only the largest connected component is retained, and its parameters are finally refined by least-squares fitting (SVD) to obtain a unit normal $\mathbf{n}$ and centroid $\mathbf{c}$.

For the ceiling and floor subsets, the dominant plane is selected as the coherent patch with the largest number of inliers, yielding $(\mathbf{n}^{\text{ceil}}, \mathbf{c}^{\text{ceil}})$ and $(\mathbf{n}^{\text{floor}}, \mathbf{c}^{\text{floor}})$, respectively. For the wall subset, an iterative RANSAC procedure is used to extract multiple planar candidates. After connected-component filtering and SVD refinement, the minimum distance from each wall plane to the floor plane is computed. Only planes satisfying $h_{\min} \le t_{\text{touch}}$ are retained as valid wall candidates. These candidates are then ranked by the number of inliers, and the top four dominant walls are selected for subsequent analysis.

Finally, the room height is computed as the average perpendicular distance from the points in the dominant ceiling patch to the floor plane. The room length and width are estimated as the distances between two pairs of parallel walls among the four selected dominant walls.

2.2.2 Ceiling/Floor Elevation Range

Based on the dominant ceiling/floor plane estimated in Section 2.2.1, we compute a corner-level elevation range using four local patches. Let the plane be $\pi : \mathbf{n}^\top \mathbf{x} + d = 0$ with $\|\mathbf{n}\|_2 = 1$, and let $\mathbf{x}_{\text{corner}}^{(i)}$ $(i = 1, \ldots, 4)$ denote the four 3D corners of the fitted room rectangle. For each corner, we define a square patch (with side length s) whose center is given by

$$\begin{aligned} \mathbf{x}_{\text{center}}^{(i)} &= \mathbf{x}_{\text{corner}}^{(i)} + o\,\widehat{\mathbf{t}}^{(i)}, \\ \widehat{\mathbf{t}}^{(i)} &= \frac{(\mathbf{I} - \mathbf{n}\mathbf{n}^\top)(\mathbf{c} - \mathbf{x}_{\text{corner}}^{(i)})}{\|(\mathbf{I} - \mathbf{n}\mathbf{n}^\top)(\mathbf{c} - \mathbf{x}_{\text{corner}}^{(i)})\|_2} \end{aligned} \tag{1}$$

where $\mathbf{c}$ is the rectangle center and o is an inward offset along the plane.

Using an orthonormal in-plane basis $(\mathbf{u}, \mathbf{v})$, we collect points $\mathcal{X}^{(i)}$ within each $s \times s$ patch centered at $\mathbf{x}_{\text{center}}^{(i)}$ to obtain a robust local estimate. The average point of each patch is computed and its signed distance to the plane is evaluated as $\delta(\mathbf{x}) = \mathbf{n}^\top \mathbf{x} + d$. Finally, the elevation

range is defined as the difference between the maximum and minimum average distances among the four corner patches.

2.2.3 Wall Flatness

We quantify wall surface flatness using a patch-level root-mean-square (RMS) deviation as a global indicator, complemented by a dense per-point distance-to-plane map $\{D_i\}$ to capture localized deviations.

Given a contiguous planar patch extracted in Section 2.2.1, denoted as $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$, we fit a plane via SVD to obtain the centroid $\mathbf{c}$ and unit normal $\mathbf{n}$ ($\|\mathbf{n}\|_2 = 1$). The absolute deviation of each point from the fitted plane is defined as

$$D_i = \left|\mathbf{n}^\top (\mathbf{x}_i - \mathbf{c})\right|, \qquad \mathbf{x}_i \in \mathcal{S} \tag{2}$$

and the patch-level RMS deviation is computed as

$$\mathrm{RMS} = \sqrt{\frac{1}{M}\sum_{i=1}^{M} D_i^2} \tag{3}$$

The set of values $\{D_i\}$ forms a dense deviation map that can be used for visualization and analysis of local surface irregularities.

2.2.4 Wall Verticality and Orthogonality

We evaluate wall geometry quality from two aspects: (i) wall-to-floor verticality and (ii) orthogonality between adjacent walls. The dominant floor plane $(\mathbf{n}_f, \mathbf{c}_f)$ is estimated from $\mathcal{P}^{\text{floor}}$ (Section 2.2.1), where $\|\mathbf{n}_f\|_2 = 1$. Four dominant wall patches $\{\mathcal{W}_k\}$ are extracted using the same decomposition procedure as described in Section 2.2.1, and each patch is refitted via SVD to obtain a unit normal $\mathbf{n}_k$ and centroid $\mathbf{c}_k$.

Wall-to-floor verticality. For each floor-connected wall patch $\mathcal{W}_k$, we define the height of a point $\mathbf{x} \in \mathcal{W}_k$ along the floor normal as $h(\mathbf{x}) = (\mathbf{x} - \mathbf{c}_f)^\top \mathbf{n}_f$. The verticality deviation is quantified by the angular deviation between the wall normal and the floor normal:

$$\theta_k = \arcsin\left(\left|\mathbf{n}_k^\top \mathbf{n}_f\right|\right) \tag{4}$$

To provide an interpretable geometric measure, we further approximate the lateral deviation along the wall height as

$$d_k \approx \theta_k \, (h_{\max} - h_{\min}), \tag{5}$$

where $h_{\min} = \min_{\mathbf{x}\in\mathcal{W}_k} h(\mathbf{x})$ and $h_{\max} = \max_{\mathbf{x}\in\mathcal{W}_k} h(\mathbf{x})$

Wall-to-wall orthogonality. Adjacency between two wall patches $\mathcal{W}_i$ and $\mathcal{W}_j$ is determined by the minimum Euclidean distance between their inlier point sets; a pair is considered adjacent if $d_{ij}^{\min} \le t_{\text{conn}}$. For each adjacent pair, the orthogonality error is defined based on their unit normals as

$$e_{ij} = \left|\arccos\left(\left|\mathbf{n}_i^\top \mathbf{n}_j\right|\right) - \frac{\pi}{2}\right| \tag{6}$$

2.2.5 Window Size

We first extract the window points $\mathcal{P}^{\text{win}} = \{\mathbf{x}_i \mid \ell_i = \ell_{\text{window}}\}$ and apply DBSCAN to remove noise, obtaining a denoised point set $\mathcal{P}^{\text{win}}_{\text{db}}$. A planar model is then fitted to $\mathcal{P}^{\text{win}}_{\text{db}}$ using RANSAC, and the largest inlier set is retained as $\widetilde{\mathcal{P}}^{\text{win}}$. Next, the normal of the window plane is estimated from $\widetilde{\mathcal{P}}^{\text{win}}$ via PCA. Among the wall planes extracted in Section 2.2.1, the one whose normal is most parallel to the window-plane normal is selected as the supporting wall plane. The points in $\widetilde{\mathcal{P}}^{\text{win}}$ are orthogonally projected onto this wall plane and then mapped to a 2D local coordinate system, yielding the 2D point set C^{2D}.

To reduce the influence of boundary artifacts, we iteratively trim the point set along both axes by removing boundary points whose distances to their nearest interior neighbors exceed a threshold τ.

Finally, we compute the 2D convex hull of the trimmed points and fit a minimum-area bounding rectangle using OpenCV `minAreaRect`. The window width and height are defined as the side lengths of this rectangle.

3 Experiments and Results

3.1 Data Collection and Preparation

The data used in this study were collected at an MC manufacturing base in Zhuhai, China. Terrestrial laser scanning was conducted using a FARO Focus S150 to acquire as-built point clouds of MC modules with different structural types, including both steel-structured and concrete-structured modules. According to the device specifications, the ranging accuracy is better than 0.7 mm within 10 m. Under the adopted scanning configuration, the point density is approximately 87,507 pts/m^2 at a distance of about 1 m from the scanner, corresponding to an average point spacing of roughly 3 mm on surfaces perpendicular to the laser beam direction. In total, 44 point cloud files were collected and organized for subsequent semantic segmentation and geometric measurement experiments.

For semantic segmentation, all point cloud files were manually annotated. The labeling scheme follows the S3DIS class definitions and is extended to better reflect the structural characteristics and functional components specific to MC modules. As a result, the annotation taxonomy was expanded to a total of 22 semantic classes: ceiling, floor, wall, beam, column, window, door, table, chair, sofa, bookcase, board, clutter, electronic hole, electronic box, electronic pipe, water pipe, basin, toilet, ceramic tile, rebar, and lamp. Here, the electronic-related classes refer to openings, boxes, and conduits associated with electrical installations commonly found in MC modules.

From the 44 collected point cloud files, three representative scenes were selected for testing, and the remaining

Table 1. Semantic segmentation results on the MC TLS test set.

Model	Training data	$mIoU_{rep}$	Ceiling	Floor	Wall	Window	Door
PointNet	S3DIS+MC	0.69	0.93	0.93	0.86	0.59	0.12
PointNet++	S3DIS+MC	0.73	0.93	0.94	0.90	0.69	0.20
PT V3	S3DIS+MC	0.72	0.90	0.93	0.90	0.49	0.40
PointNet	MC	0.73	0.93	0.90	0.86	0.57	0.38
PointNet++	MC	0.57	0.83	0.89	0.84	0.29	0
PT V3	MC	**0.89**	**0.94**	**0.95**	**0.93**	**0.89**	**0.74**

files were used for training. This split was adopted because several scans contain relatively simple, repetitive layouts; as a result, a conventional 80/20 split would yield a test set that is not sufficiently representative. The test set was therefore curated to better reflect typical MC environments and to provide a more meaningful evaluation of generalization. The selected test scenes are shown in Fig. 1 (Stage 2), labeled from left to right as M14, M13, and M92. Although the test set is small, each test scene contains multiple components and cluttered structures commonly observed in MC modules, enabling component-level geometric measurement under realistic conditions.

3.2 Semantic Segmentation in Point Clouds

We evaluate semantic segmentation on the proposed MC TLS dataset using three representative architectures: PointNet, PointNet++, and Point Transformer V3. The results are summarized in Table 1. For concise presentation, Table 1 only reports the classes whose IoU exceeds 0.5 in our experiments. Classes whose IoU remains below 0.5 across all architectures are omitted because their predictions are insufficiently distinguishable from other categories, which can lead to unreliable inputs and reduced accuracy in subsequent geometric measurement. Accordingly, we report $mIoU_{rep}$, defined as the mean IoU over the reported classes.

As shown in Table 1, the highest $mIoU_{rep}$ is achieved by Point Transformer V3 trained from scratch (0.89), with consistently high IoU on the dominant structural classes (Ceiling/Floor/Wall) and strong performance on opening-related components such as *Window* and *Door* (0.89/0.74). Under S3DIS initialization, PointNet++ benefits from pretraining, improving $mIoU_{rep}$ from 0.57 (scratch) to 0.73 (fine-tuned), suggesting that pretrained hierarchical local features can be helpful when MC supervision is limited.

In contrast, Point Transformer V3 shows degraded performance after S3DIS pretraining, with mIoU dropping from 0.89 (scratch) to 0.72 (fine-tuned). A plausible explanation is the domain gap between S3DIS and our MC scenes. Since Transformer-based models rely heavily on global context, the pretrained representations may be biased toward S3DIS-specific scene priors, causing negative transfer on MC data. Therefore, the subsequent geometric measurement experiments use predictions from Point Transformer V3 trained from scratch.

3.3 Geometric Measurement

Based on the semantic segmentation results of MC point clouds, we develop geometry-aware algorithms to measure room size, ceiling/floor elevation range, wall flatness, wall verticality and orthogonality, and window size. The quantitative results are summarized in Table 2, which compares our measurements with manual measurements from the point clouds, and reports the corresponding errors. Detailed analyses of each measurement item are provided in the following sections.

3.3.1 Room Size

For M13, the measured room length, width, and height are 2346 mm, 2569 mm, and 2769 mm, respectively, compared with the corresponding manual measurements of 2344 mm, 2568 mm, and 2769 mm, yielding absolute errors of 2 mm, 1 mm, and 0 mm. For M14, the estimated length, width, and height are 3116 mm, 3057 mm, and 2767 mm, respectively, corresponding to absolute errors of 2 mm, 2 mm, and 1 mm.

These results demonstrate that the proposed method can reliably recover global room dimensions with millimeter-level precision. Compared with Truong-Hong et al. [22], who reported a 2.9% error in ceiling area measurement, our method achieves a substantially smaller ceiling-area error of 0.1% (assuming the ceiling area is approximated by length × width).

3.3.2 Ceiling/Floor Elevation Range

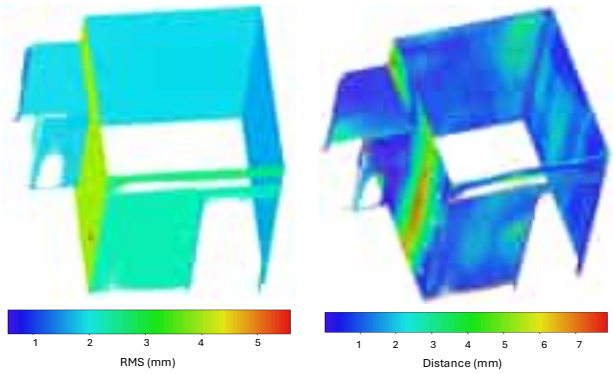
In this study, the inward offset is set to $o = 300$ mm and the patch side length to $s = 10$ mm. As shown in Table 2, the automatically measured ceiling elevation ranges are 2.8 mm for M13 and 2.3 mm for M14, corresponding to absolute errors of 0.3 mm and 0.9 mm relative to manually obtained reference values. For the floor, the measured elevation ranges yield errors of 0.2 mm (M13) and 0.0 mm (M14), demonstrating similarly high measurement accuracy for floor surfaces.

3.3.3 Wall Flatness

Figure 2a presents the RMS flatness of the extracted wall patches in room M13. All major wall patches exhibit an RMS ≤ 4 mm, indicating good overall planarity at the patch level. Among the four walls, the left wall shows the largest RMS value (4.0 mm), primarily due to irregular openings and voids that introduce larger local deviations and consequently inflate the RMS metric. In contrast, the right wall demonstrates the best flatness performance, with

Table 2. Geometric measurement results of major MC components from M13 and M14 point clouds.

Items	M13			M14		
	Measurement	Manual	Error	Measurement	Manual	Error
Length (mm)	2346	2344	2	3116	3114	2
Width (mm)	2569	2568	1	3057	3055	2
Height (mm)	2769	2769	0	2767	2766	1
Ceiling elevation range (mm)	2.8	2.5	0.3	2.3	3.2	0.9
Floor elevation range (mm)	1.3	1.1	0.2	0.9	0.9	0.0
Wall flatness (RMS, mm)	1.6, 1.9 2.3, 4.0	1.7, 1.9 1.8, 3.9	0.1, 0.0 0.5, 0.1	2.6, 2.3 2.1, 3.2	2.4, 1.8 2.2, 4.1	0.2, 0.5 0.1, 0.9
Wall verticality (mm)	0.6, 1.5 1.7, 1.1	0.1, 3.2 0.5, 1.5	0.5, 1.7 1.2, 0.4	5.2, 4.2 1.3, 0.1	5.6, 4.1 1.0, 0.1	0.4, 0.1 0.3, 0.0
Wall orthogonality (°)	0.2, 0.3 0.4, 0.3	0.2, 0.2 0.4, 0.3	0.0, 0.1 0.0, 0.0	0.1, 0.0 0.0	0.1, 0.0 0.0	0.0, 0.0 0.0
Window size (mm)	1730, 1037	1743, 1028	13, 9	1766, 2476	1750, 2469	16, 7



(a) RMS map. (b) Distance-to-plane map.

Figure 2. Wall flatness visualization for the M13 room.

an RMS of only 1.6 mm. When compared to manual measurements, the maximum and minimum absolute errors are 0.5 mm and 0.0 mm, respectively. Furthermore, the proposed method achieves substantially improved measurement accuracy over the work of Li et al. [23], who reported an average flatness measurement error of 7.5 mm for the M4 room in their dataset.

Figure 2b further shows that most wall regions have absolute point-to-plane distances within approximately 1 mm, indicating that the major wall surfaces remain close to an ideal plane. Notably, deviations exceeding 3 mm are mainly concentrated around the opening regions and electrical box areas. Using a practical acceptance criterion of 3 mm for point-to-plane deviation, room M13 is assessed as broadly compliant. Similarly, for room M14, the RMS values range between 2 and 4 mm. Compared to manual measurements, the maximum and minimum absolute errors for this room are 0.9 mm and 0.1 mm, respectively.

Overall, the TLS data and the proposed method enable both quantitative evaluation and intuitive visualization of wall flatness.

3.3.4 Wall Verticality and Orthogonality

In this experiment, we set $t_{\text{touch}} = t_{\text{conn}} = 0.15$ m. For the M13 room, the estimated verticality deviations of the four dominant wall planes are 0.6 mm, 1.5 mm, 1.7 mm, and 1.1 mm, respectively. Compared with the corresponding manual measurements, the absolute errors are 0.5 mm, 1.7 mm, 1.2 mm, and 0.4 mm, yielding a mean absolute error of 0.95 mm. This result is lower than the average verticality measurement error of approximately 1.33 mm reported by Tan et al. [13], demonstrating the higher accuracy of the proposed method. Under the commonly adopted tolerance requirement that the wall verticality deviation should be within 4 mm for walls lower than 3 m, the M13 room satisfies the verticality criterion.

We also assess the orthogonality of adjacent wall planes in room M13. The computed deviations for the four wall pairs are 0.2°, 0.3°, 0.4°, and 0.3°, with corresponding absolute errors of 0.0°, 0.1°, 0.0°, and 0.0° relative to manual measurements. As all values are well below the standard tolerance of 1°, room M13 satisfies the orthogonality requirement with a large margin.

For the M14 room, the estimated verticality deviations are 5.2 mm, 4.2 mm, 1.3 mm, and 0.1 mm, with corresponding absolute errors of 0.4 mm, 0.1 mm, 0.3 mm, and 0.0 mm, respectively. The mean absolute error is 0.2 mm. Although the estimated values are in close agreement with the manual measurements, two walls exceed the 4 mm tolerance limit, indicating that M14 does not fully satisfy the verticality criterion. For wall orthogonality in the M14 room, the computed deviations are 0.1°, 0.0°, and 0.0°, which are consistent with the manual measurements and all well below the tolerance threshold of 1°. Therefore, the M14 room satisfies the orthogonality requirement.

3.3.5 Window Size

Considering the 3 mm point density, we set the threshold $\tau = 3$ mm. For the M13 room, the measured window

height and width are 1730 mm and 1037 mm, respectively. Compared with the manual measurements, the absolute errors are 13 mm in height and 9 mm in width. For the M14 room, the measured window height and width are 1766 mm and 2476 mm, respectively, while the corresponding manual measurements are 1750 mm and 2469 mm. The resulting absolute errors are 16 mm in height and 7 mm in width. Despite these deviations, the results still show good agreement with the manual measurements, indicating that the proposed method can provide reliable window-size estimation in different room layouts.

Compared with Tang et al. [14], who reported door/window location deviations exceeding 120 mm in BIM reconstruction, our proposed method achieves substantially higher geometric accuracy in these case studies.

3.3.6 Error Analysis

This subsection analyzes the measurement errors between the proposed automatic geometric measurement method and manual measurements. Since the scanner used in this study is highly accurate, the analysis focuses primarily on other error sources.

The results show that the measurement errors are small for room dimensions and the geometric attributes of ceilings, floors, and walls. This is mainly because these components achieve high semantic segmentation performance, with IoU values all above 0.93, indicating only limited point misclassification. Moreover, these targets are relatively large, and the proposed post-processing method is robust to minor segmentation errors, thereby maintaining high measurement accuracy. The remaining errors may be partly attributed to manual measurement, which is usually performed on selected points or local regions and may therefore be less comprehensive and more sensitive to local irregularities. In contrast, the proposed method evaluates the geometry of the entire component, providing more comprehensive and stable results.

The window measurement errors are relatively larger than those of other components, mainly because of lower segmentation accuracy for windows. It should be noted that this measurement refers to the installed window within the wall opening rather than the outer structural window frame. Consequently, during semantic segmentation, the boundary between the window and the surrounding wall may be ambiguous, causing some wall points to be misclassified as window points, as observed in the length measurement of M13 and the height measurement of M14. In the current window-size measurement procedure, a rectangle is fitted to the segmented window points, making the result sensitive to such boundary misclassification and potentially leading to overestimation or underestimation. Although denoising strategies are applied to mitigate this effect, their influence cannot yet be fully eliminated.

4 Conclusion

This paper presents an automated TLS point-cloud-based geometric measurement framework tailored to MC scenarios. Using FARO-scanned MC point cloud data, we first perform learning-based semantic segmentation with PointNet, PointNet++, and Point Transformer V3 in both transfer-learning and training-from-scratch settings, and then integrate geometry-driven post-processing for quantitative measurement. The proposed framework supports a range of measurement tasks, including room-level dimensional measurement, evaluation of wall verticality and orthogonality, analysis of ceiling/floor elevation range, window size measurement, and wall surface flatness assessment. Experiments demonstrate practical accuracy for room-scale measurements and effective localization of non-conforming regions in flatness maps, enabling tolerance-based decision-making.

Limitations mainly stem from the current dataset scale and the difficulty of segmenting extremely sparse classes. Future work will expand the dataset, improve robustness to long-tailed class distributions and occlusions, and extend the framework to additional MC geometric-measurement items, such as opening squareness, door/window verticality, and geometric measurements of embedded components (e.g., their dimensions and locations), with more comprehensive validation against industrial standards.

5 Acknowledgments

This study is supported by RISUD, PolyU (P0052942), HKU Seed Fund for Collaborative Research (2407102267) and Innovation and Technology Fund (PRP/048/24LI).

References

[1] Yuanyang Qi, Xiao Li, and Ruiqi Jiang. Vision-based segmentation, measurement, and pose estimation in modular integrated construction. In *Proc. 42nd Int. Symp. Autom. Robot. Constr. (ISARC)*, pages 1097–1104, Montreal, Canada, July 2025. Int. Assoc. Autom. Robot. Constr. (IAARC).

[2] Zhiliang Ma, Yu Liu, and Jiayi Li. Review on automated quality inspection of precast concrete components. 150:104828.

[3] Guofeng Ma, Ming Wu, Zhijiang Wu, and Wenjing Yang. Single-shot multibox detector- and building information modeling-based quality inspection model for construction projects. *J. Build. Eng.*, 38:102216, 2021.

[4] Yangze Liang and Zhao Xu. Intelligent inspection of appearance quality for precast concrete components based on improved yolo model and multi-source data.

Eng. Constr. Archit. Manag., 32(3):1691–1714, 10 2023.

[5] Min-Koo Kim, Jack C.P. Cheng, Hoon Sohn, and Chih-Chen Chang. A framework for dimensional and surface quality assessment of precast concrete elements using bim and 3d laser scanning. *Autom. Constr.*, 49:225–238, 2015. 30th ISARC Special Issue.

[6] Hanbin Luo, Jianjiang Zhan, Henry J. Liu, Wei Zhou, and Weili Fang. A novel uncertainty-aware point cloud approach for geometric quality monitoring in construction. *Adv. Eng. Inf.*, 68:103642, 2025.

[7] Eberechi Ichi and Sattar Dorafshan. Effectiveness of infrared thermography for delamination detection in reinforced concrete bridge decks. *Autom. Constr.*, 142:104523, 2022.

[8] Siyu Zheng, Jiaxin Zhang, Rui Zu, and Yunqin Li. Vision transformer-enhanced thermal anomaly detection in building facades through fusion of thermal and visible imagery. *J. Asian Archit. Build. Eng.*, 24 (4):2854–2868, 2025.

[9] Ali Algadhi, Panos Psimoulis, Athina Grizi, and Luis Neves. Experimental assessment of the performance of terrestrial laser scanners in monitoring the geometric deformations in retaining walls. 15(6):1885–1900.

[10] Sander Varbla, Raido Puust, and Artu Ellmann. Quality evaluation of sizeable surveying-industry-produced terrestrial laser scanning point clouds that facilitate building information modeling—a case study of seven point clouds. *Buildings*, 14(11), 2024.

[11] Min-Koo Kim, Hoon Sohn, and Chih-Chen Chang. Automated dimensional quality assessment of precast concrete panels using terrestrial laser scanning. *Autom. Constr.*, 45:163–177, 2014.

[12] Min-Koo Kim, Qian Wang, Joon-Woo Park, Jack C.P. Cheng, Hoon Sohn, and Chih-Chen Chang. Automated dimensional quality assurance of full-scale precast concrete elements using laser scanning and bim. *Autom. Constr.*, 72:102–114, 2016.

[13] Yi Tan, Xin Liu, Shuaishuai Jin, Qian Wang, Daochu Wang, and Xiaofeng Xie. A terrestrial laser scanning-based method for indoor geometric quality measurement. *Remote Sens.*, 16(1), 2024.

[14] Shengjun Tang, Xiaoming Li, Xianwei Zheng, Bo Wu, Weixi Wang, and Yunjie Zhang. Bim generation from 3d point clouds by combining 3d deep learning and improved morphological approach. *Autom. Constr.*, 141:104422, 2022.

[15] S. Boudarbala, R. Hajji, C. Gourguechon, H. Macher, and T. Landes. Automatic detection of doors and windows using point clouds in the context of as-built bim. *ISPRS Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLVIII-2/W8-2024:37–44, 2024.

[16] Hao Xuan Zhang and Zhengbo Zou. Quality assurance for building components through point cloud segmentation leveraging synthetic data. *Autom. Constr.*, 155:105045, 2023.

[17] Xiao Wei, Jijun Wang, Chunzhong Xiao, Qasim Zaheer, Weidong Wang, Xianhua Liu, Jin Wang, and Shi Qiu. Quantification and evaluation of roughness of initial support using terrestrial laser scanning. *J. Comput. Civ. Eng.*, 39(1):04024052, 2025.

[18] Xianzhe Li, Haiming Li, Ting Jiao, Changzheng Chi, and Mingfeng Huang. Flatness detection of superflat emery floor based on terrestrial laser scanning. *Vibroeng. Proc.*, 54:186–192, April 2024.

[19] R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 77–85, 2017.

[20] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*, NIPS'17, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc.

[21] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler, faster, stronger. In *2024 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 4840–4851, 2024.

[22] L. Truong-Hong and Roderik Lindenbergh. Extracting structural components of concrete buildings from laser scanning point clouds from construction sites. *Advanced Engineering Informatics*, 51: 101490, 2022.

[23] Dongsheng Li, Jiepeng Liu, Shenlin Hu, Guozhong Cheng, Yang Li, Yuxing Cao, Biqin Dong, and Y. Frank Chen. A deep learning-based indoor acceptance system for assessment on flatness and verticality quality of concrete surfaces. *Journal of Building Engineering*, 51:104284, 2022.

DS4Net: Pseudo-Depth Injection and Spectral-Spatial State Space Modeling for Label-Efficient Crack Segmentation

Xiaohu Chen[1], Ruiqiang Xiao[2], Hongyuan Liao[2], Wei Li[1] and Mingzhu Wang[2*]

[1]School of Transportation, Southeast University, China
[2]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, China

[*]Corresponding Author: mwang489@cityu.edu.hk,

Abstract -

Crack segmentation is a critical yet challenging task in structural health monitoring, as cracks typically appear as thin, low-contrast structures that are easily confounded by complex background textures. Although recent deep learning methods have substantially improved crack segmentation accuracy in conventional RGB imagery, they often overfit dataset-specific textures, resulting in poor cross-domain generalization. As an alternative, RGB-D methods provide complementary geometric cues but are limited by costly and imperfect depth sensors. To address these limitations, we propose the Dual-Stream Spectral-Spatial State Space Network (DS4Net), a hardware-free framework that injects pseudo-depth priors estimated by a frozen depth foundation model into crack segmentation. First, instead of physical depth, we introduce a Pseudo-Depth Injection strategy driven by the zero-shot capability of Depth Anything V2, effectively transferring universal geometric priors to the crack segmentation task. Second, to resolve the conflict between global topology and local edge sharpness, we design a Dual-Path Frequency-Mamba (DFM) Block. It synergizes the linear complexity of State Space Models (Mamba) for long-range dependency modeling with a Fast Fourier Transform (FFT) branch that explicitly enhances high-frequency edge details in the spectral domain. Third, to handle cross-modal distribution discrepancy and varying modality reliability, we propose an Information-Theoretic KL-Gating mechanism, which uses Kullback-Leibler divergence to adaptively weight RGB and pseudo-depth features, thereby suppressing unreliable signals and promoting more informative cues. Experimental results show that DS4Net outperforms state-of-the-art RGB and RGB-D methods in both in-domain and cross-domain settings.

Keywords -

Crack segmentation; Pseudo-depth injection; Multimodal fusion; Visual state space model; Domain generalization

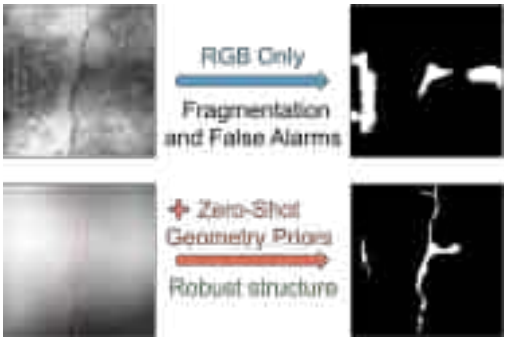


Figure 1. **Motivation**. RGB-only crack segmentation is vulnerable to environmental noise and domain shifts, causing fragmented masks and false alarms. In contrast, pseudo-depth estimated by a depth foundation model reveals more stable geometric crack structures. By injecting such zero-shot geometric priors, DS4Net achieves robust segmentation without auxiliary sensors.

1 Introduction

Critical infrastructure, such as bridges, tunnels, and pavement systems, faces an accelerating aging crisis globally. As the earliest and most prevalent symptom of structural degradation, surface cracks serve as vital indicators for Structural Health Monitoring (SHM)[1]. Traditionally, crack inspection relied on manual visual assessments, which are labor-intensive, subjective, and often hazardous [2]. In recent years, the rapid advancement of deep learning has revolutionized this field, enabling automated pixel-level crack segmentation. State-of-the-art methods, ranging from Convolutional Neural Networks (CNNs) [3, 4] to recent Vision Transformers (ViTs) [5] and State Space Models (SSMs/Mamba) [6, 7], have achieved impressive performance on standard benchmarks.

Despite these advancements, a fundamental bottleneck remains: most existing approaches rely exclusively on the RGB modality. While effective in controlled environments, purely RGB-based models suffer from inherent visual ambiguity in complex, real-world scenarios [8]. Cracks, characterized by low pixel intensity, are often visually indistinguishable from shadows, oil stains, water spots, and dark aggregate textures. This ambiguity leads to two critical issues: (1) In-domain precision loss: Even within standard datasets, models frequently generate false positives in cluttered regions (e.g., mistaking a tire mark for a crack) or fail to detect fine cracks in low-light areas. (2) Cross-domain generalization failure: RGB models tend to overfit to the color distributions and textures of the training domain (e.g., asphalt), leading to severe performance degradation when transferred to unseen materials (e.g., concrete) or varying lighting conditions. This suggests

that relying solely on appearance cues is insufficient for robust and precise segmentation. As shown in Fig 1, under cross-domain settings and challenging illumination, RGB-only crack segmenters often yield fragmented masks and false alarms.

In principle, integrating depth information can resolve this ambiguity, as cracks represent physical geometric discontinuities that are invariant to illumination and surface texture. However, the adoption of RGB-D fusion has been hindered by a practical hardware paradox. High-precision sensors such as LiDAR are prohibitively expensive for large-scale deployment, while consumer-grade depth cameras often suffer from limited resolution and sensor-camera misalignment, introducing artifacts that impair boundary segmentation [9]. Fortunately, the emergence of Vision Foundation Models, such as Depth Anything V2[10], offers a paradigm shift. Trained on massive-scale data, these models possess strong zero-shot capability to estimate fine-grained geometric structures from single RGB images. This presents a unique opportunity: **can we distill these universal, hardware-free geometric priors to improve both the accuracy and robustness of crack segmentation?** Indeed, Fig 1 suggests that pseudo-depth predicted by Depth Anything V2 highlights a more stable crack structure than RGB, making it a promising source of geometric priors for robust segmentation.

To harness this potential, we propose the Dual-Stream Spectral-Spatial State Space Network (DS4Net), a novel framework that synergizes geometric priors with efficient structural modeling. Our approach addresses two key technical challenges. First, simply concatenating RGB and pseudo-depth features is suboptimal due to their distinct characteristics: RGB contains rich texture details but high noise, while pseudo-depth offers clean structure but may lack high-frequency edge sharpness. To resolve this, we design the Dual-Path Frequency-Mamba (DFM) Block. This module combines the linear complexity of State Space Models (Mamba) for modeling long-range crack topology with a Fast Fourier Transform (FFT) branch that enhances high-frequency signals in the spectral domain, preserving fine crack edges. Second, to handle the dynamic reliability of the two modalities (e.g., RGB is unreliable in shadows; pseudo-depth is unreliable on flat markings), we introduce an Information-Theoretic KL-Gating mechanism. By dynamically measuring the Kullback-Leibler divergence between feature distributions, the network adaptively re-weights fusion, suppressing unreliable modality cues and prioritizing the more informative modality under challenging conditions.

The main contributions of this work are summarized as follows:

- We propose a hardware-free RGB-D segmentation framework that leverages zero-shot geometric priors from Foundation Models (Depth Anything V2), significantly enhancing robustness without the cost of physical sensors.
- We design the DFM Block, a hybrid unit that integrates the global modeling capability of Mamba with spectral-domain learning (FFT). This design effectively balances long-range crack connectivity with local boundary precision.
- We introduce an Information-Theoretic KL-Gating mechanism that dynamically regulates multimodal fusion based on distributional divergence.
- Extensive experiments demonstrate that DS4Net achieves State-of-the-Art (SOTA) results on standard benchmarks and exhibits superior robustness in cross-domain scenarios compared to existing CNN, Transformer, and RGB-D-based methods.

2 Related Work

2.1 Crack Segmentation using deep learning methods

Deep learning has shifted crack segmentation from handcrafted pipelines to pixel-level prediction using CNN and Transformer models. CNN approaches such as FCN [11], U-Net [3], DeepLabV3+ [4], and DeepCrack [12] learn strong local cues but may produce fragmented masks for long curvilinear cracks under occlusion or low contrast. Transformer designs such as CrackFormer [13] and SCDeepLab [2] improve connectivity via self-attention, yet their computational and memory costs hinder large-scale training and edge deployment. Both paradigms rely mainly on RGB appearance, making them sensitive to illumination changes and background clutter such as shadows, oil stains, water marks, and patch repairs, which limits cross-domain generalization and motivates the use of complementary cues beyond photometric appearance.

2.2 Visual State Space Models

Visual State Space Models, abbreviated as VSSMs, offer a linear-time alternative to self-attention, capturing global context through selective scanning [14, 15]. Vision adaptations such as VMamba [6] extend selective scanning to multidirectional two-dimensional passes and have been integrated into encoder-decoder segmenters such as U-Mamba [16] and VM-UNet [17] to improve the accuracy-efficiency balance in dense prediction. For crack segmentation, CrackMamba [7] and SCSegamba [18] adapt serialization and scanning to curvilinear topology, improving long-range continuity while remaining more deployment-friendly than Transformers. Nonetheless, existing VSSM-based pipelines operate mainly in the spatial domain and may under-represent high-frequency boundaries of extremely thin cracks, motivating explicit frequency cues.

2.3 Foundation Models for Geometry Estimation

Monocular depth estimation has evolved from supervised learning on scarce RGB-depth pairs to foundation models that learn transferable geometric priors from diverse data. MiDaS [19] improves cross-dataset generalization via mixed-domain training, while foundation depth models further strengthen zero-shot transfer. Depth Anything V2 [10] improves fine-detail recovery through synthetic-to-real distillation, helping resolve thin structural discontinuities. These priors enable hardware-free pseudo-depth for crack inspection without costly RGB-D sensors such as LiDAR [20] or active depth cameras [9], improving robustness and cross-domain generalization.

2.4 Multi-modal Fusion and Frequency Domain Learning

Frequency-domain learning uses the Fast Fourier Transform (FFT) to emphasize high-frequency edges while decoupling low-frequency illumination, benefiting crack segmentation dominated by thin boundaries [21]. Yet frequency cues are rarely integrated with VSSM or Mamba encoders, leaving the combination of linear-complexity topology modeling and explicit edge enhancement underexplored. Meanwhile, RGB and depth priors are often fused by concatenation, summation, or cross-attention [22, 8], but such designs assume both modalities are reliable, even though RGB suffers from shadows and stains and pseudo-depth fails on textureless or reflective surfaces. Measuring cross-modal distributional discrepancy can adaptively weight modalities for robust inspection.

Recent works introduce multimodal cues or frequency analysis for crack segmentation to improve robustness and boundary quality [22, 8, 21]. However, multimodal systems often require extra sensors, while VSSM or frequency designs still rely mainly on RGB. Moreover, spatial token mixing alone fails to explicitly preserve high-frequency edges. To bridge this gap, we transfer foundation depth priors, combine VSSM global modeling with spectral enhancement, and use discrepancy-aware gating to weight modalities, enabling robust crack segmentation.

3 The Proposed DS4Net

3.1 Overview of DS4Net

We propose the **Dual-Stream Spectral-Spatial State Space Network (DS4Net)** to achieve robust crack segmentation without physical depth sensors. The overall architecture of DS4Net is depicted in Fig 2. The architecture follows a hierarchical encoder-decoder design driven by two parallel streams: the RGB Stream (processing texture) and the Pseudo-Depth Stream (processing geometry injected from the frozen Depth Anything V2 foundation model). To effectively fuse these heterogeneous modalities, we introduce two key components at each encoder stage: the **Dual-Path Frequency-Mamba (DFM) Block** for intra-modal feature refinement, and the **Information-Theoretic KL-Gating** for inter-modal adaptive fusion.

3.2 Dual-Path Frequency-Mamba (DFM) Block

Cracks exhibit a unique dual characteristic: they possess global topological continuity but define their presence through high-frequency local edges. To capture these complementary cues, the DFM Block processes the pre-fused feature through two parallel branches: a spatial Mamba branch for modeling long-range topological dependencies and a spectral attention branch for enhancing high-frequency boundary cues. The outputs of both branches are then concatenated and fused by a 1×1 convolution.

3.2.1 Spatial Mamba Branch

We employ the Visual State Space (VSS) module based on the Mamba [14] architecture to capture long-range dependencies with linear complexity $O(N)$. This branch ensures that the connectivity of cracks is preserved even when partially occluded. Since the implementation follows standard SS2D scanning mechanisms [6], we omit the details here for brevity.

3.2.2 Learnable Spectral Attention Branch

Standard convolutions often struggle to decouple high-frequency crack details from low-frequency background variations, such as illumination gradients. We propose a **Learnable Spectral Attention** mechanism to explicitly filter spectral components.

Let $X \in \mathbb{R}^{H\times W\times C}$ be the input feature map. We first transform it to the frequency domain via 2D Fast Fourier Transform (FFT):

$$F(X) = \mathrm{FFT}(X) \in \mathbb{C}^{H\times W\times C} \tag{1}$$

We decouple the complex spectrum into Amplitude $\mathcal{A}(X)$ and Phase $\mathcal{P}(X)$ components:

$$\mathcal{A}(X) = |F(X)|, \quad \mathcal{P}(X) = \angle F(X) \tag{2}$$

To adaptively highlight crack-sensitive frequencies, we employ a channel-wise attention mechanism in the spectral domain. We first aggregate spectral statistics using both Average Pooling and Max Pooling along the channel dimension, followed by a 7×7 convolution layer to generate the attention masks. Formally, the Amplitude Attention Map $M_{\mathcal{A}}$ is computed as:

$$M_{\mathcal{A}} = \sigma\left(\mathrm{Conv}_{7\times 7}\left([\mathrm{AvgPool}_c(\mathcal{A}); \mathrm{MaxPool}_c(\mathcal{A})]\right)\right) \tag{3}$$

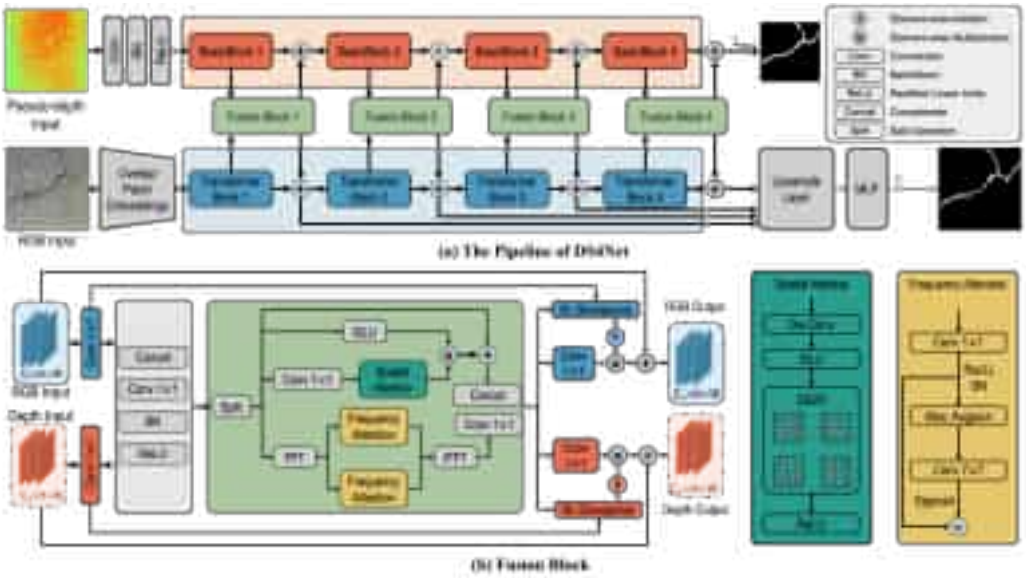


Figure 2. (a) Architecture of the proposed DS4Net, including a dual-stream encoder–decoder with an RGB stream and a pseudo-depth stream generated by a frozen depth foundation model. (b) The fusion block integrates the Dual-Path Frequency-Mamba (DFM) module for spectral–spatial refinement and employs KL-Gating to adaptively reweight cross-modal fusion.

Similarly, the Phase Attention Map $M_{\mathcal{P}}$ is computed as:

$$M_{\mathcal{P}} = \sigma\left(\text{Conv}_{7\times7}\left([\text{AvgPool}_c(\mathcal{P}); \text{MaxPool}_c(\mathcal{P})]\right)\right) \tag{4}$$

where $[\cdot;\cdot]$ denotes concatenation along the channel dimension, and σ is the Sigmoid activation function. These masks effectively learn to suppress background noise in the amplitude spectrum while preserving structural cues in the phase spectrum. The modulated features are reconstructed via Inverse FFT:

$$X_{freq} = \text{IFFT}\left((\mathcal{A}(X) \odot M_{\mathcal{A}}) \cdot e^{j(\mathcal{P}(X)\odot M_{\mathcal{P}})}\right) \tag{5}$$

This branch acts as a frequency-domain high-pass filter, explicitly sharpening crack boundaries before fusion with the parallel Mamba branch.

3.3 Information-Theoretic KL-Gating

Simply adding or concatenating RGB and Pseudo-Depth features assumes both modalities are equally reliable. However, in real-world scenarios, one modality may fail (e.g., RGB in shadows, Depth on flat stains). To address this, we propose an **Information-Theoretic Gating** mechanism that dynamically assesses the "Information Gain" of each modality at the image level.

3.3.1 Global Distribution Modeling

We interpret the feature channels as a semantic probability distribution. Let $F_{rgb}, F_{depth} \in \mathbb{R}^{B\times C\times H\times W}$ be the features from the two branches, while F_{mamba} represents the intermediate feature processed by the Mamba fusion block that serves as a context-aware consensus.

To capture the **holistic reliability**, we first apply Global Average Pooling (GAP) to compress the spatial dimensions, followed by a Softmax normalization along the channel dimension:

$$p_k = \text{Softmax}(\text{GAP}(F_k)), \quad k \in \{rgb, depth, mamba\} \tag{6}$$

where $p_k \in \mathbb{R}^{B\times C}$ represents the global channel activation distribution of modality k.

3.3.2 Divergence-Based Weight Generation

We employ the **Kullback-Leibler (KL) Divergence** [23] to measure how much unique information each modality contributes relative to the fusion consensus (p_{mamba}). A higher divergence implies that the modality contains distinct, potentially complementary information that the current consensus lacks.

$$D_{KL}(p_k||p_{mamba}) = \sum_{c=1}^{C} p_k^c \log \frac{p_k^c}{p_{mamba}^c + \epsilon} \tag{7}$$

This yields a scalar reliability score for each modality. The final fusion weights α_{rgb} and β_{depth} are generated by normalizing these divergence scores:

$$[\alpha_{rgb}, \beta_{depth}] = \text{Softmax}([D_{KL}(rgb), D_{KL}(depth)]) \tag{8}$$

3.3.3 Residual Injection

Finally, the fused features are dynamically re-weighted and injected back into the main streams via residual connections:

$$F_{out}^{rgb} = F_{in}^{rgb} + \alpha_{rgb} \cdot \Psi(F_{mamba}) \quad (9)$$

$$F_{out}^{depth} = F_{in}^{depth} + \beta_{depth} \cdot \Psi(F_{mamba}) \quad (10)$$

where Ψ denotes a projection layer (implemented as 1×1 convolution). This mechanism allows the network to prioritize the modality with higher information value for the current scene, achieving **Sample-Adaptive Fusion**.

4 Experiments and Results

4.1 Dataset

We evaluate DS4Net on a public RGB-T asphalt pavement crack segmentation benchmark [22], which provides RGB images and pixel-wise crack annotations. The benchmark contains 448 images with a resolution of 640×480, and we follow the standard split of 358 images for training and 90 images for testing using a predefined random seed to guarantee experimental consistency and reproducibility. To construct RGB-D inputs without additional sensing hardware, we employ Depth Anything V2 [10] to generate a pseudo-depth map for each RGB image, producing aligned RGB-D pairs for both training and evaluation.

To further assess cross-domain robustness, we additionally introduce the TUT dataset [24] as an unseen test set. Unlike the relatively structured benchmark, TUT features dense and cluttered backgrounds and cracks with more elaborate, intricate geometries, which better reflects challenging real-world conditions. It contains 1,408 RGB images covering diverse materials, backgrounds, and surface conditions; in our experiments, it is used exclusively for cross-domain testing without any fine-tuning.

4.2 Implementation Details

To enhance the diversity of the training data and improve the model's robustness, spatial, color, and numerical transformations were applied to the training set, such as random horizontal or vertical flips in spatial, randomly altering brightness or contrast in color. For images in the validation set, only resizing to 480×480 pixels and normalization processes were applied. The proposed DS4Net was implemented using the PyTorch framework and optimized with AdamW, incorporating a weight decay of 1e-4. A batch size of 8 and 150 training epochs were designated for training. All experiments were performed on an NVIDIA GeForce RTX 4090 for training and testing.

4.3 Evaluation Metrics

To accurately and objectively evaluate the performance of various models, we employed four widely used evaluation metrics in image segmentation: Dice, IoU (Intersection over Union), Accuracy and Precision [25]. Higher values for these metrics indicate superior segmentation performance. Simultaneously, these metrics are also used as the evaluation criteria for the RGB-D asphalt pavement crack segmentation benchmark.

4.4 Comparison with State-of-the-Art Methods

To validate the effectiveness of DS4Net, we compared it with nine representative methods on the Infrared dataset. Specifically, U-Net[3], DeepLabV3[4], DeepCrack[12], and CrackFormer[13] rely only on RGB inputs. In contrast, CRM_RGBT_Seg[26], PotCrackSeg, Evi-RoadSeg, CMNeXt[27], and IRFusionFormer take both RGB and depth images as inputs. Quantitative results are summarized in Table 1 and Table 2. Table 1 reports in-domain performance on the benchmark test set, whereas Table 2 evaluates cross-domain generalization on TUT. Qualitative visualizations are provided in Fig 3.

Table 1. In-domain quantitative comparison on the Infrared dataset. The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Type	Dice ↑	IoU ↑	Acc. ↑	Prec. ↑
Unet	RGB	0.7890	0.6515	0.9794	0.8160
DeepLabV3	RGB	0.8337	0.7149	0.9828	0.8133
DeepCrack	RGB	0.8211	0.6964	0.9831	0.8806
CrackFormer	RGB	0.8487	0.7372	0.9847	0.8458
CRM_RGBTSeg	RGB-D	0.8391	0.7286	0.9809	0.8311
PotCrackSeg	RGB-D	0.8220	0.6978	0.8222	0.8109
Evi-RoadSeg	RGB-D	0.8380	0.7212	0.9775	0.8172
CMNeXt	RGB-D	0.8589	0.7527	0.9808	0.8565
IRFusionFormer	RGB-D	0.8795	0.7850	0.9879	0.8808
Ours	**RGB-D**	**0.8982**	**0.8152**	**0.9898**	**0.9053**

Results on in-domain setting. Table 1 shows that DS4Net delivers the best performance across all four metrics, achieving a Dice score of 0.8982 and an IoU of 0.8152. It substantially outperforms the strongest RGB-only baseline, CrackFormer, with Dice and IoU of 0.8487 and 0.7372, as well as the best RGB-D competitor, IRFusionFormer, with corresponding values of 0.8795 and 0.7850. DS4Net also improves accuracy and precision, suggesting more reliable crack delineation under in-domain setting.

Results on cross-domain setting. As shown in Table 2, DS4Net achieves the strongest cross-domain performance on TUT, attaining the highest Dice score of 0.6806, IoU of 0.5158, and accuracy of 0.9790. Although some methods yield higher precision under domain shift, such as PotCrackSeg, DS4Net maintains strong precision at

Table 2. Cross-domain generalization on the TUT dataset (trained on the Infrared dataset, tested on TUT without fine-tuning). The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Type	Dice ↑	IoU ↑	Acc. ↑	Prec. ↑
Unet	RGB	0.3012	0.1773	0.9507	0.2729
DeepLabV3	RGB	0.3594	0.2191	0.9534	0.3180
DeepCrack	RGB	0.5231	0.3542	0.9687	0.5044
CrackFormer	RGB	0.4694	0.3067	0.9512	0.3576
CRM_RGBTSeg	RGB-D	0.5471	0.4095	0.9466	0.4743
PotCrackSeg	RGB-D	0.3118	0.1847	0.4721	**0.7472**
Evi-RoadSeg	RGB-D	0.5558	0.3848	0.9706	0.5317
CMNeXt	RGB-D	0.5437	0.3733	0.9537	0.3951
IRFusionFormer	RGB-D	0.5340	0.3643	0.9654	0.4648
Ours	**RGB-D**	**0.6806**	**0.5158**	**0.9790**	0.6545

0.6545 while substantially improving overlap-based metrics, which indicates more complete and consistent crack segmentation in cluttered scenes. A category-wise analysis on TUT indicates that DS4Net performs best on the road category, with a Dice score of 0.7158, while alternator, weld, and pipeline remain more challenging, with Dice scores of 0.4895, 0.5398, and 0.5488, respectively.

4.5 Ablation Study

Our DS4Net comprises two core components, the DFM block and the information theoretic KL-Gating mechanism. Accordingly, we conduct ablation studies to assess the contribution of each component under both in-domain and cross-domain settings.

Ablation protocol. We construct three variants by progressively enabling the proposed components while maintaining an identical SegFormer backbone, training schedule, and evaluation protocol. DS4Net (Backbone) denotes an RGB-only baseline that excludes the depth branch and RGB-D fusion, and is used as the reference model. DS4Net (Fixed) introduces the pseudo-depth branch and DFM fusion, but replaces adaptive gating with fixed late fusion, in which RGB and depth features are combined with constant weights of 0.5 and 0.5 at each fusion site. DS4Net (without KL-G) further employs the proposed gating network, but removes KL-based discrepancy supervision, such that the gating is learned solely from the segmentation objective. DS4Net (Full) uses the complete design, incorporating pseudo-depth priors, DFM fusion, and KL-Gating together with KL-based regularization.

Ablation analysis across settings. Quantitative comparisons in Tables 3 and 4, reported using Dice, IoU, Accuracy and Precision, show that incorporating depth cues and fusion improves upon the RGB-only backbone. However, fixed-weight fusion becomes suboptimal when modality reliability varies across spatial locations. Moreover, removing the KL loss consistently degrades performance relative to the full model, which indicates that KL-based discrepancy supervision is critical for learning robust, sample-adaptive fusion and for maintaining stable segmentation quality under domain shift.

Table 3. Ablation study of the proposed components on the Infrared dataset (in-domain). The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Dice ↑	IoU ↑	Acc. ↑	Prec. ↑
DS4Net (Backbone)	0.8518	0.7419	0.9839	0.7939
DS4Net (Fixed)	0.8834	0.7911	0.9885	0.9003
DS4Net (without KL-G)	0.8858	0.7950	0.9886	0.8916
DS4Net (Full)	**0.8982**	**0.8152**	**0.9898**	**0.9053**

Table 4. Cross-domain ablation study on the TUT dataset (trained on the Infrared dataset). All models are trained on the Infrared dataset. The best results are highlighted in **bold**, and the second-best results are underlined.

Model	Dice ↑	IoU ↑	Acc. ↑	Prec. ↑
DS4Net (Backbone)	0.5126	0.3446	0.9580	0.4048
DS4Net (Fixed)	0.6536	0.4855	0.9778	0.6452
DS4Net (w/o KL-G)	0.6669	0.5003	0.9765	0.6046
DS4Net (Full)	**0.6806**	**0.5158**	**0.9790**	**0.6545**

5 Conclusion

This paper presented DS4Net, a hardware-free dual-stream framework for robust crack segmentation that integrates foundation-model geometric priors with efficient spectral–spatial modeling. By injecting pseudo-depth generated from a frozen depth foundation model, DS4Net augments RGB appearance with geometry-aware cues without requiring additional sensing devices. To reconcile the inherent conflict between global crack continuity and local boundary sharpness, we introduced the Dual-Path Frequency-Mamba (DFM) block, which combines linear-complexity state space modeling for long-range topology with an FFT-based spectral attention branch that explicitly strengthens high-frequency edge details. Furthermore, we proposed an information-theoretic KL-Gating mechanism to dynamically regulate cross-modal fusion, enabling the network to suppress unreliable modality signals and emphasize the more informative modality under challenging conditions.

Extensive evaluations on the in-domain benchmark and cross-domain TUT dataset demonstrated that DS4Net achieves state-of-the-art performance and substantially improves generalization under domain shift. Ablation studies further verified the effectiveness of both the DFM block

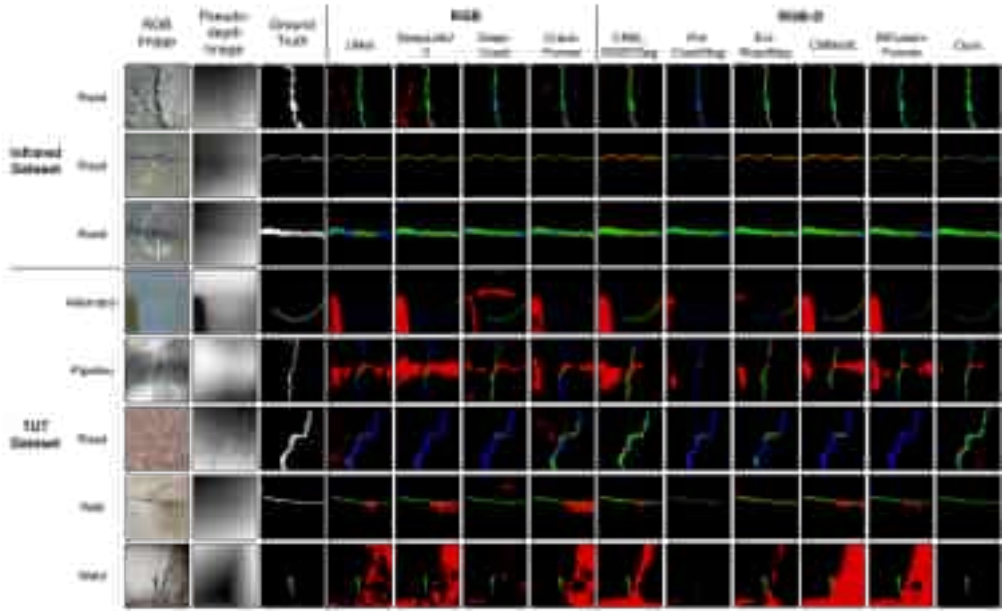


Figure 3. Qualitative results on in-domain and cross-domain scenes. Columns show RGB input, pseudo-depth prior, ground truth, and predictions. Error maps denote TP (green), FP (red), and FN (blue). DS4Net yields more continuous masks with fewer false alarms under noise and domain shift.

and KL-guided adaptive fusion, showing consistent gains when each component is enabled. In future work, we plan to further improve pseudo-depth robustness in textureless, reflective, and low-contrast scenarios, where modality reliability may degrade, explore lightweight designs for real-time edge deployment, and extend the proposed sensor-free paradigm to broader surface-defect segmentation tasks in large-scale infrastructure inspection.

References

[1] Ruiqiang Xiao and Xiaohu Chen. Irfusionformer: Enhancing pavement crack segmentation with rgb-t fusion and topological-based loss. *arXiv preprint arXiv:2409.20474*, 2024.

[2] Zhong Zhou, Junjie Zhang, and Chenjie Gong. Hybrid semantic segmentation for tunnel lining cracks based on swin transformer and convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, 38(17):2491–2510, 2023.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[5] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

[6] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. *Advances in neural information processing systems*, 37:103031–103063, 2024.

[7] Wanqiang Cai, Xudong Wang, Yifan Xue, Yingyao Ma, Jiasong Wu, Zongyuan Ge, and Bin Wang. Crackmamba with normalized soft-frangi-filter enhancement towards accurate crack segmentation. In *Proceedings of the 2025 International Conference on Multimedia Retrieval*, pages 43–51, 2025.

[8] Yingjie Wu, Shaoqi Li, and Yancheng Li. Depth-aware rgb-d concrete crack segmentation and quantification using progressive cross-modal attention. *Measurement*, page 119453, 2025.

[9] Yu-Ting Huang, Mohammad Reza Jahanshahi, Nikkhil Vijaya Sankar, and Fangjia Shen. Affordable, autonomous, and comprehensive road condition assessment using rgb-d sensors: Enhancing pavement condition evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[10] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *Advances in Neural Information Processing Systems*, 37:21875–21911, 2024.

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[12] Qin Zou, Zheng Zhang, Qingquan Li, Xianbiao Qi, Qian Wang, and Song Wang. Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Transactions on Image Processing*, 28 (3):1498–1512, 2019.

[13] Huajun Liu, Jing Yang, Xiangyu Miao, Christoph Mertz, and Hui Kong. Crackformer network for pavement crack segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 24(9):9240–9252, 2023.

[14] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*, 2024.

[15] Badri Narayana Patro and Vijay Srinivas Agneeswaran. Mamba-360: Survey of state space models as transformer alternative for long sequence modelling: Methods, applications, and challenges. *Engineering Applications of Artificial Intelligence*, 159: 111279, 2025.

[16] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.

[17] Jiacheng Ruan, Jincheng Li, and Suncheng Xiang. Vm-unet: Vision mamba unet for medical image segmentation. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.

[18] Hui Liu, Chen Jia, Fan Shi, Xu Cheng, and Shengyong Chen. Scsegamba: lightweight structure-aware vision mamba for crack segmentation in structures. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29406–29416, 2025.

[19] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3): 1623–1637, 2020.

[20] Huifang Feng, Wen Li, Zhipeng Luo, Yiping Chen, Sarah Narges Fatholahi, Ming Cheng, Cheng Wang, José Marcato Junior, and Jonathan Li. Gcn-based pavement crack detection using mobile lidar point clouds. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):11052–11061, 2021.

[21] Zhenguang Zhang, Bo Peng, and Tingyu Zhao. An ultra-lightweight network combining mamba and frequency-domain feature extraction for pavement tiny-crack segmentation. *Expert Systems with Applications*, 264:125941, 2025.

[22] Fangyu Liu, Jian Liu, and Linbing Wang. Asphalt pavement crack detection based on convolutional neural network and infrared thermography. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):22145–22155, 2022.

[23] Jiequan Cui, Beier Zhu, Qingshan Xu, Zhuotao Tian, Xiaojuan Qi, Bei Yu, Hanwang Zhang, and Richang Hong. Generalized kullback-leibler divergence loss. *arXiv preprint arXiv:2503.08038*, 2025.

[24] Hui Liu, Chen Jia, Fan Shi, Xu Cheng, Mianzhao Wang, and Shengyong Chen. Staircase cascaded fusion of lightweight local pattern recognition and long-range dependencies for structural crack segmentation. *arXiv preprint arXiv:2408.12815*, 2024.

[25] Esraa Elhariri, Nashwa El-Bendary, and Shereen A Taie. Automated pixel-level deep crack segmentation on historical surfaces using u-net models. *Algorithms*, 15(8):281, 2022.

[26] Ukcheol Shin, Kyunghyun Lee, In So Kweon, and Jean Oh. Complementary random masking for rgb-thermal semantic segmentation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11110–11117. IEEE, 2024.

[27] Jiaming Zhang, Huayao Liu, Kailun Yang, Xinxin Hu, Ruiping Liu, and Rainer Stiefelhagen. Cmx: Cross-modal fusion for rgb-x semantic segmentation with transformers. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

Integrating RAG-LLMs with BIM for Automated Fall Hazard Detection and Prevention in Building Design

Mudasir Hussain[1,2], Tan Tan[1,2*], Zhuoran Zhang[1,2], and Mei Yalan[1,2]

[1]Department of Real Estate and Construction, The University of Hong Kong, 999044, Hong Kong Special Administration Region of China
[2]National Center of Technology Innovation for Digital Construction (NCTI-DC) Hong Kong Branch, The University of Hong Kong, 999044, Hong Kong Special Administrative Region of China
mudasirh@hku.hk, tant@hku.hk, zhuoran_zhang@connect.hku.hk, yalan_mei@connect.hku.hk

Abstract –

Falls from height (FFH) remain a leading cause of construction injuries and fatalities, necessitating proactive Design-for-Safety (DfS) approaches. While Building Information Modeling (BIM) facilitates safety checks, existing Automated Rule Checking (ARC) models rely on rigid, manual rule-coding and lack the spatial reasoning required to interpret complex safety regulations. This paper proposes a Retrieval-Augmented Generation (RAG) Large Language Model (LLM) pipeline integrated with BIM for automated FFH identification and prevention. The methodology incorporates a domain-specific ontology that maps unstructured regulatory constraints to BIM geometry, enabling the precise detection of spatial hazards. Additionally, the system automates the placement of preventive measures, such as Occupational Safety and Health Administration (OSHA)-compliant guardrails, directly within the design model. By bridging the gap between semantic regulatory interpretation and geometric analysis, this approach overcomes the spatial limitations of current LLM-BIM frameworks, empowering designers to eliminate fall hazards early in the project design.

Keywords –

Retrieval augmented generation; Large language models; Building information modeling; Automated rule checking; Design for Safety; Fall hazards

1 Introduction

The construction industry remains the most hazardous sector worldwide, despite safety improvements, legislation, and technological advancements [1]. Falls from height (FFH) are the leading cause of severe injuries and fatalities at construction sites [2]. Research shows that over 50% of construction accidents stem from design decisions [3], and 60-90% can be prevented through decisions made during the design phase [4]. Consequently, there is a growing shift from reactive, site-based safety management to proactive Design-for-Safety (DfS) approaches that eliminate hazards before construction begins [5].

Building Information Modeling (BIM) offers an effective platform to implement DfS principles by enabling designers to visualize construction scenarios and identify spatial risks [6]. However, current BIM-based safety tools have key limitations, such as high initial costs and interoperability issues between software. It also relies on manual expertise, and designers lack safety knowledge. Similarly, Automated Rule-Checking (ARC) systems use rigid, manually created rules [7]. ARC systems can assess BIM models against safety requirements, but converting safety regulations from natural language to computer code requires extensive programming expertise and regular updates [8]. Furthermore, fall hazard identification is particularly challenging as regulations involve context-based judgments and semantic understanding. Terms like "unprotected edges" and "openings" are difficult to define in traditional logic-based systems. As a result, most existing ARC systems function primarily as post-design compliance checkers rather than adaptive decision-support tools.

Recent advances in Artificial Intelligence (AI), particularly Large Language Models (LLMs), offer new opportunities to address these challenges. LLMs can process complex regulatory texts and, when combined with Retrieval-Augmented Generation (RAG), dynamically reference safety standards. Studies have demonstrated the potential of LLM-based approaches for code compliance and rule extraction [9]. However, existing studies largely frame safety as a general management problem, focusing on guidance generation and inspections rather than specific prevention mechanisms. Current LLM approaches lack spatial reasoning capabilities and cannot accurately identify fall

hazards within building geometry without deep BIM integration. Consequently, FFH regulations, which rely heavily on spatial dependencies, remain inadequately addressed by existing AI-driven frameworks.

To bridge the gap between unstructured regulations and spatial hazard analysis, this paper proposes a RAG-LLM-augmented BIM pipeline for the automated identification and prevention of fall hazards in building design. The approach uses LLMs to interpret safety regulations and connects regulatory logic with BIM geometry to enable automated hazard analysis. By reducing manual rules encoding and providing early safety feedback, this method advances DfS and BIM-based safety automation.

2 Literature review

2.1 Embedding DfS rules in BIM

BIM serves as an effective platform for operationalizing DfS principles through its representation of geometric, semantic, and spatial information. Early research emphasized visualization and coordination for safety planning, while later studies explored embedding DfS knowledge to enable systematic hazard detection and prevention. Advanced implementations utilize rule-based reasoning engines to evaluate designs against encoded safety requirements. Safety ontologies further strengthen these systems by formalizing relationships between building components and risk categories [10]. Despite advances, embedding DfS into BIM faces limitations due to fragmented safety knowledge and rigid rule representations [11]. Most rule libraries are difficult to update or adapt across jurisdictions. Moreover, safety requirements are context-dependent, influenced by construction methods and temporary conditions, and are inadequately represented in static BIM models. Consequently, existing systems function primarily as post-design compliance checkers rather than adaptive decision-support tools.

2.2 Automated rule checking systems

ARC systematically evaluates BIM models against regulations using frameworks like CORENET X, FORNAX™, and Solibri Model Checker (SMC). Various studies have developed specific rule-checking algorithms for fall protection, job hazard analysis, and temporary structure assessment. Others have employed ontology-based schemas and voxelization techniques to enhance hazard detection. However, ARC effectiveness is constrained by reliance on manual rule creation, which requires domain expertise and programming skills. This dependency limits scalability, particularly for regulations involving qualitative descriptions. Fall protection requirements, such as identifying "exposed edges," demand spatial and contextual understanding beyond logic-based models. The gap between unstructured natural language regulations and computational rules highlights the need for adaptive compliance verification methods.

2.3 LLM-Enhanced Compliance Checking

LLMs present a transformative solution to current ARC limitations. Their Natural Language Processing (NLP) capabilities allow for the interpretation of regulatory text and the creation of machine-readable formats without manual coding. When combined with RAG, LLMs can dynamically reference authoritative sources to ensure accuracy. Recent research demonstrates the potential of LLMs in this domain. Zheng et al. [12] and Madireddy et al. [13] developed frameworks for interpreting building codes and checking compliance. Chen et al. [14] and Du et al. [15] focused on rule extraction and code generation. Other studies have integrated knowledge graphs and visual language models to enhance safety monitoring. Despite these advancements, existing studies largely frame safety as a general management and inspection problem. A critical limitation is that current LLM approaches lack spatial reasoning capabilities and cannot accurately identify fall hazards without deep BIM integration. Consequently, FFH regulations involving implicit spatial dependencies remain inadequately addressed. This paper proposes a RAG-LLM-augmented BIM framework to bridge this gap by aligning semantic interpretation with BIM geometry for automated hazard prevention in building design.

3 Methodology

This study proposes a RAG-LLM-augmented BIM pipeline to integrate DfS principles into early design stages. As illustrated in Figure 1, the framework formally bridges regulatory knowledge and BIM geometry through a domain-specific ontology. The workflow proceeds through five integrated phases: (1) collection of fall hazard rules and regulations; (2) fall hazard ontology development; (3) automated fall hazard identification and visualization; (4) rule-based fall prevention via OSHA-compliant guardrail generation; and (5) an intelligent advisory layer powered by RAG-LLM reasoning to provide context-aware safety decision support. The proposed framework employs deterministic, rule-based methods for fall hazard detection and guardrail generation to ensure reliable safety compliance, while the RAG-LLM component serves as an advisory layer providing regulatory knowledge and contextual guidance for user decision-making rather than directly generating geometric elements.

3.1 Fall hazard rules and regulations

To standardize safety requirements, this study analyzed fall hazard regulations from Hong Kong's Labour Department and China's Ministry of Housing and Urban-Rural Development (MOHURD), classifying hazards into three categories: elevated works, openings, and access-related works (Table 1). Although numerous FFH rules exist, this study considered only those predefined in the BIM model, as they can be applied effectively.

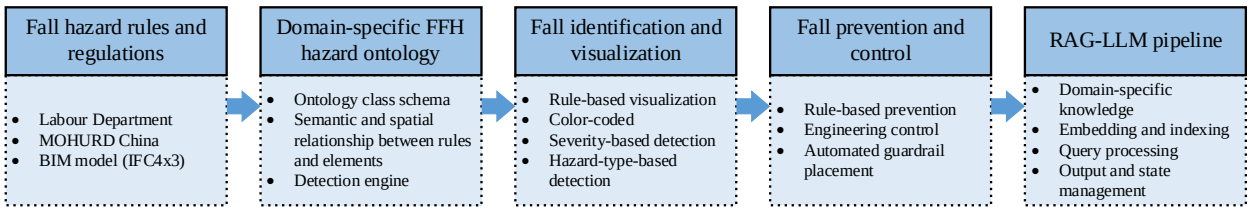


Figure 1. Overview of research methodology.

Table 1. Fall hazard rules and regulations.

Fall hazard category	Specific type	Definition	Threshold (mm)	Regulation
Elevated work hazards	Falls from height	Risk of falling from elevated work areas	≥ 2000	JGJ80-2016
	Unprotected sides and edges	Edges of floors, roofs, or platforms without fall protection	≥ 1800	Reg. 38B, CSSR Part VA
Opening hazards	Floor openings (Horizontal)	Uncovered holes or openings in floors	≥ 500	GB10000-88
	Wall openings (Vertical)	Uncovered holes or openings in walls	≥ 500	GB10000-88
Access hazards	Stairways	Use of stairways without adequate safety measures	≥ 760	Reg. 38C-F, CSSR Part VA

3.2 Fall hazard ontology development

This study uses an ontology-based approach to represent regulatory knowledge and establish semantic and spatial relationships between FFH rules and BIM elements, enabling automated reasoning. A domain-specific ontology was developed through regulatory analysis, BIM-IFC mapping, and rule-based spatial-logic encoding.

By mapping fall hazards to IFC entities (e.g., IfcSlab, IfcOpeningElement, IfcStair), semantic relationships were established through IFC property sets, aggregation hierarchies, and filling relationships. Spatial relationships were derived from bounding-box analysis and perimeter-edge detection. The ontology defines these hazard classes and executes detection through a ten-stage pipeline: IFC file loading, geometry extraction, unit normalization, opening classification, protection analysis, edge detection, stair-railing spatial analysis, threshold comparison, severity assignment, and result export. The pipeline generated JSON output containing OntologyID, SpecificHazard, DetectedValue_mm, Threshold_mm, Severity, Location coordinates, ElementContext, and Notes, supporting visualization, prevention planning, and LLM-based reasoning.

3.3 Fall identification and visualization

The fall hazard identification workflow comprises three steps: data parsing, IFC coloring, and report generation. This approach integrates ontology-driven hazard data (stored in JSON) with the IFC model to produce a color-enriched representation that enables intuitive interpretation of fall risks during design review and construction planning.

First, the hazard visualization controller processes input JSON files containing semantic and geometric attributes, including hazard type, severity, and spatial location. Second, the IFC-based colorizer embeds hazard information by classifying risks by their physical and representational characteristics within IFC. Element-based hazards (e.g., slabs, walls) are visualized through surface styles, while void-based hazards (e.g., floor

openings) are represented using proxy-based 3D markers. Finally, hazard integration is executed using IfcOpenShell for hierarchical element identification. For hazards requiring independent markers, the system generates extruded geometry as IfcBuildingElementProxy objects linked to the building's spatial hierarchy. The outputs include a color-enriched IFC model and JSON and text-based reports summarizing visualization success rates, hazard distributions, marker creation statistics, and IFC-based coloring limitations. These outputs support visual inspection and systematic assessment of fall hazards, meeting automation-focused construction research requirements.

3.4 Fall prevention

Following FFH identification and visualization, the system prioritizes engineering controls by automatically generating OSHA-compliant guardrail systems at detected and identified hazard locations. The prevention system implements three layers: 1) data, 2) logic, and 3) output, using Python libraries: IfcOpenShell for IFC analysis, NumPy for geometric computation, and SciPy for boundary extraction. For stairways, walking edges are identified to place continuous top and mid rails at 1100 mm and 550 mm heights. For elevated slabs, a wall-aware safety classification evaluates edges, and guardrails are installed only where adjacent walls are insufficient (<1070 mm). Components are created as IfcRailing objects with safety coloring and proper storey assignment, enabling proactive BIM-based safety planning.

3.5 Integrated RAG-LLM pipeline

To address the limitations of rigid rule checking, this study proposes a RAG-LLM system that introduces a knowledge-augmented reasoning layer combining BIM hazard data with expert reasoning.

Since current LLMs struggle to interpret complex regulatory requirements outside their pretraining scope, a specialized knowledge base was developed where hazard instances are converted into regulatory standards, prevention best practices, and case-based documents. The hazard instances are structured into five fall hazard types, unprotected sides and edges, floor openings, wall openings, stairways, and FFH, which are derived from regulatory classifications (Table 1) and consistently used across the ontology development, detection, and prevention stages. These knowledge base documents are transformed into dense 384-dimensional vector representations using the *all-MiniLM-L6-v2* model and stored in an in-memory vector database for real-time access.

Upon receiving a user query, the system computed semantic relevance using cosine similarity to retrieve the top-k (k=5) documents. Subsequently, the system utilizes prompt engineering to integrate these retrieved contexts with a BIM-derived hazard summary, comprising hazard frequencies, severity distributions, and geometric measurements. The final structured prompt incorporates role-specific instructions, regulatory guidelines, best practices, case-based knowledge, BIM hazard context, conversation history, and the user query. The system interfaces with a DeepSeek-based LLM via an OpenAI-compatible API, utilizing a temperature setting of 0.7 to balance analytical precision and interpretability. By functioning as a processor of retrieved safety knowledge rather than an independent source, this architecture ensures regulatory compliance and substantially reduces the risk of hallucination during safety advisory interactions. The system maintains conversation continuity by storing user-assistant exchanges in a JSON-based history.

The system implements comprehensive monitoring for evaluation and optimization. Key performance indicators include response latency, token usage, response length, and retrieval efficiency, with data logged per query and aggregated across sessions. The evaluation framework comprises test cases covering hazard prioritization, regulatory compliance, and mitigation strategies, enabling systematic comparison of embedding models, retrieval parameters, and LLM configurations. To assess classification performance, the system utilizes confusion matrix metrics (precision, recall, and F1 score) across five fall hazard types corresponding to those in the BIM model. True Positives (TP) represent fall hazards correctly identified and categorized, False Positives (FP) indicate elements incorrectly classified as fall hazards, and False Negatives (FN) represent fall hazards in the BIM model that were either unidentified or misclassified. The metrics are defined by Eqs. (1)-(3).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} - \text{score} = 2\text{x}\frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

4 Results

4.1 Fall detection engine

The system employs an ontology-driven detection engine that analyzes IFC-based BIM geometry through spatial reasoning. Fall hazards are identified using bounding box analysis, elevation stratification, proximity detection, and semantic validation of protection elements.

The ontology encodes safety regulations as machine-interpretable constraints, enabling hazard classification based on regulatory non-compliance rather than purely geometric analysis. This approach successfully detects unprotected openings, exposed edges, stair hazards, and multi-story fall risks while excluding protected elements. The results demonstrate that ontology-guided spatial logic effectively formalizes and detects fall hazards, enabling automated safety assessment during the design phase.

After running the ontology, the processing statistics showed a geometric success rate of 97.7%, with a total of 115 hazards identified. These included 75.7% elevated work hazards, 20.8% opening hazards, and 3.5% access hazards. The total processing time was 0.68 s. As illustrated in Figure 2, the total number of fall hazards was 85 for unprotected sides and edges, 16 for wall openings, 8 for floor openings, 4 for stairways, and 2 for FFH. These hazards were stored in a JSON file, and the subsequent phases, such as fall identification, fall prevention, and LLM reasoning advisory, were based on this JSON file. For fall identification, the extracted hazards were further visualized using a color-coded scheme to clearly distinguish different fall-risk categories.

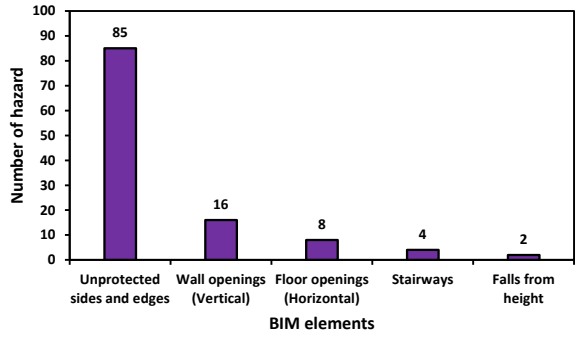


Figure 2. Element-based detection.

4.2 Fall identification and visualization

The results demonstrate that the proposed ontology-driven and BIM-based framework effectively identifies, classifies, and visualizes fall hazards in construction environments. The results support fall hazard identification through color-coded visualization, as shown in Figure 3. Each detected hazard is mapped to its corresponding IFC element or spatial location and rendered using a predefined hazard-based color scheme. Solid building components like slabs, stairways, and structural elements are colored directly, while void-related hazards (floor and wall openings) are represented using 3D markers to overcome IFC visualization limitations. This visual differentiation allows users to distinguish between unprotected edges, openings, stair-related risks, and FFH scenarios within the 3D model, transforming abstract hazard data into interpretable spatial information.

From a safety prevention perspective, the results enable proactive decision-making by making fall risks visible during design and planning stages rather than during construction. The results indicate the potential of automated hazard identification and visualization to support improved coordination and reduce reliance on manual inspections. Although these aspects are not quantitatively evaluated in this study, color-coding indicates each hazard type associated with BIM elements, allowing stakeholders to easily identify the hazard type, its location, the responsible element, and prevention measures. The color-coded BIM model provides a shared awareness platform for designers, planners, and safety officers, enabling earlier design revisions, clearer safety briefings, and informed construction sequencing. The results confirm that the proposed approach achieves accurate fall hazard identification while strengthening safety prevention strategies and supporting data-driven construction management practices.

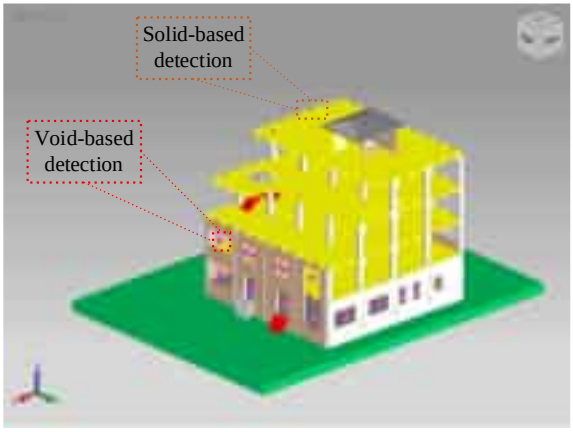


Figure 3. Rule-based fall hazard visualization.

4.3 Fall prevention

The guardrail placement demonstrates that fall prevention measures can be automatically generated and embedded into BIM models in an OSHA-compliant manner. Guardrails were placed at locations identified as unsafe or partially protected, including stair flights without railings, elevated slab perimeters, and floor openings. The placement logic evaluated existing protective elements and skipped compliant locations to avoid redundant safety

components, explaining why guardrails appear only along unprotected areas. The correctness and compliance of generated guardrails are ensured through a rule-based validation mechanism, where OSHA safety requirements, such as guardrail height, mid-rail placement, and edge protection conditions, are explicitly encoded into the system. Guardrail geometries are directly derived from these regulatory constraints, ensuring that all generated elements conform to predefined safety standards without requiring additional manual validation.

As illustrated in Figure 4, the system placed guardrails using geometric and semantic reasoning from the IFC model. For stairs, the code analyzed step geometry across elevations, identified edges, and generated continuous top and mid rails with posts at OSHA-specified spacing. For slabs, the algorithm extracted perimeter and opening loops, classified edges based on adjacent wall height relative to OSHA's 1070 mm parapet threshold, and generated guardrails where needed. Wall-aware inset logic offset guardrails to prevent clashes, reflecting constructability requirements. The results demonstrate fall prevention operationalization in construction sites. By enclosing unprotected edges, openings, and stairways with properly dimensioned guardrails, the system blocks fall paths and reduces height-related hazards. This approach translates detected risks into preventive controls, bridging hazard analysis and mitigation. The BIM integration enables safety strategy review and communication before construction.

From a management perspective, automated guardrail generation improves consistency, efficiency, and traceability. Safety measures are applied uniformly according to rules rather than manual modeling or subjective judgment, reducing missed hazards. The BIM model serves as a digital safety plan supporting inspections, method statements, and stakeholder coordination. Integration of hazard detection and prevention strengthens proactive safety management. However, these guardrail placement methods are hard-coded, rule-based, and deterministic, making them time-consuming to develop, error-prone when conditions deviate from assumptions, and difficult to generalize, motivating the need for an LLM-based approach to overcome these limitations.



Figure 4. Rule-based guardrail placement.

4.4 RAG-LLM reasoning advisor

This section evaluates the advisory and reasoning stage using confusion matrix analysis. As shown in Figure 5, a 5×5 confusion matrix assessed fall hazard classification accuracy against ontology-identified ground truth labels across five categories: unprotected edges, floor openings, wall openings, stairways, and FFH. The ground truth labels are generated through an ontology-based hazard detection process, where regulatory rules and spatial logic are explicitly encoded to deterministically classify fall hazards. These ontology-derived classifications serve as a rule-based reference for evaluating the consistency of predicted hazard categories. Strong diagonal dominance confirmed high classification consistency and validated the ontology-based hazard extraction and rule-based reasoning. Overall performance achieved precision, recall, and F1-score of 0.940 based on 113 true positives, 2 false positives, and 2 false negatives. Classification was achieved for floor openings (0.934), wall openings (0.940), and stairways (0.960), while unprotected edges (0.941) showed minimal misclassification. Compared to recent LLM-based ARC studies [13, 16], which primarily focus on regulatory text interpretation and rule extraction, the proposed approach demonstrates comparable classification performance while introducing a key advancement through integrating ontology-driven spatial reasoning with BIM geometry. Unlike prior works that treat compliance checking as a text-based or post-processing task, this framework directly links semantic regulatory understanding with geometric hazard detection and extends functionality to automated prevention (e.g., guardrail generation). This integration enables spatially grounded hazard identification and mitigation within the BIM

environment.

To evaluate practical usability, a query table (Table 2) tested structured safety-related questions about critical hazards, fall prevention recommendations, and mitigation priorities. The queries successfully retrieved hazard instances from the JSON-based repository and combined them with regulation-informed guidance from the knowledge base. This shows the system can translate visualized, color-coded fall hazards and implemented guardrail measures into actionable safety insights. Combined with previous results on automated hazard detection and preventive modeling, these findings confirm that the framework identifies and mitigates fall hazards while supporting safety decision-making and construction safety management. However, the evaluation of LLM-generated advisory responses is limited to structured query-based testing and does not include validation against expert judgment, robustness analysis under varying prompts or model parameters, or detailed error analysis. These limitations will be addressed in future work to further assess the reliability and robustness of the RAG-LLM advisory module.

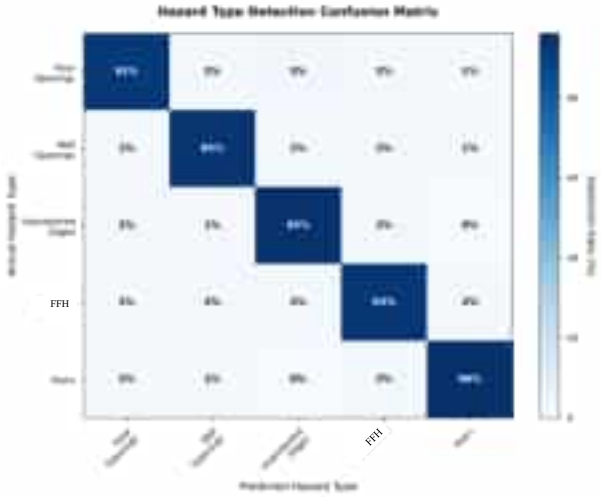


Figure 5. Confusion matrix for fall hazard classifications.

Table 2. RAG-LLM safety-related questions.

Role: How many hazards do I have? **Content**: User uploaded hazard report: 115 hazards detected.
Role: How many high-risk hazards do I have to cover? **Content**: Based on the information provided in your hazards summary, you have 75.7% elevated wor hazard. This number is specifically categorized under elevated work hazards.
Role: How do I fix slab opening hazards according to fall rules? **Content**: To address slab opening hazards in compliance with Hong Kong Labour Department and MOHURD rules, you must follow specific guidance to ensure worker safety and prevent falls through guardrails.

5 Conclusion

This study proposes a RAG-LLM-augmented BIM pipeline for automated identification and prevention of fall hazards in building design, advancing DfS beyond traditional rule-based approaches. By integrating BIM geometry with a domain-specific ontology and enhancing it through an RAG-enabled LLM, the framework bridges unstructured safety regulations and spatial building data. This integration enables precise, context-aware detection of fall hazards, like unprotected edges, floor and wall openings, and stairway risks, within BIM environments. The system demonstrates effectiveness through three functions: (1) ontology-driven hazard detection with 97.7% geometric processing success, (2) color-coded visualization of hazards within the BIM model, and (3) automated generation of OSHA-compliant guardrails at identified hazard locations. The RAG-LLM layer provides dynamic, knowledge-augmented safety advice, reducing reliance on rule coding and delivering regulation-aware recommendations.

The contributions are threefold. First, the study introduces a semantic-spatial ontology that translates safety regulations into machine-interpretable logic linked to IFC entities. Second, it operationalizes prevention by connecting hazard detection with real-time guardrail placement. Third, it incorporates RAG-LLM as an adaptive reasoning layer to enhance regulatory compliance and support proactive safety decisions.

The study identifies key limitations affecting its applicability. Validation on a single BIM case limits generalizability across diverse project types and scales. Guardrail placement relies on deterministic, rule-based methods that cannot adapt to complex geometries or regulatory changes. The ontology and LLM knowledge base derive primarily from Hong Kong and Chinese safety standards, restricting broader applicability. Future directions include expanding the regulatory knowledge base to international standards, enhancing LLM spatial reasoning through BIM-specific fine-tuning, incorporating sensor data for real-time construction-phase monitoring, and developing generative design approaches for hazard-free alternatives.

This research presents a scalable framework that shifts BIM-based safety checking from compliance verification to proactive hazard prevention. By combining semantic regulatory understanding, geometric analysis, and LLM-augmented reasoning, the pipeline

supports earlier design decisions and advances automated, human-centric safety management in digital construction.

References

[1] A. A. Fasoyinu, S. Azhar, A. Sattineni, and J. O. Toyin, "Wearable sensing devices for construction safety: Research trends, applications, challenges, and future opportunities," *Automation in Construction*, vol. 179, p. 106424, 2025/11/01/ 2025, doi: https://doi.org/10.1016/j.autcon.2025.106424.

[2] F. Pereira, M. D. González García, and J. Poças Martins, "BIM for Safety: Applying Real-Time Monitoring Technologies to Prevent Falls from Height in Construction," *Applied Sciences*, vol. 15, no. 4, p. 2218doi: 10.3390/app15042218.

[3] R. A. Haslam *et al.*, "Contributing factors in construction accidents," *Applied Ergonomics*, vol. 36, no. 4, pp. 401–415, 2005/07/01/ 2005, doi: https://doi.org/10.1016/j.apergo.2004.12.002.

[4] M. Behm, "Linking construction fatalities to the design for construction safety concept," *Safety Science*, vol. 43, no. 8, pp. 589–611, 2005/10/01/ 2005, doi: https://doi.org/10.1016/j.ssci.2005.04.002.

[5] K. Hong and J. Teizer, "Digital construction site layout planning and real-time trajectory analysis for proactive safety monitoring and control of struck-by hazards," *Automation in Construction*, vol. 177, p. 106353, 2025/09/01/ 2025, doi: https://doi.org/10.1016/j.autcon.2025.106353.

[6] M. Kasirossafar, A. Ardeshir, and L. Shahandashti Reza, "Developing the Sustainable Design with PtD Using 3D/4D BIM Tools," in *World Environmental and Water Resources Congress 2012*, (Proceedings, 2012, pp. 2786–2794.

[7] K. W. Johansen, J. Teizer, and C. Schultz, "Automated rule-based safety inspection and compliance checking of temporary guardrail systems in construction," *Automation in Construction*, vol. 168, p. 105849, 2024/12/15/ 2024, doi: https://doi.org/10.1016/j.autcon.2024.105849.

[8] W. Solihin, Z. Liu, Y. Lu, and L. Wei, "BIM-based automated rule-checking in the AECO industry: Learning from semiconductor manufacturing," *Automation in Construction*, vol. 162, p. 105406, 2024/06/01/ 2024, doi: https://doi.org/10.1016/j.autcon.2024.105406.

[9] N. Chen, X. Lin, H. Jiang, and Y. An, "Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology," *Buildings*, vol. 14, no. 7, p. 1983doi: 10.3390/buildings14071983.

[10] Z. Ma, H. Zhu, X. Xiang, Ž. Turk, and R. Klinc, "Automatic compliance checking of BIM models against quality standards based on ontology technology," *Automation in Construction*, vol. 166, p. 105656, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.autcon.2024.105656.

[11] S. Hwang, S. Jung, and S. Lee, "Automated inference of context-specific hazards in construction using BIM and Ontology," *Automation in Construction*, vol. 177, p. 106338, 2025/09/01/ 2025, doi: https://doi.org/10.1016/j.autcon.2025.106338.

[12] Z. Zheng, J. Han, K.-Y. Chen, X.-Y. Cao, X.-Z. Lu, and J.-R. Lin, "Translating regulatory clauses into executable codes for building design checking via large language model driven function matching and composing," *Engineering Applications of Artificial Intelligence*, vol. 163, p. 112823, 2026/01/01/ 2026, doi: https://doi.org/10.1016/j.engappai.2025.112823.

[13] S. Madireddy *et al.*, "Large Language Model-Driven Code Compliance Checking in Building Information Modeling," *Electronics*, vol. 14, no. 11, p. 2146doi: 10.3390/electronics14112146.

[14] L.-I. W. Junyong Chen, Minyu Chen, Xiaoying Qian, Haoze Zhu, Qiongfang Zhang, Guoqiang Li, "ARCEAK: An Automated Rule Checking Framework Enhanced with Architectural Knowledge," 2026, doi: https://doi.org/10.48550/arXiv.2501.14735.

[15] C. Du, S. Esser, S. Nousias, and A. Borrmann, "Text2BIM: Generating Building Models Using a Large Language Model-Based Multiagent Framework," *Journal of Computing in Civil Engineering*, vol. 40, no. 2, p. 04025142, 2026/03/01 2026, doi: 10.1061/JCCEE5.CPENG-6386.

[16] O. Iversen and L. Huang, "Leveraging large language models for BIM-based automated compliance checking," *Automation in Construction*, vol. 182, p. 106707, 2026/02/01/ 2026, doi: https://doi.org/10.1016/j.autcon.2025.106707.

A Decision-Oriented Synthesis of AI-Based Assembly Sequence Planning for Construction Automation

Anas Itani[1], Mohamed Al-Hussein[1], and Simaan Abourizk[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
aitani@ualberta.ca, malhussein@ualberta.ca, abourizk@ualberta.ca

Abstract

Assembly Sequence Planning (ASP) is a critical decision-making problem in construction automation, affecting productivity, cost, constructability, and system robustness. ASP is NP-hard (Non-deterministic Polynomial-time hard) due to the factorial growth in feasible assembly sequences as product complexity increases. While artificial intelligence (AI) and soft computing methods have been widely applied to ASP, existing research remains fragmented, algorithm-centric, and largely rooted in manufacturing contexts, limiting its transferability to construction automation. This paper presents a construction-oriented synthesis of AI-based ASP approaches, emphasizing decision-level insights rather than algorithmic mechanics. This paper reviews ASP modelling strategies, constraints, assembly scales, and objectives from a construction perspective. AI-based optimization methods are organized into major algorithm families and evaluated according to scalability, computational effort, and construction suitability. The paper further discusses emerging considerations such as parallel assembly, BIM-enabled digital twins, and data-driven learning approaches within a construction automation context. Key research gaps are prioritized to guide future research and development. The synthesis provides guidance for researchers and practitioners seeking to design scalable, realistic, and implementable ASP solutions for automated and industrialized construction.

Keywords –

Assembly Sequence Planning; Construction Automation; Artificial Intelligence; Prefabricated and Modular Construction; Optimization Methods

1 Introduction

The increasing adoption of automation, prefabrication, and modular construction is transforming how construction systems are designed, planned, and executed. As construction adopts manufacturing-inspired production paradigms, assembly-related decisions, such as component ordering, assembly directions, and tool usage, have become critical determinants of productivity, cost efficiency, and constructability [1,2]. Assembly Sequence Planning (ASP) addresses these challenges by determining feasible and efficient assembly sequences under technical and operational constraints.

ASP is NP-hard combinatorial optimization problem, as the number of feasible sequences grows factorially with the number of components [3,4]. In construction applications, this complexity is amplified by large-scale components, heterogeneous materials, diverse connection types, and varying degrees of automation [5]. Poor sequencing can result in excessive tool changes, frequent reorientation, unbalanced workloads, and increased human or robotic effort, reducing overall system efficiency [6,7].

Figure 1 positions ASP within the construction automation decision framework, highlighting its role as a coordinating function linking product design, production planning, and execution. Unlike manufacturing ASP (rigid, sequential, controlled), construction involves parallelism, variability, and human–robot interaction [8,9]. These characteristics increase planning complexity and limit the transferability of manufacturing-oriented solutions.

This distinction is illustrated in Figure 2, which contrasts manufacturing- and construction-oriented assembly systems in terms of component scale, assembly parallelism, and operational variability, clarifying the need for construction-specific modelling assumptions and decision criteria.

To address ASP's combinatorial complexity, AI and soft computing methods, such as genetic algorithms, ant colony optimization, particle swarm optimization, simulated annealing, and hybrid approaches, have been widely applied [12–14]. However, much of the literature remains algorithm-centric, emphasizing method-specific implementations rather than decision-level applicability. In addition, several construction-critical considerations remain underrepresented, including flexible or deformable components, ergonomic and human–machine interaction constraints, sustainability objectives, and explicit treatment of parallel or collaborative assembly processes [13]. These limitations are synthesized in Figure 3.

In this context, this paper presents a focused synthesis of AI-based ASP approaches from a construction automation perspective, emphasizing decision-level insights rather than algorithmic mechanics. The paper: (i) reframes ASP as a construction decision problem spanning design, planning, and execution; (ii) introduces assembly scale as a key

determinant of modelling and algorithm suitability; (iii) synthesizes ASP modelling approaches and constraint types relevant to construction systems; (iv) compares AI-based optimization methods according to their suitability across different assembly scales; and (v) identifies prioritized research gaps to guide future development of practical ASP solutions for construction automation.



Figure 1. Role of assembly sequence planning in construction automation



Figure 2. Comparison of manufacturing- and construction-oriented assembly systems



Figure 3. ASP limitation in construction automation research

This study adopts a structured narrative review approach. Literature was identified from Scopus, Web of Science, and Google Scholar using keywords such as 'assembly sequence planning' and 'construction automation'. The review focuses on studies (2000–2024) relevant to ASP modelling, constraints, and optimization, covering approximately 120 papers.

2 Assembly Sequence Planning as a Construction Decision Problem

ASP has traditionally been formulated as a combinatorial optimization problem within manufacturing-oriented research, where it is treated as a technical sequencing task aimed at generating feasible or near-optimal assembly orders under predefined precedence constraints [3,4,8]. While this formulation has enabled significant algorithmic advances, it captures only part of ASP's role in construction automation.

In construction-oriented systems, ASP is more appropriately viewed as a system-level decision problem spanning product design, production planning, and execution. Construction assemblies differ fundamentally from manufacturing assemblies due to larger components, heterogeneous materials, variable environments, and differing levels of automation [1,2,5]. Consequently, ASP decisions influence not only feasibility but also constructability, system efficiency, and automation readiness. Across the construction lifecycle, ASP acts as a coordinating function, shaping component segmentation and accessibility at the design stage [6,10], interacting with resource allocation and line balancing during production planning [2,11], and guiding task scheduling and human–robot coordination during execution [13].

The limited transferability of existing ASP research largely stems from manufacturing-oriented assumptions, including rigid components, sequential processes, and controlled environments [8,9,12]. These assumptions rarely hold in construction, where parallel sub-assembly, distributed production, and semi-automated operations are common, and where human–robot collaboration introduces additional ergonomic and safety constraints [2,13]. In prefabricated and modular construction, parallel assembly further increases ASP complexity by introducing concurrency, synchronization, and resource-sharing constraints [11,12], requiring ASP to address not only component sequencing but also coordination across parallel sub-assemblies.

Reframing ASP as a construction decision problem has direct implications for modelling and optimization. Models must capture system-level characteristics such as parallelism, resource coupling, and human–robot interaction, while optimization objectives must extend beyond local efficiency to include constructability, safety, and robustness [6,7]. Algorithm selection should therefore be guided by assembly characteristics and system requirements rather than algorithm popularity alone. This reframing provides the foundation for the next section, which introduces assembly scale and system characteristics as key determinants of modelling strategies and AI-based optimization suitability

in construction-oriented ASP.

3 Assembly Scale and System Characteristics in Construction-Oriented ASP

Building on the decision-oriented framing introduced in Section 2, this section operationalizes that perspective by formalizing assembly scale as a determinant of modelling and optimization suitability. Assembly scale remains an underexplored dimension in ASP research, where modelling and optimization strategies are often assumed to scale uniformly with problem size. In construction automation, however, assembly scale fundamentally influences solution space growth, constraint dominance, modelling feasibility, and algorithm suitability, making its explicit consideration essential for practical ASP solutions.

3.1 Definition of Assembly Scale in Construction Contexts

In manufacturing-oriented ASP research, assembly scale is often defined solely by the number of components. While component count remains relevant, it is insufficient to characterize construction assemblies. In construction automation, assembly scale is more appropriately defined by the combined effects of component count, presence of sub-assemblies, degree of parallelism, and level of system coordination required.

Based on the literature and construction practice [2,5,11], construction-oriented ASP problems can be classified according to assembly scale, as summarized in Table 1, providing a decision-oriented basis for selecting modelling and optimization strategies.

While assembly scale is presented conceptually, future work should define quantitative thresholds based on component count, degree of parallelism, and interaction density to enable more objective classification across construction scenarios.

Table 1. Assembly scale classification for construction-oriented ASP

Assembly Scale	Typical Characteristics	Implications for ASP
Small	Few components; mostly sequential	Exact or simple meta-heuristics feasible
Medium	Multiple sub-assemblies; mixed sequencing	Evolutionary or swarm methods effective
Large	Parallel sub-assemblies; system coupling	Hierarchical, hybrid, or memetic methods

3.2 ASP Solution Space Growth and Assembly Scale

Sequential assembly represents the baseline ASP case, where components are assembled one at a time under precedence constraints; parallel assembly generalizes this formulation by allowing concurrent sub-assemblies subject to synchronization and resource constraints

For a purely sequential assembly of n components, the theoretical upper bound of feasible sequences is:

$$S_{seq} = n! \qquad (1)$$

For a system composed of k independent sub-assemblies with component counts $n_1, n_2, \ldots, n_k$, the number of possible interleavings can be approximated as:

$$S_{par} = \frac{(n_1+n_2+\cdots+n_k)!}{n_1!+n_2!+\cdots+n_{k!}} \qquad (2)$$

Parallel sub-assembly, common in prefabricated and modular construction, therefore, expands the solution space dramatically relative to sequential assemblies, explaining the limited scalability of many manufacturing-oriented ASP approaches [12,14].

3.3 Scale-Dependent Dominance of ASP Constraints

Assembly scale strongly influences which constraints dominate ASP outcomes. In small-scale assemblies, feasibility constraints, such as precedence, geometric, and stability constraints, are dominant, as ensuring feasible assembly orders constitutes the primary planning challenge [20,21]. In medium-scale assemblies, optimization-oriented constraints, including assembly direction changes, tool changeovers, and operation-type variation, increasingly influence sequence quality and cycle time [6,7].

In large-scale construction assemblies, typical of modular and prefabricated systems, system-level constraints often dominate. These include workstation coordination, resource sharing, ergonomic limits in human–robot collaboration, and synchronization of parallel tasks [13,15]. At this scale, local sequencing decisions can significantly affect overall system performance. This transition in constraint dominance is illustrated in Figure 4.



Figure 4. Dominant ASP constraints across assembly scales

3.4 Implications for Modelling and Algorithm Selection

Assembly scale directly informs the suitability of ASP modelling approaches. Part-based modelling offers high fidelity and is effective for small-scale assemblies

but becomes computationally prohibitive as scale increases. Task-based modelling reduces problem dimensionality through operation aggregation [15], while connector-based modelling is well suited to large-scale modular systems, where standardized interfaces support hierarchical planning [18,19].

The interaction between assembly scale, modelling complexity, and constraint dominance also guides algorithm selection: small-scale assemblies can be addressed using evolutionary or physics-inspired methods, medium-scale assemblies benefit from swarm intelligence approaches, and large-scale assemblies typically require hybrid or memetic strategies that combine global search with hierarchical decomposition. By explicitly incorporating assembly scale into ASP formulation, algorithm selection becomes a decision-driven process rather than a methodological preference, providing a foundation for the comparative analysis of AI-based ASP methods in the next section.

4 ASP Modelling Approaches and Constraint Formulation

ASP performance and applicability in construction automation are strongly influenced by how the assembly problem is modelled and how constraints are formulated. Modelling choices define the structure of the search space, the types of constraints that can be represented, and the computational feasibility of optimization. In construction-oriented ASP, models must account for large-scale components, heterogeneous connections, and parallel assembly processes, which are often inadequately captured by manufacturing-focused formulations.

4.1 Overview of ASP Modelling Approaches

Existing ASP research typically adopts one of three modelling approaches: part-based, task-based, or connector-based representations [15,16,18]. These approaches differ in abstraction level and exhibit distinct advantages and limitations when applied to construction assemblies.

4.1.1 Part-Based Modelling

Part-based modelling treats individual components as the fundamental assembly units. Precedence relationships are derived from geometric interference, mechanical feasibility, and stability conditions [16,17]. This approach offers high modelling fidelity and is well suited to small-scale assemblies where component interactions remain manageable.

In construction automation, part-based models are commonly used for detailed prefabrication tasks or small modular assemblies. However, as assembly scale increases, the precedence graph grows rapidly, leading to significant computational challenges. As a result, part-based modelling becomes impractical for large-scale construction assemblies with extensive parallelism.

4.1.2 Task-Based Modelling

Task-based modelling abstracts the assembly process into sequences of operations rather than individual components. Tasks may represent fastening, positioning, or tool-specific actions [15]. This abstraction reduces problem dimensionality and enables direct incorporation of operational constraints such as tool usage and assembly direction.

Task-based models are particularly suitable for medium-scale assemblies, where multiple sub-assemblies exist but full part-level modelling is computationally prohibitive. However, this abstraction may obscure detailed geometric constraints, which can be critical for large or irregular construction components.

4.1.3 Connector-Based Modelling

Connector-based modelling organizes the assembly process around connection mechanisms such as bolts, welds, or standardized interfaces [18,19]. Components associated with a common connector are grouped, enabling hierarchical or multi-stage planning.

This approach aligns well with modular and prefabricated construction systems, where standardized connections are prevalent. Connector-based modelling supports scalable ASP by reducing effective problem size and enabling decomposition into manageable sub-problems, although it often requires more complex model formulation and multi-level optimization.

4.2 Constraint Categories in Construction-Oriented ASP

ASP constraints can be broadly classified into feasibility constraints and optimization-oriented constraints [6,15], with their relative importance varying by assembly scale, as discussed in Section 3.

4.2.1 Feasibility Constraints

Feasibility constraints determine whether an assembly sequence is physically and functionally admissible and typically include precedence constraints that enforce required assembly orderings [20], geometric constraints that prevent collisions and ensure accessibility [21], and stability constraints that maintain structural integrity during intermediate assembly states [22]. These constraints dominate small-scale assemblies and must be satisfied before optimization objectives are considered.

4.2.2 Optimization-Oriented Constraints

Optimization-oriented constraints influence the efficiency and quality of feasible sequences. In construction automation, commonly considered constraints include assembly direction changes, tool changeovers, operation-type variation, and task sequencing consistency. These factors directly affect cycle time, energy consumption, and operational complexity, particularly in robotic and semi-automated systems [6,7].

4.2.3 System-Level Constraints

In large-scale construction assemblies, system-level constraints become increasingly important. These include

workstation coordination, resource sharing across parallel tasks, ergonomic and safety constraints in human–robot collaboration, and synchronization of parallel sub-assemblies. Such constraints link ASP with higher-level planning functions and require models capable of representing system-wide interactions [13,15].

System-level constraints such as ergonomic limits, resource sharing, and workstation coordination can be integrated through multi-layered modelling approaches, where local sequencing decisions are coupled with higher-level scheduling and resource allocation models, enabling synchronized optimization across parallel sub-assemblies.

4.3 Relationship Between Assembly Scale, Modelling, and Constraints

The interaction between assembly scale, modelling approach, and constraint formulation is central to construction-oriented ASP. Part-based models are most effective when feasibility constraints dominate, task-based models balance feasibility and optimization concerns, and connector-based models support system-level constraint representation in large-scale assemblies. Table 2 summarizes the suitability of ASP modelling approaches as a function of assembly scale and dominant constraint types.

Table 2. Relationship between assembly scale, modelling approach, and dominant constraints

Assembly Scale	Dominant Constraints	Suitable Modelling Approach
Small	Feasibility	Part-based
Medium	Optimization-oriented	Task-based
Large	System-level	Connector-based/Hierarchy

4.4 Implications for Construction Automation

The analysis highlights that no single ASP modelling approach is universally applicable across construction scenarios. Effective ASP solutions must align modelling fidelity with assembly scale and dominant constraints. Overly detailed models risk computational infeasibility, while overly abstract models may fail to capture critical construction constraints.

This reinforces the need to treat modelling and constraint formulation as decision problems, guided by construction system characteristics rather than algorithmic convenience. The next section builds on this foundation by examining AI-based optimization methods under varying modelling and constraint conditions.

5 AI-Based Optimization Methods for Construction-Oriented ASP

The combinatorial nature of Assembly Sequence Planning has led to extensive adoption of AI-based and soft computing optimization methods. However, much of the literature evaluates these methods in isolation, emphasizing algorithmic performance rather than decision-level suitability for construction systems. As established in Sections 3 and 4, the effectiveness of an optimization method in construction-oriented ASP depends strongly on assembly scale, modelling abstraction, and dominant constraint types.

Accordingly, this section synthesizes AI-based ASP methods by algorithm family, focusing on their strengths, limitations, and applicability to construction automation rather than their internal mechanics.

While quantitative benchmarking across ASP studies remains challenging due to differences in problem formulation and evaluation metrics, a comparative summary of reported scalability characteristics and typical problem sizes has been incorporated to enhance analytical depth

5.1 Evolutionary Algorithms

Evolutionary algorithms, particularly Genetic Algorithms (GA), are among the most widely applied methods for ASP [27–30]. GA-based approaches encode assembly sequences as chromosomes and evolve solutions through selection, crossover, and mutation. Their flexibility enables straightforward incorporation of multiple objectives and constraint penalties.

In construction contexts, evolutionary algorithms perform well for small- to medium-scale assemblies, particularly when part-based or task-based models are used. However, their performance degrades as assembly scale increases due to premature convergence and reduced exploration efficiency [23,24]. GA-based methods are therefore poorly suited to large-scale construction assemblies involving parallel sub-assemblies and system-level constraints.

5.2 Swarm Intelligence Methods

Swarm intelligence methods, including Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO), exploit collective search mechanisms to explore large solution spaces [25–27]. ACO is especially effective for sequencing problems, as pheromone trails provide structured guidance through discrete search spaces.

In construction-oriented ASP, swarm intelligence methods are well suited to medium-scale assemblies, where optimization-oriented constraints dominate. ACO-based approaches have demonstrated greater robustness than GA in problems with complex precedence relationships [25,26]. PSO offers computational efficiency but is less effective for highly discrete or heavily constrained ASP formulations [27].

5.3 Physics- and Bio-Inspired Optimization Methods

Physics-inspired methods, such as Simulated Annealing, have been applied to ASP due to their ability to escape local optima via probabilistic acceptance of inferior solutions [25–27]. These methods are simple to implement and require fewer parameters than population-based approaches.

In construction applications, Simulated Annealing is

most effective for small-scale assemblies or as a local refinement mechanism. Its slow convergence and limited scalability restrict its applicability to large-scale assemblies with extensive parallelism and system-level constraints.

Bio-inspired methods, including Artificial Immune Systems and Shuffled Frog Leaping Algorithms, emphasize exploration and adaptability in complex search spaces [27]. While capable of handling complex constraint landscapes, these methods often incur high computational cost and require careful parameter tuning.

Hybrid and memetic algorithms combine complementary strategies, such as global search with local refinement or hierarchical decomposition [25–27]. In construction-oriented ASP, hybrid approaches show strong potential for large-scale assemblies where system-level constraints dominate. These methods align naturally with connector-based and hierarchical modelling approaches, although they require greater implementation effort and domain knowledge.

5.4 Learning-Based and Data-Driven ASP Approaches

Recent advances in machine learning have stimulated interest in learning-based ASP approaches, particularly reinforcement and imitation learning, which aim to learn sequencing policies that generalize across similar assembly configurations [28–30]. In RL-based ASP, assembly is formulated as a sequential decision-making problem, where an agent selects actions based on the current assembly state to maximize cumulative reward [28,30].

Although conceptually attractive for construction automation, current learning-based ASP research remains largely experimental. Most studies focus on simplified environments or small assemblies, with limited consideration of construction-specific challenges such as large components, parallel sub-assembly, and human–robot collaboration [29,31]. In addition, learning-based methods typically require substantial training data and careful reward design, which are difficult to obtain in construction contexts.

Related approaches, including case-based reasoning and graph-based learning, have been proposed to exploit prior solutions or structural patterns [28–30], but their applicability to large-scale construction systems remains constrained by data availability and interpretability concerns. From a decision-level perspective, learning-based methods are therefore best viewed as complementary tools, particularly suited to repetitive or data-rich contexts such as standardized modular systems or digital twin environments, rather than as replacements for established optimization methods.

Recent advancements in synthetic data generation and sim-to-real transfer techniques are increasingly addressing data scarcity and generalization challenges, suggesting that learning-based ASP approaches may become more practically viable in construction contexts than previously assumed.

5.5 Decision-Level Suitability of AI Methods for Construction ASP

The synthesis above confirms that no single AI-based optimization method is universally optimal for ASP. Algorithm suitability depends on assembly scale, modelling abstraction, and dominant constraint types. Treating algorithm selection as a decision problem is therefore essential for construction automation.

While quantitative benchmarking across ASP studies remains challenging due to differences in problem formulation and evaluation metrics, a comparative summary of reported scalability characteristics and typical problem sizes has been incorporated to enhance analytical depth.

5.6 Implications for Construction Automation

The prevalence of algorithm-centric ASP research has obscured the importance of decision-level alignment in construction automation. Algorithms that perform well on benchmark problems may be unsuitable for real construction systems if assembly scale, parallelism, and system-level constraints are ignored. Future ASP development should therefore prioritize integration, scalability, and decision transparency over incremental algorithmic novelty.

The next section synthesizes insights from Sections 3–5 to provide a comparative, construction-oriented perspective on ASP decision-making.

6 Decision-Oriented Synthesis and Research Directions

The preceding sections examined assembly scale, modelling approaches, constraint formulation, and AI-based optimization methods individually. This section synthesizes these elements to provide decision-oriented insights for construction-oriented ASP. Rather than comparing algorithms based solely on performance metrics, the synthesis emphasizes *when* and *why* specific modelling and optimization strategies are appropriate given construction system characteristics.

6.1 Alignment Between Assembly Scale, Modelling, and Optimization

A central insight from the literature is that ASP effectiveness depends on alignment between assembly scale, modelling abstraction, and optimization strategy. Mismatches between these elements are a common cause of poor scalability and limited practical applicability in construction automation.

For example, applying part-based modelling with population-based optimization to large-scale modular assemblies often leads to prohibitively large search spaces and excessive computation time [16,18]. Conversely, overly abstract models applied to small-scale assemblies may fail to capture critical geometric or stability constraints, resulting in infeasible or unsafe sequences [20,21].

Effective ASP therefore requires balancing modelling fidelity with computational tractability. Assembly scale

provides a unifying lens for achieving this balance by guiding both modelling choices and algorithm selection.

6.2 Trade-offs Between Feasibility, Optimization, and System-Level Objectives

ASP objectives shift systematically with assembly scale. In small-scale assemblies, feasibility dominates, and performance differences among feasible sequences are often marginal. In medium-scale assemblies, optimization objectives, such as minimizing tool changes, direction changes, or operation-type variation, become increasingly important [6,7].

In large-scale construction assemblies, system-level objectives, including resource coordination, ergonomic considerations, and synchronization of parallel sub-assemblies, frequently outweigh local optimization goals [13,15]. ASP solutions that focus solely on local efficiency may therefore degrade overall system performance by creating bottlenecks or underutilized resources.

These observations highlight the need to frame ASP objectives in terms of overall construction system performance rather than isolated task-level efficiency.

6.3 Decision-Oriented Synthesis Framework

Integrating the insights from Sections 3–5, Table 3 presents a decision-oriented framework linking assembly characteristics, dominant constraints, modelling approaches, and suitable AI-based optimization strategies.

Table 3. Decision-oriented synthesis of ASP approaches for construction automation

Assembly Characters	Dominant Constraints	Model Method	Optimize Method
Small, sequential	Feasible	Part-based	GA, SA
Medium, mixed parallelism	Optimize-oriented	Task-based	ACO, PSO
Large, parallel, modular	System-level	Connector-based / hierarchical	Hybrid, memetic
Repetitive, data-rich	Adaptive / dynamic	Hierarchical / learned	RL-assisted, learning-based

Rather than prescribing specific algorithms, this framework supports informed decision-making by aligning ASP strategies with construction system requirements.

6.4 Implications for Practical Deployment

From a practical perspective, the synthesis explains why many ASP approaches reported in the literature have seen limited adoption in real construction automation systems. Algorithm-centric solutions that neglect assembly scale, modelling abstraction, and system-level constraints often fail to scale beyond laboratory examples.

For industrialized construction applications, such as off-site modular fabrication and robotic prefabrication facilities, ASP solutions must integrate with production planning, resource allocation, and execution control. Decision-oriented ASP frameworks are therefore more likely to support practical deployment than isolated optimization algorithms.

6.5 Research Gaps and Future Directions

Another critical gap is the limited integration of human–robot collaboration and ergonomic constraints in ASP models. Construction automation often operates in semi-automated settings where humans and robots interact closely, yet safety limits, ergonomics, and task handovers are rarely embedded explicitly in sequence planning. Incorporating these factors is essential for improving both feasibility and acceptance of automated assembly solutions.

Existing ASP formulations also commonly assume rigid components and deterministic conditions, assumptions that rarely hold in construction. Flexible or semi-deformable components, tolerance variability, and environmental uncertainty remain difficult to represent, highlighting the need for advanced modelling approaches, including hybrid geometric and data-driven representations.

While learning-based ASP methods, such as reinforcement and imitation learning, have gained interest, their practical applicability remains limited by data scarcity, training cost, and interpretability challenges. In the near term, these methods are more likely to augment traditional optimization, through solution initialization, adaptation, or digital twin integration, rather than replace established approaches.

Finally, scalability and the lack of construction-specific benchmarks remain major obstacles. Many studies rely on small or idealized examples, limiting insight into industrial-scale performance. Developing standardized datasets and benchmark problems (e.g., modular frames or panelized façade systems) would significantly improve comparability and accelerate translation to practice.

Addressing these gaps requires a shift from algorithm-centric experimentation toward integrated, system-aware, and decision-oriented ASP frameworks capable of supporting real-world automated and industrialized construction.

Future work should focus on validating the proposed decision-oriented framework through pilot case studies and real-world construction scenarios, particularly in modular and prefabrication contexts, to assess its practical applicability and scalability.

7 Conclusions

This paper presented a focused synthesis of AI-based Assembly Sequence Planning from a construction automation perspective, reframing ASP as a decision problem spanning design, production planning, and execution rather than an algorithm-centric optimization task.

A central contribution is the introduction of assembly scale as a unifying concept for understanding ASP complexity in construction. Distinguishing between small-,

medium-, and large-scale assemblies clarifies how solution space growth, constraint dominance, and algorithm suitability vary systematically, explaining the limited scalability of many manufacturing-oriented ASP approaches in construction contexts involving parallel sub-assembly and system-level coordination.

The synthesis further demonstrated that ASP modelling approaches and constraint formulations are inherently coupled. Part-based, task-based, and connector-based models involve distinct trade-offs in fidelity and scalability, and no single AI-based optimization method is universally optimal; effective ASP solutions instead arise from decision-level alignment between assembly scale, modelling abstraction, and optimization strategy.

By positioning learning-based methods alongside classical meta-heuristics, the paper identified reinforcement learning and related approaches as complementary tools rather than replacements for established frameworks. Overall, this work advances ASP research toward practical applicability, scalability, and decision support, providing a structured foundation for future construction-oriented ASP development.

References

[1] Bock, T., "The future of construction automation: Technological disruption and the upcoming ubiquity of robotics," *Automation in Construction*, vol. 59, pp. 113–121, 2015.

[2] Pan, W., and Gibb, A. G. F., "Decision-making for industrialized housing systems," *Journal of Construction Engineering and Management*, vol. 138, no. 11, pp. 123–134, 2012.

[3] De Fazio, T. L., and Whitney, D. E., "Simplified generation of all mechanical assembly sequences," *IEEE Journal of Robotics and Automation*, vol. 3, no. 6, pp. 640–658, 1987.

[4] Baldwin, D. F., Abell, T. E., Lui, M. C., De Fazio, T. L., and Whitney, D. E., "An integrated computer aid for generating and evaluating assembly sequences," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 1, pp. 78–94, 1991.

[5] Yuan, Z., Sun, C., and Wang, Y., "Design for assembly in prefabricated construction," *Automation in Construction*, vol. 88, pp. 13–22, 2018.

[6] Lu, Y., Wang, X., and Chen, K., "Multi-objective assembly sequence planning using genetic algorithms," *International Journal of Advanced Manufacturing Technology*, vol. 28, no. 7–8, pp. 721–730, 2006.

[7] Pan, Z., Polden, J., Larkin, N., Van Duin, S., and Norrish, J., "Recent progress on programming methods for industrial robots," *International Journal of Production Research*, vol. 50, no. 13, pp. 3603–3619, 2012.

[8] Homem de Mello, L. S., and Sanderson, A. C., "AND/OR graph representation of assembly plans," *IEEE Transactions on Robotics and Automation*, vol. 6, no. 2, pp. 188–199, 1990.

[9] Wolter, J. D., Chakrabarty, S., and Tsao, T. C., "Automated assembly planning: A survey," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 3, pp. 453–464, 1994.

[10] Chen, L., and Liu, Y., "Assembly sequence planning based on CAD models," *Computer-Aided Design*, vol. 33, no. 6, pp. 403–417, 2001.

[11] Lu, N., "Design for manufacture and assembly for prefabricated housing," *Automation in Construction*, vol. 18, no. 2, pp. 211–219, 2009.

[12] Rashid, M. M., Fattah, A., and Ahmed, S., "Assembly sequence planning using ant colony optimization," *International Journal of Production Research*, vol. 49, no. 18, pp. 5377–5406, 2011.

[13] Villani, V., Pini, F., Leali, F., and Secchi, C., "Survey on human–robot collaboration in industrial settings," *Mechatronics*, vol. 55, pp. 248–266, 2018.

[14] Yu, J., and Wang, X., "Assembly sequence planning using particle swarm optimization," *International Journal of Advanced Manufacturing Technology*, vol. 64, no. 9–12, pp. 1437–1450, 2013.

[15] Mok, H. S., Kim, Y. S., and Lee, J. H., "Constraint-based assembly sequence planning," *International Journal of Production Research*, vol. 39, no. 17, pp. 4007–4023, 2001.

[16] Tseng, M. M., and Su, C. J., "Assembly planning using part-based modelling," *International Journal of Production Research*, vol. 45, no. 21, pp. 4827–4849, 2007.

[17] Shan, H., Wang, Y., and Liu, J., "Memetic algorithms for assembly sequence planning," *Applied Soft Computing*, vol. 9, no. 2, pp. 651–660, 2009.

[18] Lu, Y., and Wang, X., "Hierarchical assembly planning for modular products," *International Journal of Advanced Manufacturing Technology*, vol. 83, no. 5–8, pp. 1001–1016, 2016.

[19] Sacks, R., Girolami, M., and Brilakis, I., "Building information modelling, digital twins and smart construction," *Data-Centric Engineering*, vol. 1, e14, 2020.

[20] Dul, J., and Neumann, W. P., "Ergonomics contributions to company strategies," *Applied Ergonomics*, vol. 40, no. 4, pp. 745–752, 2009.

[21] Chen, Y., Xu, Y., and Wang, Z., "Reinforcement learning for robotic assembly planning," *Robotics and Computer-Integrated Manufacturing*, vol. 68, 2021.

[22] Mnih, V., Kavukcuoglu, K., Silver, D., et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.

[23] Goldberg, D. E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.

[24] Holland, J. H., Adaptation in Natural and Artificial Systems, MIT Press, 1992.

[25] Dorigo, M., and Stützle, T., *Ant Colony Optimization*, MIT Press, 2004.

[26] Dorigo, M., Maniezzo, V., and Colorni, A., "Ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 26, no. 1, pp. 29–41, 1996.

[27] Kennedy, J., and Eberhart, R., "Particle swarm optimization," *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, 1995.

[28] Sutton, R. S., and Barto, A. G., *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2018.

[29] Silver, D., Huang, A., Maddison, C. J., et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489, 2016.

[30] Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A., "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[31] Zhang, J., Teizer, J., and Pradhananga, N., "Digital twin–enabled construction automation," *Automation in Construction*, vol. 136, 2022.

Cognitive Scan-to-BIM: LLM-Driven Hyperparameter Optimization for Robust MEP Primitive Fitting in Complex Construction Environments

Muhammad Talha[1], Rafiq Ahmad[1], Yitong Li[1], Omair Shafiq[2], Qipei (Gavin) Mei[1]
[1]University of Alberta, Edmonton, Canada
Email: {tmuhamm1, rafiq.ahmad, yitong2, qipei.mei}@ualberta.ca
[2]University Malaysia Pahang Al Sultan Abdullah, Kuantan, Malaysia
Email: pdz24005@adab.umpsa.edu.my

***Abstract*—Generating accurate Digital Twins of existing buildings remains a persistent bottleneck in Scan-to-BIM workflows, especially for Mechanical, Electrical, and Plumbing (MEP) systems. Although recent deep learning methods can segment MEP components from point clouds with good accuracy, the next step for converting segmented point sets into metrically consistent cylindrical primitives for downstream BIM integration remains brittle and heavily dependent on manually tuned geometric hyperparameters. On active construction sites, especially timber frames, severe clutter, occlusion, and variable point density make such tuning labor-intensive and unstable. This paper presents a novel approach called "Cognitive Fitting," in which a Large Language Model (LLM) is integrated into a closed-loop optimization process as a proposal engine for hyperparameter search. The LLM does not generate geometry directly. Instead, it receives structured geometric feedback, including coverage and dispersion for candidate ranking together with fragmentation-related diagnostic signals for iterative reasoning, and proposes improved clustering, fitting, and merging parameters. The proposed framework was validated on a challenging residential timber dataset from Canada containing heavily occluded wiring and piping. Results show the strongest gains on the thin and fragmented *Wires* class, where coverage improved from 66.19% to 71.81% relative to a budget-matched non-LLM search condition, while the number of merged output primitives decreased from 341 to 302, indicating improved continuity recovery. For *Pipes*, the gains were smaller but still consistent, with the proposed method achieving the best overall objective score while reducing dispersion relative to the fixed classical baseline. These results show that LLM-guided reasoning can reduce manual hyperparameter tuning in post-segmentation MEP primitive fitting, particularly for thin and fragmented structures in cluttered timber environments.**

***Index Terms*—Scan-to-BIM, Large Language Models (LLM), Hyperparameter Optimization, MEP Reconstruction, Digital Twins, Automation**

I. Introduction

A paradigm shift is underway in the Architecture, Engineering, and Construction (AEC) industry, aligned with Industry 4.0. Central to this shift is the Digital Twin concept. These are not static models. They are dynamic, living systems for full lifecycle asset management [1, 2]. Creating them starts with Scan-to-BIM. This is the process of transforming terrestrial laser scanning (TLS) data into semantic Building Information Models. TLS hardware today captures geometry rapidly and with millimeter accuracy. The bottleneck has moved downstream. Converting raw point clouds into parametric BIM objects is now the major constraint. Studies indicate this phase can consume 70% of the total modeling effort on a project, all due to manual work [3].

This bottleneck tightens dramatically around MEP systems. Structural elements like walls are large and planar. MEP components are not. They have small cross-sections. Their topologies are complex and curvilinear. They are highly susceptible to occlusion by other building elements [4]. Timber construction compounds these issues. Temporary supports, insulation, and general site clutter further obscure pipes and wires. Standard automated fitting algorithms struggle here. They frequently fail to reconstruct continuous runs. Instead, they output fragmented geometry. Practitioners often call it "spaghetti." This output requires extensive manual rework, defeating the purpose of automation.

We contend the core problem is not the algorithms themselves. Efficient Random Sample Consensus (RANSAC) for shape detection and density-based clustering like Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) are mathematically sound [5, 6]. The problem is their parameter sensitivity. RANSAC, which finds shapes by randomly sampling points and checking for consensus, needs a distance threshold. A threshold ideal for a large pipe will completely miss a thin wire. HDBSCAN, which groups points based on local density, has settings like minimum cluster size. Clustering settings tuned for a dense mechanical room will catastrophically over-segment sparse attic wiring. Today, finding the right set of parameters for a given scan is manual. It is a trial-and-error process that leans heavily on the modeler's tacit knowledge. There is no systematic method.

This study proposes an automated solution. We introduce an optimization framework driven by a Large Language Model (LLM). Our hypothesis is straightforward. The heuristic reasoning a BIM engineer applies to diagnose poor fits and adjust

tolerances can be emulated by an LLM operating as a proposal engine inside a closed-loop geometric search process. In our framework, the LLM does not generate geometry, parse BIM entities, or replace the fitting algorithms. Instead, it receives structured geometric feedback from the pipeline, reasons about likely failure modes, and proposes adjusted hyperparameters for the next candidate evaluation. The geometric engine then evaluates those candidates objectively and retains the best-scoring result.

This distinction is important. The contribution of this paper is not "LLM-generated reconstruction." It is LLM-guided hyperparameter search for post-segmentation primitive fitting in cluttered residential timber scenes. The system is therefore scene- and class-specific: given a semantically segmented point cloud class such as pipes or wires, it searches for better clustering, fitting, and merging settings without requiring manual trial-and-error from an expert modeler.

The diagram illustrates the closed-loop cycle of our framework. At its core is an LLM Agent, instantiated in our experiments with Qwen-2.5-14B under local inference. This agent proposes a set of hyperparameters, θ_t. A Geometric Engine then takes these parameters and the input point cloud. It runs HDBSCAN for clustering and a cylinder fitting stage for primitive estimation, followed by rule-based merging and refinement. Next, an Evaluator module scores the resulting geometry. It computes a scalar objective score J from coverage and dispersion, while also recording additional diagnostic signals such as primitive count and fragmentation. The scalar score is used for candidate ranking, whereas the auxiliary signals are formatted as plain-text feedback and passed to the LLM to support reasoning about the next parameter set, θ_{t+1}. This cycle repeats, pushing the search toward parameter settings that generate cleaner, more continuous, and metrically valid primitives from the segmented point cloud.

To make the feedback loop concrete, the text sent to the LLM may include a short summary such as: "Iteration 7. Parameters used: `min_cluster_size`=45, `ransac_dist`=0.008. Results: Coverage = 58%, Dispersion = 5.2 mm, Merged Primitive Count = 310. The output is still highly fragmented, though fitting error is acceptable." The LLM then reasons over this summary and proposes a new candidate, for example by relaxing clustering conservatism or widening a merge tolerance. This iterative proposal-and-evaluation dialogue is the core of the Cognitive Fitting approach.

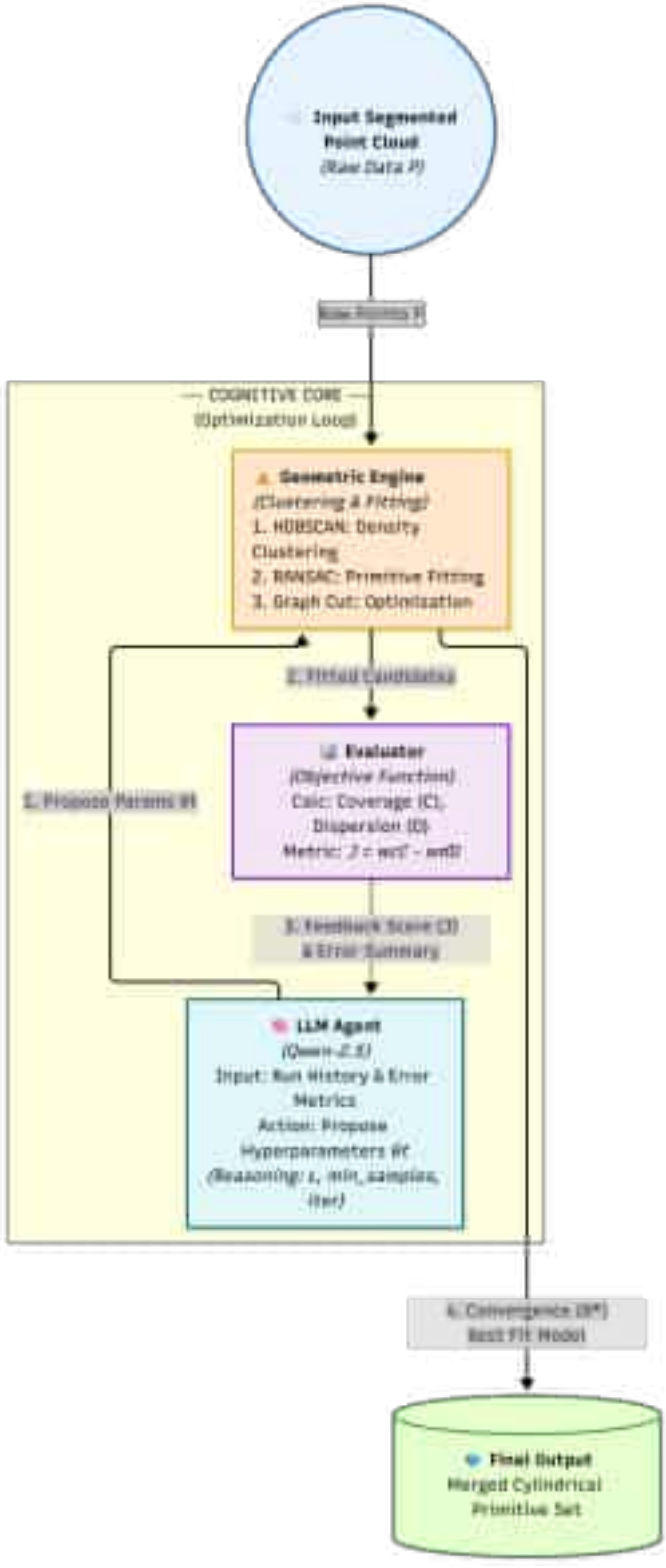


Fig. 1. Architecture of the Cognitive Scan-to-BIM Framework.

II. Related Work

A. Scan-to-BIM: From Manual to Automated Methods

Initial Scan-to-BIM techniques depended on user guidance. They used primitive fitting or constructive solid geometry (CSG) [1]. These methods worked for basic structures. They cannot deliver the semantic depth modern Digital Twins require. The field changed with deep learning. Current state-of-the-art relies on networks like PointNet++ and Point Transformers [7, 8]. They segment point clouds with high per-point accuracy. But segmentation is just the beginning. The

crucial next step is reconstruction. Turning labeled points into parametric IFC entities still uses classical geometric fitting algorithms. And these algorithms have a well-known weakness. They are fragile. Their performance degrades badly with real-world noise, occlusion, and uneven point density [9]. You can have perfect segmentation and still get a useless model if the fitting step fails.

B. Why MEP Reconstruction is Especially Hard

MEP systems break the assumptions of most reconstruction pipelines. Their cylindrical, curvilinear geometry is fundamentally different from planar elements. Agapaki and colleagues made this distinction clear, noting industrial pipes demand unique handling [10]. Some recent work tries to solve this with more AI. Yue *et al*. (2025) developed deep learning networks to complete occluded pipes [11]. These show potential for better connectivity. But they come with a heavy cost. They require large, labeled, domain-specific datasets for training. For residential construction, such datasets are rare and expensive to create. The alternative is procedural methods like RANSAC or Hough Transforms [5]. They need no training data. Their flaw is brittleness. Success hinges on a few hyperparameters. A poor choice for a distance threshold or minimum cluster size can ruin the entire output, merging objects or missing them entirely [12]. This creates a clear gap. We need an approach that adapts like deep learning but does not need massive training sets. Our framework targets this gap directly.

C. LLMs as Optimization Engines

Large Language Models are now being applied beyond text. Researchers are testing them as reasoning tools for complex tasks. Romera-Paredes *et al*. (2024) demonstrated LLMs solving combinatorial puzzles [13]. In engineering, Yang *et al*. (2023) used them to tune optimization parameters [14]. More relevant to us, new studies position LLMs as optimization agents. Work by Wang *et al*. (2025) and van Stein *et al*. (2025) shows LLMs can navigate parameter spaces more efficiently than random search [16, 17]. They use semantic cues to guide the process. This suggests hyperparameter tuning is not just a random walk. It can be a reasoning task.

So, when these three areas are taken together, a specific research gap comes into focus. On one side you have powerful but data-hungry deep learning methods for MEP reconstruction. On the other you have lightweight but brittle geometric algorithms that need expert tuning. The tuning process itself is a kind of heuristic reasoning problem. But the application of LLMs as reasoning agents for this specific task in AEC for tuning geometric reconstruction hyperparameters in the face of real-world clutter has not been explored. Previous LLM work in optimization is general. No one has plugged an LLM into the Scan-to-BIM loop to deal with the conditions of an active construction site. That is the gap this paper addresses.

III. Methodology

The MEP reconstruction task is framed as a black-box optimization problem. The goal is to maximize the geometric quality of the final fitted primitives. The proposed framework eliminates static configuration. A dynamic LLM Agent iteratively tunes the hyperparameters for standard geometric algorithms.

A. Data and Pre-processing

Validation data comes from a Leica BLK360 G1 scanner. It captured an under-construction residential timber building in Canada. This site represents a worst-case test for automation. The raw scans have high geometric noise from wood grain texture. Stud framing creates severe occlusion. The mix includes rigid pipes and flexible electrical wires, each presenting different challenges. The raw point clouds were pre-processed. This involved statistical outlier removal and voxel down-sampling. The data is then semantically segmented into two target classes: 'Pipes' and 'Wires'. These labeled clouds are the input to the system.

B. Cognitive Fitting Engine Components

The pipeline has three core modules, all parameterized by the LLM.

Adaptive Clustering with HDBSCAN: HDBSCAN is used [6]. It handles variable density better than DBSCAN, which is critical for TLS data. In the final reported search space, the primary HDBSCAN parameter optimized by the LLM is *min_cluster_size*, which controls the smallest group of points that forms a valid object. Additional HDBSCAN controls such as *min_samples* and *cluster_selection_epsilon* were explored in preliminary tuning and then fixed before the main reported experiments.

Shape Fitting with RANSAC: Each cluster is processed by an Efficient RANSAC algorithm [5] to find cylindrical models. The key parameter here is *ransac_dist*. This is the maximum allowed distance from a point to the cylinder surface to be considered an inlier. Set it too high and the cylinder swallows noise. Set it too low and valid surface points get rejected.

Graph-Based Merging: Initial cylinders are often broken. A graph traversal step stitches compatible ones into continuous runs. In the final reported search space, the primary merge parameters optimized by the LLM are *merge_angle* (how much cylinder axes can deviate) and *merge_gap* (maximum spatial gap to connect). Additional compatibility guards, including radius-related constraints, were tuned in preliminary runs and then fixed before the main experiments.

C. The LLM Feedback Loop

This is the novel part of the system. The automated reasoning loop.

For each iteration, t:

1) The pipeline runs with the current parameters, θ_t. Performance metrics are computed: **coverage** (C), the percentage of input points assigned to accepted primitives; **dispersion** (D), the average point-to-cylinder fitting error; and structural continuity indicators such as raw and merged primitive counts. These signals are useful because thin MEP structures may exhibit acceptable



(a) Raw Cluttered Environment



(b) Isolated 'Wires' Class



(c) Isolated 'Pipes' Class

Fig. 2. Input Data Characteristics. Vertical breakdown showing (a) the heavy timber clutter, (b) the disjointed nature of the 'Wires' class, and (c) the spatial interference in the 'Pipes' class. Note: Views are zoomed to highlight density.

local fitting error while still remaining globally over-fragmented.

2) A compact plain-text diagnostic summary is produced and provided to the LLM together with recent search context. The LLM was executed locally using Qwen-2.5-14B-Instruct through LM Studio with fixed low-temperature decoding to reduce stochastic variance. The prompt followed a fixed structured template containing the current parameter set, the latest evaluation metrics, and a short instruction to return only the next bounded candidate configuration — acting as a constrained proposal policy rather than an unconstrained conversational assistant. All candidate parameters were clamped to physically meaningful search ranges before evaluation. A salvage stage using a secondary lightweight DBSCAN pass was applied to residual noise before final merging; this used fixed settings across all compared methods and was not part of the LLM-tuned search space.
3) Candidate solutions are ranked using the scalar objective

$$J(\theta) = w_c \times C - w_d \times \left(\frac{D}{0.01}\right), \qquad (1)$$

where the normalization by 0.01 reflects a 10 mm MEP prefabrication tolerance. In all experiments $w_c = 1.0$; $w_d = 0.15$ for wires (favoring continuity) and 0.20 for pipes (favoring precision). These weights were fixed *prior* to running the optimization loop and were not adjusted by observing search outcomes, avoiding circularity. This formulation follows the standard weighted-sum scalarization approach in multi-objective optimization, where weights encode *a priori* preference trade-offs between competing objectives rather than universal constants [15]. In the present study, the selected weights reflect an engineering preference for continuity in wires and geometric precision in pipes. Fragmentation-related signals and primitive counts are treated as auxiliary diagnostic context for the LLM and as secondary evaluation criteria; they are not directly optimized by Eq. (1).

This formulation was chosen deliberately for the present feasibility study. More sophisticated formulations, such as explicit multi-objective optimization or multimodal diagnostic feedback, are promising extensions left outside the scope of this first study.

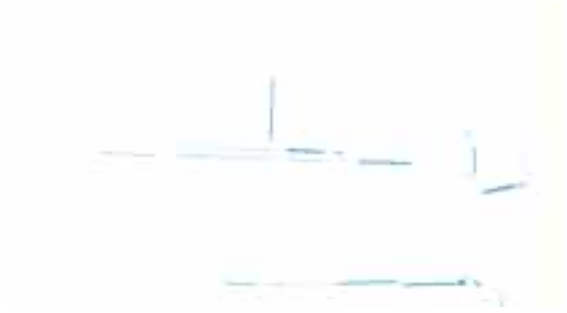
TABLE I
HYPERPARAMETER SEARCH SPACE

Parameter	Unit	Range	Description
hdb_min_cluster	points	[8, 1200]	Minimum points to form a cluster
ransac_dist	meters	[0.001, 0.03]	Max point-to-cylinder distance
merge_angle_deg	degrees	[2, 15]	Max axis deviation for merging
merge_gap	meters	[0.01, 0.12]	Max spatial gap to connect cylinders

Note: The active LLM-tuned search space reported in this paper consists of the four primary parameters above. Additional clustering and merge-guard parameters, including *min_samples*, *cluster_selection_epsilon*, radius-compatibility, axis-offset, and minimum-overlap constraints, were explored in preliminary runs and then fixed prior to the main reported search.

IV. RESULTS

The proposed framework was compared with two baselines. **Baseline A** uses a standard DBSCAN+RANSAC pipeline with fixed, expert-suggested parameters. **Baseline B** is a budget-matched non-LLM search condition using the same autotuning scaffold but with LLM proposals disabled. The search budget was matched by using the same total round count for each class (24 rounds indexed from 0 to 23 for wires, and 32 rounds indexed from 0 to 31 for pipes), thereby isolating the value of LLM-guided search from generic iterative search. We acknowledge that stronger baselines such as Bayesian optimization or TPE would further strengthen this comparison and are identified as future work. Note that Baseline A uses standard DBSCAN whereas the proposed method uses HDBSCAN; the comparison therefore reflects practitioner-style



(a) Fragmented Wires Output (Before fix)



(b) Continuous Wires Run (After LLM fix)

LLM Agent Reasoning Step	
Observation:	Too many pieces (312 fragments)
But:	Fitting is accurate (3.1mm error)
Diagnosis:	`merge_gap=0.02m` too small
Fix:	Increase to `merge_gap=0.08m`
Result:	θ_5: {merge_gap: 0.08m, ...}

The Cognitive Loop:

$\boxed{\theta_t} \xrightarrow{\text{Run}} \boxed{\text{Metrics}} \xrightarrow{\text{LLM thinks}} \boxed{\theta_{t+1}} \xrightarrow{\text{Run}} \ldots$

Fig. 3. Example of parameter tuning driven by the LLM. The agent observes a diagnostic summary indicating high fragmentation despite acceptable fitting error. It therefore adjusts merge-related parameters, and the next iteration produces a more continuous wire run. Fragmentation is used here as reasoning context, not as a direct term in the scalar objective.

fixed-parameter fitting against the full proposed framework, while Baseline B provides the controlled ablation of the LLM guidance component specifically.

A. Quantitative Performance

The data shows a consistent benefit from LLM-guided parameter proposals, with the largest gains observed on the more challenging *Wires* class.

TABLE II
PERFORMANCE COMPARISON SUMMARY

Method	Class	Coverage (%)	Dispersion (mm)	J-Score
Baseline A (DBSCAN)	Wires	56.03	1.54	0.5372
Baseline B (Budget-Matched Non-LLM Search)	Wires	66.19	1.33	0.6420
Ours (LLM-Guided)	**Wires**	**71.81**	**1.40**	**0.6971**
Baseline A (DBSCAN)	Pipes	75.08	4.42	0.6625
Baseline B (Budget-Matched Non-LLM Search)	Pipes	74.15	3.65	0.6685
Ours (LLM-Guided)	**Pipes**	**74.53**	**3.63**	**0.6727**

For wires, the proposed method achieved the clearest advantage. The best LLM-guided solution reached 71.81% coverage with 1.40 mm dispersion and a best objective score of $J = 0.6971$. Relative to the budget-matched non-LLM search condition, which achieved 66.19% coverage and $J = 0.6420$, the proposed method recovered more wire points while preserving low fitting error. The slight increase in dispersion relative to Baseline B (1.33 to 1.40 mm) reflects the coverage–dispersion trade-off encoded in the objective; the net J-score gain of 0.0551 confirms that the additional coverage outweighs the marginal fitting cost. In addition, the number of merged output primitives decreased from 341 to 302, indicating improved continuity recovery and reduced over-fragmentation in the final reconstructed runs. Notably, fragmentation was not a term in the optimized objective J; this improvement therefore appears as an emergent consequence of the LLM's use of auxiliary diagnostic signals rather than direct optimization pressure.

For pipes, the gains were smaller but still consistent. The best LLM-guided solution reached 74.53% coverage with 3.63 mm dispersion and $J = 0.6727$. Compared with the budget-matched non-LLM search condition, the improvement in objective value was modest ($0.6685 \rightarrow 0.6727$), suggesting that the pipe-fitting search landscape is less ambiguous than the wire case. However, compared with the fixed DBSCAN baseline, the proposed method achieved substantially lower dispersion (4.42 mm to 3.63 mm) while maintaining comparable coverage. Notably, Baseline B achieved slightly lower coverage than Baseline A for Pipes (74.15% vs 75.08%), suggesting that undirected search can occasionally degrade fixed-parameter performance on less ambiguous classes within a limited budget.

Taken together, these results support the central claim of the paper. The LLM does not replace the geometry engine; rather, it improves the proposal stage of the search process in those settings where manual tuning would otherwise depend heavily on expert intuition.

B. Convergence Analysis

The search trajectories provide additional insight into how the proposed method behaves. On the *Wires* class, the optimization progressed from an initial score of $J = 0.6215$ in the first evaluated round to a final best score of $J = 0.6971$ by the end of the 24-round search budget. This is not a single-step jump. It is a progressive improvement across the search horizon, indicating that the proposal mechanism moved the search toward increasingly useful regions of the parameter space rather than relying on isolated luck.

By contrast, the budget-matched non-LLM search condition improved more gradually and plateaued earlier, reaching a final best score of $J = 0.6420$. The gap between the two trajectories widened in the later rounds, which is consistent with the quantitative results reported for the *Wires* class. This supports the interpretation that LLM-guided proposals are particularly useful when the reconstruction problem is dominated by fragmentation, continuity recovery, and difficult local trade-offs between clustering and merging.

For *Pipes*, the gains were smaller, which is consistent with the weaker separation observed in the final scores. Taken together, these convergence patterns suggest that the proposed method is most beneficial in the harder thin-class regime, where manual tuning would otherwise rely heavily on qualitative expert judgment.

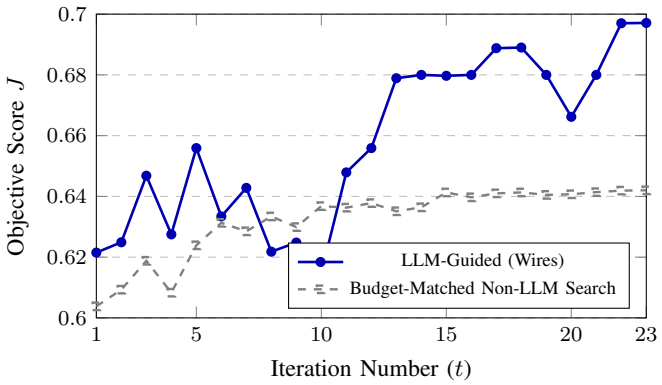


Fig. 4. Measured convergence on the *Wires* class. The LLM-guided search reaches a higher final objective value than the budget-matched non-LLM search condition and exhibits a more directed improvement trajectory over the 23 evaluated rounds.

C. Qualitative Results

Visual inspection supports the numbers. In complex areas like junctions, baseline outputs were often broken or missed curves. The LLM-optimized results show cleaner, more continuous centerlines that navigate occluded spaces more plausibly.



(a) Input Points (Fragmented & Noisy)



(b) Baseline Output (Broken 'Spaghetti')



(c) Cognitive Fitting (Continuous Runs)

Fig. 5. Qualitative Results: Wires. By stacking the views, the continuity improvements in (c) over the baseline (b) become visible. The LLM successfully bridged gaps caused by occlusion.

V. Discussion

The presented results support the core hypothesis of the paper. It was posited that a portion of the heuristic reasoning normally applied by a BIM engineer during primitive-fitting calibration could be approximated by an LLM operating inside a closed-loop search process. The results indicate that this is plausible. The LLM-guided method consistently outperformed the fixed classical baseline and also improved upon the budget-matched non-LLM search condition. This suggests that the search is not purely random and not purely geometric. There is value in a proposal mechanism that can interpret compact diagnostic summaries and steer the next candidate toward more promising regions of the parameter space.

At the same time, the experiments clarify where the benefit is strongest. The largest gains occur on wires rather than pipes. The remaining uncovered wire points are likely a mixture of severe occlusion, sparsity in partially observed runs, and upstream segmentation imperfections, rather than solely a failure of the fitting loop itself. This is an important finding. Wires are the more difficult class because they are thinner, more fragmented, less regular, and more easily confused with clutter or partial observations. The stronger improvement on wires therefore indicates that the proposed framework is particularly relevant for the thin-class frontier in Scan-to-BIM automation, where manual tuning remains most labor-intensive. A natural question is whether a Bayesian optimizer would achieve similar gains within the same iteration budget. We cannot rule this out; comparison with Bayesian optimization or TPE under equivalent budgets remains an important next step. However, the budget-matched non-LLM condition serves a distinct purpose: it isolates whether LLM reasoning over geometric diagnostics provides value beyond generic iterative search of the same space, and the positive result for *Wires* suggests that the LLM's interpretation of structured textual feedback, particularly fragmentation indicators, steers the search in ways that purely numerical methods with no domain-semantic context may not replicate as efficiently under



(a) Input Points (High Occlusion)



(b) Baseline Output (Missing Junctions)



(c) Cognitive Fitting (Recovered Connections)

Fig. 6. Qualitative Results: Pipes. The baseline (b) fails to connect segments separated by structural framing. The proposed method (c) successfully merges these into coherent runs.

small iteration budgets.

The study also makes clear what the current method does *not* yet solve. First, the scalar objective remains intentionally simple. It trades off coverage against dispersion, while fragmentation-related measures are used only as auxiliary diagnostics and secondary quality indicators. This simplicity was useful for a first feasibility study, but it also means the objective is not a complete description of reconstruction quality. Second, the present implementation provides the LLM only with structured text feedback. This isolates the reasoning contribution, but it limits the context available to the agent. Certain failure modes such as false merges in spatially adjacent bundles may be systematically harder to diagnose from scalar summaries alone. Visual residual maps, rendered overlays, or registered RGB cues could plausibly improve diagnosis of failure modes such as under-coverage versus false merging. Third, the current evaluation is based on a single real residential timber project. While this is a practically relevant and difficult testbed, broader cross-site validation is still needed before strong generalization claims can be made.

There are also practical considerations. The weighting in the scalar objective was selected heuristically to reflect class-specific priorities in this dataset. The low-temperature local inference setting improves reproducibility, but prompt sensitivity and repeated-run variance were not exhaustively characterized in the present paper. Finally, the use of an LLM introduces computational overhead that would not exist in a purely geometric optimization loop. A detailed runtime decomposition was not included in the present study and remains an important target for future experimental validation. As a measured reference on the hardware used in this study (RTX 4080, Qwen-2.5-14B-Instruct via LM Studio), each geometric evaluation step required a mean of 30.7 s per candidate across 116 evaluations, while each LLM inference call required a mean of 3.2 s. The full 24-round wire search completed in approximately 61 minutes, this is a one-time autotuning cost per scene class rather than a per-model overhead. For this reason, Cognitive Fitting should be viewed as a targeted autotuning stage for difficult classes and difficult scenes, rather than as a universal replacement for all simpler reconstruction tasks.

Despite these limitations, the practical implications are favorable. The framework automates a stage that is normally dominated by repeated manual trial-and-error. It does so without replacing the underlying geometry engine and without requiring large new training datasets. In that sense, it offers a pragmatic middle ground between brittle classical fitting and fully data-hungry end-to-end learning.

VI. Conclusion and Future Work

This paper presented a Cognitive Scan-to-BIM framework that uses a Large Language Model as a proposal engine for hyperparameter search in post-segmentation MEP primitive fitting. The central contribution is not direct geometry generation by the LLM. Instead, the contribution lies in coupling structured geometric evaluation with LLM-guided parameter proposals inside a closed-loop optimization process.

On the challenging residential timber dataset, the method produced its strongest gains on the *Wires* class, where the best result reached 71.81% coverage with 1.40 mm dispersion and improved continuity relative to both fixed classical fitting and a budget-matched non-LLM search condition. For *Pipes*, the improvements were smaller but still consistent, with the LLM-guided method achieving the best overall objective score while reducing dispersion relative to the fixed baseline. This class-dependent behavior is itself an important result: it indicates that LLM-guided search is most beneficial in thin, fragmented, and topologically ambiguous regimes.

Several limitations remain. The current scalar objective is a practical surrogate rather than a complete multi-objective quality model. Fragmentation is used as diagnostic context but is not yet directly optimized. The LLM receives only textual feedback, not visual or multimodal evidence. The evaluation is limited to one real project dataset, and prompt sensitivity as well as runtime overhead deserve more systematic analysis.

Immediate future work will focus on four directions. First, the objective function will be extended toward a more explicit multi-objective formulation that better reflects continuity and topology preservation. Second, multimodal diagnostic feedback will be investigated by augmenting scalar summaries with rendered residuals and overlap visualizations. Third, broader validation across additional site conditions will be conducted to assess transferability beyond a single timber project. Fourth, the reconstructed primitives will be integrated into a direct Scan-to-Revit workflow through Dynamo or an equivalent API-based bridge, enabling automatic generation of native BIM elements from the optimized cylinder sets.

Overall, the results show that LLM-guided reasoning can be integrated into a classical geometric reconstruction pipeline in a technically meaningful way. This does not eliminate the need for robust geometry processing, but it does reduce the dependence on manual hyperparameter tuning in one of the most difficult parts of Scan-to-BIM automation.

REFERENCES

[1] P. Tang, D. Huber, B. Akinci, R. Lipman, and A. Lytle, "Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques," *Automation in Construction*, vol. 19, no. 7, pp. 829–843, 2010.

[2] R. Volk, J. Stengel, and F. Schultmann, "Building Information Modeling (BIM) for existing buildings—Literature review and future needs," *Automation in Construction*, vol. 38, pp. 109–127, 2014.

[3] Q. Wang and M.-K. Kim, "Applications of 3D point cloud data in the construction industry: A fifteen-year review," *Advanced Engineering Informatics*, vol. 39, pp. 306–319, 2019.

[4] E. Agapaki and I. Brilakis, "CLOI: An Automated Benchmark Framework for Generating Geometric Digital Twins of Industrial Facilities," *Journal of Construction Engineering and Management*, vol. 147, no. 11, p. 04021145, 2021.

[5] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for point-cloud shape detection," *Computer Graphics Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[6] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *The Journal of Open Source Software*, vol. 2, no. 11, p. 205, 2017.

[7] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems*, 2017.

[8] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point Transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 16 259–16 268.

[9] F. Bosché and C. T. Haas, "Automated retrieval of 3D CAD model objects in construction range images," *Automation in Construction*, vol. 17, no. 4, pp. 499–512, 2008.

[10] E. Agapaki, A. Glyn-Davies, S. Mandoki, and I. Brilakis, "CLOI: A Shape Classification Benchmark Dataset for Industrial Facilities," in *Proceedings of the ASCE International Conference on Computing in Civil Engineering 2019*, 2019, pp. 66–73.

[11] H. Yue, Q. Wang, Y. Yan, and G. Huang, "Deep learning-based point cloud completion for MEP components," *Automation in Construction*, vol. 175, p. 106218, 2025.

[12] M. Bassier, B. Van Genechten, and M. Vergauwen, "Classification of sensor independent point cloud data of building objects using random forests," *Journal of Building Engineering*, vol. 21, pp. 468–477, 2018.

[13] B. Romera-Paredes *et al.*, "Mathematical discoveries from program search with large language models," *Nature*, vol. 625, pp. 468–475, 2024.

[14] C. Yang, X. Wang, Y. Lu, H. Liu, Q. V. Le, D. Zhou, and X. Chen, "Large language models as optimizers," *arXiv preprint arXiv:2309.03409*, 2023.

[15] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization," *Structural and Multidisciplinary Optimization*, vol. 41, no. 6, pp. 853–862, 2010.

[16] W. Wang, J. Peng, M. Hu, W. Zhong, T. Zhang, S. Wang, Y. Zhang, M. Shao, and W. Ni, "LLM Agent for Hyper-Parameter Optimization," *arXiv preprint arXiv:2506.15167*, Jun. 2025.

[17] N. van Stein, D. Vermetten, and T. Bäck, "In-the-loop Hyper-Parameter Optimization for LLM-Based Automated Design of Heuristics," *ACM Transactions on Evolutionary Learning and Optimization*, 2025.

A Traceable Clause-Level Retrieval and Evidence-Based Question Answering System for Tunnel Specifications

Yafei Sun[1], Xuesong Shen[1, *], Sisi Zlatanova[2], Khalegh Barati[1], Milad Mousavi[1], and James Linke[3]

[1]School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia
[2]School of Built Environment, University of New South Wales, Sydney, NSW 2052, Australia
[3]GeoAI, Sydney, NSW 2019, Australia
yafei.sun@unsw.edu.au, x.shen@unsw.edu.au, s.zlatanova@unsw.edu.au, khalegh.barati@unsw.edu.au, milad.mousavi@unsw.edu.au, j.linke@geoai.au

Abstract

Tunnel engineering specifications are critical normative documents for compliance verification and decision support. Existing text-based retrieval methods do not sufficiently exploit the hierarchical structure between clauses, making it difficult to reliably locate evidence and prioritize core clauses. This paper proposes a clause-level retrieval and evidence localization method for tunnel specifications that explicitly leverages document structure as retrieval evidence. A two-stage framework is adopted, where text-based candidate retrieval is followed by structure-aware reranking. A lightweight structural knowledge graph is constructed to encode clause hierarchy and document membership, enabling the derivation of interpretable document-level aggregation and clause-level structural cues. These features provide an explainable structural bias for ranking without training reranking models or requiring additional annotation. Experiments on actual tunnel specifications with 50 engineering queries show improved top-ranking accuracy and evidence coverage over lexical and semantic baselines, while keeping millisecond-level retrieval latency. The results indicate that modeling structural cues is an effective foundation for evidence-based question answering.

Keywords –

Tunnel specifications; Clause retrieval; Structural knowledge graph; Structure-aware reranking.

1 Introduction

Tunnel engineering is characterized by strong compliance requirements and high complexity, involving the coordination of multiple systems such as electrical and ventilation systems [1]. Misinterpretation or omission of any key clauses can lead to significant safety consequences. Thus, tunnel specifications and technical standards documents play a crucial role as "traceable compliance evidence" throughout the entire operations management process [2]. In current practice, engineers typically rely on manual browsing or keyword search within document management systems to locate relevant clauses, a process that becomes increasingly time-consuming as specification sets grow in size and complexity. The risk of overlooking a critical clause or retrieving a similar but incorrect provision is amplified when related requirements are distributed across multiple hierarchical levels within a single document. Given the volume and depth of such documents, quickly locating the specific clauses that directly support decision-making remains a persistent challenge. Traditional information retrieval based on chapters or documents is insufficient to meet the needs of refined compliance retrieval and specification question answering (QA) [3, 4]. Therefore, clause-level retrieval and evidence location have become foundational for tunnel intelligent applications.

Existing document retrieval methods for specific domains often rely on keyword matching or semantic similarity calculations, treating clauses as independent text units [5, 6]. When dealing with tunnel specifications, these methods lack the application of explicit hierarchical organization and document-level aggregation patterns. At the clause level, simply ranking based on text similarity easily leads to scattered evidence clauses and core clauses being placed at the end; while retrieval granularity at the document or chapter level is insufficient to directly support the output of locatable and citable clause-level evidence. Specification clauses carry implicit structural context—such as document membership, hierarchical depth, and clause adjacency [7]—that, if left unexploited, limits the interpretability and traceability of retrieval results [8]. However, how to transform the specification structure into a computable and interpretable ranking criterion without introducing costly annotation or complex semantic modeling remains the core problem facing tunnel specification retrieval.

To address this issue, this paper proposes a clause-

level retrieval method for tunnel specifications that explicitly models document hierarchy as structural evidence for ranking. The method adopts a two-stage pipeline: (i) text-based candidate recall over clause content, followed by (ii) structure-aware reranking within the candidate set. We construct a lightweight structural knowledge graph (KG) with clauses as nodes to encode document membership and hierarchical position, from which interpretable structural cues are derived to bias ranking toward coherent and citable evidence. This design avoids complex semantic KG and trained reranking models, while capturing the "structure-as-evidence" property of specification texts.

The main contributions of this paper are as follows:

- A clause-level retrieval and evidence localization framework for tunnel specifications that outputs traceable clauses.
- A lightweight structural KG with clauses as nodes to encode document membership and hierarchy as a unified source of provenance information.
- A structure-aware reranking scheme, which uses document-level aggregation and clause-level structural cues.

The remainder of this paper is organized as follows. The following sections review related work, describe the proposed method, present the experimental results, and conclude with key findings and future directions.

2 Related Works

This section reviews prior work on tunnel specification retrieval, retrieval and reranking methods for domain-specific documents, and summarizes the key research gaps motivating this study.

2.1 Tunnel Specification Retrieval

Existing research mainly focuses on the digital management of specification texts, keyword retrieval, and rule-based clause location [9, 10]. Some works have integrated with engineering management systems to support the querying, browsing, and citation of specification clauses. However, these methods often use documents or chapters as the basic retrieval unit, emphasizing information organization and retrieval efficiency, and making limited use of the explicit hierarchical organization and clause structure within the specification texts. For specification of QA and compliance verification scenarios addressing specific engineering problems, these methods often struggle to directly output locatable and citable clause-level evidence, thus limiting their applicability in complex tunnel engineering practices.

2.2 Retrieval and Reranking Methods

In the broader processing of technical documents, information retrieval research often employs a two-stage retrieval paradigm [11, 12]. This involves initial candidate recall through keyword matching or semantic similarity models, followed by finer-grained reranking on the candidate set to improve ranking performance. This paradigm has been widely validated in balancing retrieval efficiency and result quality. Regarding retrieval tasks involving long documents and normative texts, some research has begun to focus on the role of document structure information (such as chapters, levels, or metadata) in evidence localization. It has attempted to improve retrieval or QA performance through structure-aware mechanisms [13-16]. However, it should be noted that research explicitly leveraging document structure for clause-level retrieval in normative texts is still relatively limited. Existing studies mainly explore the use of structural information as auxiliary context or organizational metadata, such as for text segmentation, navigation, or result presentation, rather than treating document structure as a computable and interpretable ranking signal. In addition, the retrieval granularity in most existing works is typically defined at the document or chapter level, which poses challenges when directly addressing fine-grained, hierarchically organized clauses in specification documents.

2.3 Research Gaps

Despite these initial efforts, research on structure-aware clause-level retrieval for normative engineering documents remains limited, leaving several open challenges. The review of existing research reveals several shortcomings in retrieval methods for tunnel engineering specifications. First, retrieval granularity is generally limited to the document or chapter level, failing to directly support the location and citation of clause-level evidence required in engineering practice. Second, while structure-aware retrieval is gaining attention, most methods rely on learned reranking models or complex semantic modeling, failing to explicitly incorporate structural cues into the ranking decision process in a low-cost, highly interpretable manner. Third, in applications such as retrieval-augmented generation, structural information is primarily used for text segmentation or organization, lacking research on systematically modeling document-level aggregation and clause-level hierarchical cues on candidate clause sets. Unlike approaches that rely on learned reranking models or complex semantic knowledge graph construction, this study treats structural cues as directly computable ranking features derived from document metadata, without model training or manual ontology design. This positions the approach as an interpretable and low-cost

alternative suited to engineering applications where deployment simplicity and model transparency are practical priorities.

3 Research Methodology

This section introduces the proposed clause-level retrieval method and system framework, including the overall process, clause-level structural representation, and structure-aware reranking strategy.

As shown in Figure 1, the system takes natural language queries as input, performs candidate recall and structure-aware reranking on a clause-level corpus, and outputs the ranked relevant clauses and their evidence fragments. The overall process adopts a two-stage design.

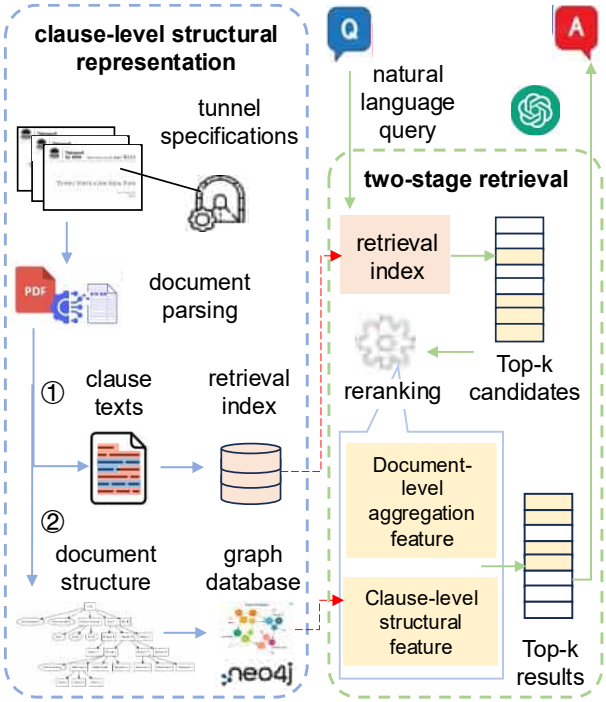


Figure 1. Framework of the proposed clause-level retrieval and evidence-based question answering system.

3.1 Clause-Level Structural Representation

The first step is parsing the specification document into a corpus with clauses as the basic unit. Each clause corresponds to an explicit entry in the specification. It is associated with structural meta-information such as its text content, clause number, the specification document it belongs to, and its hierarchical depth. The clause number and hierarchical depth are automatically inferred from the original chapter structure, used to characterize the relative position of the clause in the overall specification. Based on the above information, a lightweight structural knowledge representation is built, in which clauses are treated as nodes, and their hierarchical and membership relationships are explicitly encoded as node attributes. This structural representation can be regarded as a structural KG of the specification, serving as a unified structural basis in the subsequent ranking stage, without relying on complex semantic relation extraction or manual ontology modeling. Engineering specifications frequently contain explicit internal cross-references between clauses (e.g., "refer to Clause 6.3"). The current implementation does not parse such references at the indexing stage. However, the document-level aggregation mechanism in the reranking stage partially addresses this limitation: when multiple related clauses from the same specification appear in the candidate set, their shared document membership reinforces their ranking, indirectly capturing within-document co-referencing patterns. Explicit cross-reference parsing and graph-based linking represent a direction for future structural enrichment.

3.2 Candidate Retrieval

During the candidate retrieval phase, the system performs text-based retrieval over the clause corpus and returns the Top-k candidate clauses, narrowing the search space while maintaining high recall for subsequent structure-aware reranking. Each candidate corresponds to a complete clause text at the smallest granularity, ensuring a uniquely locatable and citable unit. All subsequent reranking operations are conducted on this unified Top-k candidate set.

3.3 Structure-Aware Reranking

Specification documents exhibit a highly explicit hierarchical structure, with relevant evidence typically distributed in “clause groups” within the same document or adjacent levels, rather than appearing as isolated clauses. When ranking solely based on text similarity, models often struggle to leverage this structured evidence distribution, leading to core clauses with stronger responsiveness failing to consistently appear at the top.

To address this, the reranking strategy incorporates document-level and clause-level structural cues from the KG into the initial text relevance ranking, applying a structured evidence bias that guides results toward more representative clauses without altering the candidate set.

The reranking process utilizes two complementary structural signals: first, document-level aggregation signals, where multiple candidate clauses originate from the same specification, indicating stronger overall relevance within the current query context. Therefore, candidate clauses within the same document are weighed to reinforce the “evidence clustering within the document” pattern. Second, clause-level structural signals,

leveraging attributes such as the hierarchical depth of clause nodes in the structural KG to characterize the functional differences of clauses in normative expression. Thus, they can be seen as structural constraints alongside textual relevance in the ranking process. Through the incorporation of document-level and clause-level structural cues, the reranking mechanism adjusts the relative ordering of candidate clauses in a structure-aware manner without changing the underlying retrieval results.

4 System Evaluations

To evaluate the proposed method, this paper conducts experimental analysis from aspects such as components ablation and external comparison and discusses the retrieval effectiveness and answer quality.

4.1 Experimental Setup

The corpus used is the design compliance specifications from New South Wales, Australia, covering multiple subsystems in tunnels, including electrical and lighting systems. A total of six specifications were selected as the corpus source for the system's retrieval and QA. The documents involved are shown in Table 1.

Table 1. Summary of specifications used in evaluation.

Code	Name	Scope	Clause number	Pages
D&C R153	Tunnel and underpass main switchboard, distribution boards and control panels	electrical boards	33	13
QA R154	Tunnel and underpass electrical services works	electrical services	46	13
QA R158	Road tunnel and underpass lighting	lighting equipment	45	17
D&C R163	Tunnel ventilation axial fans	ventilation equipment	61	27
D&C TS931	Tunnel electrical boards	electrical boards	35	16
D&C R164	Tunnel jet fans	ventilation equipment	65	27

The evaluation experiments were run in a WSL2 (Ubuntu) environment on a Windows host machine. During the experiments, the WSL environment had 12 logical processors (Intel® Core™ i7-8700 @ 3.20 GHz) and approximately 7.7 GB of memory (plus 2.0 GB swap). No graphics processing unit acceleration or model training was used during the experiments; retrieval and reranking were performed on the local central processing unit, and QA generation was completed online through OpenAI large language model interface. This configuration demonstrates the lightweight nature and reproducibility of the proposed method in engineering applications.

The system is evaluated from three aspects: retrieval effectiveness, system efficiency, and QA quality. Retrieval is assessed using standard information retrieval metrics, including Hit@k and Mean Reciprocal Rank (MRR). Hit@k indicates whether at least one relevant evidence clause appears in the Top-k results; MRR measures the reciprocal rank of the first relevant clause. The system efficiency is reported as average query latency. QA quality is evaluated against a small set of reference answers. Over generation, evidence-attribution constraints are enforced to ensure traceability and auditability. Unless otherwise stated, k is set to 5 in all evaluation experiments; the term Top-k is used to describe the general retrieval framework, while Top-5 refers to the specific cutoff adopted for quantitative evaluation.

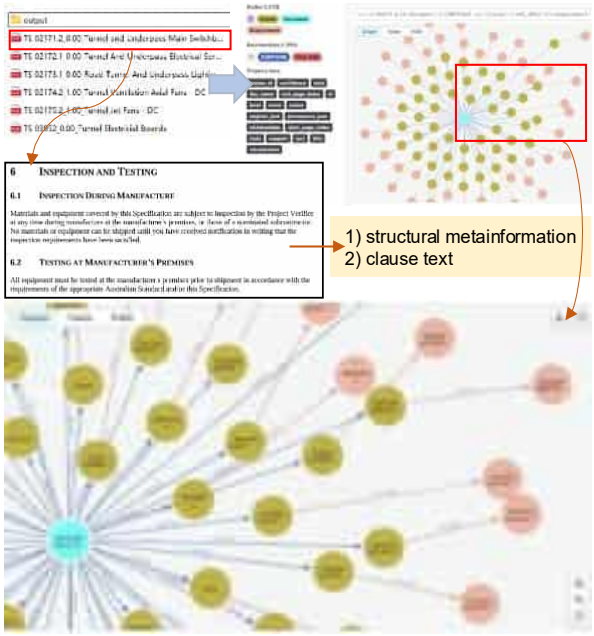


Figure 2. Clause-level KG scaffold constructed from specification documents.

4.2 Results

This subsection evaluates the proposed approach from three aspects: end-to-end functional performance, key component ablation, and comparative analysis.

4.2.1 Functionality Demonstration

This part uses representative cases to demonstrate key system outputs. The proposed system automatically parses specification PDFs into a clause-level index and constructs a KG scaffold encoding clauses, requirement entities, and their hierarchical relationships, as shown in Figure 2.

For the retrieval phase, taking specification QA R154 as an example, the query "What are the requirements for cable routing in tunnel electrical systems?" is selected. The system returned the Top-5 evidence clauses, as shown in Figure 3. Among them, Clause 6.2 "Cable Routing" was consistently ranked first, and its content directly specifies aspects such as cable route selection and support; the remaining returned clauses provide relevant information from aspects such as conduit specifications.

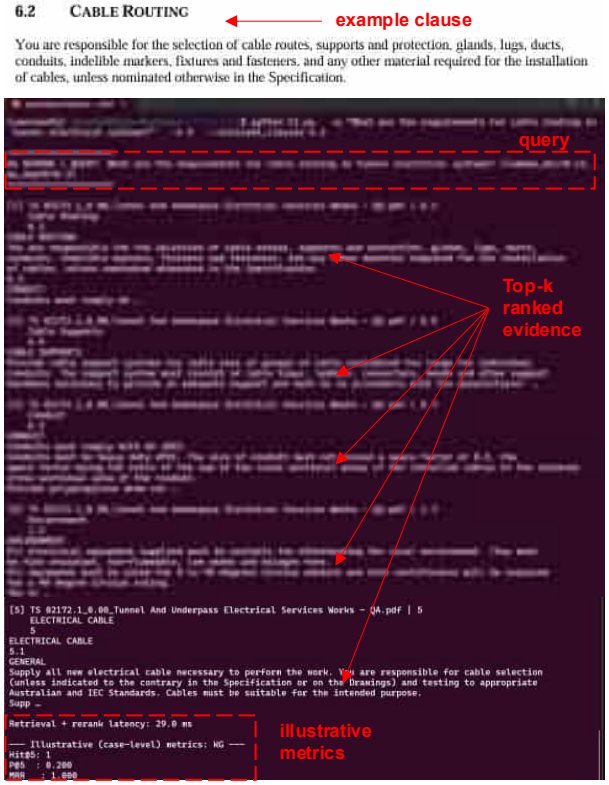


Figure 3. Example clause retrieval results.

Based on this representative case, exemplary (case-level) retrieval metrics are further reported: the system successfully covered the core relevant clauses in the Top-5 results (Hit@5 = 1), and the first relevant clause was ranked first (MRR = 1.00). In addition, the response latency for one retrieval and reranking was approximately 29.0ms, indicating that the system can efficiently return a set of evidence highly relevant to the specification requirements in interactive use scenarios.

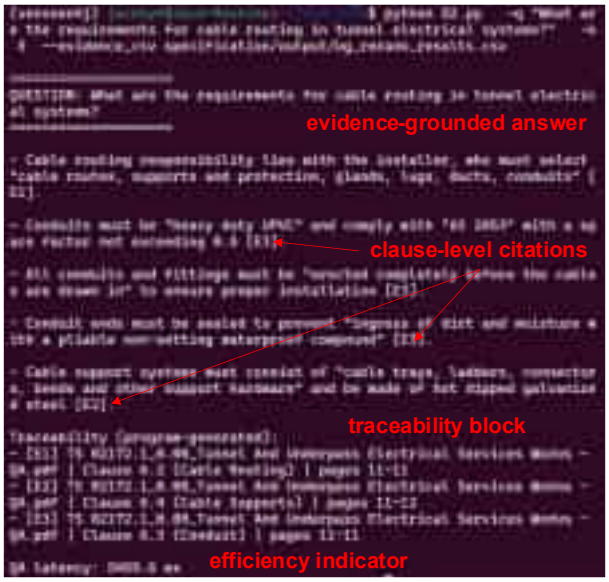


Figure 4. Evidence-based QA output.

During the QA phase, the system generates an answer from the Top-k evidence clauses, as shown in Figure 4. Each conclusion must cite its source clause, and a traceability list maps answers to specific documents, clause numbers, and page numbers, forming a verifiable "answer-evidence" closed loop. The end-to-end QA latency for this example was approximately 5.46 seconds.

4.2.2 Ablation Experiment

This subsection evaluates the impact of two types of rearrangement features derived from the KG structure design, document aggregation features and clause depth features, on the ranking performance of evidence clause retrieval through component ablation experiments.

The experiment was built upon 50 engineering queries designed around the typical information requirements of tunnel specifications. These queries were divided by topic and sampled evenly (e.g., electrical and mechanical, fire protection and safety) to avoid bias caused by queries focusing on a single chapter or term. Each query was expressed as a natural language question, consistent with the clause-level evidence retrieval task. In terms of the experimental procedure, the system first used BM25 [17], a widely used keyword-based ranking function, to generate a candidate clause list. Then, under the same candidate set and parameter settings, the BM25 baseline, a complete model incorporating only a single KG structural feature, and a complete model incorporating both types of KG structural cues were compared. For clarity, we denote the document-level aggregation feature as Doc, and the clause-level hierarchical depth feature as Depth in the ablation variants. Evaluation used Hit@1 and Hit@5, where @1 checks whether the top-ranked clause is relevant and @5 checks whether at least one relevant clause appears

among the top five results; all automatic judgments were manually reviewed and confirmed.

Table 2 shows a comparison of retrieval performance across different ablation variants. The BM25 baseline achieves a Hit@5 score of 0.84, indicating that keyword-based candidate recall is highly usable given the highly structured nature of the specification clause titles. However, its Hit@1 score is only 0.56, suggesting that clauses that directly answer the query are not always prioritized. Introducing document aggregation features derived from the KG structure design improves Hit@1 to 0.62, demonstrating that document-level aggregation using document-clause attribution relationships helps prioritize valid evidence within the same specification document.

Table 2. Ablation experiment results

Variant	Hit@1	Hit@5
BM25	0.56	**0.84**
Depth	0.54	0.82
Doc	0.62	0.80
Full	**0.62**	0.80

In contrast, introducing only clause depth features has limited effect, and the complete model does not improve over the Doc-only variant, indicating that document-level aggregation is the primary source of ranking gain in this task setting. Hit@5 shows little variation across variants, suggesting that the main benefit of structural cues lies in improving top-ranking quality rather than expanding candidate coverage.

4.2.3 Comparative Experiment

This subsection compares the proposed method against lexical, semantic vector, and pretrained embedding baselines under a unified setting to examine the effectiveness of structured evidence bias. A more challenging and application-oriented query set consisting of 50 natural-language specification questions is adopted. These queries were formulated based on domain consultation to reflect typical information needs in tunnel specification compliance workflows, covering electrical, lighting, ventilation, and control panel subsystems. While the queries were authored by the research team rather than extracted from actual engineer search logs, they were designed to span diverse subsystem topics and avoid single-chapter concentration. Integrating real-world search logs from engineering practice is recognized as a direction for future validation.

Compared with the ablation queries, these questions exhibit weaker lexical overlap with clause titles/headings, longer semantic expressions, and less explicit structural cues, making purely keyword-based retrieval substantially harder.

In the comparative experiments, the lexical retrieval method used the standard BM25 model as the baseline; the semantic retrieval method was based on latent semantic analysis, representing clause texts as low-dimensional semantic vectors and ranking them according to cosine similarity [18, 19]; our proposed method, based on the initial BM25 ranking results, introduced document-level aggregation features and clause-level features to perform structure-aware reranking of candidate clauses. All methods returned results under a unified Top-k setting, evaluated using the same weakly supervised judgment rule as the ablation experiments, with all relevance labels manually verified.

As shown in Table 3, our proposed method outperforms the comparative methods on both evaluation metrics. Compared to the Lexical retrieval baseline, the structure-aware reranking strategy achieves a stable improvement in the Hit@1 metric, indicating that introducing document-level and clause-level structural cues helps to prioritize clauses with greater judgment value. On the Hit@5 metric, our proposed method achieves near-saturation retrieval coverage, demonstrating that structured evidence bias can effectively reduce the omission of key clauses.

Table 3. Comparative experiment results

Method	Hit@1	Hit@5
Lexical retrieval	0.16	0.82
Latent semantic vector retrieval	0.12	0.60
Pretrained embedding (text-embedding-3-small)	0.04	0.26
Pretrained embedding (MiniLM-L6-v2)	0.06	0.30
Ours (structure-aware reranking)	**0.18**	**1.00**

To examine whether pretrained text embedding models offer advantages in this setting, two additional baselines were evaluated: OpenAI text-embedding-3-small and all-MiniLM-L6-v2 [20]. Clause texts were encoded into dense vectors and retrieval was performed via cosine similarity ranking. As shown in Table 3, both pretrained models achieved the lowest performance among all methods, falling below even the unsupervised LSA baseline. This result indicates that general-purpose pretrained embeddings, without domain-specific adaptation, do not transfer effectively to normative specification retrieval in the current zero-shot setting, where clause-level functional distinctions are encoded in domain-specific terminology rather than general semantic proximity. Together with the LSA result, these findings confirm that purely semantic approaches—whether corpus-specific or pretrained—are insufficient to capture the functional differences between specification clauses in this retrieval task. Overall, the

results support the effectiveness of structural information in normative clause retrieval tasks.In addition, the Hit@1 performance of the BM25 baseline drops significantly compared to the ablation setting, indicating that lexical matching alone is insufficient for evidence-oriented clause retrieval when queries deviate from structured specification terminology.

4.3 Discussion

The experimental results confirm that the structured organization of specification documents provides natural contextual constraints for retrieval, while purely semantic approaches are insufficient to characterize clause-level functional differences. Incorporating structural information into the ranking process improves both coverage and ranking reliability.

From a practical standpoint, the improvement in Hit@1 from 0.56 to 0.62 in the ablation setting means that for an additional 6% of queries, the most relevant clause is ranked first, reducing the need for engineers to manually inspect lower-ranked results. In the comparative evaluation, the proposed method achieved Hit@5 = 1.00, meaning that for all 50 queries, at least one relevant clause appeared within the top five results. In compliance verification workflows, where overlooking a key regulatory clause can carry safety and legal consequences, this near-complete evidence coverage directly supports the practical goal of reducing clause omission risk in engineering decision-making.

To assess QA generation quality beyond case-level demonstration, ROUGE scores [21] were computed for 44 of the 50 queries for which both retrieval results and matching ground-truth clause labels were available. The scores were computed by comparing system-generated answers against the full text of their ground-truth relevant clauses. The mean ROUGE-L precision reached 0.468, indicating that approximately 47% of generated answer content can be traced back to the source clause text. The corresponding ROUGE-L recall (0.053) and F1 (0.090) were lower, which reflects the expected asymmetry between concise bullet-point answers and full-length reference clauses—the system is designed to extract key provisions rather than reproduce entire clause content. These results confirm that the evidence-grounding constraints imposed during answer generation are effective in maintaining content fidelity to the source specifications.

Overall, the experiments provide consistent evidence supporting the design rationale and practical applicability of the proposed system in specification clause retrieval scenarios. Moreover, while this study focuses on tunnel specifications, the approach can be extended to other road-infrastructure normative documents with similar hierarchical structures to improve evidence-oriented operations management [22].

5 Conclusions

This paper developed a provenance-aware, clause-level retrieval and evidence localization system for tunnel engineering specifications. The proposed approach adopts a two-stage pipeline and a lightweight structural KG. The main contribution of this study lies in explicitly treating document structure as retrieval evidence rather than auxiliary context. Two structural signals, i.e. document-level aggregation and clause-level hierarchical cues, are incorporated into the ranking process without relying on trained reranking models or additional annotation, enabling a lightweight yet explainable retrieval mechanism suitable for engineering applications.

Experimental results demonstrate that the proposed method improves top-ranking reliability, with document-level aggregation increasing Hit@1 from 0.56 to 0.62 and the full approach achieving near-saturated Hit@5 coverage in comparative evaluations. Overall, the results confirm that explicitly modeling structural cues enhances ranking consistency while maintaining low computational overhead.

Future work will explore richer structural signals, extend the evaluation to broader categories of normative documents, and conduct more rigorous assessments of evidence-based QA quality under real-world engineering scenarios.

Acknowledgment

This research is supported by funding from the Australian Research Council (ARC) Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), and the ARC Research Hub for Human-Robot Teaming for Sustainable and Resilient Construction (IH240100016).

References

[1] Bjelland, H., Gehandler, J., Meacham, B., Carvel, R., Torero, J. L., Ingason, H., & Njå, O. (2024). Tunnel fire safety management and systems thinking: Adapting engineering practice through regulations and education. Fire Safety Journal, 146, 104140.

[2] Chen, L., Shi, P., Tang, Q., Liu, W., & Wu, Q. (2020). Development and application of a specification-compliant highway tunnel facility management system based on BIM. Tunnelling and Underground Space Technology, 97, 103262.

[3] Lee, S. H., Choi, S. W., & Lee, E. B. (2023). A question-answering model based on knowledge graphs for the general provisions of equipment purchase orders for steel plants

maintenance. *Electronics*, *12*(11), 2504.

[4] Du, H., Feng, Y., Li, C., Li, Y., Lan, Y., & Zhao, D. (2023, July). Structure-discourse hierarchical graph for conditional question answering on long documents. In *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 6282-6293).

[5] Yousefi, S., & Dami, S. (2024, February). Medical Documents Search Engine in the Comprehensive Hospital System Using Ontology-Based Semantic Similarity Measurement. In *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)* (pp. 1-6). IEEE.

[6] Yang, Y., Tang, W., Lu, Y., Wang, J., & Jiang, Y. (2024, October). Similarity Measurement Model for Standard Clauses Based on RAG. In International Conference on Computer Engineering and Networks (pp. 319-330). Singapore: Springer Nature Singapore.

[7] Chen, H., Yang, Y., Li, Y., Zhang, M., & Zhang, M. (2025). DISRetrieval: Harnessing Discourse Structure for Long Document Retrieval. *arXiv preprint arXiv:2506.06313*.

[8] Ma, Y., Wu, Y., Ai, Q., Liu, Y., Shao, Y., Zhang, M., & Ma, S. (2023). Incorporating structural information into legal case retrieval. *ACM Transactions on Information Systems*, *42*(2), 1-28.

[9] Ling, J., Li, X., Li, H., An, Y., Rui, Y., Shen, Y., & Zhu, H. (2024). Hybrid NLP-based extraction method to develop a knowledge graph for rock tunnel support design. *Advanced Engineering Informatics*, *62*, 102725.

[10] Yu, G., Wang, Y., Mao, Z., Hu, M., Sugumaran, V., & Wang, Y. K. (2021). A digital twin-based decision analysis framework for operation and maintenance of tunnels. *Tunnelling and underground space technology*, *116*, 104125.

[11] Dang, V., Bendersky, M., & Croft, W. B. (2013, March). Two-stage learning to rank for information retrieval. In *European Conference on Information Retrieval* (pp. 423-434). Berlin, Heidelberg: Springer Berlin Heidelberg.

[12] Ren, R., Ma, J., & Zheng, Z. (2025). Large language model for interpreting research policy using adaptive two-stage retrieval augmented fine-tuning method. *Expert Systems with Applications*, *278*, 127330.

[13] Jia, R., Zhang, B., Méndez, S. J. R., & Omran, P. G. (2025, May). StructRAG: Structure-Aware RAG Framework with Scholarly Knowledge Graph for Diverse Question Answering. In *Companion Proceedings of the ACM on Web Conference 2025* (pp. 2567-2573).

[14] Wu, Chengke, et al. "Retrieval augmented generation-driven information retrieval and question answering in construction management." *Advanced Engineering Informatics* 65 (2025): 103158.

[15] Xu, L., Feng, C., Zhang, K., Zhengyong, L., Xu, W., & Meng, F. (2025, October). Equipping Retrieval-Augmented Large Language Models with Document Structure Awareness. In *Findings of the Association for Computational Linguistics: EMNLP 2025* (pp. 24608-24631).

[16] Wang, S., Zhou, Y., & Fang, Y. (2025). BookRAG: A Hierarchical Structure-aware Index-based Approach for Retrieval-Augmented Generation on Complex Documents. *arXiv preprint arXiv:2512.03413*.

[17] Akkuş, Ş. G., & Emekci, H. (2025, August). Ranking Matters: A Comparative Study of BM25, BERT-E5, and Hybrid (BM25+ BERT-E5) Retrival Systems on the SQuAD 2.0 Dataset. In *2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)* (pp. 1-6). IEEE.

[18] Li, B., Wen, H., Wang, S., Hu, T., Liang, X., & Luo, X. (2025). An Optimized Semantic Matching Method and RAG Testing Framework for Regulatory Texts. *Electronics*, *14*(14), 2856.

[19] Collini, E., Kurniadi, F. I., Nesi, P., & Pantaleo, G. (2025). Context-Aware Retrieval Augmented Generation Using Similarity Validation to Handle Context Inconsistencies in Large Language Models. *IEEE Access*.

[20] Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*. 2019.

[21] Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out*. 2004.

[22] Sun, Y., Shen, X., Zlatanova, S., Barati, K., & Linke, J. (2026). Text-based automatic knowledge graph construction for road infrastructure operations management. *Automation in Construction*, *182*, 106733.

Automated BIM-Ready 3D Reconstruction from Muon Flux Tomography Data: An IFC-Based Integration Pipeline for Bridge Inspection

Soheila Kooakalani[1], Ioannis Brilakis[1], Sander Sein[2], Tengiz Pataraia[2], Ya Wen[1], Mudan Wang[1], Xiaofang Wen[1]

[1]Department of Engineering, University of Cambridge, Cambridge CB3 0FA, UK.
[2]GScan Ltd, The Bradfield Centre, 184 Cambridge Science Park, Milton Road, Cambridge CB4 0GA, UK.
sk2268@cam.ac.uk, ib340@cam.ac.uk, sander.sein@gscan.eu, tangiz.pataraia@gscan.eu, yw710@cam.ac.uk, mw923@cam.ac.uk, xw436@cam.ac.uk

Abstract –

Ageing bridge infrastructure requires scalable, accurate, and minimally intrusive inspection methods that can reveal subsurface conditions while producing outputs compatible with digital engineering workflows. Muon Flux Technology (MFT) enables passive, non-destructive internal imaging of reinforced concrete elements, producing high-value geometric outputs that can support infrastructure inspection and asset management. A persistent barrier to operational deployment is the limited interoperability between processed MFT-derived outputs and standard Building Information Modelling (BIM) ecosystems. This paper presents a deterministic, programmatic pipeline that converts processed MFT-derived geometries into Industry Foundation Classes 4 (IFC4)-compliant models using a geometry-to-IFC tessellation workflow. Each detected object is provided as an individual Polygon File Format (PLY) file containing either a triangle mesh or a point cloud. Where only point data are available, the workflow reconstructs a surface via Ball Pivoting, applies mesh cleaning for topological stability, and exports the result as an IFC4 tessellated shape. A minimal IFC spatial hierarchy (Project-Site-Building-Storey) is authored to ensure broad compatibility across BIM tools and consistent unit handling. The resulting IFC outputs enable downstream coordination, archival, and integration of MFT-derived 3D information within open BIM workflows, supporting objectives related to data conversion, automation, integration, and standards compliance.

Keywords –

Muon tomography; Non-destructive testing; BIM integration; IFC; Bridge inspection.

1 Introduction

Ageing bridge infrastructure requires inspection methods that can reveal internal condition while minimizing traffic disruption and intrusive testing. Conventional non-destructive testing (NDT) methods, such as ground-penetrating radar, X-ray, and ultrasound, often face practical limitations such as limited penetration depth, operational complexity, cost, and restricted integration into digital workflows. Surface-based capture is widely adopted but does not provide subsurface insight critical for reinforced concrete condition assessment.

Muon Flux Technology (MFT) provides a complementary sensing modality. Muon-based imaging supports 3D internal imaging and can enable detection of internal features such as steel element geometry, voids, and other anomalies by measuring naturally occurring atmospheric muons and their scattering or attenuation through objects.

The M-TABS project aims to advance muon flux tomography for bridge assessment and to incorporate the resulting information stream into established formats such as 3D Building Information Modelling (BIM), thereby enabling integration into broader infrastructure maintenance practices. In this context, the paper addresses the automated conversion of processed MFT-derived outputs into BIM-compatible representations using industry-standard data formats [1]. Although the broader MFT-to-BIM vision is intended for bridge inspection in operational settings, the present paper focuses specifically on the deterministic conversion of processed MFT-derived geometries into IFC4-compliant BIM outputs. Accordingly, validation in this study is limited to processed data generated under simulation-based and controlled specimen conditions.

Recent work on construction digital transformation

increasingly positions scan-to-BIM and digital twins as enabling layers for maintenance-focused decision support, where interoperability and traceable data exchange are recurring challenges [2–6]. Artificial intelligence (AI) techniques, including machine learning and deep learning, have been increasingly adopted to support automated analysis, prediction, and representation learning in complex engineering workflows [7–11]. Scan-to-BIM pipelines that convert reality-capture point clouds into semantic BIM have rapidly evolved from rule-based modelling to AI-assisted segmentation and automated Industry Foundation Classes (IFC) generation, and recent syntheses provide conceptual frameworks and best-practice guidance for reproducible adoption [12]. In parallel, the release of domain-specific datasets and benchmarks is beginning to standardise evaluation of scan-to-BIM methods and support generalisable learning-based approaches.

Beyond general reviews, open-source and benchmark-validated pipelines have been proposed to accelerate large-scale point-cloud-to-IFC conversion [13]. Domain-specific work has also demonstrated automated IFC generation for infrastructure assets such as railways by combining topology priors, parametric assemblies, and mesh-based reconstruction [14], and semi-automated IFC bridge modelling from point clouds has been explored for maintenance-oriented workflows [15]. Complementary efforts and related venues report deep learning-based Scan-to-BIM frameworks capable of reconstructing complex indoor elements from incomplete point clouds, underlining the continued push toward higher automation and robustness [16].

Muon imaging offers a complementary, passive modality for probing internal steel and density anomalies. It has demonstrated feasibility for reinforced concrete specimens and rebar-scale targets [17,18]. Yet, practical uptake in the built environment depends not only on reconstruction quality, but also on integrating outputs into established information-management ecosystems.

A key research gap therefore lies in the information integration layer: muon-tomography reconstructions are naturally produced as volumetric grids, point clouds or triangle meshes that are not readily expressible as parametric BIM objects and there is limited guidance on representing such internal, irregular geometries in openBIM formats while preserving metric integrity, provenance, and auditability. Although IFC4 supports tessellated geometry via IfcTessellatedFaceSet/IfcTriangulatedFaceSet for exchanging meshes in a standards-compliant manner [19], the use of these representations for NDT-derived internal reconstructions remains underexplored.

This paper addresses that gap by providing a deterministic Polygon File Format (PLY)-to-IFC4 conversion workflow for MFT outputs, designed to be reproducible and standards-compliant. The novelty is in (1) targeting subsurface muon-tomography reconstructions rather than surface scan-to-BIM, (2) encoding internal steel geometries and deviation maps using IFC tessellation to enable tool-agnostic exchange, and (3) emphasising auditability through explicit units, consistent spatial structure, and repeatable processing steps aligned with downstream inspection and digital-twin workflows. The main contributions are as follows: A deterministic, programmatic PLY-to-IFC4 conversion pipeline that supports both triangle meshes and point clouds; A robust preprocessing strategy, including surface reconstruction for point clouds, and mesh cleaning for stable IFC tessellation; Standards-based IFC export within a minimal IFC spatial structure, with explicit unit definition to prevent scaling ambiguity.

2 Background and Motivation

MFT uses naturally occurring muons in background radiation to compare incoming and outgoing particle distributions after they pass through a target. Variations in attenuation and scattering reveal internal geometry and material characteristics. Combined with AI-based reconstruction and classification, MFT enables subsurface 3D inspection for identifying embedded steel properties, cross-sections, and defects such as corrosion, voids, and cracking, as outlined in Table 1 against XRF. In the M-TABS project, this approach is applied to reinforced concrete bridge components, with the goal of producing 3D BIM outputs that fit existing inspection and asset-management workflows for operators and infrastructure owners. Figure 1 presents the GScan muon flux detector hardware used to collect the measurement data in this study. Muons are generated when cosmic rays interact with the atmosphere and, due to their high mass, can penetrate significant thicknesses of materials such as steel and concrete. In MFT, the process involves measuring the incoming particle flux, observing how particles are scattered or absorbed as they pass through an object, and measuring the exiting flux and scattering features such as $\Delta\theta$. These measurements are then used with AI to reconstruct internal structures and infer material composition, as different materials produce distinct attenuation and scattering signatures, as illustrated in Figure 2. A key technical challenge is converting tomography-native outputs into BIM-native representations. Tomography generates voxel-based or unstructured geometric data, while BIM represents structured, semantic objects within a spatial hierarchy. IFC, as the main interoperability standard, uses tessellated geometry in IFC4 to encode irregular surfaces as triangulated meshes, preserving geometric fidelity and compatibility with BIM tools, but with reduced parametric flexibility and larger file sizes.

Table 1 Comparison of XRF and muon tomography + AI across key operational criteria.

Criterion	XRF	Muon
Penetration depth	Limited	~10 m
Ability to penetrate shielded materials	Moderate	High
Chemical composition identification	Yes	Yes
Material agnostic / number of materials	High	Very high
Scanning time per m³	Rapid	Extended
Environmental insensitivity	Very low	High
System size / portability	Portable	1.75 × 0.75 m
Resolution / accuracy	Millimetre-scale	~1 mm



Figure 1. muon flux detector hardware.

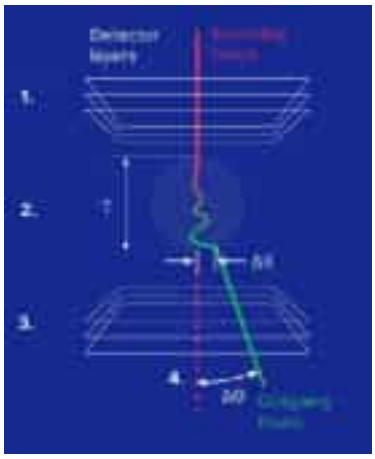


Figure 2. Muon tomography imaging principle.

3 Process Context and interdependencies

3.1 Data Acquisition

The first step establishes the foundational data assets required for method development and benchmarking across the project by creating a simulation-based dataset derived from an infrastructure asset. In addition to generating representative synthetic measurements, this step is explicitly responsible for delivering (1) a Universal Simulation Environment and (2) a Benchmark Data Library that provide consistent, controlled inputs for subsequent development and validation in the following steps. In practical terms, first step outputs must support two complementary needs: (a) physics-consistent scene and sensor configuration to enable reproducible simulation studies, and (b) ground-truth geometric or volumetric representations that can be used to benchmark reconstruction fidelity, segmentation accuracy, and downstream model conversion. The attached artefacts exemplify these two output classes: Simulation environment or scene definition, Benchmark volumetric ground truth. Overall, first step outputs operationalize the project's data-first dependency chain: they enable second step to develop and validate processing algorithms under controlled conditions, and they provide the reference geometries and volumetric benchmarks that ultimately support conversion of processed outputs into interoperable BIM deliverables. Figure 3 provides an example first step benchmark environment and associated ground-truth geometry.

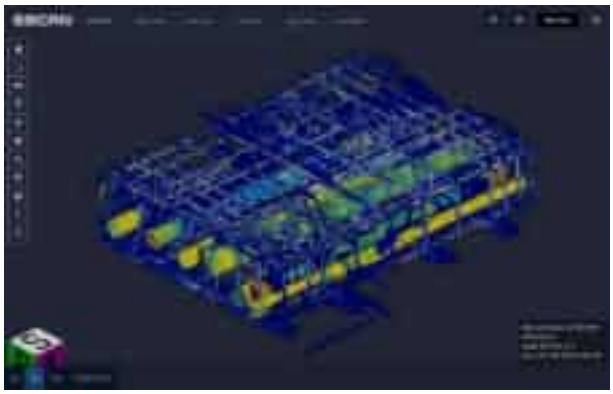
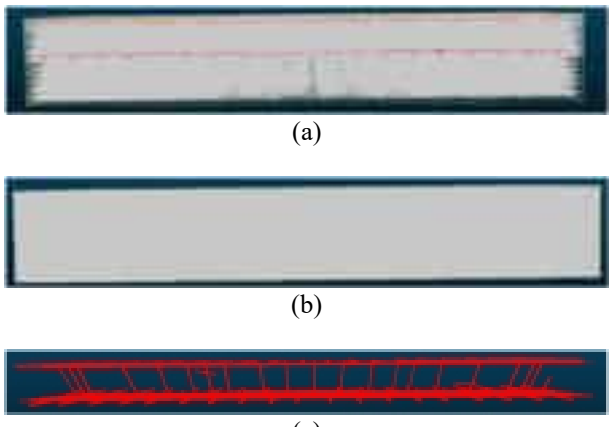


Figure 3. Simulation environment and benchmark volumetric ground truth.

3.2 Data Processing

The next step focuses on transforming muon-derived reconstructions into engineering-consumable, object-level outputs that can be directly ingested by downstream BIM integration workflows. In addition to improving the quality and usability of processed results, second step provides the intermediate representations required for

IFC generation: (1) classified material outputs, (2) instance-level segmentation of reinforcement, and (3) geometry and location descriptors suitable for object placement and semantic assignment. The outputs illustrate a concrete processing chain starting from a volumetric steel representation and culminating in object-level data products: Voxel-to-point conversion and material classification: A vowelized MFT volume is converted into a point-based (PLY) format and classified by material (e.g., concrete and rebar), producing the basic candidate point clouds required for further modelling and export. In the voxel-to-point conversion stage, each retained voxel was represented by a single point located at its centroid. This choice provides a compact object-level representation for downstream classification and export, but it also smooths boundary detail relative to explicit voxel-surface extraction. Consequently, the final mesh quality depends on the voxel resolution and on the subsequent surface reconstruction stage, which together govern the trade-off between data compactness and geometric fidelity. Representative outputs from this conversion and classification step are shown in Figure 4.



(a)

(b)

(c)

Figure 4. Voxel-to-point conversion and material classification outputs from MFT-derived volumes: (a) Point cloud; (b) Concrete classification; (c) Rebar classification.

Instance segmentation to isolate individual rebars: This stage evaluates multiple geometric instance segmentation methods to extract individual rebars from classified point clouds, a prerequisite for correct IFC export. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) leads to over-segmentation, cylinder fitting under-detects rebars, and region growing improves detection but still suffers from fragmentation and missed elements. Overall, these results show that purely geometric approaches struggle with connected, irregular, or partially observed rebars, limiting reliable one-object-per-element IFC generation. This limitation is important for the downstream IFC conversion stage. The geometry-to-IFC workflow assumes that each input PLY corresponds to a single, correctly isolated physical object. Therefore, when upstream segmentation produces merged rebars, fragmented instances, or incomplete separations, the downstream export does not resolve these errors; instead, it deterministically converts the provided geometry into a tessellated IFC element. In this sense, segmentation quality directly governs semantic reliability, object granularity, and the engineering interpretability of the resulting BIM model. Figure 5 illustrates representative qualitative results of instance segmentation.

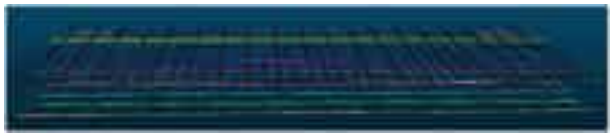


Figure 5. Instance segmentation of steel data.

Few-shot learning-based instance segmentation: This method is introduced to address the limitations of purely geometric approaches, trained on a small manually labelled rebar dataset. This approach improves isolation of individual rebars and more robustly handles connected and irregular shapes, enabling consistent object-level outputs suitable for BIM/IFC generation. A representative output from the few-shot approach is shown in Figure 6.

Although quantitative instance segmentation metrics such as Intersection over Union (IoU), Precision, Recall, and F1-score would strengthen the upstream evaluation, they are not reported in this study because no benchmark dataset with ground-truth instance labels was available for this stage of the MFT processing pipeline. The segmentation assessment is therefore presented qualitatively, with emphasis on the observed failure modes of geometric methods, including over-segmentation, fragmentation, and missed elements, and on the improved object isolation achieved by the few-shot learning approach. Establishing a benchmark dataset with instance-level ground truth remains an important next step for rigorous quantitative evaluation of segmentation performance. Geometry information extraction for downstream modelling: extracts per-instance geometric and location attributes into a structured tabular format, producing instance-level descriptors that support semantic classification, spatial placement, validation, and auditability in downstream BIM/IFC workflows.

Collectively, these outputs provide the two key input families required for IFC authoring: Object-level geometries, including one PLY per segmented rebar instance and meshes derived from those instances, and Instance metadata that support semantic assignment and consistent placement.

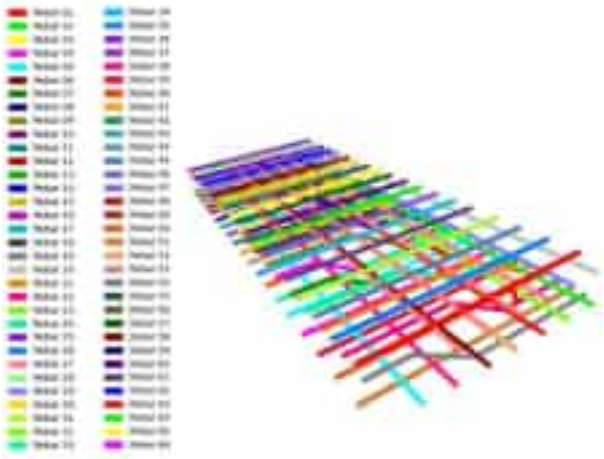


Figure 6. Few-shot learning-based instance segmentation result.

3.3 3D Model Generation and BIM Integration

The next step converts processed outputs into BIM-compatible 3D representations and ensures the resulting models adhere to industry standards for data exchange and integration. Its objectives include (1) data conversion to BIM-compatible formats, (2) automation of 3D model generation from processing outputs, (3) integration into BIM software for accurate 3D representations, and (4) standards alignment for model generation and exchange. The deliverable of this step is an updated front-end for GScan's MFT, which consumes outputs from the previous steps and supports downstream pilot implementation and evaluation in later steps. This paper addresses a specific and critical third step capability: open-standard IFC4 generation from processed, object-level geometries (exported as PLY). The presented methodology implements a deterministic, programmatic geometry-to-IFC tessellation workflow using IFC4 constructs to maximize interoperability.

The third step consumes the output of the second step in the form of per-object geometry files, PLY point clouds and reconstructed meshes for individual reinforcement instances, along with associated instance-level geometry and location descriptors**.** These artefacts enable a direct, automated export pathway: each segmented instance is converted into a discrete IFC element, represented using IFC4 tessellated geometry (IfcTriangulatedFaceSet) and inserted into a minimal IFC spatial hierarchy with explicitly units (metres).

4 Methodology: IFC generation method

The IFC model was generated through a deterministic, programmatic geometry-to-IFC tessellation workflow. Processed MFT-derived geometries were converted into BIM-compatible objects by constructing a minimal IFC spatial hierarchy and embedding each object's surface representation as an IFC4-compliant tessellated shape. The implementation used Open3D for geometry handling and IfcOpenShell for IFC authoring.

Each input object was provided as an individual PLY file that may contain either (1) a triangle mesh (faces + vertices) or (2) a point cloud (vertices only). For each PLY:

Mesh detection: The file was first read as a triangle mesh. If valid faces were present, the mesh was used directly.

Surface reconstruction (point clouds only): When no faces were present, the point cloud was converted into a triangle mesh using Ball Pivoting Surface Reconstruction. Surface normals were estimated prior to reconstruction. For a point cloud with N points, the average nearest-neighbor spacing was computed as:

$$\bar{d}_{nn} = \frac{1}{N}\sum_{i=1}^{N} d_i^{nn} \quad (1)$$

where d_i^{nn} is the Euclidean distance from point i to its nearest neighboring point. The Ball Pivoting radii were then defined as:

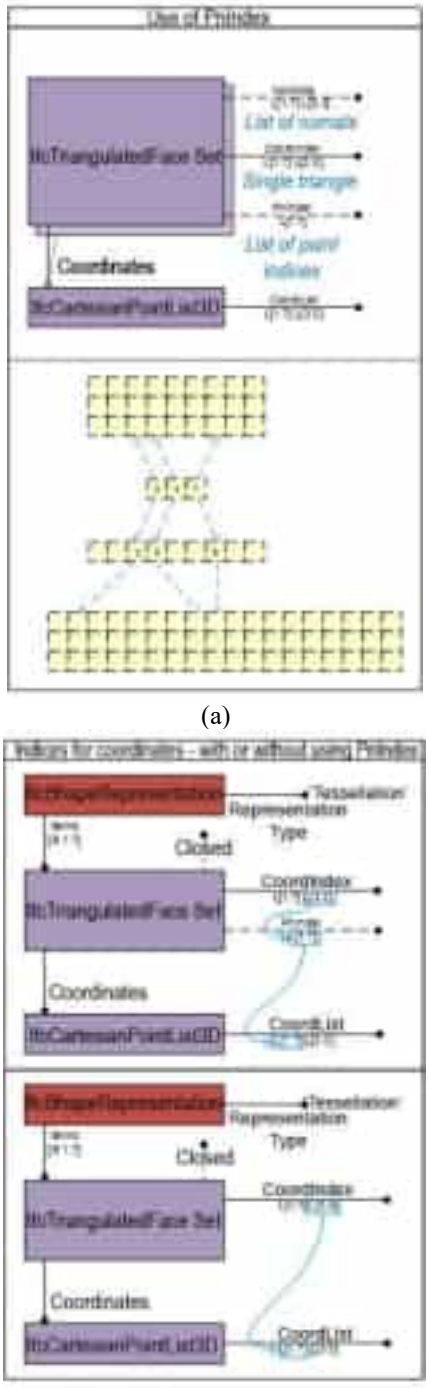
$$R = \{k_1\bar{d}_{nn}, k_2\bar{d}_{nn}, k_3\bar{d}_{nn}\} \quad (2)$$

where k_1, k_2, and k_3 are fixed scale multipliers used consistently across the dataset. In our implementation, $k_1 = [1.5]$, $k_2 = [2.5]$, and $k_3 = [3.5]$. This radius-selection strategy preserves geometric scale consistency across point clouds with different sampling densities.

Mesh cleaning: The resulting mesh (either imported or reconstructed) was cleaned using standard topological repairs, including removal of duplicated and degenerate triangles, non-manifold edge mitigation, and recomputation of vertex normal, to ensure stable downstream tessellation.

A valid IFC4 container structure was created to ensure compatibility with standard BIM tools. The following spatial entities were instantiated: IfcProject (root), IfcSite, IfcBuilding, and IfcBuildingStorey. Length units were explicitly assigned as meters in "UnitsInContext" to prevent viewer-dependent scaling issues. This structure provides the minimum required context so that imported elements appear correctly in IFC viewers and BIM authoring platforms.

Each mesh was encoded as IFC tessellated geometry using IfcTriangulatedFaceSet, which is the IFC4 entity intended for representing arbitrary triangular surfaces: Mesh vertices were written to IfcCartesianPointList3D (CoordList), preserving the original coordinate system after any required scaling or normalisation. Mesh faces were written to IfcTriangulatedFaceSet (CoordIndex), as illustrated in Figure 7 [19]. Because IFC indexing i**s** 1-based**,** face indices were shifted accordingly during export.



(a)

(b)

Figure 7. (a) IfcCartesianPointList3D; (b) IfcTriangulatedFaceSet.

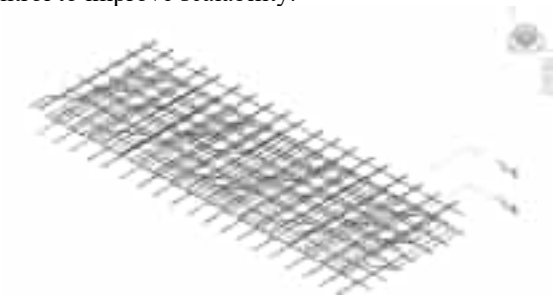
The tessellated geometry was placed into a standard shape representation: (1) IfcShapeRepresentation with: RepresentationIdentifier = "Body", RepresentationType = "Tessellation", Items = [IfcTriangulatedFaceSet]; (2) Wrapped in IfcProductDefinitionShape and assigned to the element's Representation. This approach produces broad interoperability across IFC viewers. For each exported object, an IFC product instance was created and inserted into the spatial structure: (1) Element class assignment: Objects were exported either as IfcReinforcingBar (when the object corresponded to rebar) or as a generic IfcBuildingElementProxy when semantic classification was not available or not required; (2) Placement: Each element was assigned an IfcLocalPlacement (default orientation and origin placement, or a project-defined local reference frame if applied); (3) Containment: Elements were associated with a bridge-relevant spatial container, including IfcBridge and IfcFacilityPart, via IfcRelContainedInSpatialStructure, ensuring correct visibility, hierarchy, and navigation within BIM and digital engineering environments.

5 Results and Discussion

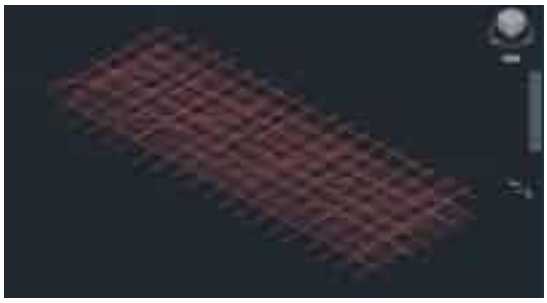
The exported IFC contains a valid spatial hierarchy, one IFC product per input PLY object, a tessellated body representation for each object, and consistent placement, ensuring that the objects appear as expected in standard BIM viewers. The symmetric Chamfer distance between each input geometry and the tessellated geometry re-extracted from the exported IFC was used to assess geometric fidelity. The average Chamfer mean was 0.61 mm, and the average Chamfer p95 was 2.53 mm. When the input was already a mesh, the export was effectively lossless, with a mean Chamfer distance of 0.02 mm and a p95 of 0.09 mm, confirming high-fidelity tessellation-based IFC encoding. For point-cloud inputs requiring Ball Pivoting reconstruction, the average errors increased to a mean of 1.20 mm and a p95 of 4.97 mm, indicating that the dominant source of error arises from surface reconstruction rather than from the IFC encoding itself. In terms of efficiency, end-to-end processing averaged 2.60 s per element. The exported IFC size was 27,774 KB, indicating that the conversion achieves interoperability with only a modest increase in file size. For rebar-representative objects, IfcReinforcingBar provides a semantically meaningful class assignment; otherwise IfcBuildingElementProxy ensures a safe fallback without overstating semantics. While this export strategy provides a robust and conservative pathway for geometry-to-IFC conversion, the present implementation offers only limited semantic enrichment. The current BIM output primarily encodes geometry, placement, and basic class assignment, and does not yet attach richer property information such as reconstruction confidence, uncertainty bounds, estimated bar diameter, or provenance metadata from the upstream MFT processing stages. As a result, the exported model is currently more suitable for visualization, exchange, coordination, and

auditability than for advanced engineering querying or asset-management analytics. Future work should therefore extend the exported entities with infrastructure-relevant property sets and provenance-aware metadata once such attributes can be derived and validated consistently. Figure 8 shows a representative exported model visualised in Autodesk Revit and AutoCAD. The reported results therefore demonstrate the feasibility of standards-compliant geometry-to-IFC conversion under controlled inputs, but they should not yet be interpreted as full evidence of robustness under active bridge operating conditions.

A practical challenge in processing muon-derived outputs is variability in representation: some objects may be exported as watertight meshes, while others may be sparse point clouds. The two-path approach (direct mesh handling vs deterministic surface reconstruction) addresses this variability while maintaining a single IFC export strategy. Mesh cleaning is critical: repairing degenerate triangles and mitigating non-manifold conditions substantially reduces the likelihood of unstable or invalid tessellation in downstream BIM tools. The use of IfcTriangulatedFaceSet was chosen for robustness and compatibility with irregular MFT-derived geometries, but it may be inefficient for large infrastructure models because high-resolution meshes can increase file size and reduce viewer performance. The paper should therefore state that this choice prioritizes reliable geometry preservation over efficiency, while noting that future work will explore parametric rebar modelling, mesh decimation, and level-of-detail control to improve scalability.



(a)



(b)

Figure 8. Representative exported IFC visualized in: (a) Revit; (b) AutoCAD.

6 Conclusion

This paper presented an IFC generation method that converts processed MFT-derived geometries, including PLY meshes and point clouds, into IFC4-compliant BIM deliverables through a deterministic geometry-to-IFC tessellation workflow. Using Open3D for geometry handling and IfcOpenShell for IFC authoring, the approach constructs an IFC spatial hierarchy, explicitly defines meter units, and exports each object as a standard Body/Tessellation representation using IfcTriangulatedFaceSet. The result is a robust, interoperable pathway for integrating muon-derived 3D information into open BIM workflows, supporting objectives related to data conversion, automation, integration, and standards compliance.

An important next step is validation using data acquired from an actively operating bridge, where operational noise, environmental variability, attenuation uncertainty, and site constraints may affect the reliability of the broader pipeline. Future work can extend this foundation in four directions: (1) Richer semantic mapping: expanding beyond IfcReinforcingBar and proxy entities to additional infrastructure-relevant classes and property sets, including provenance metadata and confidence descriptors where available from upstream processing; (2) Parametric reconstruction when feasible: for objects approximating canonical forms, such as cylindrical reinforcement, exporting swept solids or other parametric representations alongside tessellated geometry to improve editability and downstream analytics; (3) Scalability optimization: introducing mesh decimation strategies and level-of-detail management to control file size and improve import performance for large-scale scans; (4) Infrastructure-specific spatial context: where project requirements demand, align the spatial structure with bridge/facility conventions and integrate with wider asset information containers.

Acknowledgement

The financial support of Innovate UK under the project “Muon Tomography for Assessing Bridge Structures” (Project Reference: 10132366) and the D-HYDROFLEX project under the European Union’s Horizon Program (Grant Agreement No. 101122357) is gratefully acknowledged.

References

[1] S. Kookalani, E. Parn, I. Brilakis, S. Dirar, M. Theofanous, A. Faramarzi, M.A. Mahdavipour, Q. Feng, Trajectory of building and structural design automation from generative design towards the integration of deep generative models and

optimization: A review, Journal of Building Engineering 97 (2024) 110972. https://doi.org/10.1016/J.JOBE.2024.110972.

[2] M. Abdelrahman, E. Macatulad, B. Lei, M. Quintana, C. Miller, F. Biljecki, What is a Digital Twin anyway? Deriving the definition for the built environment from over 15,000 scientific publications, Build. Environ. 274 (2025) 112748. https://doi.org/10.1016/J.BUILDENV.2025.112748.

[3] W. Hu, Y. Ou, H. Liu, P. Ni, C. Chang, Integrating digital twin technologies for maintenance 4.0 in the building industry: A review and conceptual framework, Build. Environ. 288 (2026) 113997. https://doi.org/10.1016/J.BUILDENV.2025.113997.

[4] H. Alavi, S. Kookalani, F. Rahimian, N. Forcada, Integrated Building Intelligence, (2024). https://doi.org/10.1007/978-3-031-68865-2.

[5] S. Kookalani, S. Green, P. Luo, H. Alavi, E. Parn, Z. Sun, I. Brilakis, Mapping digital twin applications in infrastructure and the built environment across research types, methods, sectors, phases, and scales, Autom. Constr. 182 (2026) 106778. https://doi.org/10.1016/J.AUTCON.2026.106778.

[6] P. Luo, E. Parn, S. Kookalani, I. Brilakis, TailorAlert: Large Language Model-Based Personalized Alert Generation System for Road Infrastructure Management with Digital Twins, Computing in Civil Engineering 40 (2025). https://doi.org/https://doi.org/10.1061/JCCEE5.CPENG-7025.

[7] S. Kookalani, B. Cheng, S. Xiang, Shape optimization of GFRP elastic gridshells by the weighted Lagrange ε-twin support vector machine and multi-objective particle swarm optimization algorithm considering structural weight, Structures 33 (2021) 2066–2084. https://doi.org/10.1016/j.istruc.2021.05.077.

[8] S. Xiang, B. Cheng, S. Kookalani, An analytic solution for form finding of GFRP elastic gridshells during lifting construction, Compos. Struct. 244 (2020). https://doi.org/10.1016/j.compstruct.2020.112290.

[9] S. Kookalani, S. Nyunn, S. Xiang, Form-finding of lifting self-forming GFRP elastic gridshells based on machine learning interpretability methods, Structural Engineering and Mechanics 84 (2022) 605–618. https://doi.org/10.12989/sem.2022.84.5.605.

[10] H. Liu, H. Su, S. Kookalani, M. Cao, Y. Lei, W. Wu, Attention-enhanced spatiotemporal deep learning-based automatic warning model with uncertainty estimation in dam safety, Struct. Health Monit. (2025). https://doi.org/10.1177/14759217251358537.

[11] S. Kookalani, B. Cheng, J.L.C. Torres, Structural performance assessment of GFRP elastic gridshells by machine learning interpretability methods, Frontiers of Structural and Civil Engineering 16 (2022) 1249–1266. https://doi.org/10.1007/s11709-022-0858-5.

[12] E. Noroozinejad, A. Hajirasouli, G.M. Morrison, M. Kagioglou, J. Patil, M. Kalantari, Automatic Scan-to-BIM—The Impact of Semantic Segmentation Accuracy, Buildings 2025, Vol. 15, Page 1126 15 (2025) 1126. https://doi.org/10.3390/BUILDINGS15071126.

[13] S. Zbirovský, V. Nežerka, Open-source automatic pipeline for efficient conversion of large-scale point clouds to IFC format, Autom. Constr. 177 (2025) 106303. https://doi.org/10.1016/J.AUTCON.2025.106303.

[14] M.R.M.F. Ariyachandra, I. Brilakis, Leveraging railway topology to automatically generate track geometric information models from airborne LiDAR data, Autom. Constr. 155 (2023) 105068. https://doi.org/10.1016/J.AUTCON.2023.105068.

[15] Y. Schatz, B. Domer, Semi-automated creation of IFC bridge models from point clouds for maintenance applications, Front. Built Environ. 10 (2024) 1375873. https://doi.org/10.3389/FBUIL.2024.1375873/BIBTEX.

[16] M. Mahmoud, Z. Zhao, W. Chen, M. Adham, Y. Li, Automated Scan-to-BIM: A deep learning-based framework for indoor environments with complex furniture elements, Journal of Building Engineering 106 (2025) 112596. https://doi.org/10.1016/J.JOBE.2025.112596.

[17] E. Niederleithinger, S. Gardner, T. Kind, R. Kaiser, M. Grunwald, G. Yang, B. Redmer, A. Waske, F. Mielentz, U. Effner, C. Köpp, A. Clarkson, F. Thomson, M. Ryan, D. Mahon, Muon Tomography of the Interior of a Reinforced Concrete Block: First Experimental Proof of Concept, Journal of Nondestructive Evaluation 2021 40:3 40 (2021) 65-. https://doi.org/10.1007/S10921-021-00797-3.

[18] K. Doi, S. Hiromoto, T. Shinohara, al -, L. Chen, X. Liu, Y. Lin, S. Hiromoto -, M. Dobrowolska, J. Velthuis, A. Kopp, M. Perry, P. Pearson, Towards an application of muon scattering tomography as a technique for detecting rebars in concrete, Smart Mater. Struct. 29 (2020) 055015. https://doi.org/10.1088/1361-665X/AB7A3F.

[19] 8.8.3.47 IfcTriangulatedFaceSet - IFC4.3.2.0 Documentation, (2024). https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/HTML/lexical/IfcTriangulatedFaceSet.htm#Figure-8.8.3.47.B-Triangulated-face-set.

Real-time Animal Intervention Detection in Roadways Using Quantum Enhanced YOLOv11 framework for CCTV Images and Inclement Visibility Conditions

Prasanna Venkatesan Ramani, Sugeerthi M S and Naresh Kumar V

School of Civil Engineering, Vellore Institute of Technology, India

prasanna.venkatesan@vit.ac.in, sugeerthi.2023@vitstudent.ac.in, nareshkumar.2023@vitstudent.ac.in

Abstract -

Stray animals colliding with moving vehicles is one of the most serious road safety hazards that exists today. Many of these collisions occur at roads that run directly next to important urban spaces like institutional corridors, where stray animals often invade the highways. Since the use of physical barriers and sophisticated sensor technologies can be prohibitively expensive and logistically impractical, this paper proposes a simple yet scalable computer-vision-based approach for detecting stray animals, such as dogs and cows, crossing the roads using existing CCTV cameras installed along highways as part of the regular road surveillance systems. A database of animal intervention images was extracted from the CCTV video recordings along a state highway that runs adjacent to the University campus. Each image contained either a cow or a dog that had invaded the travel lane of the roadway. By using a lightweight YOLOv11 nano object detection model to detect animals within the visual field of the camera, each of the images was then passed through a hybrid classifier that utilised both CNNs and QNNs to refine the object detection and improve the accuracy of the detection under low-light and other challenging operating conditions. The model accomplished a precision of 86.7%, a recall of 63.4% and mAP50 of 79.6%. Through experimentation, it has been demonstrated that the proposed methodology will converge to a stable solution and provide increased robustness to variations in lighting and other environmental factors.

Keywords -

Animal Detection, Road Safety, Computer Vision, CCTV, Hybrid Quantum Computing, CNN, Smart Traffic Control.

1 Introduction

In the context of Indian roads, vehicle collisions with stray animals like dogs and cattle cause road accidents that are almost persistent and underreported [1].

Recent studies have shown that collisions between animals and vehicles are a major cause of highway crashes and other vehicle accidents, especially in areas where there is a large population of loose roaming (free-roaming) animals. These incidents have several repercussions, including loss of life of humans and animals, economic losses due to property damage and vehicle repair, traffic congestion and disruption of traffic flow, damaged roadway infrastructure, and increased risk for drivers, pedestrians and local residents, particularly on highways and rural roads [2].

Field observations obtained from National and State highways show that dogs and cows are frequently seen crossing the roads, which leads to single or multiple vehicle collisions. This situation requires smart technological solutions, such as automated detection of animal intervention and alert mechanisms to ensure a safe and comfortable environment for road users. The current technological framework of the Indian traffic management system primarily comprises surveillance CCTV cameras that are installed, inspected, monitored, controlled and documented by the highway administration authorities. The fundamental objective of this study is to reframe animal detection as a civil infrastructure resilience problem and propose a scalable solution that takes advantage of the existing edge technologies, instead of deploying any new sophisticated technologies that require additional expensive investments. This study utilised the real-world video footage obtained from the existing CCTV cameras installed in and around the highway adjacent to the university campus. From this, about 453 frames corresponding to different times of the day were generated and labelled as the baseline dataset. This hybrid YOLOv11-CNN-QNN architecture is proposed for the real-time detection of animals using real-world datasets collected using traffic surveillance CCTV camera images under low-light conditions and occlusions, focusing on the detection of animal-related disruptions to road traffic rather than the classification of animals.

2 Background and Literature Review

Animal–Vehicle Collision (AVC) is an omnipresent and rising global concern regarding road safety, which has a multitude of effects like loss of life, severe injury, loss of revenue and loss of biodiversity. Both in the developed world and the developing world, there is a reasonable body of empirical evidence showing that AVCs constitute a substantial proportion of all road accidents, notably on highways and country roads where speeds are higher, and animal crossings are less predictable [3, 4, 5]. Moreover, AVCs occur at a greater frequency during night, and low-visibility conditions when the driver's perception and response time are significantly impaired, and when standard surveillance/warning systems are ineffective [2].

In developing countries, AVCs are further complicated by issues like the presence of feral/stray animals, lack of compliance with laws governing animal control, and inadequate infrastructure to mitigate wildlife conflicts [4]. Literature evidence specifically shows that common domestic animals like cattle and dogs cause the majority of the severe AVCs that occur close to human habitation and/or near agricultural land adjacent to highways [3, 5]. Although the magnitude of AVCs is being recognized globally, the mitigation efforts still remain piecemeal, and many of the current AVC solutions focus on isolated areas rather than providing a scalable system-wide solution.

2.1 Conventional AVC Mitigation Methods

Conventional AVC mitigation methods utilize physical/infrastructure-based solutions, i.e., wildlife fencing, underpasses, overpasses, warning signs and lighting along the roadway [3, 6]. However, these types of AVC mitigation methods are expensive to install, require significant planning and need a commitment to maintenance, limiting their feasibility for large-scale deployment throughout vast highway networks in low- and middle-income countries. Furthermore, these static AVC mitigation methods are unable to accommodate the dynamic nature of animal movement patterns, seasonal migrations and diurnal behaviors [1]. Therefore, AVC mitigation efforts will require dynamic and responsive solutions to supplement conventional infrastructure-based AVC mitigation methods.

2.2 Sensor-Based and Hardware-Dependent AVC Detection Systems

Some studies have proposed AVC detection using sensor-based systems incorporating infrared, thermal imaging, ultrasonic sensors, RFID and GPS-based tracking [7]. Of these, thermal imaging has been of particular interest due to its capability of detecting warm-bodied animals in low light and nighttime conditions. However, while thermal imaging systems can be effective in controlled environments, they have several disadvantages, including limited spatial resolution, inability to differentiate among species, sensitivity to weather conditions and high cost [8, 2]. Sensor-based AVC detection systems are also subjected to significant challenges in terms of installing, calibrating and maintaining hardware, especially along lengthy sections of highways [7]. This has prompted researchers to explore software-centric, vision-based AVC detection methods utilizing existing infrastructure.

2.3 Computer Vision and Deep Learning for AVC Detection

Recent advances in computer vision and deep learning have allowed for substantial improvements in real-time object detection, and several studies have demonstrated the effectiveness of YOLO-based architectures for AVC detection, particularly when coupled with image enhancements to increase visibility in low-light conditions [8, 2]. In addition, vision-based safety monitoring approaches using existing cameras have shown great potential for detecting hazards in infrastructure systems[9]. However, deep learning-based AVC detection systems have several challenges. First, AVC detection performance is heavily reliant upon the quality and diversity of the training dataset used to train the detector, which is often limited in terms of AVC detection in regionalized contexts [8, 10]. Second, most of the studies focus on identifying specific species, limiting the model's ability to generalize to unseen/rare species [10]. From a road safety perspective, identifying the species may not be necessary for AVC mitigation since the primary goal is timely hazard detection, regardless of whether or not the species is identified. Recent studies have shown the efficiency of deep learning-based vision systems in object detection in unstructured and real-world environments, addressing problems such as occlusions [11]. Finally, deployment of AVC detection systems in realistic environments, including low-quality CCTV footage, variable lighting, occlusion, and environmental noise, impairs the accuracy of AVC detection relative to laboratory testing [2]. These issues illustrate the disparity between AVC detection system performance in experiments versus AVC detection system reliability in operational use.

2.4 Road Safety Analysis and Computer Vision Infrastructure

More recently, the field of computer vision has extended beyond the application of detecting animals using video feeds (e.g., AVC) to include the application of computer vision in larger roadway safety assessments (e.g., roadside feature extraction, crash risk prediction) [12]. The majority of this research has focused on static, non-transient hazards (e.g., infrastructure hazards), whereas animal intrusion represents a high-priority, time-sensitive hazard that requires immediate detection and response [12]. Therefore, integrating the capability to detect animals into intelligent transportation systems presents a current research gap [12].

2.5 Research Gap and Motivation

Sensor-based solutions rely on specific hardware and are difficult to deploy in large scale [7]. Although deep learning-based solutions hold a good promise,

many struggle to overcome the issues of data availability, poor lighting, and lack of generalization in real-world settings [2, 10]. Additionally, over-emphasizing the classification of specific species limits the robustness of such solutions in multi-species traffic environments [4]. Therefore, these aspects provide the motivation to develop an intelligent, vision-based animal intervention framework that utilizes existing roadside CCTV infrastructure, focuses on generic animal presence detection, and performs well under real- world constraints.

3 Methodology

Figure 1 explains the flow of work from data collection to frame extraction as the basic input data stage. Post, the data was split into three categories as train, validate and test with the proportion of 80:10:10, respectively and was augmented for better performance. Annotation was succeeded by a hybrid CNN and QNN reinforcement for animal detection and classification to outperform in inclement lighting conditions with the available dataset.

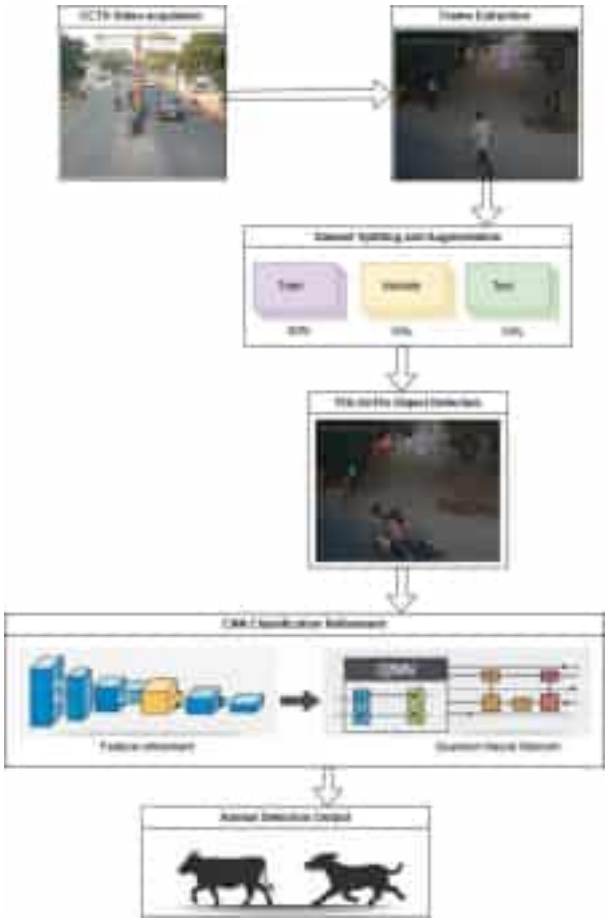


Figure 1. Methodology

3.1 Study Location and Data Collection

Data collection was based upon visual evidence collected from a stretch of state highway with documented recurrent animal crossings of the highway; this location is also proximate to the University campus, located near Vellore, Tamil Nadu, India, increasing the likelihood of animal-vehicle collisions due to increased human settlement influences. Video from existing roadside CCTV cameras was used as the sole source of data for this research to support compatibility with existing infrastructure. Video recording occurred at different times of the day, at night, during dawn, dusk and at low light levels, to provide a diverse range of lighting conditions and traffic scenarios, and the time of collection of data is roughly between 05:00 PM and 7:00 AM for about 5 consecutive days. Based on the CCTV's capabilities, the usual range is between 30 and 50m. To cover a distance of 500m, we would require 12-17 cameras, depending on the overlapping strategy. This model of monitoring systems does exist in real time, and the dataset collected is from the existing infrastructure system established on highways at Katpadi - Vellore.



Figure 2. Data Collection Dog Frame



Figure 3. Data Collection Cow Frame

3.2 Dataset Creation and Labelling

A total of 453 relevant and clear image frames were created from the collected CCTV footage and manually

validated to confirm their relevance and clarity. These images depicted a variety of animals crossing or occupying the highway, including stray dogs and cows, which represent two of the most common types of animals. Labelling of objects was completed using the RoboFlow. To minimise label noise and ensure accurate positioning of bounding boxes, all images were labelled using a standard protocol. In addition, all of the images with bounding boxes were reviewed manually to ensure that the images remained of high quality.

To increase the robustness of the model developed for this research, data augmentation techniques were applied to the training subset. Data augmentation techniques included horizontally flipping the images, scaling, and altering brightness to increase the ability of the model to perform well in various lighting conditions, and to simulate a variety of viewpoints while preventing distortion to the images' realistic characteristics.



Figure 4. Annotated Images

3.3 Object Detection Utilizing YOLOv11(Nano)

YOLOv11(nano), a lightweight version of the YOLOv11 architecture, was selected for the detection of animals because of its ability to provide real-time detection and be deployed on edge devices for the roadside surveillance system. The lightweight version provides a desirable trade-off between the accuracy of detection and the computational requirements of the model, making it suitable for deployment on roadside surveillance systems. The dataset annotated to detect animals in the CCTV frames was used to train the YOLOv11 (nano) model to detect animals in the frames. Since the main objective of the study is rooted in road safety, the focus was on developing a reliable method to detect animal presence in the images, rather than determining the specific type of animal. The proposed YOLO-based detection framework



Figure 5. Annotated Images

with the addition of a post-detection refinement step using CNN and QNN modules improves the reliability of the detection framework by reducing the rate of false positives, especially in low-light conditions.

3.4 CNN-Based Feature Enhancement

In order to improve the reliability of detection, especially in difficult visual environments, a Convolutional Neural Network (CNN) was implemented as a secondary classification module. The cropped region from the YOLO bounding box was input into the CNN for enhanced feature extraction and classification of the cropped region. The CNN architecture was designed to be simple so as to prevent overfitting, considering the small size of the dataset. The architecture of the CNN included stacked convolutional layers to learn spatial features, followed by pooling and fully connected layers for classification. This layer was intended to enhance discrimination of true animal detections versus background artefact detections, and therefore enhance overall precision and recall.

3.5 Quantum Neural Network (QNN)-Assisted Classification

Further improvements to classification performance under data-limited conditions were achieved through the implementation of a Quantum Neural Network (QNN) as an auxiliary learning model. The QNN was implemented as a hybrid quantum-classical model, whereby classical features derived from detected image regions were encoded into a quantum feature space. The motivation for utilizing a QNN is that it can potentially capture the complex relationships of the features and thus improve generalization when the training data is limited. The QNN was trained and tested using simulated quantum

circuits in the Google Colab environment, to ensure reproducibility and independence of physical quantum hardware. Unlike previous studies that replaced classical deep learning models, the QNN was used as a performance-enhancing module to refine classification confidence in conjunction with the CNN. Methodology performance was evaluated on the basis of detection reliability, not species-level detail, since the purpose of the method is to provide an early warning of potential animal hazards to vehicles on roads.

4 Results and Discussion

4.1 Training and Validation Convergence Analysis

The Training and Validation Loss Curves for Box Loss, Classification Loss, and Distribution Focal Loss are shown in Figure 6. They depict a consistent decrease in training loss throughout the epochs. The continuous decline in the box loss clearly illustrates that the YOLOv11n Detector learned and adapted effectively to the task of detecting cows and dogs on the roadways. Validation loss followed the trend in the training loss, showing very little evidence of overfitting, even with a very small number of frames used to train the model. The smooth decline in classification loss shows that the two-Class problem is well defined for the Architecture Chosen; some oscillation was expected due to the size of the dataset (453 Annotated Frames), real-world variability of CCTV imagery, such as occlusion, illumination changes, and motion blur. Overall, the trends illustrated in the graphs confirm that the Baseline Detector achieved reliable localization capabilities and served as a reasonable starting point for subsequent CNN and QNN-based refinements.

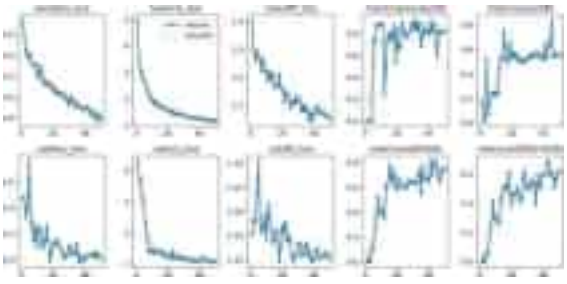


Figure 6. Training and validation Loss Curves

4.2 Precision – Recall Characteristics

The Precision–Recall Curve from Figure 7 provides insights into detection behavior between classes. Cow class exhibits extremely High Recall (99.5%); Cow detection is extremely important for road safety because missing a large animal like a cow represents a significant risk of collision. While the cow class has a very high Recall, it has Lower Precision (65.2%).

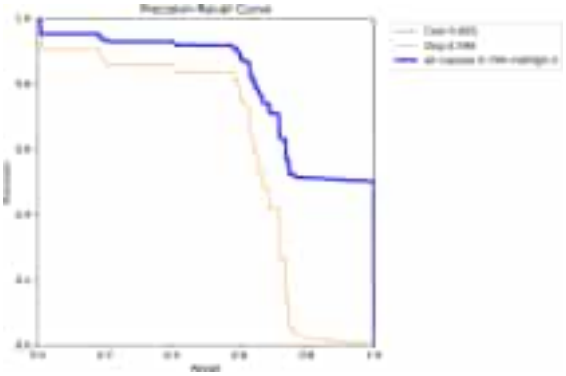


Figure 7. Precision - Recall Curve

This indicates there are many false positives that were detected due to objects in the background that have similar shapes or colors under low-light conditions. The dog class has Higher Precision (73.4%) but Lower Re-call (61.6%) than the Cow Class; Dog Class Detection is more difficult than Cow Class detection because dogs appear smaller and move faster than cows, and therefore have less distinctive visual features. In Addition, dogs often appear Partially Occluded or further away from the camera than cows, which results in Reduced Recall. Although dogs have a lower recall, the dog class has a higher Mean Average Precision (mAP_{50}) (59.6%) than the cow class (45.1%), indicating better localization quality when detections are made. These results are presented in Table1. Combined-Class Precision – Recall Curve shows the model performs effectively when both classes are considered together ((mAP_{50}) = 0.796). This supports the idea of using a Generic Animal Intervention approach, where priority is given to detecting animal presence rather than perfect class separation.

Table 1. Class Separation

Class	Precision	Recall	mAP_{50}
Cow	65.2%	99.5%	45.1%
Dog	73.4%	61.6%	59.6%

4.3 Confidence-Based Performance Analysis

The Recall – Confidence Curve from Figure 8 illustrates that the recalls decrease steadily as confidence thresholds increase. This behavior is consistent with what would be expected of a detector. At lower confidence thresholds, recalls remain high, especially for the cow class, confirming that the model is sensitive to animal presence. This type of behavior is advantageous for Early Warning Systems, where the detection of animals, even at lower

levels of confidence may provide enough time to take action before a potential collision occurs.

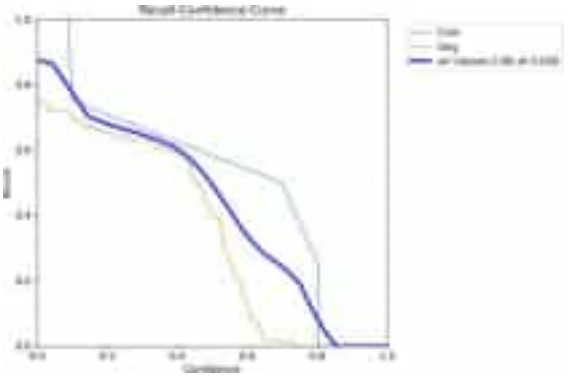


Figure 8. Recall - Confidence Curve

Figure 9 explains the Precision–Confidence Curve that shows a strong increase in precision as the confidence threshold increases, reaching near perfection for combined classes at higher confidence thresholds. This indicates that false positives are progressively filtered out as strict confidence criteria are applied. For system designers deploying the system, this allows them to tune confidence thresholds based on Operational Requirements – favoring either higher recall for safety-critical alerts or higher precision to reduce unnecessary warnings. Figure 9 represents the Precision – Confidence Curve showing a strong increase in precision as the confidence threshold increases, reaching near perfection for combined classes at higher confidence thresholds. This indicates that false positives are progressively filtered out as strict confidence criteria are applied. For system designers deploying the system, this allows them to tune confidence thresholds based on Operational Requirements – favoring either higher recall for safety-critical alerts or higher precision to reduce unnecessary warnings.

4.4 F_1-Confidence Trade-Off and Optimal Operating Point

Figure 10 shows the F1-Confidence curve, which exhibits a clear Optimal Operating Region, where the maximum combined F1-Score of approximately 0.73 occurred at a confidence threshold around 0.33. This operating point represents a balance between Precision and Recall, making it suitable for real-time road safety applications. Importantly, the cow class maintains a higher F1-Score across a wide range of confidence thresholds than the dog class, reflecting more consistent detection performance. The dog class showed a sharper decline at higher confidence- thresholds, again

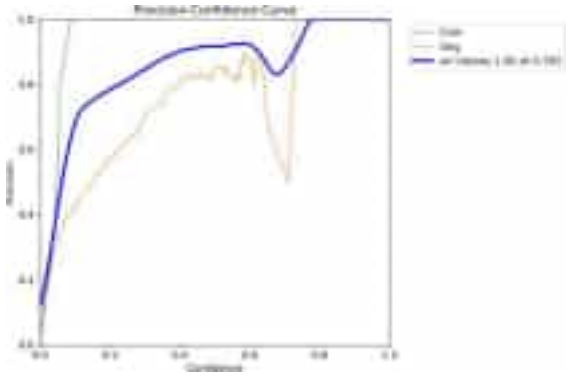


Figure 9. Precision - Confidence Curve

highlighting the difficulty of detecting small animals in unconstrained highway environments.

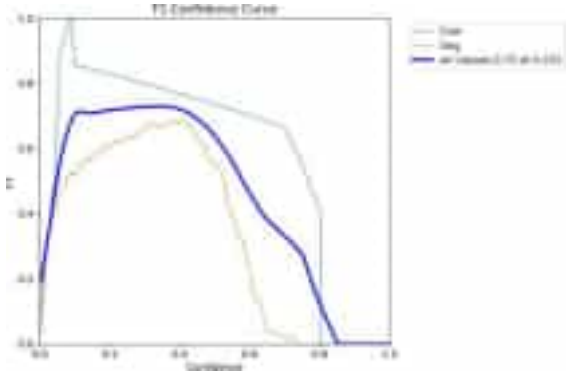


Figure 10. F_1- Confidence Curve

4.5 Impact of CNN and QNN-Based Refining

CNN and QNN Modules played a key role in stabilizing classification confidence and improving overall detection reliability under data-limited conditions. The CNN-enhanced spatial feature refinement for regions of interest identified by the detector resulted in fewer misclassifications caused by background clutter. The QNN contributed to the improvement of generalization by capturing nonlinear relationships between features, leading to smoother confidence distributions. This hybrid enhancement is reflected in the improved Precision-Confidence and F1-Confidence behaviors, especially for ambiguous cases and Low-Light Scenarios. The CNN-QNN Combination does not replace the YOLO Detector, but instead acts as a complementary refinement layer, strengthening the decision confidence without adding excessive computational overhead.

The ablation results shown in Table 2 indicate that the YOLOv11n+CNN+QNN configuration shows better performance in general, as it attains the highest mAP_{50} metric at 79.6% and precision at 86.7%. This shows better robustness and feature representation ability. Although there is a moderate decline in recall, it provides a good trade-off and could be a promising approach.

Table 2. Ablation Study

Configuration	Precision	Recall	mAP_{50}
YOLOv11n	0.713	0.7668	0.7651
YOLOv11n+CNN	0.8653	0.7951	0.7651
YOLOv11n+QNN	0.7401	0.7853	0.7651
YOLOv11n+CNN+QNN	0.867	0.634	0.796

Table 2 indicates that the YOLOv11n+CNN+QNN configuration shows better performance in general, as it attains the highest mAP@50 metric at 79.6% and precision at 86.7%. This shows better robustness and feature representation ability. Although there is a moderate decline in recall, it provides a good trade-off and could be a promising approach.

4.6 Implications for Road Safety and Animal Intervention

As far as road safety is concerned, these results show that the framework has chosen to maximize the large animal recall that will have the greatest impact in reducing serious AVCs. Although small animals like dogs are difficult to detect, the level of performance obtained for detection and early warning/risk reduction is adequate. The results support the key premise of this research, which is that animal intervention can be achieved through existing edge devices like CCTV systems that are in place currently, and through hybrid learning models and light-weight solutions that do not rely upon special sensor technology or large-scale datasets. The fact that the system's performance was tunable for recall, precision, and confidence provides the adaptability required to deploy this type of system on the highways.

Figure 11 presents the proposed system architecture for the implementation of this AVC mitigation framework. The layers involved are elaborated in the figure for the execution of this model. The physical layer includes the hardware present at the site. These devices are connected to the digital layer, where the detection and classification are performed using YOLOv11n and Quantum Computing. The service and user layer is a place where real-time monitoring and a subsequent alert mechanism are triggered.

The proposed system contributes to the infrastructure value chain by assisting real-time monitoring and alerting the intrusion events by animals. It helps the authorities to monitor and label animal intervention zones.

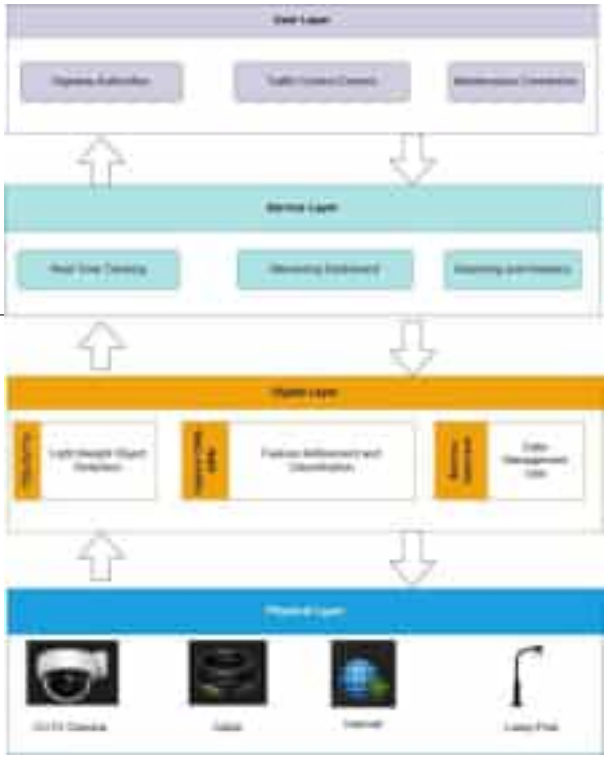


Figure 11. Proposed System Architecture

For the road users and pedestrians, this system, when deployed, serves as an alert mechanism, improving safety and effective usage of existing infrastructure like CCTV cameras.

5 Conclusion and Future Work

This paper demonstrated an effective and cost-effective method to apply Artificial Intelligence (AI) to prevent animal-induced vehicle collisions by using the existing road surveillance systems and lightweight hybrid deep learning models. The study used images taken from highway CCTV cameras and adapted a novel process using object detection and classification algorithms based on YOLOv11n, CNN and QNN for better detecting stray animals on roads with limited amounts of training data and inclement lighting conditions. The experimental results showed that the hybrid model converged steadily, detected over 90% of the larger animals and achieved a mean average precision of about 0.8 when evaluating at a 50% threshold (mAP_{50}). This validates the use of a hybrid approach. It was chosen to identify animals present in the image field of view and then classify them as either an animal or a non-animal because this is a better way to alert the drivers and road safety authorities about potential collisions.

5.1 Limitations and Observed Trends

Although there were encouraging results from the research, the differences in performance between cows and dogs illustrated the effects of object size, motion and visual distinction on classification performance. In addition, though the system performed well given the dataset size used in the research, it may not be possible to achieve optimal levels of accuracy due to the limitations imposed by the dataset size, especially for smaller animals. However, since the hybrid learning model converged consistently, regardless of whether the system was tested for different metrics, it appears to be both robust and scalable based on increased amounts of data. Overall, the authors found that using AI to augment existing surveillance systems can significantly reduce the risk of collisions between vehicles and animals, without needing additional high-level cameras or other hardware. It is recommended that future work in this area explore the following aspects:

1. Expanding the number of animals represented in the database (e.g., buffalo, horse, cat, camel), as well as different environments along roads.
2. Simplifying the detection of any type of animal into a "one animal" category to improve robustness and make it easier to deploy.
3. Deploying the system on the edge and developing a real-time alerting capability to notify traffic control centres.

References

[1] C. R. Mohanty, R. V. Radhakrishnan, M. Jain, P. K. Sasmal, U. Hansda, S. K. Vuppala, and S. K. Doki. A study of the pattern of injuries sustained from road traffic accidents caused by impact with stray animals. *Journal of Emergencies, Trauma, and Shock*, 14(1): 23–27, 2021.

[2] R. Khadka, C. S. Negi, M. Khatri, P. Pali, S. Dhanadi, K. D. Bhatta, and Y. R. Bhatta. Impact of stray animals on road safety: A case study of the east-west highway section in Bheemdatta municipality area, Nepal. *Journal of Engineering Technology and Planning*, 6(1):115–124, 2025.

[3] D. C. Wilkins, K. M. Kockelman, and N. Jiang. Animal–vehicle collisions in Texas: How to protect travellers and animals on roadways. *Accident Analysis & Prevention*, 131:157–170, 2019. doi:10.1016/j.aap.2019.05.030.

[4] A. K. Subedi, A. Rashidi, and N. Markovic´. Assessing roadside safety with computer vision: FHWA ratings as the key predictor of rural road departure crashes and severity. *Journal of Advanced Transportation*, 2025(1):5559576, 2025.

[5] K. Parkavi, A. Ganguly, A. Banerjee, S. Sharma, and K. Kejriwal. Enhancing road safety: Detection of animals on highways during night. *IEEE Access*, 2025.

[6] F. Ascensa˜o, A. Kindel, F. Z. Teixeira, R. Barrientos, M. D'Amico, L. Borda-de Á gua, and H. M. Pereira. Beware that the lack of wildlife mortality records can mask a serious impact of linear infrastructures. *Global Ecology and Conservation*, 19:e00661, 2019.

[7] C. A. Aguilar-Lazcano, I. E. Espinosa-Curiel, J. A. R´ıos-Mart´ınez, F. A. Madera-Ram´ırez, and H. Pe´rez-Espinosa. Machine learning-based sensor data fusion for animal monitoring: A scoping review. *Sensors*, 23(12):5732, 2023.

[8] L. Balc˘iauskas. Conservation and management of forest wildlife. *Forests*, 16(7):1134, 2025.

[9] J. Seong, H. S. Kim, and H. J. Jung. Tracking algorithm for hazardous area in construction sites using PTZ cameras. Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC), 42 (1): 81-88, 2025.

[10] K. Khatri, C. S. Asha, and J. M. D'Souza. Detection of animals in thermal imagery for surveillance using a GAN and object detection framework. In *2022 International Conference for Advancement in Technology (ICONAT)*, pages 1–6. IEEE, 2022.

[11] Y. Aykin, H. Sachs and N. Gerzen. Enhancing robotic vision through deep learning techniques: From detection to construction applications. In Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC), 42(1), 187-194, 2025.

[12] V. Sibanda, K. Mpofu, J. Trimble, and N. Zengeni. Design of an animal detection system for motor vehicle drivers. *Procedia CIRP*, 84:755–760, 2019.

VLM-Based Component-Level Task Planning for Adaptive Bricklaying: Integrating 3D Models and 2D Shop Drawings

Nayun Kim[1,2], Florian Noichl[1,2] and André Borrmann[1,2]

[1]Chair of Computing in Civil and Building Engineering,
Technical University of Munich, Germany
[2]TUM Georg Nemetschek Institute, Germany

nayun.kim@tum.de, florian.noichl@tum.de, andre.borrmann@tum.de

Abstract -

Construction robots require component-level task specifications that include per-unit dimensions, 6D poses, and assembly sequences respecting construction constraints. However, existing BIM-based robot task planning pipelines rely on either ad-hoc heuristic decomposition or exhaustive pre-modeling, both of which are brittle under design changes and fail to capture the assembly semantics encoded in 2D shop drawings. We propose a two-phase framework that integrates design-stage 3D models and 2D shop drawings to generate component-level task specifications for downstream robotic execution. The framework separates task formalization from task compilation, with two intermediate representations (IRs) serving as the interface between phases. In the formalization phase, a Vision-Language Model (VLM) extracts assembly logic from 2D shop drawings into an Assembly Representation (AsmRep), while geometric constraints are obtained from a design-stage IFC 3D model as a Building Element Representation (ElemRep). In the compilation phase, a deterministic IR compiler integrates both representations to generate robot-executable task specifications, including per-unit 6D poses, dimensions, and assembly sequences. Experiments on bricklaying demonstrate that foundation VLMs can reliably formalize drawing-encoded rules such as layout lines, anchor conventions, and bond patterns, and that varying only the AsmRep under identical ElemRep produces different decompositions and sequences. A preliminary simulation study further shows that the generated task specification can be consumed by a downstream robotics layer. These results support the feasibility of the proposed framework and its integration with downstream robotic execution layers.

Keywords -

Assembly Sequence Planning; Robot Task Planning; VLM; Construction Robotics; Bricklaying;

1 Introduction

Construction robots require component-level task specifications that are executable on site: the dimensions and 6D pose of each unit, and an assembly order that respects construction constraints. Unlike factory automation, construction is executed in a changing environment: dimensions drift, tolerances accumulate, and design details evolve across phases and trades [1]. As a result, an assembly specification is not just an order of operations; it encodes where work is referenced (e.g., layout lines), how boundary conditions are handled, and how residual errors are absorbed.

Building Information Modeling (BIM) is the dominant digital representation for 3D building geometry in construction. Therefore, many construction robotics studies start from BIM to derive executable task specifications, such as decomposing assemblies, assigning placements, and generating sequences, directly from 3D models. However, geometry-centric planning often shows a limitation in applying to the practice. Existing pipelines either use ad-hoc heuristic decomposition[2, 3], which is rigid and hard to generalize, or rely on exhaustive pre-modeling and manual scripting [4, 5, 6], which is labor-intensive and brittle under inevitable revisions. Crucially, both strategies face practical limitations on large-scale assemblies and dynamic sites. As the number of components increases, these approaches become increasingly difficult to maintain, and they remain brittle under design revisions and site deviations because the required assembly semantics is rarely encoded explicitly.

Construction tasks are developed progressively across project phases by multiple stakeholders. To communicate the maturity and intended use of model elements, the Level of Development (LOD) specification is widely used. LOD 300 typically provides measurable design-stage geometry (size, shape, location, and orientation), while LOD 400 represents the detailed components for fabrication and installation in 3D models [7]. However, the procedural knowledge required for robot execution, such as assembly logic, reference conventions, and sequencing, is rarely encoded explicitly in 3D BIM models. Instead, this semantic information is commonly delivered to the field through 2D shop drawings [8]. While shop drawings are a primary carrier of such semantics, they express it implicitly through

symbolic, human-centric drafting conventions rather than explicit procedural rules, making them not directly usable for generating per-unit placements and sequential robot actions.

Recent progress in vision-language models (VLMs) suggests a path to interpret such drawing content [9, 10, 11]. Yet, using VLMs for long-horizon geometric planning remains unreliable and difficult to constrain, especially when the desired output must be consistent, verifiable, and reproducible at the component level [12]. This motivates a different role for VLMs in construction robotics: not as end-to-end planners, but as formalizers that extract and structure drawing-encoded assembly semantics into a constrained intermediate representation (IR).

These observations raise three research questions: **(RQ1)** What information from design-stage 3D models and 2D shop drawings is necessary for robot assembly, and how can it be structured into intermediate representations? **(RQ2)** How can VLMs reliably extract implicit assembly semantics from 2D shop drawings? **(RQ3)** How can extracted semantics be compiled deterministically into robot-executable task specifications?

To address these questions, we present a two-stage pipeline that separates (1) **task formalization**, which extracts assembly rules from 2D shop drawings and geometric constraints from design-stage 3D models and encodes them into two intermediate representations: the Assembly Representation (AsmRep) and the Building Element Representation (ElemRep); and (2) **task compilation**, in which a deterministic compiler compiles both IRs into a robot-executable assembly task specification (per-unit 6D poses, dimensions, and assembly sequences). Experiments on bricklaying show that, under identical BIM-driven geometry, varying only assembly rules produces different decompositions and sequences. This indicates that our system explicitly captures assembly logic, enabling scalable component-level task planning and providing a foundation for rapid re-planning in a dynamic construction environment.

2 Related works

Existing BIM-based robotics pipelines largely treat task planning as geometric translation from 3D models. Early work mapped IFC to robotic description formats to reconstruct the building environment for robot operation [5]. Subsequent studies attempted to derive component-level information: Brinkhoff and Kolani [2] geometrically slices LOD 300 walls into brick via scripts, Gao et al. [3] generated assembly coordinates with heuristic spatial sorting, and Wong Chong et al. [4] extracted coordinates for wood frame assembly from BIM. To improve site adaptability, Oyediran et al. [13] integrated 4D BIM scheduling and Fingrut et al. [14] employed computer vision for adaptive bricklaying correction. More recently, Kim et al. [6] used an LLM for dialogue-based robot planning, where BIM-derived context serves as a world model.

Despite these advancements, these BIM-based planning efforts demonstrate critical limitations in scalability and robustness. Explicitly modeling every component at an LOD 400 level becomes practically impossible as the scale of construction increases due to massive computational and modeling overhead. Conversely, heuristic-driven decomposition from LOD 300 models lacks the robustness needed for complex geometric scenarios, such as intricate joints or openings. These limitations motivate extracting assembly semantics from 2D drawings and compiling it deterministically with 3D geometry, rather than treating planning as geometric translation alone.

In parallel, VLMs have been applied to interpret assembly instructions in robotics. Manual2Skill [9] extracts hierarchical assembly graphs from furniture manuals to acquire robotic skills, VLM-MSGraph [10] constructs multi-hierarchical scene graphs to guide assembly sequences, and Kyaw et al. [11] utilized VLMs for component decomposition of 3D meshes. However, these systems primarily operate on general-purpose reasoning, such as identifying that a chair requires legs, and typically rely on unique identifiers for a small set of distinct parts. Such methods do not scale to repetitive construction tasks, which involve the massive repetition of identical units where assigning unique IDs to every component is infeasible. Furthermore, interpreting construction drawings requires sophisticated spatial reasoning across multiple views (plans, sections, elevations) and conventions that exceed the general functional cues used in furniture or toy assembly. Moreover, end-to-end VLM planning remains unreliable for long-horizon reasoning; recent work suggests VLMs are more effective as formalizers that translate multimodal inputs into structured representations [12].

These observations motivate our approach: formalized assembly semantics from 2D drawings via constrained VLM extraction and compile them deterministically with 3D model geoemtry.

3 Methodology

We propose a component-level assembly sequence planning framework that synthesizes geometric contexts from 3D Models with assembly rules from 2D architectural drawings to generate robot-executable assembly information.

As illustrated in Figure 1, the framework operates in two phases: (1) **task formalization**, where different levels of construction data are parsed into two structured intermediate representation (IRs), a Building Element Representation (ElemRep) encoding spatial constraints, and

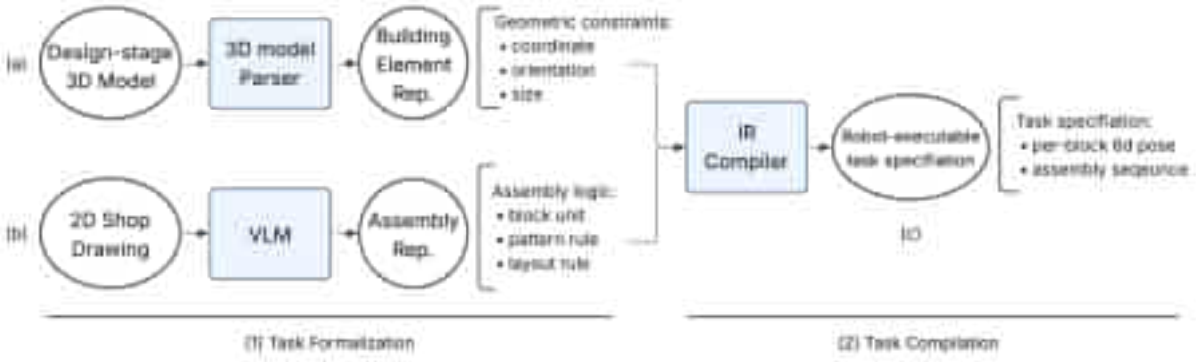


Figure 1. System overview. We separate (1) task formalization into intermediate representations (IRs): (a) Building Element Representation (ElemRep) from design-stage 3D models and (b) Assembly Representation (AsmRep) from 2D shop drawings, and (2) task compilation into (c) robot-executable assembly information via an IR compiler.

an Assembly Representation (AsmRep) encoding assembly rules; and (2) **task compilation**, where a deterministic compiler translates these IRs into a final component assembly sequence, including per-unit 6D poses, dimensions, and assembly sequence. This separation leverages the VLM's strength in semantic formalization while avoiding its unreliability in long-horizon geometric planning.

3.1 Task Formalization from 3D Models

Derived from a design-stage 3D model, the ElemRep encapsulates the geometric location and boundary of the element. We formally define the ElemRep G as a tuple comprising three primary components:

$$\mathcal{G} = (E, O, R) \tag{1}$$

- **Target Element** (E): This component defines the volumetric element to be constructed (e.g., the wall). It is represented as $E = (\mathbf{p}_e, \mathbf{s}_e)$, where $\mathbf{p}_e \in SE(3)$ denotes the global 6D pose of the element's origin, and $\mathbf{s}_e \in \mathbb{R}^3$ represents its volumetric dimensions (length, height, thickness).
- **Opening Entities** (O): This set captures geometric definitions for voids within the target element, defined as $O = \{o_1, \ldots, o_n\}$. Each opening o_k is a tuple $(\mathbf{p}_{o,k}, \mathbf{s}_{o,k})$, specifying the global pose and dimensions required for boolean subtraction.
- **Reference Alignment** (R): When the target element is constrained by an adjacent structure, this component provides the spatial referencing data. It is defined as $R = (\mathbf{c}_{ref}, \mathbf{v}_{ref}, d_{offset})$, containing the reference centroid $\mathbf{c}_{ref} \in \mathbb{R}^3$, the normalized axis direction $\mathbf{v}_{ref}$, and the scalar offset distance d_{offset} relative to the target element.

3.2 Task Formalization from 2D Drawings

Derived from a 2D shop drawing via VLM parsing, the AsmRep S captures the explicit assembly rules for interior masonry. We define $\mathcal{S}$ as a tuple comprising:

$$\mathcal{S} = (B, P, L) \tag{2}$$

- **Block** (B): Identifies the physical attributes of the finishing unit. It is represented as $B = (\mathbf{d}_{unit}, id)$, where $\mathbf{d}_{unit} \in \mathbb{R}^3$ denotes the nominal dimensions (l_u, w_u, h_u) and id specifies the material type.
- **Pattern** (P): Specifies the topological arrangement of units. It is represented as $P = (\tau, \alpha, \mathbf{d}_{joint})$, where τ denotes the bond type (e.g., running bond, stack bond), $\alpha \in [0, 1)$ is the row-wise offset ratio, and $\mathbf{d}_{joint}$ defines the visual joint thickness.
- **Layout** (L): Defines the reference axis and boundary conditions for grid distribution. It is represented as $L = (\lambda_{align}, \lambda_{anchor}, \delta, C_{edge})$. The first three parameters deterministically resolve the pattern origin: λ_{align} selects identifies the semantic reference layout line (e.g., wall center, opening center, structure finish face), δ applies a scalar offset distance from this layout line to account for finish thickness or design shifts, and λ_{anchor} determines which specific feature of the unit (e.g., brick center or edge, joint center) locks onto the layout line. Finally, to manage dimensional remainder at boundaries, C_{edge} defines the edge treatment policy $(\mathbf{s}_{min}, J_{range}, \epsilon_{adj})$, specifying the minimum allowable cut size $\mathbf{s}_{min}$ and the permissible joint adjustment range J_{range} for absorbing cumulative residuals.

3.3 Task Compilation via IR Compiler

Given ElemRep $\mathcal{G}$ and AsmRep $\mathcal{S}$, the compiler deterministically computes a robot-executable **task specification** $\mathcal{T}$:

$$\mathcal{T} = (\mathcal{U}, \pi) \tag{3}$$

where $\mathcal{U} = \{u_i\}_{i=1}^{n}$ is **the set of component units** and $\pi = (i_1, \ldots, i_n)$ is a permutation of $\{1, \ldots, n\}$ that defines **the assembly sequence**. Each unit is defined as:

$$u_i = (\mathbf{p}_i \in SE(3),\ \mathbf{s}_i \in \mathbb{R}^3) \tag{4}$$

with 6D pose $\mathbf{p}_i$ and dimensions $\mathbf{s}_i$. All decisions follow fixed rules, so identical ElemRep and AsmRep $(\mathcal{G}, \mathcal{S})$ yield an identical task specification $\mathcal{T}$.

The IR compiler first maps the target element E and openings O from ElemRep $\mathcal{G}$ into a wall-local coordinate frame defined by the wall pose in E. All layout decisions are then made in the wall-local (x, z) plane, where x follows the wall direction and z corresponds to course height. Finally, each unit extent is lifted back to a global 6D pose in $SE(3)$ using the inverse wall transform, and an assembly sequence π is assigned using a deterministic ordering rule.

Step 1: Coordinate framing From the ElemRep $\mathcal{G}$, the compiler reads the global 3D pose and dimensions of the target element E and each opening $o \in O$, and transforms them into a common wall-local coordinate frame defined by the wall pose in E. In this wall-local frame, the compiler derives the wall bounds in the (x, z) domain, where x follows the wall direction and z corresponds to course height, and maps openings into the same domain.

Step 2: Layout line and course grid setup Using the layout term $L = (\lambda_{\text{align}}, \lambda_{\text{anchor}}, \delta, C_{\text{edge}})$ from the AsmRep $\mathcal{S}$, the compiler deterministically resolves a horizontal reference layout line x_{ref}. If $\lambda_{\text{align}} =$ *WALL_CENTER*, x_{ref} is obtained from the wall extents in E; if $\lambda_{\text{align}} =$ *OPENING_CENTER*, x_{ref} is obtained from the designated opening geometry in O; and if $\lambda_{\text{align}} =$ *STRUCTURE_FINISH_FACE*, x_{ref} is obtained from the referenced structure face and offset in R. Finally, λ_{anchor} selects the anchored unit feature (e.g., unit center or edge), which determines a unique grid origin x_0. The scalar offset δ then shifts x_{ref} to account for finish thickness or design intent.

The compiler then defines the horizontal module $m_x = l_u + j_h$ from the unit length in B and head-joint thickness in P, and the vertical module $m_z = h_u + j_b$ from the unit height in B and bed-joint thickness in P. Vertically, courses are generated from the wall base upward at increments of m_z without vertical cutting.

Step 3: Pattern instantiation For each course, the bonding pattern $P = (\tau, \alpha, \mathbf{d}_{\text{joint}})$ deterministically defines a row-wise horizontal offset (e.g., alternating offsets for running bond). The compiler then instantiates nominal 1D unit spans along x, anchored at the grid origin x_0 with the course-dependent offset. Depending on λ_{align}, the compiler uses either *grid mode* for center-aligned reference lines (wall/opening center), where tiling is symmetric about x_{ref} and clipped to $[x_{\min}, x_{\max}]$, or *linear mode* for face-aligned reference lines (structure finish face), where tiling proceeds in a fixed direction from x_{ref} and is clipped to the same bounds. This step yields per-course nominal spans prior to applying the edge policy.

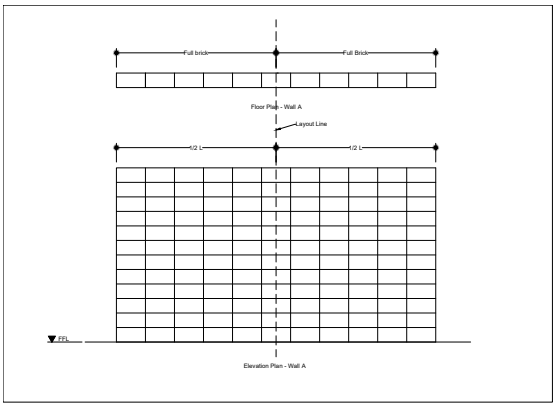
Step 4: Edge policy The compiler then applies the edge policy $C_{\text{edge}} = (\mathbf{s}_{\min}, J_{\text{range}}, \epsilon_{\text{adj}})$ to handle wall-end remainders ρ (remaining length after clipping, with head-joint thickness j_h from P): if $\rho \le \epsilon_{\text{adj}}$, the residual is absorbed by adjusting head joints within J_{range}; if $\rho \ge j_h + \mathbf{s}_{\min}$, a cut unit is appended; if neither condition holds (i.e., $\epsilon_{\text{adj}} < \rho < j_h + \mathbf{s}_{\min}$), the compiler redistributes the layout over the preceding k neighboring units to resolve the remainder. Here, k is chosen deterministically to remove ambiguity.

Step 5: Opening subtraction Openings O extracted from $\mathcal{G}$ are then subtracted from the generated spans: the compiler splits any course overlapping an opening at its boundaries and removes interior segments. Resulting fragments are checked against the minimum brick size $\mathbf{s}_{\min}$. Segments meeting this constraint are retained as cut units, while shorter ones are dropped.

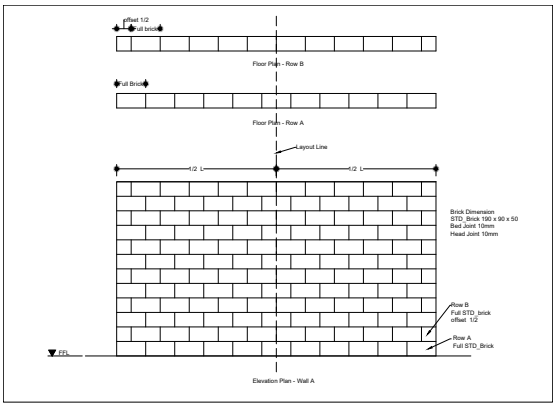
Step 6: Sequencing Finally, the compiler assigns an assembly sequence π: units are placed bottom-to-top by course index, and within each course, in increasing distance from the reference layout line. This places units near the reference line first, letting tolerances be absorbed at the boundaries instead of accumulating toward the layout line.

4 Experiments

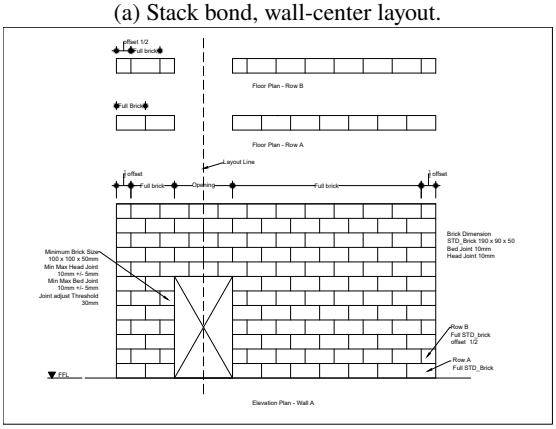
We evaluate the proposed pipeline on controlled bricklaying scenarios to test our two-phase system. First, we measure whether a VLM foundation model can formalize assembly semantics in 2D drawings into an AsmRep, with emphasis on visually implied rules (e.g., layout line, anchor, and bond type). Second, we perform a compiler sensitivity study: We fix geometry (ElemRep) and vary assembly semantics (AsmRep) to see how assembly semantics affect the decomposition and sequencing of task specification.



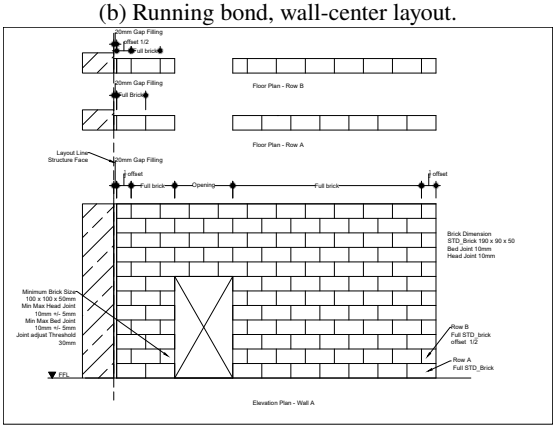
(a) Stack bond, wall-center layout.



(b) Running bond, wall-center layout.



(c) Running bond with an opening and opening-center layout.



(d) Running bond with an opening and a finish-face layout.

Figure 2. The syntactic drawings for experiment 1: Bricklaying drawing variants used for AsmRep extraction.

Table 1. Field-wise VLM extraction accuracy across drawing types and models. (k correct/n valid runs)

Field	Type	(a) stack Gemini 3	(a) stack GPT-4o	(b) running Gemini 3	(b) running GPT-4o	(c) running_opening Gemini 3	(c) running_opening GPT-4o	(d) running_opening_ref Gemini 3	(d) running_opening_ref GPT-4o
bond_type	Cat.	10/10	0/10	10/10	10/10	10/10	10/10	10/10	10/10
offset_ratio	Num.	10/10	0/10	10/10	10/10	10/10	10/10	10/10	10/10
head_joint	Num.	–	–	10/10	10/10	10/10	10/10	10/10	10/10
bed_joint	Num.	–	–	10/10	10/10	10/10	10/10	10/10	10/10
layout_line	Cat.	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
anchor_type	Cat.	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
brick_dims	Num.	–	–	10/10	10/10	10/10	10/10	10/10	10/10
opening_detected	Cat.	–	–	–	–	10/10	10/10	10/10	10/10
min_brick_size	Num.	–	–	–	–	10/10	10/10	10/10	10/10
joint_threshold	Num.	–	–	–	–	10/10	10/10	10/10	10/10
offset_distance	Num.	–	–	–	–	–	–	10/10	7/10

4.1 Experiment 1: VLM AsmRep Extraction

4.1.1 Setup

We evaluate whether foundation VLM can reliably formalize drawing-encoded assembly semantics into the schema-constrained AsmRep. We compare Gemini 3 with GPT-4o under an identical schema and prompting setup. Both models were configured with temperature set to 0 and structured JSON output mode enforced via a Pydantic Schema that defines three field groups: brick definition, pattern logic, and wall mapping rule. Each model was executed 10 times independently. We report field-level extraction accuracy as the number of runs producing the correct value for each AsmRep field, evaluated against a manually defined ground truth per drawing. The full prompt specification is available in the git repository. [1]

We created four synthetic brickwork drawings (Figure 2) with progressively increasing semantic complexity to isolate specific extraction challenges: (a) stack bond with wall-center layout, containing only bond pattern and layout line without explicit numeric parameters; (b) running bond with wall-center layout, introducing offset logic and joint dimensions; (c) running bond with an opening and opening-center layout, adding opening detection and edge-policy parameters; and (d) running bond with an opening

[1] https://github.com/narchitect/assembly-seqeunce

and structure finish-face layout, further adding a reference entity and offset distance. This progression allows us to evaluate VLM reliability as drawing complexity and the number of extractable semantic fields increase.

4.1.2 Result

Table 1 reports field-level extraction accuracy across 10 independent runs per configuration. Gemini 3 achieved perfect extraction across all applicable fields and all four drawings. GPT-4o performed comparably on drawing (c) and (d) with running bonding, but failed on implicit visual conventions in the stack bonding drawing (a). It misidentified stack bond as running bond in all 10 runs and also returned null for offset distance in 3/10 runs, confirming that VLM outputs are not fully deterministic even at temperature 0. These results suggest that the primary bottleneck in extraction is not numeric recovery but the interpretation of visually encoded drafting conventions, such as bond type and offset logic.

4.2 Experiment 2: IR Compiler Task Compilation

4.2.1 Setup

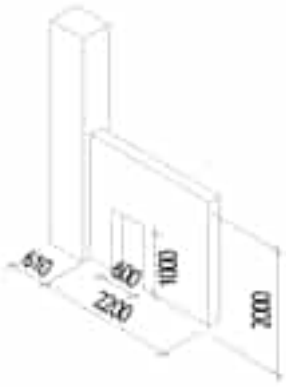


Figure 4. A sample design-stage 3D model for IR compilation experiment

For task compilation, the IR compiler takes both an ElemRep and an AsmRep as inputs. We construct the ElemRep by exporting a design-stage 3D model to IFC (Figure 4) and parsing it with IfcOpenShell to extract the wall, a rectangular opening, and an adjacent reference structure, including their global 6D poses and dimensions. This ElemRep is reused across all runs.

We compile four AsmRep configurations against this identical geometry to examine how assembly semantics affect task specification generation. Configuration (a) and (b) differ in bond pattern (stack bond vs. running bond) under the same wall-center layout condition, allowing us to observe the effect of bond type on decomposition. Configurations (b), (c), and (d) share the same running bond but differ in layout line (wall center, opening center, and structure finish face), allowing us to isolate the effect of layout semantics under a fixed bond pattern.

4.2.2 Result

Figure 3 shows that, under an identical geometry (ElemRep), changing only the assembly semantics (bonding type and layoutline) term in the AsmRep produces different brick decompositions and assembly sequences.

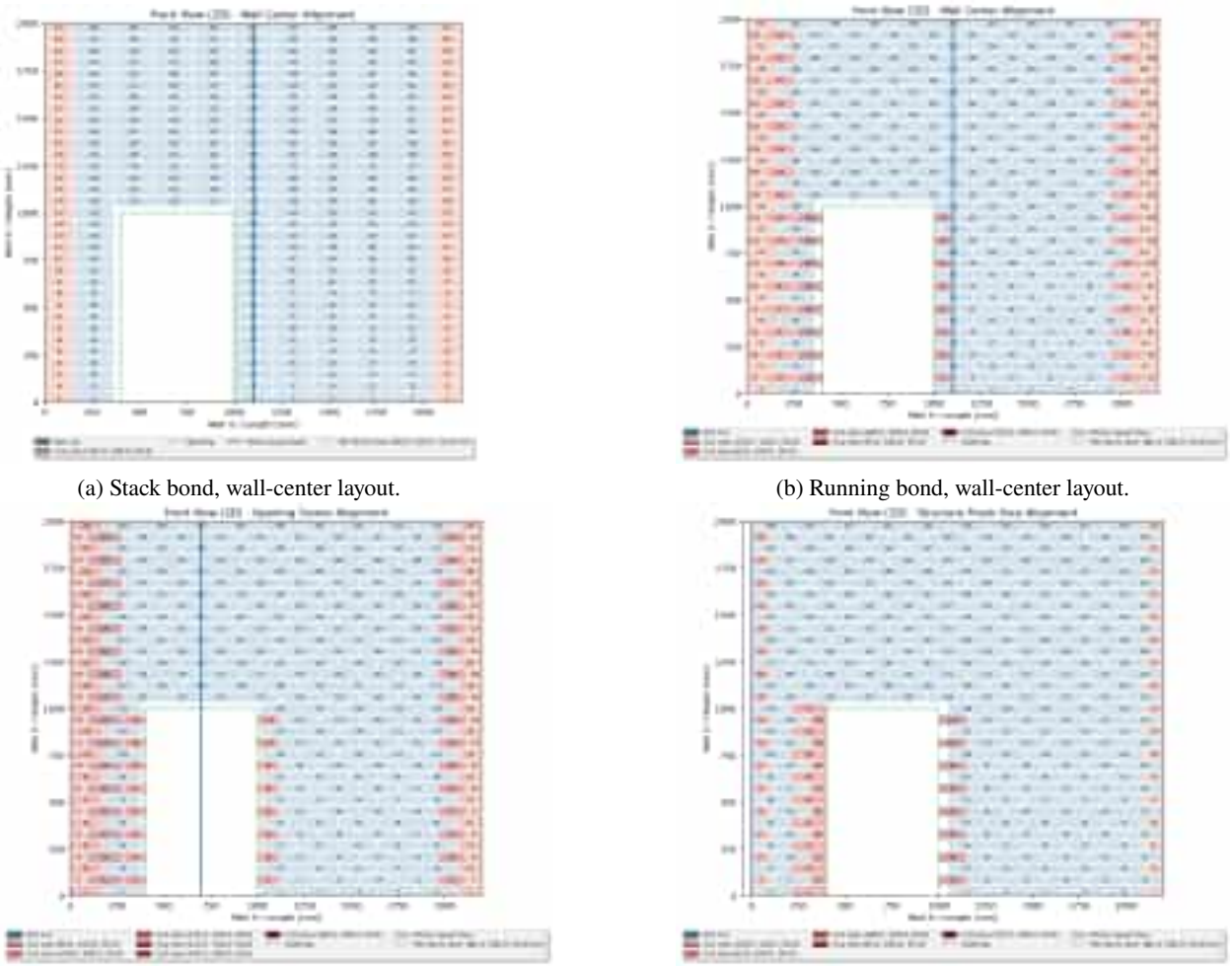
Comparing configurations (a) and (b), the bond pattern changes the decomposition substantially. Stack bond (Figure 3a) maintains the same grid phase across courses and therefore produces more regular opening-adjacent fragmentation. In contrast, running bond (Figure 3b) introduces alternating row offsets that interact with opening boundaries, producing more varied cut-brick sizes and void patterns around jambs and lintels.

Within the running-bond configurations, Figures 3b, 3c, and 3d show that layout line alone also affects decomposition and sequencing. In the wall-center layout (Figures 3b), the overall pattern is symmetric with respect to the wall, but small fragments can appear near the opening boundaries. In the opening-center layout (Figure 3c), the bricks are arranged more symmetrically around the opening, although this often leads to uneven cuts at the wall ends. In the structure-finish-face layout (Figure 3d), the brick pattern is aligned with the referenced structure, and the remaining adjustment is shifted to the opposite side. These differences also appear in the generated sequence, because the bricks are placed in a fixed order by course and then by distance from the reference line.

4.3 Preliminary Simulation Validation



Figure 5. Compiled task specification executed in Isaac Sim: UR10e sequentially placing bricks according to the generated per-unit targets



(a) Stack bond, wall-center layout.

(b) Running bond, wall-center layout.

(c) Running bond, opening-center layout.

(d) Running bond, finish-face layout.

Figure 3. The results of experiment 2: Compiled task specifications for four different AsmRep configurations under identical ElemRep. red color: cut bricks; numbers: sequence IDs

To provide an execution-level validation of the compiled task specification, we integrated the generated output into a robot simulation environment. The simulation was set up in NVIDIA Isaac Sim with a UR10e manipulator and a Robotiq 2F-140 gripper. RMPflow was used for collision-aware motion generation. A finite-state pick-and-place controller sequentially executed the per-unit targets from the compiled task specification, following an approach $\rightarrow$ grasp $\rightarrow$ lift $\rightarrow$ move $\rightarrow$ place $\rightarrow$ release $\rightarrow$ retreat cycle. As shown in Figure 5, the robot can execute sequential brick placements at the compiled target poses in simulation. Although this does not yet constitute a full validation of autonomous bricklaying, it shows that the proposed task specification can be consumed by a downstream robotics execution layer.

5 Discussion

The experiments support our central premise that assembly semantics encoded in drawings affects component-level decomposition and sequencing even under identical BIM geometry. This motivates the proposed separation between semantic formalization and deterministic compilation: VLMs translate drawing conventions into a constrained AsmRep, while the IR compiler produces reproducible, verifiable per-unit task specifications by integrating the semantic IR with the geometric IR.

RQ1 (IR specification): The results (Figure 3) indicate that the combination of an ElemRep (element bounds, openings, and optional reference structures) and an AsmRep (layout line, anchor convention, bond pattern, and edge policy) is sufficient to generate component-level brick poses, dimensions, and assembly sequences, in the evalu-

ated cases.

RQ2 (VLM Task Formalization): The results of Experiment 1 (Table 1) show that both Gemini 3 and GPT-4o perform well overall under the evaluated drawing conditions, but GPT-4o is less consistent in extracting implicit visual assembly semantics. This suggests that the main bottleneck in VLM-based semantic extraction is not the recovery of explicit numeric or textual information, but the interpretation of visual conventions such as bond pattern, layout line, and anchor type. Although the current study is limited to synthetic drawings with controlled conventions, the results indicate that foundation VLMs already have practical potential to formalize construction-relevant drawing semantics without task-specific fine-tuning in this controlled setting.

RQ3 (Deterministic compilation): Experiment 2 (Figure 3) shows that, under an identical geometry (ElemRep), varying the assembly semantics (AsmRep) changes the resulting decomposition and assembly sequence in predictable ways. The comparison between stack bond and running bond shows that bond pattern alone affects row-phase behavior and opening-adjacent fragmentation. Within the running-bond cases, changing only the layout line further changes the grid phase relative to the opening, which leads to different cut-brick sizes, void patterns, and sequences under the same minimum cut-length rule. These results show that the compiler makes the effect of assembly semantics explicit and reproducible, rather than leaving it implicit in geometric modeling alone.

A preliminary robot simulation validation further showed that the compiled task specification can be integrated with a downstream robotics layer in simulation. The present study should be read as a controlled feasibility study rather than a robustness benchmark. Within the bricklaying domain, our results show that explicit modeling of assembly semantics and compilation logic enables structured and reproducible task specification generation across controlled layout variations and edge conditions. Future work should evaluate the framework on real drawings, more complex bricklaying cases, and different assembly tasks.

6 Conclusion

This paper presented a two-phase framework for component-level task planning in adaptive bricklaying by integrating design-stage 3D models with 2D shop drawings. The main contributions are two intermediate representations, ElemRep and AsmRep, and a deterministic IR compiler that combines them to generate per-unit brick poses, dimensions, and assembly sequences in a reproducible manner. The results show that foundation VLMs can formalize key assembly semantics from drawings, and that varying assembly semantics under identical geometry leads to different decompositions and sequences. A preliminary robot simulation further showed that the generated task specification can be integrated with a downstream robotics layer.

7 Acknowledgments

We gratefully acknowledge the support of the German Research Foundation (DFG) for funding the project under grant FOR 5672, subproject A "Generic framework for simulating and executing robotized construction", grant number 517965147, as well as the support of the TUM Innovation Network "CoConstruct".

References

[1] J. Buchli, M. Giftthaler, N. Kumar, M. Lussi, T. Sandy, K. Dörfler, and N. Hack. Digital in situ fabrication - challenges and opportunities for robotic in situ fabrication in architecture, construction, and beyond. *Cement and Concrete Research*, 112:66–75, 2018. doi:10.1016/j.cemconres.2018.05.013.

[2] M. J. S. Brinkhoff and M. R. Kolani. Deriving fabrication information from BIM for automated robotic task planning. In *Proceedings of Forum Bauinformatik*. Technische Universität Hamburg, 2024. doi:10.15480/882.13543.

[3] Y. Gao, J. Meng, J. Shu, and Y. Liu. BIM-based task and motion planning prototype for robotic assembly of COVID-19 hospitalisation light weight structures. *Automation in Construction*, 140: 104370, 2022. doi:10.1016/j.autcon.2022.104370.

[4] O. Wong Chong, J. Zhang, R. M. Voyles, and B.-C. Min. BIM-based simulation of construction robotics in the assembly process of wood frames. *Automation in Construction*, 137:104194, 2022. doi:10.1016/j.autcon.2022.104194.

[5] K. Kim and M. Peavy. BIM-based semantic building world modeling for robot task planning and execution in built environments. *Automation in Construction*, 138:104247, 2022. doi:10.1016/j.autcon.2022.104247.

[6] K. Kim, P. Ghimire, and P.-C. Huang. Framework for LLM-Enabled Construction Robot Task Planning: Knowledge Base Preparation and Robot–LLM Dialogue for Interior Wall Painting. *Robotics*, 14(9):117, 2025. doi:10.3390/robotics14090117.

[7] BIM Forum. LOD Specification 2025, 2025. URL `https://bimforum.org/wp-content/uploads/2026/01/LOD-Spec-2025-Part-I-Official.pdf`.

[8] O. Disney, M. Roupé, M. Johansson, J. Ris, and P. Höglin. Total BIM on the construction site: a dynamic single source of information. *ITcon*, 28:519–538, 2023. doi:10.36680/j.itcon.2023.027.

[9] C. Tie, S. Sun, J. Zhu, Y. Liu, J. Guo, Y. Hu, H. Chen, J. Chen, R. Wu, and L. Shao. Manual2Skill: Learning to read manuals and acquire robotic skills for furniture assembly using vision-language models, 2025. arXiv:2502.10090.

[10] S. Li, Z. Yan, Z. Wang, and Y. Gao. VLM-MSGraph: Vision language model-enabled multi-hierarchical scene graph for robotic assembly. *Robotics and Computer-Integrated Manufacturing*, 94: 102978, 2025. doi:10.1016/j.rcim.2025.102978.

[11] A. H. Kyaw, R. Gupta, D. Shah, A. Sinha, K. Mathewson, S. Pender, S. Chitta, Y. Koga, F. Ahmed, L. Sass, and R. Davis. Text to robotic assembly of multi component objects using 3D generative ai and vision language models, 2025. arXiv:2511.02162. NeurIPS 2025 (Creative AI Track).

[12] M. He, Y. Zheng, Y. Liu, Z. An, B. Cai, J. Huang, L. Zhou, F. Liu, Z. Li, and L. Zhang. Vision language models cannot plan, but can they formalize?, 2025. arXiv:2509.21576.

[13] H. Oyediran, W. Turner, K. Kim, and M. Barrows. Integration of 4D BIM and robot task planning: Creation and flow of construction-related information for action-level simulation of indoor wall frame installation, 2024. arXiv:2402.03602.

[14] A. Fingrut, S. Altun, and Y. Li. Integration of robotics and computer vision for responsive and autonomous bricklaying in construction. *Technology|Architecture + Design*, 8(2):271–281, 2024. doi:10.1080/24751448.2024.2405406.

Deep Learning-Based Multi-Task Recognition of Objects and Text in Engineering Drawings

Jyun-Yu You[1] and Ying-Hua Huang [2]

[1] Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, Taiwan

[2] Department of Civil and Construction Engineering, National Yunlin University of Science and Technology, Taiwan

D11310206@yuntech.edu.tw, huangyh@yuntech.edu.tw

Abstract –

Engineering drawings remain a critical medium for conveying design intent in architectural and construction projects. Despite the increasing adoption of Building Information Modeling (BIM), two-dimensional drawings are still widely used in practice, particularly during early design stages and interdisciplinary coordination. The interpretation of such drawings largely relies on manual processes, which are time-consuming and prone to human error, especially when drawings contain dense graphical elements and complex textual annotations.

This study proposes an automated framework that integrates deep learning–based object detection and optical character recognition (OCR) to extract graphical symbols and textual information from architectural floor plans. A tile-based detection strategy combined with global non-maximum suppression (NMS) is employed to address large-format drawings and multi-scale objects. Detected text regions are further processed using an OCR module to extract text for subsequent analysis. Experimental results demonstrate the feasibility of the proposed pipeline and highlight key challenges related to complex layouts, rotated text, and domain-specific terminology. Overall, the proposed framework establishes a structured foundation for future semantic interpretation and intelligent engineering applications.

Keywords –

Engineering drawings, Floor plan recognition, Object detection, OCR, Semantic interpretation

1 Introduction

Engineering drawings remain a primary medium for conveying design intent in building projects. Despite the increasing adoption of Building Information Modeling (BIM), a substantial portion of design information is still generated and exchanged in two-dimensional floor plans. These drawings typically contain dense geometric components and domain-specific textual annotations, making automated interpretation particularly challenging.

Existing studies have explored object detection or text recognition in architectural drawings using deep learning techniques. However, most approaches focus on single-task recognition and lack an integrated framework capable of jointly extracting geometric objects and textual information from complex floor plans. This limitation restricts their applicability to downstream tasks such as information structuring and BIM model generation.

To address this gap, this study proposes a deep learning-based multi-task recognition framework for architectural floor plans. While the core components leverage established architectures like YOLO and OCR, the contribution of this work lies in integrating these components into a unified pipeline tailored to the characteristics of floor plans, including large-format images, dense layouts, mixed graphical-textual content, and Chinese annotations. This integrated design provides a practical basis for structured information extraction and subsequent engineering-oriented interpretation.

2 Literature Review

2.1 Automated Drawing Analysis

Early studies on automated engineering drawing analysis primarily relied on rule-based and traditional image processing techniques, including heuristic rules, line tracing, and template matching [1]. Due to significant variability in drawing styles, symbol conventions, and annotation density, these methods generally exhibited limited robustness and poor generalization [2], [3].

With advances in deep learning, convolutional neural networks (CNNs) have become the dominant approach for the analysis of engineering drawings. Architectures

such as U-Net, Mask R-CNN, and YOLO have demonstrated promising performance in detecting symbols and structural components in complex line-based drawings [4]-[6]. However, engineering drawings differ fundamentally from natural images, as they are characterized by dense geometric structures, high-contrast linework, scale variation, and the close interleaving of graphical and textual information, which continue to pose challenges for reliable automated recognition.

Within this category, compared with P&ID diagrams, architectural floor plans exhibit less standardized symbol libraries and higher layout variability.

2.2 Research Focus and Limitations of Existing Studies

A large body of existing research has focused on piping and instrumentation diagrams (P&ID), in which diagram structures and symbol libraries are relatively standardized, enabling high detection accuracy using deep learning–based methods [7]. In contrast, architectural floor plans have received comparatively limited attention.

Most current studies on architectural drawings address isolated tasks, such as object detection, room boundary extraction, or room type recognition [3], [5], [8]. Recent surveys classify automated floor plan analysis into room boundary detection, room type recognition, object detection, and text detection and recognition [9].

Nevertheless, most approaches remain single-task oriented, and few studies attempt to integrate multiple recognition tasks within a unified framework. As a result, many systems provide only low-level outputs, such as bounding boxes or masks, without cross-task integration or semantic interpretation.

2.3 Text Recognition in Engineering Drawings

Textual annotations in engineering drawings convey critical information, including dimensions, material specifications, and construction notes. Although recent studies have explored text detection and OCR in engineering diagrams [6], [10], recognition performance remains sensitive to text orientation, font variation, background noise, and image resolution.

These challenges are further amplified in drawings containing Chinese annotations. Chinese characters exhibit high visual complexity and lack explicit word boundaries, while engineering texts often involve domain-specific terminology and mixed-language expressions. Consequently, even when character-level recognition is achieved, the absence of contextual understanding frequently limits the interpretability of extracted text [11], [12] , rendering text recognition alone insufficient for downstream engineering applications.

2.4 Research Gap and Motivation

The reviewed literature reveals several limitations. Architectural floor plans remain underrepresented compared to P&ID diagrams, and most existing studies focus on single-task recognition with limited integration of graphical and textual information. Moreover, text recognition for Chinese engineering annotations remains insufficiently addressed, and few approaches link recognition outputs to higher-level engineering semantics or application-oriented representations.

These gaps motivate the development of an integrated framework that jointly considers object detection and text recognition. Accordingly, this study proposes a scalable detection–recognition pipeline for architectural floor plans, aiming to bridge the gap between visual extraction and intelligent engineering applications.

3 Proposed Framework

The proposed framework consists of four main stages: image tiling, object detection, result merging, and text recognition. The overall workflow is designed to accommodate high-resolution drawings and diverse annotation characteristics commonly found in engineering practice. The complete processing pipeline and the interactions between individual modules are illustrated in Figure 1.

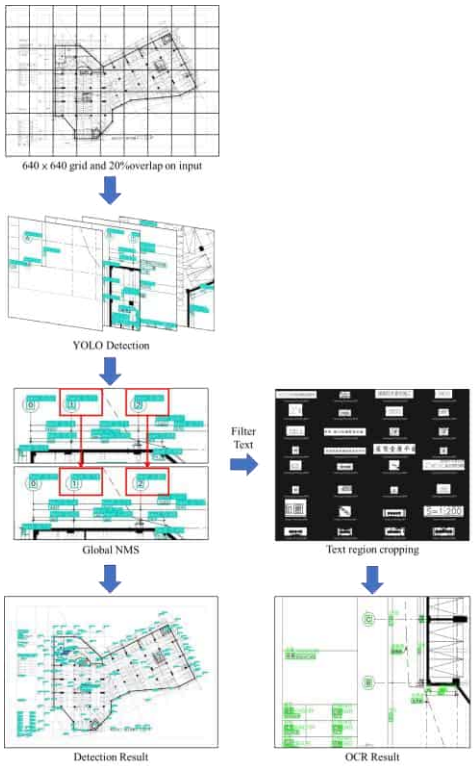


Figure 1. Overall framework

3.1 Image Tiling Strategy

Architectural floor plans are typically provided at high resolutions, which exceed the input size constraints of common deep learning models. To address this issue, a tile-based strategy is adopted, in which each input drawing is divided into fixed-size image patches (640 × 640 pixels). An overlap is introduced between adjacent tiles to mitigate boundary truncation effects and preserve contextual information for objects located near tile edges.

The selection of the tile size and overlap ratio was guided by a trade-off between detection accuracy and computational efficiency. Larger tiles preserve more contextual information but increase memory consumption, whereas smaller tiles may lead to object truncation and fragmented detections. An overlap between adjacent tiles was therefore introduced to mitigate boundary effects and reduce missed detections near tile edges.

3.2 Object Detection Using YOLO

A YOLO-based object detection model is employed to identify graphical symbols and text regions within each image tile. Detection is performed independently on each tile, producing bounding boxes, class labels, and confidence scores. The detected bounding boxes are then mapped back to the original image's coordinate system.

To eliminate redundant detections arising from overlapping tiles, a global non-maximum suppression (NMS) process is applied. This step retains the most representative bounding boxes based on confidence scores while removing highly overlapping predictions, thereby ensuring robust object localization across large-format drawings.

3.3 Text Region Filtering and Cropping

Among all detected objects, bounding boxes classified as text are selected for subsequent processing. Each text region is cropped from the original drawing, and padding is applied when necessary to ensure sufficient context. This selective filtering allows the OCR module to focus exclusively on candidate text regions, reducing computational cost and false recognition.

3.4 OCR-Based Text Recognition

The cropped text regions are processed by an OCR module to recognize text. The recognized text, confidence scores, and spatial information are recorded and associated with their corresponding bounding boxes. These outputs serve as intermediate results for future semantic interpretation and structured information extraction.

4 Dataset Construction

4.1 Data Sources

The dataset used in this study consists of 118 architectural floor plans collected from nine real-world building projects, including residential buildings, townhouses, educational facilities, and public infrastructure. The drawings exhibit substantial variation in scale, layout, annotation density, and visual style, reflecting realistic engineering conditions. Detailed statistics of the dataset, including project distribution, drawing characteristics, and data splits, are summarized in Table 1.

Table 1. Overview of the dataset

Item	Description
Number of projects	9
Number of floor plans	118
Drawing types	Residential and public buildings
Image resolution	3500–5000 px (short edge)
Annotation types	Symbols + text

As shown in Table 1, the dataset covers multiple building types and exhibits substantial variation in resolution and annotation density, which reflects realistic engineering drawing conditions rather than controlled benchmark datasets.

4.2 Annotation Strategy

Textual annotations are labeled on a line-by-line basis using horizontal rectangular bounding boxes. Blurred or partially occluded text is annotated as long as its textual nature can be reasonably identified. Graphical symbols are annotated separately to support multi-class object detection.

4.3 Dataset Generation Workflow

Annotated full-size drawings are divided into overlapping tiles following the image tiling strategy. Bounding boxes fully contained within a tile are retained, while partially truncated boxes are selectively preserved for specific classes, such as text. Tiles without annotations are discarded to reduce dataset imbalance. The resulting dataset is randomly split into training, validation, and testing sets using a 7:2:1 ratio.

5 Experimental Results

Preliminary experiments were conducted using multiple YOLO variants, including models from the YOLOv8 and YOLOv11 families. Performance was

evaluated using standard object detection metrics, including mean Average Precision (mAP) at IoU thresholds of 0.50 and 0.50–0.95, as well as bounding box precision. Quantitative detection performance for all evaluated models is reported in Table 2. Representative qualitative detection results under different drawing conditions are presented in Figure 2.

Table 2. Detection performance for all classes

Model	mAP@50	mAP@50–95
YOLOv8n	0.624	0.469
YOLOv8s	0.942	0.752
YOLOv8m	0.957	0.793
YOLOv8l	0.957	0.802
YOLOv8x	0.961	0.781
YOLOv11n	0.629	0.473
YOLOv11s	0.947	0.767
YOLOv11m	0.966	0.791
YOLOv11l	0.972	0.815
YOLOv11x	0.945	0.787

The results in Table 2 indicate an overall performance improvement with increasing model capacity up to a certain scale, particularly under the mAP@50–95 metric, suggesting enhanced robustness for multi-scale and densely distributed objects. This trend is especially relevant for architectural drawings, where object sizes and spatial distributions vary significantly within a single layout.

Compared with mAP@50, the mAP@50–95 metric provides a stricter evaluation of localization robustness, which is particularly relevant for architectural floor plans with dense and closely spaced elements. The observed performance gains of mid- to large-scale models suggest an enhanced ability to capture multi-scale contextual cues, albeit with increased computational cost, indicating that model selection should balance accuracy and practical deployment constraints.



Figure 2. Qualitative detection results

Figure 2 illustrates representative detection results under different drawing conditions, including dense annotation regions, overlapping symbols, and varying line thicknesses. While most objects are correctly localized, failure cases are mainly observed in regions with rotated text and severe overlap.

The results indicate that the tile-based detection strategy combined with global non-maximum NMS effectively improves detection coverage for large-format drawings. As summarized in Table 2 and illustrated in Figure 2, both textual regions and graphical symbols can be detected with acceptable accuracy across diverse drawing layouts. Nevertheless, several challenges remain, particularly in handling complex layouts and non-standard text orientations.

The OCR results demonstrate the feasibility of extracting textual content from detected regions using an off-the-shelf OCR model, although recognition accuracy varies with font style, text orientation, and background complexity. In practice, the OCR module performs reliably for short, horizontally aligned annotations with relatively clean backgrounds, while errors are more frequently observed for rotated or densely packed text. These observations suggest that OCR reliability is strongly influenced by detection quality, highlighting the importance of tighter integration between detection and recognition modules. Representative examples of Chinese text recognition results, illustrating typical successful recognition and failure cases, are shown in Figure 3.

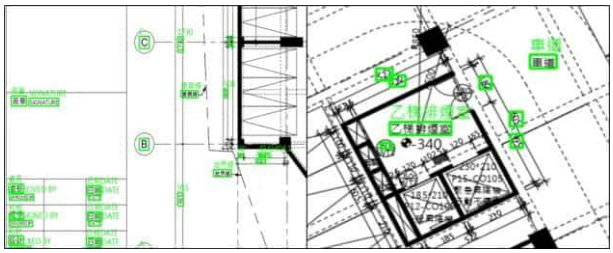


Figure 3. Examples of Traditional Chinese text recognition results on floor plans

To provide a more in-depth evaluation of the proposed framework, two supplementary analyses were conducted. First, an ablation study was performed to examine the effect of the tile-based strategy and overlap settings on object detection performance. Second, a quantitative OCR comparison was conducted to assess how different text extraction workflows influence the number of recognized text regions and the average OCR confidence. The corresponding results are summarized in Table 3 and Table 4, respectively.

Table 3. Ablation study on the effect of tiling and overlap ratio for YOLOv11l

Method	mAP@50	mAP@50–95
YOLOv11l without tiling	0.045	0.013
YOLOv11l tiling (no overlap)	0.882	0.699
YOLOv11l tiling (10% overlap)	0.972	0.815
YOLOv11l tiling (15% overlap)	0.992	0.858

As shown in Table 3, the tile-based strategy substantially improves detection performance compared with direct inference on the full-size drawing. Without tiling, YOLOv11l achieves only 0.045 mAP@50 and 0.013 mAP@50–95, indicating that large-format floor plans cannot be reliably processed in a single pass under the current model setting. Introducing tiling without overlap already leads to a significant improvement, while adding overlap further enhances detection robustness. Among the tested settings, a 15% overlap achieves the best performance, suggesting that overlap helps preserve contextual information near tile boundaries and reduces missed detections for objects located across adjacent patches.

Table 4. Quantitative comparison of text extraction workflows for OCR pre-processing

Method	Direct OCR on full drawings	OCR on tiled drawings	YOLO-based text detection + OCR
Number of detected boxes	153/241	173/241	233/241
Average OCR confidence	0.785	0.835	0.823

Table 4 presents a quantitative comparison of OCR performance under three text extraction workflows. Direct OCR applied to full drawings detects 153 out of 241 text instances, whereas OCR on tiled drawings improves the number of detected text regions to 173. The proposed YOLO-based text detection followed by OCR achieves the highest text coverage, with 233 out of 241 text regions successfully identified. Although the average OCR confidence of the tiled OCR approach is slightly higher than that of the YOLO-based pipeline, the latter provides substantially better text localization coverage. This result indicates that the main advantage of the proposed framework lies in improving text region recall before recognition, which is particularly important for engineering drawings containing dense and spatially distributed annotations.

6 Discussion

The experimental findings suggest that integrating object detection and OCR provides a viable approach for automated information extraction from engineering drawings. The tile-based strategy effectively addresses input size limitations while preserving detection performance across different drawing scales.

Nevertheless, the results also highlight inherent challenges associated with engineering drawings. Complex layouts, heterogeneous annotation styles, and domain-specific terminology pose difficulties for both detection and recognition models. These observations underscore the importance of domain-adaptive training strategies and robust post-processing mechanisms.

At the current stage, the proposed framework is not intended to achieve full automation. Instead, it aims to provide a structured basis upon which higher-level semantic interpretation and intelligent engineering applications can be developed.

From an application perspective, the current framework can serve as a front-end information extraction module for engineering workflows that require rapid access to drawing content. In particular, detected symbols and recognized annotations can be converted into structured records to support space inventory, drawing review, preliminary quantity takeoff, and BIM-assisted data entry. For example, text labels associated with rooms or components can be extracted together with their spatial locations, allowing practitioners to reduce the manual effort required for locating, transcribing, and organizing drawing information. While the full integration of these downstream applications remains outside the immediate scope of this paper, the current results demonstrate that the proposed pipeline can provide an operational basis for construction-specific decision support and further semantic interpretation. However, these use cases currently remain at the structured information extraction stage and have not yet been validated through deployment in a full production workflow.

7 Future Work

Future research will focus on several key directions:

1. OCR Model Fine-Tuning: Constructing a domain-specific Chinese annotation dataset with diverse fonts, sizes, orientations, and background noise to improve recognition robustness.

2. Hyperparameter Optimization: Systematically analysing the influence of detection-related hyperparameters, including input resolution, overlap ratio, confidence thresholds, and NMS strategies.
3. Extended Ablation and Sensitivity Analysis: Further examining the influence of model architectures, overlap ratios, confidence thresholds, and post-processing settings on detection and recognition performance across more diverse drawing conditions.
4. Semantic Interpretation: Integrating large language models to transform recognized graphical and textual information into structured engineering semantics for downstream applications such as quantity takeoff and cost estimation.

8 Conclusion

This study presents a compact and integrated deep learning framework for the joint recognition of geometric objects and textual information in architectural floor plans. By combining tile-based object detection with OCR-based text recognition, the proposed approach effectively addresses key challenges associated with large-format engineering drawings and heterogeneous annotations. While certain limitations remain, particularly in handling highly complex layouts and domain-specific textual variations, the proposed framework establishes a scalable and extensible foundation for future research on semantic interpretation and intelligent engineering drawing analysis.

Acknowledges

This research was supported by the National Science and Technology Council, Taiwan.

References

[1] F. C. A. Groen, A. C. Sanderson, and J. F. Schlag, "Symbol recognition in electrical diagrams using probabilistic graph matching," *Pattern Recognit. Lett.*, vol. 3, no. 5, pp. 343–350, Sep. 1985, doi: 10.1016/0167-8655(85)90066-2.

[2] C. F. Moreno-García, E. Elyan, and C. Jayne, "New trends on digitisation of complex engineering drawings," *Neural Comput. Appl.*, vol. 31, no. 6, pp. 1695–1712, Jun. 2019, doi: 10.1007/s00521-018-3583-1.

[3] Y. Zhao, X. Deng, and H. Lai, "A Deep Learning-Based Method to Detect Components from Scanned Structural Drawings for Reconstructing 3D Models," *Appl. Sci.*, vol. 10, no. 6, Art. no. 6, Jan. 2020, doi: 10.3390/app10062066.

[4] E. Elyan, L. Jamieson, and A. Ali-Gombe, "Deep learning for symbols detection and classification in engineering drawings," *Neural Netw.*, vol. 129, pp. 91–102, Sep. 2020, doi: 10.1016/j.neunet.2020.05.025.

[5] A. Rezvanifar, M. Cote, and A. B. Albu, "Symbol Spotting on Digital Architectural Floor Plans Using a Deep Learning-based Framework," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 2419–2428. doi: 10.1109/CVPRW50498.2020.00292.

[6] P. Schönfelder, F. Stebel, N. Andreou, and M. König, "Deep learning-based text detection and recognition on architectural floor plans," *Autom. Constr.*, vol. 157, p. 105156, Jan. 2024, doi: 10.1016/j.autcon.2023.105156.

[7] M. Gupta, C. Wei, and T. Czerniawski, "Semi-supervised symbol detection for piping and instrumentation drawings," *Autom. Constr.*, vol. 159, p. 105260, Mar. 2024, doi: 10.1016/j.autcon.2023.105260.

[8] P. N. Pizarro, N. Hitschfeld, I. Sipiran, and J. M. Saavedra, "Automatic floor plan analysis and recognition," *Autom. Constr.*, vol. 140, p. 104348, Aug. 2022, doi: 10.1016/j.autcon.2022.104348.

[9] Z. Xu, N. Jha, S. Mehadi, and M. Mandal, "Automatic floor plan analysis: Datasets, methods, and applications (2000–2025)," *Autom. Constr.*, vol. 178, p. 106378, Oct. 2025, doi: 10.1016/j.autcon.2025.106378.

[10] L. Jamieson, C. F. Moreno-Garcia, and E. Elyan, "Deep Learning for Text Detection and Recognition in Complex Engineering Diagrams," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–7. doi: 10.1109/IJCNN48605.2020.9207127.

[11] M. T. Khan, Z. Yong, L. Chen, J. M. Tan, W. Feng, and S. K. Moon, "Automated Parsing of Engineering Drawings for Structured Information Extraction Using a Fine-tuned Document Understanding Transformer," Sep. 03, 2025, *arXiv*: arXiv:2505.01530. doi: 10.48550/arXiv.2505.01530.

[12] Z. Zhang *et al.*, "From Chaotic OCR Words to Coherent Document: A Fine-to-Coarse Zoom-Out Network for Complex-Layout Document Image Translation," in *Proceedings of the 31st International Conference on Computational Linguistics*, O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert, Eds., Abu Dhabi, UAE: Association for Computational Linguistics, Jan. 2025, pp. 10877–10890. Accessed: Dec. 24, 2025. [Online].

Available: https://aclanthology.org/2025.coling-main.723/

Neural Implicit Representations for IFC Objects: A Coordinate-Preserving Approach for Volumetric Spatial Reasoning

Sang Du[1,2], Lei Hou[1,2] , Guomin(Kevin) Zhang[1,2] and Haosen Chen[1,2]

[1]School of Engineering, RMIT University, Melbourne, 3000, Victoria, Australia
[2]Centre for Future Construction, RMIT University, Melbourne, 3000, Victoria, Australia
sang.du@student.rmit.edu.au, lei.hou@rmit.edu.au, kevin.zhang@rmit.edu.au, haosen.chen@rmit.edu.au.

Abstract –

Recent advances in three-dimensional (3D) generative Artificial Intelligence (AI) hold significant potential for automating architectural design. However, existing methods cannot effectively process Industry Foundation Classes (IFC) data due to its complex structure. There is a pressing need for a unified representation that preserves the object-level spatial information for generative tasks. This paper proposes an auto-decoder approach that converts IFC object geometry into a unified, generation-ready representation, enabling volumetric spatial reasoning. First, an IFC export strategy that retains each object in its original global coordinates frame is introduced. Second, a sampling method that pairs 3D points with their distance to the nearest surface is applied. Third, a neural network that jointly optimises per-object representation vectors and the decoder is presented, yielding unified representations for IFC objects. Finally, an octree-based decoding method that enables efficient geometry reconstruction is employed. Comparative experiments demonstrate that the proposed method significantly outperforms voxel-based baselines in reconstruction fidelity. Furthermore, these unified latent vectors successfully enable downstream AI models to utilise IFC spatial information for complex volumetric reasoning , achieving an overall accuracy of 96.8% in classifying multi-constraint spatial interactions. This is validated by evaluating the classification of multi-constraint spatial interactions between slab openings and door clearances, which achieved an overall accuracy of 96.8%. This method converts complex geometry data into a unified representation, thereby enabling the use of IFC in generative AI.

Keywords –

BIM, IFC, 3D geometry representation, signed distance function, generative AI

1 Introduction

Three-dimensional (3D) generative Artificial Intelligence (AI) has rapidly advanced in recent years, demonstrating a remarkable ability to learn from existing data and synthesise novel 3D models [1], [2]. Particularly, in digital entertainment, it has been applied to learn and create coherent virtual worlds that reflect real-world plausibility [3], [4]. This ability to learn spatial constraints and patterns is particularly valuable for the built environment. For example, architectural designs require content to be spatially coherent [5]. 3D generative AI, therefore, holds significant potential for automating aspects of architectural design. For instance, current generative studies in architectural design still focus on generating floor plans or façade images [6], [7].

However, establishing such an AI system requires access to high-quality architectural design data [8]. The Industry Foundation Classes (IFC) standard is a vendor-neutral, ISO standardised Building Information Modelling (BIM) schema that couples precise building geometry with rich semantic attributes [9]. After two decades of global adoption, extensive design data has accumulated in IFC format [10], [11]. The vast amount of data contained in IFC models has supported a range of AI-based studies, such as intelligent construction management, automated risk assessment, automated compliance checking, and autonomous construction robotics [12], [13]. Despite the promise, current 3D generative AI applications in BIM are rarely based on IFC, thereby overlooking the enriched 3D geometry data it offers. For example, Text2BIM [14] generated 3D models using design software operation commands from a large language model, while FloorDiffusion [15] created floor plans through training on a dedicated 2D residential floor plan dataset. Although IFC natively contains the same geometric and spatial information, both applications bypass the IFC model and instead

utilise other data sources.

The problem stems from the fact that 3D generative AI has difficulty processing IFC's complex geometry data directly [8]. IFC employs multiple paradigms for geometry representation, with each paradigm relying on distinct mathematical formulations and modelling procedures [9]. By contrast, 3D generative AI must satisfy data format requirements from two aspects. First, general AI learning requires a unified, fixed-dimensional data format [16]. Second, generative AI requires a data format that supports interpolation. Interpolation is the process of determining unknown values between two known values [17]. This mismatch constitutes a representation gap that prevents direct use of IFC data in 3D generative AI.

To bridge this gap, existing research in the Architecture, Engineering, Construction (AEC) domain has explored various approaches. Among the possible alternatives, 3D geometry embedding emerges as a promising solution. This method converts 3D geometry into unified, fixed-length vectors for AI use while preserving fine-grained geometric detail. Broadly, such methods can be grouped into explicit and implicit embeddings [18]. Explicit embeddings are effective for discriminative tasks such as classification, but do not emphasise generative aspects, whereas implicit embeddings achieve more substantial generative performance. In this study, we focus on signed distance function (SDF) based implicit embeddings, which describe a shape with points in space and its distance to the shape [19]. This method yields a unified, fixed-length 3D geometry representation that is suitable for generative tasks. However, existing SDF-based methods typically normalise the object's spatial information to facilitate shape learning [20]. In the built environment, maintaining spatial coherence requires careful preservation of fundamental object properties such as shape, size, location, and orientation [5]. Consequently, the loss of spatial information limits the applicability of existing SDF methods to generative tasks, such as automated architectural design.

To bridge the representation gap between IFC data and 3D generative AI, this study presents an auto-decoder approach that converts IFC object geometry into a unified, generation-ready representation while preserved spatial information. Our contributions include:

- Proposing an SDF-based auto-decoder that embeds IFC geometry into a unified representation for AI utilisation, preserving accurate spatial information and thereby enabling volumetric spatial reasoning.
- Developing a complete BIM-to-representation pipeline that automates data preparation, model training, and result visualisation, thereby enabling practical deployment.

2 Literature Review

3D geometry embedding is a method that turns 3D shapes into fixed-length vectors. These vectors allow AI models to process geometric data easily. Researchers group these methods into explicit and implicit categories. Each category has different strengths and weaknesses for building design tasks.

Explicit embedding methods encode the visible parts of an object. One common method is multi-view projection [21], [22]. It turns 3D elements into sets of 2D images. This approach uses simple data formats. However, it discards important volumetric and topological information. Another method uses point clouds [23], [24], [25]. Point cloud encoders capture a shape's global structure. But these encoders are sensitive to noise and uneven sampling. They also lack explicit connections between points. Voxel-grid methods divide 3D space into regular cubes [26], [27]. These grids work well with standard 3D neural networks. But they have low resolution because of high memory costs. Fine geometric details are often lost in voxel grids. Mesh encoders work directly with vertices and faces [28], [29], [30]. They keep high-quality surfaces and topology. However, their performance depends on the quality of the mesh triangulation. Generally, explicit methods are well-suited for tasks such as classification. Their reconstruction quality is usually limited by the resolution of the input data.

Implicit methods represent 3D objects as continuous fields in space. They use neural networks to approximate these fields. This produces compact codes that enable high-quality reconstruction. Signed Distance Function (SDF) methods are a popular choice [20], [31], [32]. They learn the distance from any point in space to the object's surface. This helps the model capture very small variations in shape. Neural Radiance Fields (NeRFs) generate realistic images from different viewpoints [33], [34]. But their geometry is hidden within density values. They do not align well with the specific parts of a building model. Occupancy networks decide if a point is inside or outside an object [35], [36]. They can handle a wide range of shapes and topologies.

A significant problem exists with current implicit methods. Most of these methods use a step called normalisation. They move every object to the centre of a small unit cube. This step helps the AI learn general shapes more easily. But it removes the building's real-world scale and position. Architectural design needs this spatial context. Designers must know the exact location of parts to check joints and assembly rules. Normalised methods cannot distinguish between different interfaces or adjacent components.

To address these critical limitations, it is essential to develop a representation that maintains spatial integrity while providing high-quality visual information to

support human judgment. This study utilises an SDF-based auto-decoder to overcome the deficiencies of existing implicit methods by preserving the original spatial setting and project coordinates of components. In addition to ensuring spatial accuracy, this approach enables high-fidelity 3D reconstruction for direct visual inspection. By delivering both precise coordinates and viewable geometry, the proposed method makes building data significantly more robust and intuitive for AI-driven design tasks.

3 Methodology

The methodology described in the sources presents a comprehensive pipeline for transforming Industry Foundation Classes (IFC) 3D geometric data into AI-compatible vector representations. This approach, implemented as a web-based tool, focuses on capturing the rich spatial and semantic context of architectural components to support downstream AI-driven design tasks. The pipeline consists of four integrated stages as in Figure 1: component extraction, spatial sampling, auto-decoder implementation, and geometric reconstruction. Unlike conventional computer vision methods that often strip away real-world metadata, this framework is specifically designed to operate within the original project coordinate frame, preserving scale, position, and orientation throughout the process.

The decision to bypass traditional unit cube normalisation is driven by the inherent position-dependent nature of architectural design. While standard normalisation facilitates shape learning in general computer vision, it discards the absolute elevation and relative topological relationships, such as the specific clearance between a door and a slab opening, that are fundamental to BIM. In large-scale models, maintaining geometric consistency across heterogeneous components is vital. If global coordinates are lost, downstream tasks like automated clash detection and volumetric spatial reasoning become mathematically impossible. Therefore, preserving the original project coordinate frame is a necessary design choice to ensure that the AI-driven reasoning reflects real-world construction constraints and spatial logic.

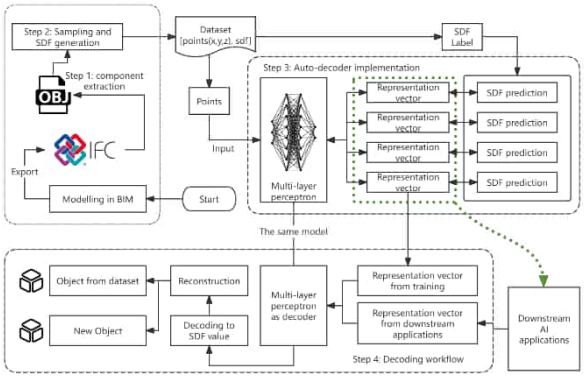


Figure 1. Overall architecture of the proposed method

The first stage, component extraction, involves isolating primary architectural entities—such as walls, columns, beams, slabs, doors, and windows—from the IFC model. Because neural representations based on SDFs require watertight, closed surfaces, the extracted geometry is converted to OBJ meshes and repaired where necessary. To handle the complexity of heterogeneous modelling practices, the workflow employs geometric proxy substitution. For instance, complex multi-part assemblies, such as windows and doors, are replaced with simplified cuboid proxies that preserve their functional volumes and spatial bounds. A critical aspect of this stage is the application of a 0.1 global scale factor, which serves as a linear unit change to ensure that coordinates and distance values remain within a stable numerical range for deep learning optimisation. Most importantly, the original coordinates and orientations are stored alongside these meshes, distinguishing this method from standard pipelines that normalise objects into an abstract unit cube and lose their real-world context.

Following extraction, the sampling and SDF generation phase converts the watertight meshes into training data. This process utilises a two-stage, surface-guided strategy to ensure the neural network learns both the precise shape of the component and the "free space" surrounding it. In the first stage, points are sampled uniformly across the object's surface to ensure total coverage. In the second stage, local jittering is applied around these surface points to create a dense cloud of near-field and far-field samples. Each sampled 3D point is paired with its SDF value, which represents the shortest distance to the component's surface; negative values indicate the interior, and positive values indicate the exterior. By sampling up to a 10-metre radius in real-world coordinates, the model is explicitly supervised to recognise that no geometry exists in the surrounding envelope, effectively preventing the AI from "hallucinating" surfaces in empty regions.

The core of the technical framework is the auto-decoder implementation. Unlike traditional encoder-decoder architectures that require an explicit network to generate features, the auto-decoder directly optimises object-specific representation vectors alongside the neural network weights during training. The architecture is based on a Multilayer Perceptron (MLP) with eight layers and 512 hidden units per layer. The network takes a 128-dimensional shape vector and a 3D spatial coordinate as input and predicts the corresponding SDF value. By sharing the same network weights across a family of similar architectural components, the model is forced to encode geometric variations, such as changes in window size or column thickness, within the latent vector space. This structured latent space enables the model to generalise across unseen architectural designs while maintaining the integrity of the spatial relationships learned from the training data.

The final stage is the reconstruction process, which identifies the zero-level set (the boundary where the SDF value is zero) to recover the explicit 3D surface from the learned vectors. Since searching a large spatial domain for these boundaries is computationally expensive, the methodology employs an adaptive octree refinement strategy. This hierarchical approach begins with a coarse grid and recursively subdivides space into smaller blocks only in regions near the predicted surface, reaching a maximum depth of five levels. Areas identified as "empty" or far from the surface are pruned to save memory and processing time. Finally, the Marching Cubes algorithm is applied to generate a triangulated mesh from the identified zero-crossing points. This bidirectional conversion ensures that the compact neural representations can be translated back into standard 3D formats for visual inspection and integration into conventional architectural modelling workflows. The following experiment validates the method using a common spatial coordination problem.

4 Experiment

4.1 Comparative Analysis with Voxel-based Representation

To justify the selection of the SDF-based auto-decoder, we conducted a comparative experiment against 3D-VAE, a representative voxel-based variational auto-encoder. A subset of 50 wall and slab components from the IFCNet dataset was utilised for evaluation. To ensure a fair comparison of representation capacity, both methods were tested under normalised conditions.

As summarised in Table 1, the proposed implicit representation achieves reconstruction errors approximately one order of magnitude lower than the voxel baseline across Chamfer distance, Hausdorff distance, and surface normal error. The qualitative comparison of these results is illustrated in Figure 2, which depicts representative reconstructions for wall and slab components.

Table 1. Reconstruction errors of the proposed method vs. 3D-VAE voxel baseline.

Scenario	Ours vs GT	3D-VAE vs GT
Chamfer Distance (mm, mean)	2.26	12.45
Hausdorff Distance (mm, mean)	8.17	54.87
Surface Normal Error (deg, mean)	7.65	19.54

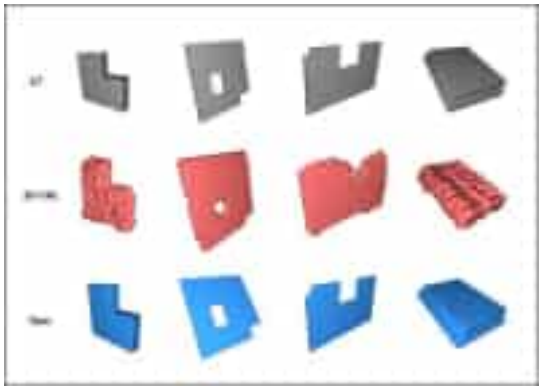


Figure 2. Qualitative comparison of per-object reconstructions. From top to bottom: ground-truth (GT), the 3D-VAE voxel baseline, and ours.

The visual evidence in Figure 2 demonstrates that the implicit representation better preserves sharp profiles and clean surfaces. By contrast, the voxel-based baseline exhibits significant aliasing and boundary artifacts due to its fixed resolution. These results indicate that while voxelization offers faster end-to-end processing, its sensitivity to resolution limits its use for high-fidelity IFC design checking. The proposed implicit representation yields smoother, more faithful surfaces, which is essential for preserving spatial integrity in BIM-to-AI workflows.

4.2 Downstream AI demonstration

To verify that the latent vectors produced by the proposed IFC-to-SDF pipeline can be used directly by downstream AI models, we designed a small demonstration experiment. The objective is to show that a simple neural network can learn implicit design rules solely from latent vectors and perform cross-component,

multi-constraint design checking. The test case is derived from a common architectural design scenario involving the coordination of building services and structural elements. Vertical service penetrations in the floor slab require specific openings, particularly for water pipes, which must be carefully positioned relative to nearby wall features. Potential conflicts arise when door swings overlap with these service openings and access routes. Furthermore, openings placed too close to the base of a wall can violate the minimum spatial clearances required for the structural integrity and precise installation of adjacent components. This scenario represents a typical cross-component, multi-constraint interaction, where traditional rule-based design checks often struggle to account for all possible spatial variations. By contrast, SDF-based latent vectors enable data-driven learning of these complex relationships, as they represent 3D objects as continuous fields that capture both the geometry and the surrounding "free space". This approach allows the neural network to identify spatial constraints without relying on handcrafted rules or explicit geometric coordinates.

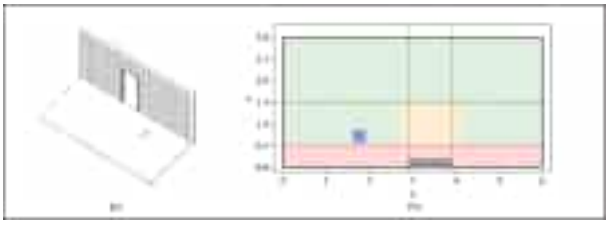


Figure 3. (a) Revit model of the slab–wall–door configuration used in the experiment. A rectangular vertical opening is placed on the slab surface. (b) Corresponding plan-view representation used for SDF-based rule classification. The red region marks the clash zone (objects with $y < 0.5$ m), the yellow region denotes the near-clearance band around the door in the mid-height range, and the green area represents the safe-far zone. The blue rectangle indicates the position of the opening sampled for this instance.

The illustrative geometry, shown in Figure 3(a), consists of a single slab, a single wall, and one door. We define three design rules for any given door position on the wall, as illustrated in Figure 3(b). In the global y direction, we measure the distance from the nearest edge of the floor opening to the wall. When the distance is less than 0.5 m, the configuration is labelled as a spatial clash case and shown in red. When the distance exceeds 1.5 m, the configuration is labelled as safe. For intermediate cases where the y-direction distance lies between 0.5 m and 1.5 m, an additional horizontal condition is applied. We check whether the horizontal projections of the door and opening overlap. We also measure their separation along the x direction. When the projections do not overlap, and their separation exceeds 0.2 m, the configuration is labelled as safe and shown in green. All remaining cases are labelled as attention required cases and shown in yellow. The yellow region indicates cases where the horizontal projections overlap, or the separation is less than or equal to 0.2 m. Such cases require additional design judgment.

The dataset for the experiment is generated procedurally. The door height is fixed at 2.1 m. The door width is sampled uniformly between 0.8 m and 1.2 m, yielding 50 distinct door variants with random positions along the wall. The floor opening size ranges from 0.2 m × 0.2 m to 0.6 m × 0.6 m. We sampled 50 distinct opening variants at random locations on the slab. All 50 door variants are combined with all 50 opening variants, yielding 2,500 unique door and opening configurations. Each configuration is labelled automatically according to the three design rules described above, producing three classes of outcomes: clash, near-clearance, and safe. The labelled dataset is then split into training, validation, and test subsets with a 0.7/0.15/0.15 ratio.

To test whether the latent vectors capture the relevant spatial relationships, we use a simple MLP classifier. The input consists only of the slab's latent codes, including its opening and the door. We adopt a two-tower architecture in which one MLP branch processes the door latent vector, and the other processes the slab latent vector. The outputs of the two branches are concatenated and passed through additional fully connected layers to predict the three design classes. After training, the classifier achieves high accuracy on the held-out test set. Representative quantitative results are reported in Table 2. Qualitative examples of correctly classified configurations are shown in Figure 4.

Table 2. Classification performance of the slab–door rule-learning classifier on the test set.

Label	Total sample	Correct predictions	Accuracy (%)
Safe	273	268	98.2
Near clearance	59	52	88.1
Clash	45	45	100.0
Overall	377	365	96.8

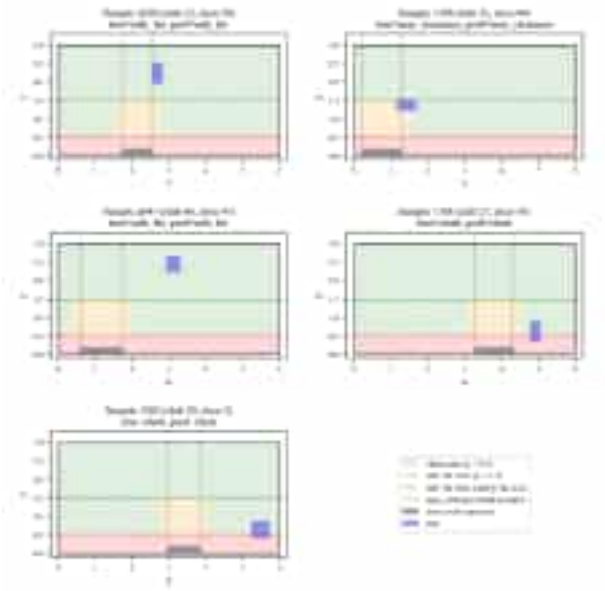


Figure 4. Representative plan-view samples from the slab–door dataset showing the rule-based labelling and MLP predictions. Each subplot depicts a slab with a vertical hole (blue) and a door segment along the wall (grey). The red band marks the clash zone ($y < 0.5$), the light green region marks the safe_far zone ($y > 1.5$), and the mid-height band is split into a near_clearance region (yellow) and a safe_far region when the door is sufficiently far in x. Vertical dashed lines indicate the x-gap thresholds used in the labelling rules. Titles report the sample index and slab/door IDs, together with the ground-truth label and the model prediction, illustrating that the classifier recovers the intended geometric rules for these examples.

The experimental results indicate that a downstream MLP can learn the spatial relationships encoded in the latent vectors. The model discovers the geometric correlation between door and slab opening positions without access to explicit geometric coordinates or handcrafted features. The model generalises well to unseen combinations of door and opening geometries and achieves high classification accuracy for new spatial configurations. In practical applications, similar pipelines can be used when as-built building projects are available as IFC models. The pipeline can encode building components and their penetrations into latent vectors and train AI models that learn from past adjustment decisions made to satisfy actual building construction. Such trained models could then provide data-driven feedback on new building designs, supporting designers with limited construction experience in checking clearances and avoiding cross-component conflicts.

5 Discussion: Pathways to Generative AI

The neural representation proposed in this study establishes a foundational framework that provides a technical bridge to Generative AI within the BIM domain. This application methodology follows a systematic pipeline that transforms irregular architectural data into a machine-learnable format. By mapping discrete IFC entities into a continuous feature space, the framework enables advanced generative architectures to interact effectively with complex structural information.

The first stage of this process involves standardising heterogeneous data through feature encoding to overcome the inherent irregularities of IFC geometries. Our approach utilises an auto-decoder architecture to compress diverse building entities into a series of fixed-length latent vectors. This transformation is essential because it converts complex symbolic data into a uniform format that mainstream machine learning models can process efficiently.

The coordinate-preserving nature of the resulting latent space serves as a critical enabler for various generative tasks. Since spatial positioning and geometric features are embedded directly within the vector distribution, generative models such as Diffusion Models or Generative Adversarial Networks can be trained to recognise underlying spatial assembly rules. These models operate directly within the latent space to predict optimal layouts or suggest design iterations while maintaining strict adherence to the global coordinate system.

Finally, the trained neural decoder functions as a restorative bridge back to the physical domain once a new latent vector is generated or predicted. The decoder interprets the synthesised vector and reconstructs it into a high-fidelity 3D geometric model. Because the latent space is conditioned on precise coordinate data, the reconstructed entities automatically maintain their intended spatial orientations and topological relationships. This closed-loop methodology ensures that AI-generated outputs remain fully compatible with standard BIM workflows and engineering requirements.

6 Conclusion

This study addresses the representation gap between complex IFC geometry and the data needs of AI-driven automated architectural design. The researchers developed an SDF-based auto-decoder that converts building components into compact latent vectors while preserving their original spatial context. Experiments show that the method is accurate enough for automated design coordination. Preserving real-world coordinates is essential, as standard normalisation results in significant accuracy loss. Additionally, the model supports the

generation of design variants through smooth interpolation of size, shape, and position. While the current experiment focuses on rectangular shapes and faces computational challenges for large projects, it successfully bridges the gap between IFC entities and neural representations. By isolating variables in a controlled setting, the study demonstrates the fundamental validity of coordinate preservation. Future research will focus on improving efficiency, conducting more robust evaluations on complex real-world scenarios, and creating new datasets to help AI learn complex assembly rules despite geometric noise.

References

[1] L. Qiu *et al.*, "Richdreamer: A generalizable normal-depth diffusion model for detail richness in text-to-3d," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 9914–9925. doi: 10.1109/CVPR52733.2024.00946.

[2] A. Sengupta, T. Alldieck, N. Kolotouros, E. Corona, A. Zanfir, and C. Sminchisescu, "Diffhuman: Probabilistic photorealistic 3d reconstruction of humans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1439–1449. doi: 10.1109/CVPR52733.2024.00143.

[3] K. Sargent *et al.*, "Zeronvs: Zero-shot 360-degree view synthesis from a single image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 9420–9429. doi: 10.1109/CVPR52733.2024.00900.

[4] Y. Yang *et al.*, "Holodeck: Language guided generation of 3d embodied ai environments," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16227–16237. doi: 10.1109/CVPR52733.2024.01536.

[5] P. Zierke, "The History of spatial coherence," *Zeszyty Naukowe Politechniki Poznańskiej. Architektura, Urbanistyka, Architektura Wnętrz*, 2023, doi: 10.21008/j.2658-2619.2023.17.25.

[6] G. Luo *et al.*, "Controllable and flexible residential floor plan layout design based on multi-agent deep reinforcement learning with layout prior size and similar experience abandon," *Advanced Engineering Informatics*, vol. 68, p. 103702, 2025, doi: 10.1016/j.aei.2025.103702.

[7] M. Deng, V. J. Gan, Y. Tan, A. Joneja, and J. C. Cheng, "Automatic generation of fabrication drawings for façade mullions and transoms through BIM models," *Advanced Engineering Informatics*, vol. 42, p. 100964, 2019, doi: 10.1016/j.aei.2019.100964.

[8] S. Du, L. Hou, G. Zhang, Y. Tan, and P. Mao, "BIM and IFC Data Readiness for AI Integration in the Construction Industry: A Review Approach," *Buildings*, vol. 14, no. 10, p. 3305, 2024, doi: 10.3390/buildings14103305.

[9] BuildingSMART International, "Industry Foundation Classes (IFC4) – ISO 16739-1:2018," International Organization for Standardization, 2018. [Online]. Available: https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD2_TC1/HTML/

[10] Y. Pan and L. Zhang, "Integrating BIM and AI for smart construction management: Current status and future directions," *Archives of Computational Methods in Engineering*, vol. 30, no. 2, pp. 1081–1110, 2023, doi: 10.1007/s11831-022-09830-8.

[11] R. Safour, S. Ahmed, and B. Zaarour, "BIM Adoption around the World," *International Journal of BIM and Engineering Science*, vol. 4, no. 2, pp. 49–63, 2021, doi: 10.54216/IJBES.040202.

[12] G. Chen, A. Alsharef, A. Ovid, A. Albert, and E. Jaselskis, "Meet2Mitigate: An LLM-powered framework for real-time issue identification and mitigation from construction meeting discourse," *Advanced Engineering Informatics*, vol. 64, p. 103068, 2025, doi: 10.1016/j.aei.2024.103068.

[13] D. Chen, L. Chen, Y. Zhang, S. Lin, M. Ye, and S. Sølvsten, "Automated fire risk assessment and mitigation in building blueprints using computer vision and deep generative models," *Advanced Engineering Informatics*, vol. 62, p. 102614, 2024, doi: 10.1016/j.aei.2024.102614.

[14] C. Du, S. Esser, S. Nousias, and A. Borrmann, "Text2BIM: Generating building models using a large language model-based multi-agent framework," *arXiv preprint arXiv:2408.08054*, 2024, doi: 10.48550/arXiv:2408.08054.

[15] J. Shim, J. Moon, H. Kim, and E. Hwang, "FloorDiffusion: Diffusion model-based conditional floorplan image generation method using parameter-efficient fine-tuning and image inpainting," *Journal of Building Engineering*, vol. 95, p. 110320, 2024, doi: 10.1016/j.jobe.2024.110320.

[16] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.

[17] W. H. Press, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, Cambridge, 2007.

[18] O. Poursaeed, M. Fisher, N. Aigerman, and V. G. Kim, "Coupling explicit and implicit surface representations for generative 3d modeling," in *European Conference on Computer Vision*, Springer, 2020, pp. 667–683. doi: 10.1007/978-3-030-58607-2_39.

[19] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, pp. 303–312. doi: 10.1145/237170.237269.

[20] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA: IEEE, Jun. 2019, pp. 165–174. doi: 10.1109/CVPR.2019.00025.

[21] L. Xu, X. Feng, J. Liu, H. Qi, and Y. Wang, "Automatic defect localization method based on BIM projected image," *Advanced Engineering Informatics*, vol. 69, p. 103874, 2025, doi: 10.1016/j.aei.2025.103874.

[22] X. Zhou *et al.*, "BIM product recommendation for intelligent design using style learning," *Journal of Building Engineering*, vol. 73, p. 106701, 2023, doi: 10.1016/j.jobe.2023.106701.

[23] A. Nichol, H. Jun, P. Dhariwal, P. Mishkin, and M. Chen, "Point-e: A system for generating 3d point clouds from complex prompts," *arXiv preprint arXiv:2212.08751*, 2022, doi: 10.48550/arXiv:2212.08751.

[24] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660. doi: 10.1109/CVPR.2017.16.

[25] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 5, pp. 1–12, 2019, doi: 10.1145/3326362.

[26] A. Brock, T. Lim, J. M. Ritchie, and N. Weston, "Generative and discriminative voxel modeling with convolutional neural networks," *arXiv preprint arXiv:1608.04236*, 2016, doi: 10.48550/arXiv.1608.04236.

[27] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Ieee, 2015, pp. 922–928. doi: 10.1109/IROS.2015.7353481.

[28] Y. Siddiqui *et al.*, "Meshgpt: Generating triangle meshes with decoder-only transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 19615–19625. doi: 10.1109/CVPR52733.2024.01855.

[29] R. Hanocka, A. Hertz, N. Fish, R. Giryes, S. Fleishman, and D. Cohen-Or, "Meshcnn: a network with an edge," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–12, 2019, doi: 10.1145/3306346.3322959.

[30] Y. Feng, Y. Feng, H. You, X. Zhao, and Y. Gao, "Meshnet: Mesh neural network for 3d shape representation," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, pp. 8279–8286. doi: 10.1609/aaai.v33i01.33018279.

[31] C. Xu, M. Mielle, A. Laborde, A. Waseem, F. Forest, and O. Fink, "Exploiting semantic scene reconstruction for estimating building envelope characteristics," *Building and Environment*, vol. 275, p. 112731, 2025, doi: https://doi.org/10.1016/j.buildenv.2025.112731.

[32] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *arXiv preprint arXiv:2106.10689*, 2021, doi: 10.48550/arXiv:2106.10689.

[33] M. Tancik *et al.*, "Block-nerf: Scalable large scene neural view synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 8248–8258. doi: 10.1109/CVPR52688.2022.00807.

[34] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021, doi: 10.48550/arXiv.2003.08934.

[35] J. Chibane, T. Alldieck, and G. Pons-Moll, "Implicit functions in feature space for 3d shape reconstruction and completion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6970–6981. doi: 10.1109/CVPR42600.2020.00700.

[36] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4460–4470. doi: 10.1109/CVPR.2019.00459.

Flood Impact Assessment on Highway Networks Considering Socioeconomic Vulnerability: Evidence from Harris County

Beixuan Dong[1], Lingzi Wu[2] and Xinming Li[1, *]

[1]Department of Mechanical Engineering, University of Alberta, Canada
[2]Department of Construction Management, University of Washington, USA
beixuan1@ualberta.ca, lingwu@wu.edu, xinming1@ualberta.ca

Abstract –

Flooding can severely disrupt highway networks, causing substantial mobility loss and delayed recovery. Existing studies that quantify flood impacts on transportation networks primarily focus on structural or functional features, largely overlooking the role of socioeconomic vulnerability. Additionally, how socioeconomic vulnerability influences flood impacts and recovery prioritization across traffic-demand groups remains unknown. This study proposed a unified disaster impact index (DII) that integrates predicted traffic performance with surrounding socioeconomic vulnerability to assess flood impacts on highway networks. A graph neural network (GNN)-based model was employed to predict traffic flow and speed under flood conditions, enabling a performance-based evaluation of network disruption. Using Harris County during Hurricane Harvey as a case study, flood impacts on 18 major highways were quantified and compared across high-traffic-, medium-traffic-, and low-traffic-demand groups. Results showed that incorporating socio-economic features did not alter the identification of the most critical highways, but rather reshaped prioritization within the medium-traffic-demand highway group to emphasize highways that serve socially vulnerable communities.

Keywords –

Flood impacts; socioeconomic vulnerability; data analysis; GNN; highway networks

1 Introduction

Flooding is one of the most disruptive natural hazards affecting transportation systems, leading to extensive mobility loss, prolonged traffic delays, and significant socioeconomic consequences [1, 2]. Highway networks, as the backbone of urban mobility and emergency response, are particularly vulnerable when flood overwhelms surface drainage and inundate critical road segments [3]. During Hurricane Harvey, several major highway closures occurred due to flooding, severely constraining regional accessibility and delaying emergency operations [4]. These challenges underscore the growing need to quantify the impacts of disasters on highway networks and identify the most critical roads for effective restoration.

Existing research has examined the impact of flooding on transportation systems from both network structural and functional features. Studies focusing on network structural features typically employ graph-based metrics, such as network efficiency and centrality, to assess changes in connectivity and accessibility. Functional studies, on the other hand, focus on performance degradation such as reduced traffic flow, increased travel time, and loss of capacity to assess the severity of disruptions. However, most existing studies fail to consider the socioeconomic aspects of surrounding communities, limiting the practicality of disaster impacts on transportation networks' recovery.

Recent research has begun to incorporate socioeconomic vulnerability into disaster planning and assessment, recognizing that communities with higher vulnerability tend to face longer recovery times and weaker resilience [5]. These studies have assessed disaster impacts on different regions by AHP [6] and prioritized the recovery of small bridge and road networks considering socioeconomic vulnerability [7]. However, there are two main gaps. First, socioeconomic vulnerability is not integrated with actual functional performance, rendering it static and failing to quantify disaster impacts with a dynamic nature. Second, how socioeconomic vulnerability influences flood impacts and recovery prioritization across traffic-demand groups remains unknown.

To address the research gaps, this study quantified flood impacts integrating network structural, functional, and socioeconomic features on highway networks and explored how socioeconomic vulnerability influences the

restoration priorities of different traffic demand highway groups. We developed a disaster impact indicator (DII) that integrates predicted traffic performance from a graph neural network-multi-layer perceptron (GNN-MLP) model with socioeconomic vulnerability of surrounding communities. Using Harris County during Hurricane Harvey as the case study, we quantified flood impacts on 18 highways and examined how integrating social vulnerability can shift recovery priorities across high-traffic-, medium-traffic-, and low-traffic-demand highway groups. The contributions of this study are (1) developing a unified DII that captures network structural, functional, and socioeconomic features to assess flood impacts; (2) investigating how socioeconomic vulnerability influences recovery prioritization across traffic-demand groups.

The remainder of this paper is organized as follows. Section 2 reviews related work on disaster impact assessment for transportation networks. Section 3 introduces the methodology, including the construction of DII. Section 4 presents the case study, results, and discusses key findings. Section 5 concludes with implications for transportation disaster management.

2 Literature Review

This section introduces previous research on disaster impact assessment for transportation networks, focusing on the input features used in these studies (subsection 2.1). We then summarized their input features, strengths, and limitations in Table 1, and finally articulated research gaps and our research objectives (subsection 2.2).

2.1 Disaster Impact Assessment for Transportation Networks

Existing studies on the impacts of disasters on transportation networks have focused on the structural and functional features of these networks. Studies focusing on network structural features typically rely on network topology and connectivity. These studies have been applied to identify urban roads most affected by flash floods [8] and detect system-critical routes under terrorist attacks and volcanic eruptions [9]. However, these methods are simple and intuitive as they directly link the intensity of physical hazard to network vulnerability. Furthermore, they tend to have low fidelity as they ignore dynamic traffic conditions and fail to reflect the broader impact on surrounding communities.

Studies focusing on functional features typically use traffic speed [10], and traffic flow [11, 12] to quantify functional degradation of transportation networks under disaster conditions. These methods provide more quantitative and operationally meaningful estimates of network performance loss. However, they largely overlook the social aspects of transportation systems, treating road segments as isolated infrastructure elements without considering the vulnerable communities they serve.

A small number of studies have attempted to integrate network structural and functional features. They highlight network vulnerabilities across different hazard types and provide a more comprehensive view of operational disruptions [12]. Despite this progress, they still fail to incorporate the influence of socio-economic features from surrounding communities.

More recently, several studies have incorporated socioeconomic features into disaster assessment frameworks. One work has proposed a decision-making framework for transportation recovery planning that recognizes the importance of socioeconomic features [5]. Other studies have developed decision-support tools, such as an AHP-based framework [6], to assess regional earthquake risks considering socio-economic features. However, these studies fail to include real-time assessment, focusing on a macro-level regional risk assessment rather than the road-level assessment. Therefore, Merschman et al. [7] proposed a post-disaster bridge repair framework that incorporates social vulnerability to identify the most resilient recovery strategy. Although they have assessed the resilience for each bridge repair sequence, they did not explore how socio-economic features affect the impact of floods on roads with different travel demands.

Table 1 Representative studies on disaster impact assessments for transportation networks

Ref.	Input features	Strength	Limitation
[8]	Network structural features	Simple, intuitive; directly relates hazard intensity to risk	Low-fidelity; ignores dynamic traffic and impacts to surrounding communities
[10]	Functional features	Provides quantitative estimates of functional performance	Ignores impacts to surrounding communities
[12]	Network structural and functional features	Highlights structural vulnerabilities, applicable to various hazards	Ignores impacts on surrounding communities
[5]	Network structural, functional, and socio-economic features	Provide a decision-making framework for disaster recovery planning	There is no quantitative analysis of the influence of socio-economic

			features
[6]	Physical and socio-economic features	Provide a decision support tool for disaster assessing earthquake risk in different regions	There was no real-time assessment, nor was there an assessment at the road segment level
[7]	Network structural, functional, and socio-economic features	Propose a post-disaster bridge repair planning framework considering social vulnerablity	The selection process for the 8 social vulnerability factors used to construct SVI was subjective.
Our study	**Network structural features, functional features, and socio-economic features**	**Propose a unified disaster impact indicator that quantifies flood impacts**	**Since CDC statistics are collected once a year, social features are not dynamic**

2.2 Research Gaps and Objectives

In summary, there are two main research gaps. First, existing studies lack quantitative and real-time assessments of disaster impacts that integrate socio-economic and dynamic traffic performance. This limitation reduces the reliability of these methods for time-sensitive decision-making during emergency response and recovery planning. Second, although the importance of socioeconomic vulnerability has been increasingly recognized, its role in influencing flood impacts on transportation networks with different travel demand levels remains insufficiently explored. As a result, recovery strategies derived from existing methods may unintentionally fail to prioritize roads that are critical for socially vulnerable communities.

To fill these gaps, this study developed a unified indicator that integrates predicted traffic performance with surrounding socioeconomic vulnerability to assess the flood impacts and support recovery planning. Specifically, we developed a DII that integrates predicted traffic performance from a GNN-MLP model with socioeconomic vulnerability of surrounding communities to quantified flood impacts on 18 highways and examined how integrating social vulnerability can shift recovery priorities across highway groups with different travel demand.

3 Methodology

The workflow of the methodology was structured in three parts as shown in Figure 1: (1) data collection (subsection 3.1); (2) GNN-based model for traffic prediction (subsection 3.2); and (3) dynamic impact indicator for flood assessments (subsection 3.3).

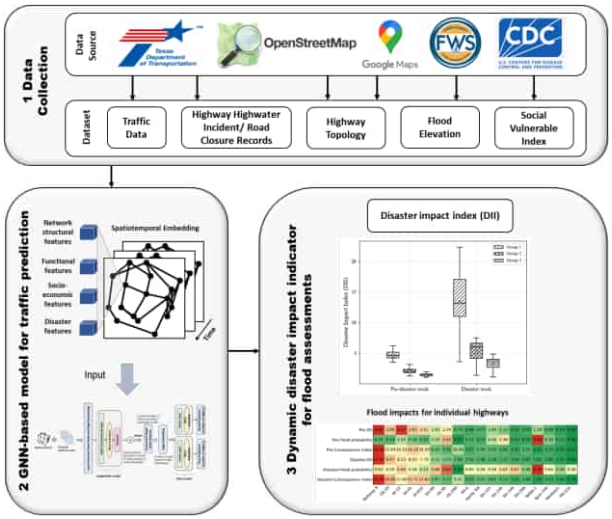


Figure 1. Workflow for flood impact assessment of highway networks

3.1 Data Collection

We collected four types of data: (1) network structural features, (2) functional features, (3) socio-economic features, and (4) disaster features. Table 2 introduces these data types, variables, and sources. flooding condition. Subsections 3.1.1-3.1.4 describe in detail the collection and processing of these four types of features. Section 3.1.5 introduces the road closures and network structure updates under the influence of floods.

Table 2 Input data types, variables, and sources

Date type	Variables	Data sources
Network structural features	Node degree centrality, node betweenness centrality, edge betweenness centrality	Texas Highway Freight Network [13]
Functional features	Traffic flow, traffic speed	Texas Department of Transportation [14], Houston TranStar [15]
Socio-economic features	Social vulnerability index	Centers for Disease Control and Prevention [16, 17]
Disaster features	Flood elevation,	Harris County Flood Warning System [18],

	flood probability	Texas Department of Transportation [14]

3.1.1 Network Structural Features

Network structural features capture the topological characteristics of the highway network, including node degree centrality, node betweenness centrality, and edge betweenness centrality. We collected the network structural features from the Texas Department of Transportation (TxDOT) and transformed them into a CSV format by manually identifying the node coordinates using Google Maps. We then constructed a directed graph $G = (V, E)$ with 546 nodes and 557 edges to encode the spatial topology of the highway network for subsequent spatiotemporal learning using GNN. Tensors $T_s \in \mathbb{R}^{J \times M_s}$ and $T_{ns} \in \mathbb{R}^{J \times N_s}$ represent the network structural features on the edges and nodes, respectively.

3.1.2 Functional Features

Functional features capture the functional performance of each highway segment, including traffic flow and speed. We collected the functional features from the TxDOT. Traffic flow data were later aggregated to characterize highway-level traffic demand and to support the classification of highways into different demand groups in the case study analysis. A speed value of zero can occur under two conditions: (1) a true zero, where a highway segment is closed, and no vehicle can pass; and (2) a false zero, where a highway segment is open, yet no vehicle travel through the segment during the recorded time. To reflect the actual condition, we kept the true zeros and used third-party speed data from Houston TranStar to supplement the false zeros. Tensor $T_f \in \mathbb{R}^{J \times M_f}$ represent the functional features on the edges.

3.1.3 Socio-economic Features

We used the social vulnerability index (SVI) to represent socio-economic features. The SVI data were collected from the Centers for Disease Control and Prevention (CDC) [17]. The SVI is the overall percentile rankings of 15 variables: below poverty, unemployed, income, no high school diploma, aged 65 or older, aged 17 or younger, civilian with a disability, single-parent households, minority, speak English "less than well," multi-unit structures, mobile homes, crowding, no vehicle, and group quarters. To integrate these data into the network, census-tract-level SVI values were mapped to graph nodes by assigning each node the SVI of the census tract in which it is located. This enables the representation of socio-economic vulnerability at the node level through the tensor $T_{sv} \in \mathbb{R}^{I \times N_{sv}}$.

3.1.4 Disaster Features

Disaster features used in this study were collected from the Harris County Warning System, including flood elevation and flood probability. The flood elevation is the water-surface elevation. Flood probability is an elevation threshold—when the water level reaches this height, it may cause flooding of surrounding roads. To map the disaster features on each highway segment, we first discretized the highway network into grids. Then, we mapped the flood elevation and flood probability from each monitoring station into these grids. Each highway segment was assigned the average value of the grids that it intersected. Tensors $T_d \in \mathbb{R}^{J \times M_d}$ represent the disaster features on the edges.

3.1.5 Road Closures and Network Structure Updates

We used the high-water incidents and road closures data provided by TxDOT to update the road network structure due to flooding. Each record contains the coordinates, detection time, and cleared time for each event. There are two flooding conditions of each grid: (1) flooded, with one or more flood or closure events; (2) unaffected, with no flood or closure event. Highway segments that intersected one or more flooding grids were marked as inundated. Accordingly, we updated the network structure based on the inundated segments.

3.2 GNN-Based Model for Traffic Prediction

We used a GNN model for generating node representations and a multi-layer perceptron (MLP) model for traffic prediction. We first conducted a rigorous hyperparameter selection to optimize six GNN-based variants and identify the most effective configuration based on root mean square error (RMSE), including GCN, STGCT, DCRNN, GAT, ChebNet and GraphSAGE. Then, we selected the GraphSAGE-MLP architecture for our case study, as shown in Figure 2.

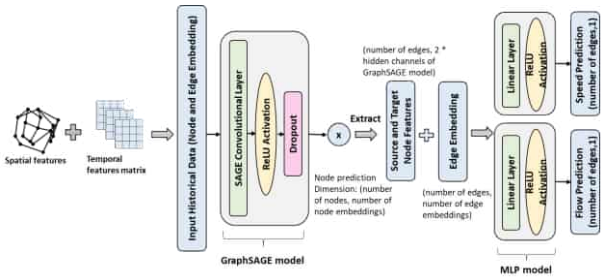


Figure 2 The architecture of the GraphSAGE-MLP case study model

Historical data, including node features (e.g., structural and social features) and edge features (e.g., disaster and functional features), is pre-processed as a

spatiotemporal matrix and used as model input. First, node features are iteratively updated through neighbourhood aggregation in SAGE convolutional layers, followed by nonlinear regularization using the Corrected Linear Unit (ReLU) activation function and dropout to obtain node prediction results. Next, we extract the embedding vectors of source and target nodes and concatenate them with edge-specific features to generate comprehensive edge embedding vectors. These embedding vectors are then fed into an MLP model, where independent linear layers predict traffic speed and flow, respectively.

3.3 Dynamic Disaster Impact Indicator for Flood Assessment

We proposed a disaster impact indicator (DII) to assess the impact of floods, which can identify the highway segments most severely affected by floods. The concept of DII was proposed by Zhang et al. [19], referring to the probability of disaster occurrence and the consequences of segment disruption. To incorporate the real functional performance and the social influence from the surrounding vulnerable communities, we extended Zhang et al.'s definition by integrating functional and socio-economic features. The DII for a highway segment e is formulated as:

$$DII(e) = DP(e) \times CI(e) \quad (1)$$

where $DP(e)$ represents the probability of a highway segment e is flooded, derived from the disaster tensor $T_d \in R^{J \times M_d}$. $CI(e)$ refers to the consequence index, which measures the degree of disruption caused by highway segment e being flooded. The $CI(e)$ is defined as:

$$CI(e) = \frac{(S_i + S_j)}{2} \times F(e) \quad (2)$$

where S_i and S_j capture the social vulnerability at the starting and ending nodes of highway segment e, respectively. Both values are derived from the socio-economic feature tensor $T_{sv} \in R^{I \times N_{sv}}$. Specifically, S_i and S_j correspond to the SVI of the census tracts where node i and j are located. The functional feature $F(e)$ quantifies function performance road segment e, and is derived from the functional feature tensor $T_f \in R^{J \times M_f}$. The functional feature $F(e)$ is calculated as:

$$F(e) = \alpha F_f(e) + \beta F_s(e) \quad (3)$$

where traffic flow $F_f(e)$ and speed $F_s(e)$ are commonly-used indicators for evaluating functional performance. $F_f(e)$ characterizes the traffic-carrying capability of a highway, while $F_s(e)$ indicates its service quality. We normalized $F_f(e)$ and $F_s(e)$ to the range [0,1] to eliminate scale-related bias. The weighting parameters α and β, each ranging from 0 to 1, specify the relative emphasis placed on flow versus speed and can be adjusted according to analytical needs.

4 Case Study

We applied the proposed DII to assess flood impacts on the Harris County highway network during Hurricane Harvey from August 25 -31, 2017. Figure 4 shows the 18 highways within the study area, each highlighted with a distinct colour. To facilitate comparative analysis, these highways were grouped by their traffic demand calculated as the total traffic flows during the study period.

- Group 1 (high-traffic-demand) highways: IH-10, Beltway 8, US-59, IH-45, IH-610, and US-290.
- Group 2 (medium-traffic-demand) highways: SH-99, SH-225, Hardy Toll Road, SH-146, SH-288, and SH-249.
- Group 3 (low-traffic-demand) highways: Spur-330, SH-6, US-90, NASA-1, Westpark Tollway, and FM-529.



Figure 4 18 highways within the study area

4.1 Data Collection

Network structural, functional, socio-economic, and disaster features were collected and processed as introduced in Section 3.1. The dataset spans the period from August 18 to 31, 2017, with a temporal resolution of 15 minutes. Specifically, data from the pre-disaster period (August 18–24) and the initial phase of the disaster (August 25–27) were used as inputs to predict traffic conditions and flood impacts during the subsequent four days (August 28–31). In addition, high-water incidents and road closures were used to update the highway network structure during the flood.

4.2 GNN-Based Model for Traffic Prediction

We used the GraphSAEG-MLP model introduced in

Section 3.2 to predict the traffic flow and traffic speed. The model was trained using historical data from Aug. 18–27 to predict traffic conditions during Aug. 28–31, with approximately 71% of the data used for training and 29% for testing. The model was optimized using the Adam optimizer with a learning rate of 0.0001. Hyperparameter tuning was conducted for key model components, including the number of epochs, GraphSAGE layers, hidden channels, and dropout rate. The final model configuration consisted of 25 training epochs, 1 GraphSAGE layer, 16 hidden channels, and a dropout rate of 0.3.

The RMSE values of traffic prediction for each model were presented by a pair of box plots, where speed was represented in blue and flow was represented in red (Figure 3). The lines inside the box plots mark the median RMSE, and the green triangles represent the mean values. The y-axis was capped at 1.7 to enhance readability by preventing GraphSAGE-LSTM-MLP's outliers (exceeding 2.4) from distorting the scale. We found that GraphSAGE-MLP exhibited the lowest RMSE values for both traffic speed and flow with a compact interquartile range, indicating both high accuracy and low variance.

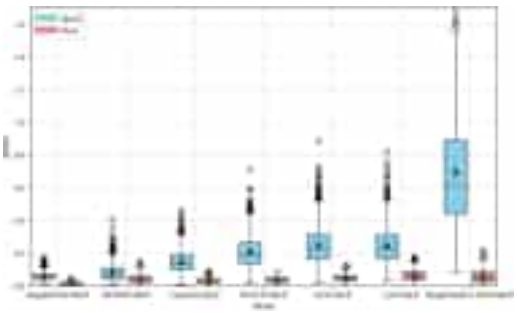


Figure 3 RMSE distribution for predicted speed and flow across models

4.3 Dynamic Disaster Impact Indicator for Flood Assessment

We calculated the DII for the pre-disaster week (August 18–24) using observed traffic data as a baseline. During the disaster week (August 25–31), observed data were used for the early disaster stages (August 25–27), while GNN-predicted traffic data were incorporated to assess flood impacts for subsequent periods (August 28–31).

First, we compared the DII across the three highway groups to evaluate how disaster impacts vary with traffic-demand levels. Second, we examined individual highways by their flood probability, consequence index, and DII to identify highways with high flooding probability, high disruption consequence, or both. Third, we explored how socioeconomic vulnerability affects flood impacts for highways with different traffic demand levels.

Figure 4 shows the boxplots of the DII across the three highway demand groups. Compared with the pre-disaster week, all groups experienced an expansion of the interquartile range (IQR) during the disaster week, indicating greater variability in impact severity among individual highway segments within the same group. Group 1 (high-traffic-demand) highways exhibited the largest DII values during the disaster, with a mean increase of 186.45%, indicating that suffered the most severe disruptions. This pattern aligns with most high-water events observed along these highways, highlighting the direct impact of flooding on them. In contrast, Group 3 (low-traffic-demand) highways showed the lowest DII values and the smallest relative increase from pre-disaster week (115.87%), implying relatively limited direct exposure to flooding. Examination of high-water records further indicates that the observed impacts on these highways were largely indirect, driven by suppressed travel demand in adjacent areas and reduced upstream network connectivity.

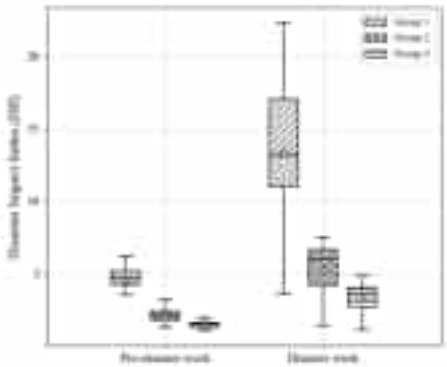


Figure 4 DII for pre-disaster week and disaster week by highway groups

We visualized the average flood probability, consequence index, and DII values of each highway during the pre-disaster and disaster weeks as a heatmap (Figure 5) to further examine the varying impacts of flood. These three indicators respectively capture a highway's exposure to flooding, its potential to disrupt network-wide mobility, and its overall disaster impact. As shown in Figure 5, highways are ranked in descending order of their DII values during the disaster week. Each row is colour-scaled independently—red indicating higher values and green indicating lower—to facilitate visual comparison.

The results revealed that highways such as Beltway 8, IH-10, and IH-45 already had high DIIs before the disaster, suggesting they are particularly vulnerable to potential impacts even before the flood. This is due to their locations in flood-prone (Beltway 8) or low-lying areas (IH-45) and their critical roles in regional traffic

(IH-10). This result highlights the need for flood prevention infrastructure and pre-disaster traffic management plans for these highways. Beltway 8, US-59, and IH-10 exhibited the highest DIIs during the disaster, indicating that these highways experience the most significant impacts during the flood. This underscores the need to prioritize these highways in disaster response and recovery, including timely road clearance and emergency access restoration to minimize cascading network disruptions.

Interestingly, despite having the lowest flood probability during the disaster (only 0.3), US-290 showed relatively high DII and consequence index values. This is likely because US-290's strategic location as a connector between high-traffic regions makes it vulnerable to network-level disruptions, even under minor local flooding. This result highlights that flood probability alone does not determine impact severity. Therefore, disaster preparedness and resilience planning should place greater emphasis on the potential consequences of disruptions.

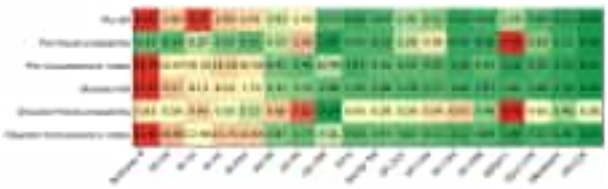


Figure 5 Average flood probability, consequence index, and DII for each highway during pre-disaster and disaster weeks

To quantify the influence of incorporating socio-economic vulnerable features, DII-based highway rankings during the disaster week were examined with and without the inclusion of SVI (Figure 6). The horizontal axis uses distinct font colours to denote highway demand groups, with blue, orange, and green representing Groups 1, 2, and 3, respectively. In both cases, Group 1 (high-traffic-demand) highways consistently had the highest DII, confirming the robustness of the overall priority. However, noticeable ranking shifts were observed among Group 2 (medium-traffic-demand) and Group 3 (low-traffic-demand) highways. Within Group 2, most highways experienced a decline in ranking after incorporating SVI, with only a few exceptions, such as SH-225 and SH-288, showing improvements. In contrast, Group 3 highways exhibited more pronounced heterogeneity. For instance, NASA-1 experienced a substantial drop in ranking, whereas US-90 and SH-6 rose by three positions, indicating that certain low-demand highways gained priority when socio-economic vulnerability was considered.

This pattern suggests that incorporating socio-economic vulnerability does not merely reshape rankings within individual demand groups but instead induces a cross-group reallocation of recovery priority. Specifically, priority shifts away from medium-demand highways toward low-demand highways that serve more socially vulnerable communities. This highlights that equity considerations can elevate the importance of otherwise lower-demand infrastructure when their surrounding populations are more vulnerable. From a broader perspective, these findings indicate that socio-economic features introduce a redistribution mechanism in recovery prioritization, moving beyond efficiency-driven decisions based solely on traffic demand. In practice, this implies that planners should pay attention to low-demand highways that may serve vulnerable communities, as they can become critical candidates for earlier intervention despite their lower functional prominence.

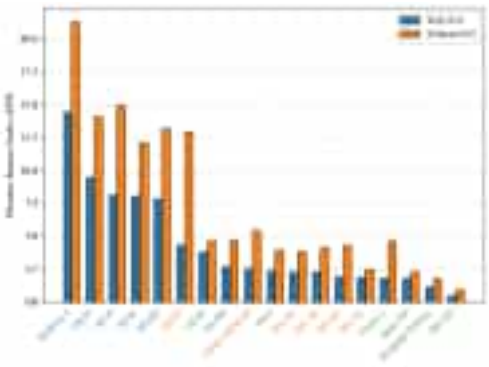


Figure 6 Highway DII rankings with and without SVI during the disaster week

5 Conclusions

This study proposed a DII that integrates predicted traffic performance with surrounding socioeconomic vulnerability to quantitatively assess the impacts of floods on highway networks. We first collected four types of features and then fed them into a GraphSAGE-MLP model for traffic prediction. Subsequently, the DII was calculated using the predicted traffic flow data. Using the flooding caused by Hurricane Harvey in Harris County as a case study, we quantified the impact of the flood on individual highways and demonstrated how socioeconomic vulnerability influences the flood impact on highway groups with different travel demands.

Several key findings emerge from this study. First, high-traffic-demand highways are more severely affected by floods. Second, assessing flood impact requires considering not only flood probability but also the consequences of the damage. Third, incorporating socio-economic features preserved the identification of critical highways while refining the prioritization of medium-traffic-demand highways to emphasize highways that

serve socially vulnerable communities.

The academic contributions of this study are (1) developing a unified DII that captures network structural, functional, and socioeconomic features to assess flood impacts; (2) investigating how socioeconomic vulnerability shapes recovery prioritization across traffic-demand groups. However, this study has some limitations. First, the analysis relies on data spanning only one week before and one week during the flood. Second, the SVI employed in this study is static and does not capture potential temporal variations in social conditions or behavioral responses throughout the disaster period. Future work could incorporate dynamic social indicators (e.g., mobility data) to better represent evolving population behavior and vulnerability during disaster events, thereby improving the realism of recovery prioritization.

References

[1] M. Diakakis, N. Boufidis, J. M. S. Grau, E. Andreadakis, and I. Stamos, "A systematic assessment of the effects of extreme flash floods on transportation infrastructure and circulation: The example of the 2017 Mandra flood," *International journal of disaster risk reduction*, vol. 47, pp. 101542, 2020.

[2] B. Dong, S. Ding, L. Wu, and X. Li, "Short-term natural disaster impacts on transportation infrastructure: a systematic review," *Natural Hazards*, pp. 1-42, 2025.

[3] Y. Chen, Q. Wang, and W. Ji, "Rapid assessment of disaster impacts on highways using social media," *Journal of Management in Engineering*, vol. 36, no. 5, pp. 04020068, 2020.

[4] A. Gori, I. Gidaris, J. R. Elliott, J. Padgett, K. Loughran, P. Bedient, P. Panakkal, and A. Juan, "Accessibility and recovery assessment of Houston's roadway network due to fluvial flooding during Hurricane Harvey," *Natural hazards review*, vol. 21, no. 2, pp. 04020005, 2020.

[5] M. Zamanifar, and T. Hartmann, "Decision attributes for disaster recovery planning of transportation networks; A case study," *Transportation research part D: transport and environment*, vol. 93, pp. 102771, 2021.

[6] V. S. Harandi, K. A. Hosseini, and B. Mansouri, "Assessing the effect of physical and social parameters on seismic risk of urban transportation networks," *International Journal of Disaster Risk Reduction*, vol. 101, pp. 104269, 2024.

[7] E. Merschman, M. Doustmohammadi, A. M. Salman, and M. Anderson, "Post-disaster decision-making framework for roadway networks considering social vulnerability," *Travel Behaviour and Society*, vol. 38, pp. 100910, 2025.

[8] J. Yin, D. Yu, Z. Yin, M. Liu, and Q. He, "Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China," *Journal of hydrology*, vol. 537, pp. 138-145, 2016.

[9] O. Woolley-Meza, D. Grady, C. Thiemann, J. P. Bagrow, and D. Brockmann, "Eyjafjallajökull and 9/11: The Impact of Large-Scale Disasters on Worldwide Mobility," *PLoS ONE*, vol. 8, no. 8, 2013.

[10] M. Pregnolato, A. Ford, S. M. Wilkinson, and R. J. Dawson, "The impact of flooding on road transport: A depth-disruption function," *Transportation research part D: transport and environment*, vol. 55, pp. 67-81, 2017.

[11] H. A. Elfenbein, "7 Emotion in organizations: a review and theoretical integration," *Academy of management annals*, vol. 1, no. 1, pp. 315-386, 2007.

[12] A. Alipour, and B. Shafei, "Assessment of Postearthquake Losses in a Network of Aging Bridges," *Journal of Infrastructure Systems*, vol. 22, no. 2, pp. 12, Jun, 2016.

[13] Texas Department of Transportation. "TxDOT Texas Highway Freight Network," 12/09/2024.

[14] Texas Department of Transportation. 06/21/2024, 2024; https://www.txdot.gov/.

[15] Houston TranStar. 10/16/2024, 2024; https://traffic.houstontranstar.org/speedcharts/.

[16] Centers for Disease Control and Prevention. "CDC's Social Vulnerability Index (SVI) – 2016 overall SVI, census tract level," 08/01/2024, 2024; https://www.arcgis.com/home/item.html?id=62b3e305b730423782c64b9696242c5e#overview.

[17] Centers for Disease Control and Prevention. "CDC Social Vulnerability Index 2018 - USA," 08/01/2024, 2018; https://www.arcgis.com/home/item.html?id=cbd68d9887574a10bc89ea4efe2b8087.

[18] Harris County Flood Warning System. 09/23/2024, 2024; https://www.harriscountyfws.org/.

[19] Q. Zhang, H. Yu, Z. Li, G. Zhang, and D. T. Ma, "Assessing potential likelihood and impacts of landslides on transportation network vulnerability," *Transportation research part D: transport and environment*, vol. 82, pp. 102304, 2020.

BIM-Align: An Automatic Framework for BIM to As-Built Alignment with Multi-View Imagery and 3D Gaussian Splatting

Guangan Chen, Michiel Vlaminck, Gianni Allebosch, Wilfried Philips, Hiep Luong

IPI-URC, imec, Ghent University, Belgium

{guangan.chen, michiel.vlaminck, gianni.allebosch, wilfried.philips, hiep.luong}@ugent.be

Abstract -

Alignment between Building Information Models (BIM) and as-built representations is essential for automated construction progress monitoring. Existing pipelines that rely on images often require manual selection of correspondences or additional site instrumentation. Recent advances in differentiable scene representations, especially 3D Gaussian Splatting (3DGS), open new possibilities for automatic global BIM to as-built alignment through dense supervision from multi-view images. This paper presents *BIM-Align*, an automatic framework for global BIM to as-built alignment with multi-view imagery and 3DGS. The proposed framework initially estimates camera poses using Structure-from-Motion (SfM) and reconstructs a dense point cloud using multi-view stereo (MVS). Subsequently, the target building is extracted from the dense point cloud using DBSCAN clustering, and an initial BIM to as-built alignment is computed with classical geometric methods. The initial alignment is then refined with a coarse-to-fine strategy, where the coarse and fine stages each perform 3DGS optimization with a separate learnable Sim(3) transformation. Experimental results on construction site images show that the proposed approach achieves a mean reprojection RMSE within 2.13 pixels and a mean 3D alignment error within 9.18 cm of a manual baseline, demonstrating the feasibility of automatic BIM to as-built alignment from images with 3DGS.

Keywords -

3D Gaussian Splatting; Construction Site Monitoring; BIM to As-Built Alignment

1 Introduction

Timely and accurate construction progress monitoring is essential for project control and early detection of cost and schedule deviations. Conventional approaches rely on manual site inspections, photographs, and periodic reports, which require substantial labor and time. These limitations motivate automated methods that continuously capture the as-built state and perform comparisons with the as-planned BIM to quantify progress and identify deviations [1, 2, 3, 4, 5].

A common pipeline for automated construction progress monitoring consists of three stages: (i) data acquisition and as-built point cloud reconstruction, (ii) BIM to as-built alignment, and (iii) progress and deviation analysis. Since progress and deviation metrics are computed by comparing BIM and the as-built point cloud, accurate alignment is key to reliable progress monitoring. In practice, the quality and completeness of the as-built point cloud have a significant impact on alignment performance, motivating the selection of appropriate data acquisition modalities. For instance, terrestrial laser scanning and mobile LiDAR can directly capture dense point clouds with high accuracy and have been widely used in construction monitoring [2, 5, 6]. However, LiDAR-based methods require expensive sensors and specialized operation, which limits their practicality for frequent monitoring. In contrast, multi-view 3D reconstruction based on SfM and MVS recovers 3D geometry from standard RGB imagery captured by conventional cameras, which enables data capture at low cost and high frequency [7]. Early work by Golparvar-Fard *et al.* employed a custom implemented SfM pipeline to reconstruct sparse as-built point clouds from unordered construction site imagery for BIM to as-built alignment [8, 9, 10, 11]. The subsequent availability of robust, off-the-shelf SfM-MVS pipeline tools, such as VisualSFM [12], COLMAP [13], and Agisoft [14], has further lowered the barrier to acquiring as-built point clouds, enabling recent studies to generate dense as-built point clouds with improved geometric completeness for BIM to as-built alignment [2, 3, 15].

Despite the advancements in multi-view 3D reconstruction, accurate BIM to as-built alignment remains a major challenge for automated construction progress monitoring. Existing approaches can be categorized into local and global alignment methods, where local methods align individual building elements independently, whereas global methods estimate a unified transformation for the entire building. For local alignment, Omara *et al.* [16] proposed a rigid alignment strategy that aligns individual BIM elements with corresponding point cloud segments. The authors deployed 12 cameras to capture images of a specific region of a building to reconstruct an as-built point cloud. Subsets of the point cloud corresponding to specific structural components, such as reinforced concrete

columns, are independently aligned with predefined BIM geometries using rigid transformations with fixed scale. However, in complex building scenes with numerous structural elements, local alignment requires a large number of images for individual elements.

For global alignment, Golparvar-Fard *et al.* [8, 9, 10, 11] proposed a BIM to as-built alignment approach based on a sparse point cloud derived from SfM. In their methods, a set of correspondences between the sparse point cloud and the BIM is manually annotated. A Sim(3) transformation is then computed with the correspondences for BIM to as-built alignment. In practice, SfM-derived sparse point clouds often suffer from incomplete scene coverage. As a result, certain regions present in the BIM lack corresponding points in the sparse point cloud, reducing the reliability of manually selected point correspondences for alignment. To address this limitation, Chen *et al.* [3] employed a dense point cloud generated by MVS for global BIM to as-built alignment using manually selected correspondences. However, these point-picking methods require manual correspondence selection, which is time-consuming and may introduce human error.

Other approaches implicitly achieve global BIM to as-built alignment by enforcing metric consistency during photogrammetric reconstruction, rather than performing an explicit geometric alignment step. Mahami *et al.* [15] placed coded photogrammetric targets on a construction site to fix the datum and absolute scale of point clouds derived from MVS. An as-built mesh was then reconstructed in Autodesk Revit [18] from the scaled point cloud, after which building elements were manually traced. Subsequently, the as-built mesh was compared with a BIM for construction progress monitoring. Similarly, Tuttas *et al.* [19] incorporated photogrammetric ground control points into the bundle adjustment [7], thereby generating point clouds directly in the BIM coordinate system. Masood *et al.* [20] leveraged GPS-based georeferencing to obtain global BIM to as-built alignment by referencing both representations to a common Universal Transverse Mercator (UTM) coordinate system. However, while these approaches demonstrate the feasibility of global BIM to as-built alignment under varying levels of manual intervention during data acquisition, achieving global BIM to as-built alignment from multi-view images without manual correspondences, additional site instrumentation, or external georeferencing remains an open challenge in practical construction environments.

Recent advances in differentiable scene representations, particularly 3D Gaussian Splatting (3DGS) [21], provide a new opportunity to address automatic global BIM to as-built alignment by leveraging dense supervision from multi-view images. 3DGS achieves high quality novel view synthesis by optimizing a set of 3D Gaussians from multi-view images and camera poses with a photometric loss. The Gaussian mean positions are typically initialized from a point cloud obtained by SfM. When the Gaussian mean positions are frozen during 3DGS training, initialization misalignment leads to misalignment in rendered views and increased photometric loss. To mitigate this effect, correcting the initial misalignment during 3DGS training reduces the impact of initialization errors on the photometric loss. In the BIM to as-built alignment setting, we assume that the Gaussian mean positions are initialized from a point cloud sampled from the BIM, while the as-built point cloud defines the reference coordinate frame. Under this formulation, BIM to as-built alignment is equivalent to correcting the misalignment between BIM and the as-built point cloud during 3DGS training. Building on this observation, this study investigates the feasibility of using 3DGS for automatic global BIM to as-built alignment. To this end, we propose an automatic framework. The proposed framework first estimates camera poses and reconstructs a dense point cloud using SfM and MVS, respectively. Subsequently, the target building is extracted from the dense point cloud, serving as the as-built point cloud. An initial BIM to as-built alignment is then estimated using classical geometric registration methods. Afterward, the initial alignment is refined by training 3DGS models initialized from the BIM in a coarse-to-fine manner.

Our main contributions are:

- We propose *BIM-Align*, an automatic framework for global BIM to as-built alignment from multi-view imagery that combines geometric initialization with 3DGS-based refinement.
- We introduce a coarse-to-fine 3DGS optimization strategy, in which BIM to as-built alignment is refined through a global learnable Sim(3) transformation with frozen Gaussian mean positions.
- We provide an evaluation on collected aerial images of a construction site using alignment metrics in image and 3D domains, together with a manual baseline based on annotated correspondences.

2 Method

Given a set of N multi-view images $\mathcal{I} = \{I_i\}_{i=1}^{N}$ of a construction site and a BIM mesh B, our objective is to automatically estimate a Sim(3) transformation that aligns the BIM with the as-built scene reconstructed from the images. The proposed framework consists of three stages and is illustrated in Figure 1. Initially, we estimate the camera poses $\Pi = \{\pi_i\}_{i=1}^{N}$ of the images and reconstruct a dense point cloud of the construction site. Subsequently, an as-built point cloud is extracted from the dense point

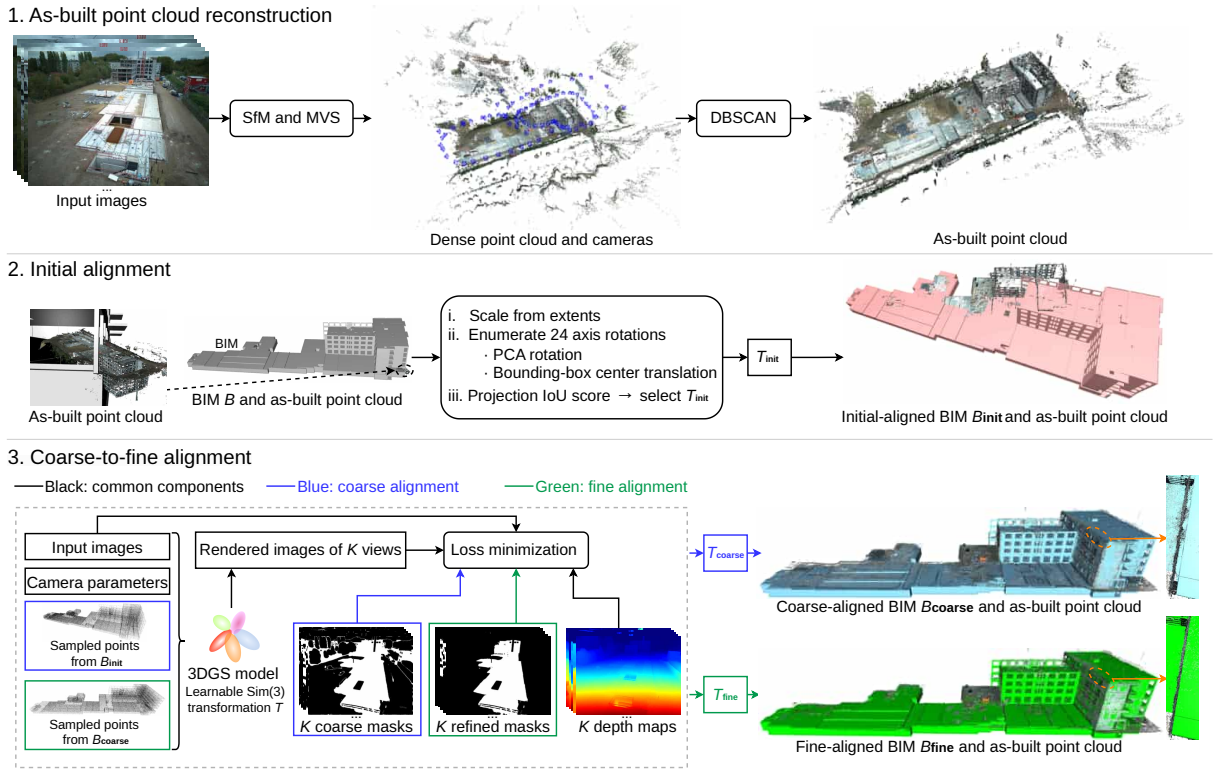


Figure 1. Overview of the proposed *BIM-Align* framework. (1) As-built point cloud reconstruction: SfM and MVS are used to estimate camera poses and reconstruct a dense point cloud, respectively. The as-built point cloud is extracted from the dense point cloud using DBSCAN clustering [17]. (2) Initial alignment: an initial Sim(3) transformation T_{init} is estimated using classical geometric alignment techniques. (3) Coarse-to-fine alignment: starting from the initial alignment, two separate 3DGS optimizations are performed. In the coarse stage, a learnable Sim(3) transform is optimized to estimate T_{coarse}. In the fine stage, the optimization is initialized from the coarse result and a separate Sim(3) transform is optimized to obtain T_{fine}. In addition, depth maps are used during 3DGS training in both phases. Specifically, coarse alignment employs coarse masks, the input point cloud sampled from B_{init}, and $K = 20$ views per iteration, whereas fine alignment employs refined masks, the input point cloud sampled from B_{coarse}, and $K = 1$ view per iteration. For visualization in stages (2) and (3), the as-built point cloud is cropped to retain only the building.

cloud. Next, an initial BIM to as-built alignment is computed using classical geometric alignment techniques, followed by a coarse-to-fine refinement using 3DGS [21].

2.1 Preliminary: 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [21] is an explicit scene representation proposed for novel view synthesis, which models a scene as a collection of anisotropic 3D Gaussians and supports efficient differentiable rendering. Each Gaussian is parameterized as $G = \{\mu, q, s, o, c, sh\}$. The mean position μ is initialized from an input point cloud and serves as the 3D anchor of each Gaussian. The rotation quaternion q defines the Gaussian orientation, s is a scaling vector that controls spatial extent, o represents opacity, c denotes color, and sh represents spherical harmonic coefficients encoding view-dependent appearance. A scene is represented by a set of Gaussians $\mathcal{G} = \{G_j\}_{j=1}^{H}$, where H denotes the number of Gaussians. Given the image set $\mathcal{I}$ and camera poses Π, the 3DGS model renders each view as $\hat{I}_i = \mathcal{R}(\mathcal{G}, \pi_i)$ using a differentiable renderer $\mathcal{R}(\cdot)$. Scene reconstruction is achieved by optimizing the Gaussian parameters by minimizing a photometric loss $\mathcal{L}$, leading to the following objective:

$$\min_{\mathcal{G}} \sum_{i=1}^{N} \mathcal{L}(\hat{I}_i, I_i) \, .$$

In our study, 3DGS is employed as an optimization backbone for global BIM to as-built alignment rather than for novel view synthesis.

2.2 As-built point cloud reconstruction

Given the image set $\mathcal{I}$, we employ a standard SfM-MVS pipeline implemented in COLMAP [13], where SfM is used to estimate the camera poses Π and reconstruct a sparse point cloud, and MVS is subsequently applied to generate a dense point cloud. Camera intrinsic parameters are estimated implicitly through COLMAP's built-in camera models and bundle adjustment. Subsequently, DBSCAN clustering [17] is applied to the dense point cloud to identify the building of interest. As the input images are captured with a focus on the building, the largest cluster is selected as the approximate as-built point cloud.

2.3 BIM to as-built initial alignment

An initial BIM to as-built alignment provides an appropriate initialization for subsequent refinement using 3DGS. Since the BIM and the COLMAP reconstruction differ in scale and coordinate frame, direct 3DGS optimization is unreliable. We therefore estimate an initial Sim(3) transformation T_{init}.

We first estimate the global scale $s_{\text{init}} \in \mathbb{R}^+$ by matching the overall spatial extents of the BIM and the as-built point cloud, and keep the scale fixed when estimating rotation and translation. To account for ambiguous BIM axis conventions [4], we enumerate 24 axis-aligned rotation hypotheses corresponding to the rotational symmetry group of a cube. For each hypothesis, rotation is further estimated by aligning the dominant principal directions of the BIM and the as-built point cloud using PCA [22]. Translation is computed by aligning their bounding-box centers. This yields a set of candidate Sim(3) transformations $\{T_{\text{init}}^k\}_{k=1}^{24}$.

To select the most plausible candidate, we evaluate each T_{init}^k using an overlap score based on orthographic projections. Specifically, we sample points from the transformed BIM mesh, project both the sampled points and the as-built point cloud onto the XY, XZ, and YZ planes, and rasterize the projections into binary footprint images. The consistency score is computed as the sum of the Intersection-over-Union (IoU) values of the three projection pairs, and the candidate with the highest score is selected as the final initial transformation T_{init}. The initially aligned BIM is denoted as $B_{\text{init}} = T_{\text{init}}(B)$.

2.4 BIM to as-built coarse alignment

We refine the initial BIM to as-built alignment with joint optimization of a BIM-initialized 3DGS model and a learnable global Sim(3) transformation T_{coarse}. Specifically, we uniformly sample P points from the initially aligned BIM B_{init} to construct an input point cloud $B_{\text{init}}^{\text{pcd}}$, which anchors the Gaussian mean positions in 3D space. To preserve the BIM-derived Gaussian structure, Gaussian densification and pruning are disabled throughout training. In addition, each sampled point is initialized with default spherical harmonic coefficients, resulting in a constant initial appearance.

The learnable Sim(3) transformation is parameterized as $T_{\text{coarse}} = (s_{\text{coarse}}, R_{\text{coarse}}, \mathbf{t}_{\text{coarse}})$, where $s_{\text{coarse}} \in \mathbb{R}^+$ denotes the scale factor, $R_{\text{coarse}} \in \text{SO}(3)$ is a rotation matrix, and $\mathbf{t}_{\text{coarse}} \in \mathbb{R}^3$ denotes the translation. To preserve the BIM-derived Gaussian structure, the mean position μ of each Gaussian initialized from $B_{\text{init}}^{\text{pcd}}$ is frozen during 3DGS training. The transformed mean position is obtained by applying the transformation T_{coarse} to the mean: $\mu' = s_{\text{coarse}} R_{\text{coarse}} \mu + \mathbf{t}_{\text{coarse}}$. Given the transformed Gaussian mean positions $\{\mu'\}$, the rendered image for view i is given by $\hat{I}_i = \mathcal{R}(\mathcal{G}(\{\mu'\}), \pi_i)$, where $\mathcal{G}(\{\mu'\})$ denotes the Gaussian set with updated mean positions, while all remaining parameters follow the standard 3DGS formulation [21]. Consequently, the BIM to as-built alignment is refined by jointly optimizing T_{coarse} and the remaining Gaussian parameters to enforce photometric consistency between rendered and observed images across all input views:

$$\min_{\mathcal{G},\, T_{\text{coarse}}} \sum_{i=1}^{N} \mathcal{L}(\mathcal{R}(\mathcal{G}(\{\mu'\}), \pi_i), I_i) .$$

To further regularize the optimization, we incorporate depth and mask supervision during 3DGS training. Depth supervision provides additional geometric constraints beyond photometric consistency, which can be ambiguous under weak texture, repetitive patterns, or occlusions. For each image, we predict a monocular inverse depth map using DepthAnything [23] and resolve the scale and shift ambiguity via per-image affine calibration to the SfM-derived point cloud using visible sparse 3D points. Views with abnormal calibration are discarded, and the calibrated inverse depth is used to supervise the rendered inverse depth. In addition, mask supervision suppresses the influence of non-building regions in the photometric loss, which is necessary since $B_{\text{init}}^{\text{pcd}}$ contains only building geometry. To generate masks, we use the prompt-based semantic segmentation model SAM3 [24]. In construction sites, buildings are often incomplete and only partially observed, especially at early stages where only structural slabs may be visible. As a result, extracting complete building masks using generic prompts such as *building* becomes unreliable. We therefore adopt an inverse strategy that segments typical background objects by querying SAM3 with the text prompts *dirt*, *bush*, *tree*, *car*, *sky*, *crane*, *container*, *road*, *fence*, and *metal*. A building mask is obtained by taking the complement of all segmented background regions. As a result, a set of coarse building masks $\mathcal{M}^0 = \{M_i^0\}_{i=1}^N$ for the N input images is created.

The optimization objective for a viewpoint π_i in our

3DGS training is a weighted combination of RGB and depth losses [21]:

$$\mathcal{L}^i = \lambda_{\text{rgb}} \left(\mathcal{L}^i_{\text{rgb}} \odot M^0_i \right) + \lambda_{\text{depth}} \left(\mathcal{L}^i_{\text{depth}} \odot M^0_i \right). \quad (1)$$

Here, the RGB loss $\mathcal{L}^i_{\text{rgb}}$ is a weighted combination of ℓ_1 loss and D-SSIM [25] loss, given by

$$\mathcal{L}^i_{\text{rgb}} = (1 - \lambda_{\text{D-SSIM}})\, \mathcal{L}^i_1 + \lambda_{\text{D-SSIM}}\, \mathcal{L}^i_{\text{D-SSIM}}, \quad (2)$$

where $\mathcal{L}^i_1 = \|\hat{I}_i - I_i\|_1$ and $\mathcal{L}^i_{\text{D-SSIM}} = 1 - \text{SSIM}(\hat{I}_i, I_i)$. The depth loss $\mathcal{L}^i_{\text{depth}}$ is defined as the absolute difference between the rendered and calibrated inverse depth maps. The operator $\odot$ denotes element-wise multiplication. Moreover, the weights λ_{rgb}, λ_{depth}, and $\lambda_{\text{D-SSIM}}$ control the weight of each component.

During coarse alignment, large initial misalignment causes the photometric objective to be dominated by mismatched pixels and incorrect visibilities, resulting in high variance and ambiguous gradients. As a result, single-view supervision is prone to 3DGS training failure. To stabilize the optimization, we compute the mean loss over multiple training views at each iteration. Specifically, we randomly sample $K \geq 1$ distinct indices from the training view index set $\{1, \ldots, N\}$, denoted as $\mathcal{S}$, where $\mathcal{S} \subset \{1, \ldots, N\}$ and $|\mathcal{S}| = K$. Thus, the total loss is given by

$$\mathcal{L}_{\text{all}} = \frac{1}{K} \sum_{i \in \mathcal{S}} \mathcal{L}^i. \quad (3)$$

As a result, the coarse alignment phase yields an optimized transformation T_{coarse} and a coarsely aligned BIM $B_{\text{coarse}} = T_{\text{coarse}}(B_{\text{init}})$.

Despite the segmentation capability of SAM3, the coarse building masks still contain non-building regions, providing insufficient constraints for fine-grained refinement at the coarse phase. We therefore introduce a fine alignment phase.

2.5 BIM to as-built fine alignment

The fine alignment phase leverages refined building masks to further suppress the impact of non-building regions during optimization. Based on the coarse alignment results, a geometric prior for the building region in each image is obtained by back-projecting the coarsely aligned BIM B_{coarse} into each input view, yielding BIM-projected masks $\mathcal{M}^1 = \{M^1_i\}_{i=1}^N$. However, residual misalignment can distort the projected building boundaries, causing boundary erosion and the loss of thin structures in the mask. To address this issue, we refine $\mathcal{M}^1$ using instance masks predicted by SAM3 on the input images. For each input image I_i, let $O_i = \{O^v_i\}_{v=1}^{V_i}$ denote the set of instance masks, where V_i is the number of instances segmented in I_i. We retain an instance mask O^v_i if the fraction of its pixels that lie inside M^1_i exceeds a threshold β. Subsequently, the retained instance masks are merged and combined with the BIM-projected mask M^1_i with set union to produce an intermediate refined mask M^2_i:

$$M^2_i = \left(\bigcup_m O^m_i \right) \cup M^1_i, \quad (4)$$

where O^m_i denotes the m-th retained instance mask in image I_i after overlap filtering. Despite this refinement, M^2_i may still contain clutter tightly adjacent to the building due to residual misalignment. To remove such clutter, we further refine M^2_i by intersecting it with the semantic segmentation mask M^0_i, yielding the final refined mask:

$$M^3_i = M^2_i \cap M^0_i. \quad (5)$$

The set of final refined masks is denoted as $\mathcal{M}^3 = \{M^3_i\}_{i=1}^N$.

In the fine alignment phase, we mostly follow the same optimization procedure as in coarse alignment. In contrast to the coarse phase, we initialize Gaussians using a point cloud sampled from B_{coarse} and compute the loss using the refined masks $\mathcal{M}^3$. Moreover, we calculate the loss using a single randomly sampled view per iteration, with $K = 1$. The fine alignment phase yields a refined Sim(3) transformation T_{fine} and a finely aligned BIM $B_{\text{fine}} = T_{\text{fine}}(B_{\text{coarse}})$.

3 Experiments

3.1 Dataset

We collected a dataset of aerial photographs of a construction site. The dataset contains 87 RGB images (960×720) captured from multiple viewpoints, yielding $N = 87$ input images in our experiments. In addition to the image data, we provide a corresponding as-planned BIM. To ensure consistency between the BIM and the as-built scene for alignment, we incorporate construction progress information and remove all unbuilt elements from the as-planned BIM. The resulting BIM B therefore represents the actual construction state at the time of image acquisition and serves as the BIM input to the subsequent automatic alignment framework in all experiments in our study.

We evaluate BIM to as-built alignment accuracy using reprojection error. For each image, we manually annotate a set of 2D points at distinctive geometric features and select the corresponding 3D reference points from the BIM at the same structural locations to form annotated 2D-3D correspondences. In total, we use 419 correspondence pairs across 87 images for evaluation, with a varying number of correspondences per image.

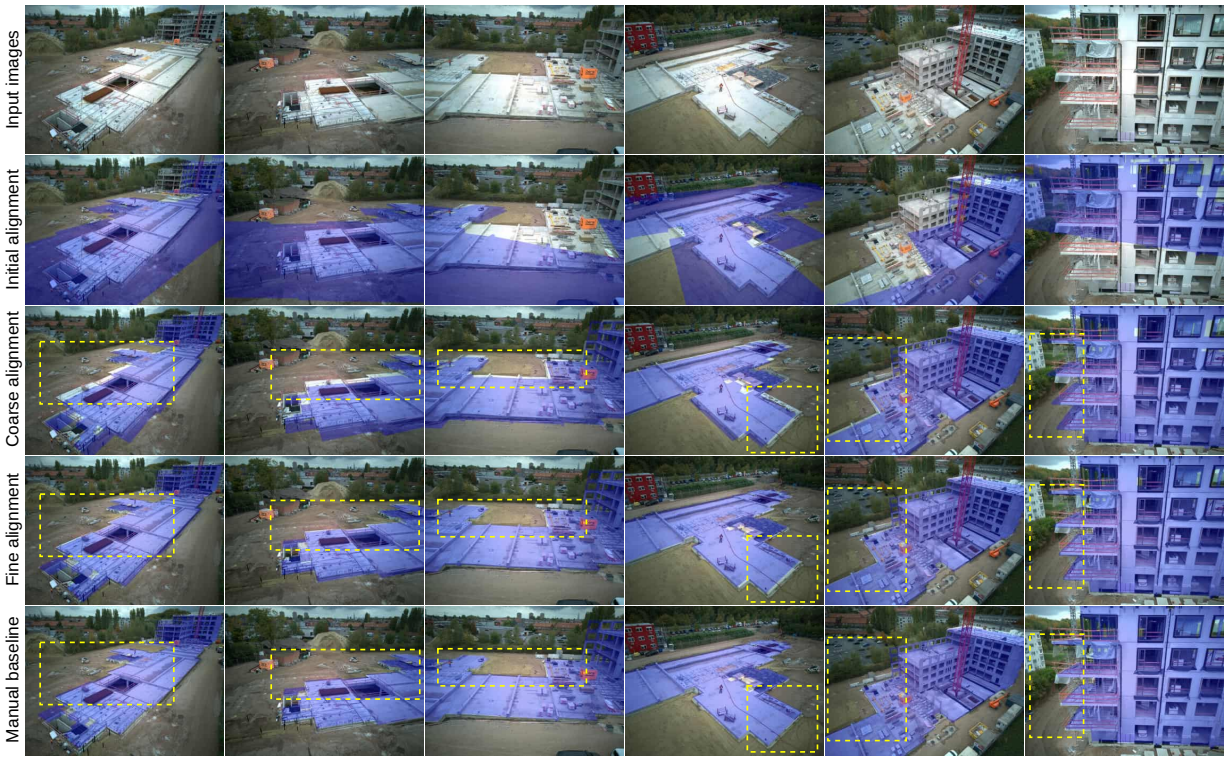


Figure 2. Qualitative comparison of BIM to as-built alignment at different stages. The aligned BIMs are back-projected into input images using the estimated camera poses. The projected BIM geometry is overlaid in blue in the visualized images. From top to bottom, the figure shows input images, initial alignment, coarse alignment, fine alignment, and the manual baseline. For visualization, BIM elements corresponding to an underground parking garage, which are not observable in the input images, are removed prior to projection.

3.2 Metrics

Alignment accuracy is evaluated in both image space and 3D space. Given an annotated 2D point p in image I_i with corresponding 3D point q on the BIM B, let T denote the estimated transformation at a given alignment stage. For the initial, coarse, and fine stages, T is defined as T_{init}, $T_{\text{coarse}}T_{\text{init}}$, and $T_{\text{fine}}T_{\text{coarse}}T_{\text{init}}$, respectively. In image space, q is transformed by T and projected onto I_i using camera pose π_i, yielding a 2D point p'. The 2D alignment error is defined as the Euclidean distance between the annotated point p and the projected point p'. In 3D space, the annotated point p is backprojected into the coordinate frame of BIM B, yielding a 3D point q''. The 3D alignment error is defined as the Euclidean distance between the corresponding BIM point q and the backprojected point q''. Since the BIM coordinate system is not expressed in physical units, the 3D distance is converted to centimeters using a scale ratio α. For each image, point-wise errors are aggregated using RMSE, and the overall performance across all images is reported using the mean ($\text{RMSE}_{\text{mean}}$), median ($\text{RMSE}_{\text{med}}$), and 95th percentile ($\text{RMSE}_{\text{p95}}$).

3.3 Implementation

In both the coarse and fine alignment phases, we uniformly sample $P = 500{,}000$ points from B_{init} and B_{coarse}, respectively, to construct the input point clouds for 3DGS training. The learnable Sim(3) transformation is optimized in a continuous parameter space. To ensure positivity, the scale is optimized in log-space and the rotation is parameterized and optimized as a unit quaternion during training. During the coarse alignment phase, the learning rates for the transformation parameters are initialized to 10^{-1} and exponentially decayed to 10^{-2} over 500 iterations [21]. The number of views used for loss computation per iteration is set to $K = 20$, balancing training stability and GPU memory constraints. In the fine alignment phase, the overlap threshold β for refined building mask generation is set to 0.9. The learning rates for the transformation parameters are initialized to 10^{-3} and exponentially decayed to 10^{-4} over 30,000 iterations. During fine alignment, a single view is used to compute the photometric loss per iteration, with $K = 1$. For metric evaluation in 3D space, we use a scale ratio of $\alpha = 99.36$ to convert

distances in the BIM B coordinate system to centimeters.

3.4 Manual baseline

We compare our method against a manual point picking baseline. Specifically, we manually annotate 30 corresponding points distributed over the BIM B and the as-built point cloud, and compute the Sim(3) transformation using [26].

3.5 Results

Table 1 reports a quantitative comparison of different alignment stages using 2D reprojection error in pixel units and 3D reprojection error in metric units (cm). For evaluation, the coarse alignment is run five times. Each coarse result is then used to initialize one corresponding fine-alignment run, yielding five paired coarse-to-fine trials. We report the mean and standard deviation over these five trials.

Table 1. Quantitative results of different alignment methods measured by 2D reprojection error (px) and 3D reprojection error (cm). Lower values indicate better performance.

Method	2D reprojection error (px)			3D reprojection error (cm)		
	$\text{RMSE}_{\text{mean}} \downarrow$	$\text{RMSE}_{\text{med}} \downarrow$	$\text{RMSE}_{\text{p95}} \downarrow$	$\text{RMSE}_{\text{mean}} \downarrow$	$\text{RMSE}_{\text{med}} \downarrow$	$\text{RMSE}_{\text{p95}} \downarrow$
Ours (initial alignment)	332.52	272.46	824.43	1275.61	1131.08	2829.79
Ours (coarse alignment)	21.22 ± 3.75	18.97 ± 3.02	39.62 ± 7.03	77.85 ± 13.53	77.40 ± 13.00	138.23 ± 28.02
Ours (fine alignment)	7.11 ± 0.72	6.46 ± 0.76	15.23 ± 1.27	28.61 ± 2.83	21.16 ± 2.44	62.67 ± 5.08
Manual baseline	**4.98**	**4.46**	**10.08**	**19.43**	**15.69**	**38.37**

As shown in Table 1, the initial alignment shows severe global misalignment. Nevertheless, it provides a feasible starting point for subsequent refinement with 3DGS. From the initial to the coarse and fine stages, both the 2D and 3D reprojection errors consistently decrease, demonstrating the effectiveness of the proposed framework in progressively minimizing misalignment. Compared with the manual baseline, the fine alignment achieves competitive performance in both 2D and 3D metrics. In particular, the fine alignment reaches a 2D $\text{RMSE}_{\text{mean}}$ of 7.11 px, which is 2.13 px higher than the manual baseline, and a 3D $\text{RMSE}_{\text{mean}}$ of 28.61 cm, which is 9.18 cm higher than the manual baseline. The larger difference in 3D RMSE_{p95} suggests that the 3D metric is more sensitive to residual misalignment, where small 2D deviations can lead to larger spatial errors after backprojection. These quantitative results demonstrate the feasibility of using 3DGS-based coarse-to-fine refinement for global BIM to as-built alignment, as evidenced by the consistent reduction in both 2D and 3D errors and the small gap to the manual baseline.

Qualitative results are shown in Figure 2, where aligned BIMs are back-projected into input images using the estimated camera poses. For visualization, we show the qualitative results of the paired trial that achieved the lowest mean reprojection error after fine alignment. A clear coarse-to-fine improvement is observed, and the fine alignment produces projections that closely approach the manual baseline. These visualizations further show that the proposed framework produces visually consistent projections across viewpoints, indicating the effectiveness of 3DGS for global BIM to as-built alignment.

3.6 Ablation study

We conducted ablation studies to analyze the impact of key design parameters in our method and report the average performance over five runs for each experiment. In this section, only 2D reprojection error in pixel units is reported.

Table 2. Ablation study on the number of views K used to compute the photometric loss during the coarse alignment stage, evaluated using 2D reprojection error in pixel units only. Results for $K = 5$ are reported as N/A due to training failure.

K	$\text{RMSE}_{\text{mean}} \downarrow$	$\text{RMSE}_{\text{med}} \downarrow$	$\text{RMSE}_{\text{p95}} \downarrow$
5	N/A	N/A	N/A
10	30.51	28.44	55.71
20	**21.22**	**18.97**	**39.62**

Effect of the number of views in coarse alignment. Table 2 evaluates the impact of the number of views K used to compute the photometric loss during the coarse alignment stage, evaluated using 2D reprojection error in pixel units. When $K = 5$, the combination of large initial misalignment and insufficient multi-view constraints leads to 3DGS training failure. This setting is therefore reported as N/A. Increasing K improves both 3DGS training stability and alignment accuracy, as more views provide more robust multi-view photometric constraints. In particular, K is evaluated up to 20 due to GPU memory constraints, with $K = 20$ achieving the lowest errors across all metrics.

Table 3. Ablation study on the mask type used in the fine alignment stage, evaluated using 2D reprojection error in pixel units only. $\mathcal{M}^1$: BIM-projected masks; $\mathcal{M}^3$: refined masks. Fine alignment is initialized from the best coarse-alignment run with $K = 20$.

Masks	$\text{RMSE}_{\text{mean}} \downarrow$	$\text{RMSE}_{\text{med}} \downarrow$	$\text{RMSE}_{\text{p95}} \downarrow$
$\mathcal{M}^1$	13.87	13.50	25.68
$\mathcal{M}^3$	**6.68**	**6.21**	**14.27**

Effect of mask quality in fine alignment. Among five coarse-alignment runs with $K = 20$, the run with the lowest $\text{RMSE}_{\text{mean}}$ is used to initialize the fine alignment in this ablation study. Table 3 studies the impact of different mask types in the fine alignment stage using 2D reprojection error in pixel units. Using BIM-projected masks $\mathcal{M}^1$ yields limited improvement due to residual misalignment. The proposed refined masks $\mathcal{M}^3$, obtained by further refining $\mathcal{M}^1$ with instance segmentation masks and semantic segmentation masks, improve performance further and

achieve the lowest mean, median, and tail errors. These results highlight the importance of accurate masks for fine alignment.

4 Limitations

The proposed framework has several limitations. First, the coarse-to-fine refinement depends on a sufficiently accurate initial alignment, particularly in orientation. The initialization may become unreliable when DBSCAN extracts only a partial building point cloud or when the target building exhibits strong geometric symmetry. Future work may improve this stage using unmanned aerial vehicle (UAV) GPS priors or semantic structural cues. Second, the refinement stage assumes relatively clean and visible building surfaces. Scaffolding, safety nets, and other temporary structures may introduce mismatches between the images and the BIM geometry, which can affect the optimization. Third, the proposed framework assumes as input a BIM that is already consistent with the observed construction state. Thus, the automation applies to the alignment process under this input condition, and extending the method to operate directly on complete BIMs remains future work. Finally, the current evaluation is limited to a single construction site, a specific construction stage, and UAV imagery. Accordingly, the robustness and transferability of the framework across different construction completion levels, capture strategies, and site layouts remain to be evaluated.

5 Conclusion

This study investigated the feasibility of using 3DGS for automatic global BIM to as-built alignment from multi-view imagery, given a BIM that is consistent with the observed construction state. The proposed *BIM-Align* framework combines classical geometric initialization with coarse-to-fine 3DGS refinement under the supervision of refined building masks. Experimental results show that our method achieves satisfactory alignment and approaches a manual baseline in a fully automatic manner, indicating the effectiveness of 3DGS for global BIM to as-built alignment. Further validation under more diverse and challenging construction scenarios is needed in future work.

6 Acknowledgment

This work was financially supported by the VLAIO COOCK-project 'AIDA' (HBC.2024.0591), a collaboration between imec, Buildwise, and Embuild. We gratefully acknowledge all project partners for their valuable contributions to technical discussions and data collection. We also thank the ibens project team of the Hoboken site for granting access to the construction project and for their practical support during data acquisition. In addition, this research received partial funding from the Flemish Government under the 'Onderzoeksprogramma Artificiele Intelligentie (AI) Vlaanderen' program.

References

[1] Mani Golparvar-Fard, Feniosky Peña-Mora, and Silvio Savarese. D^{4}AR–A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Electronic Journal of Information Technology in Construction*, 14(13):129–153, 2009.

[2] Muhammad Sami Ur Rehman, Muhammad Tariq Shafiq, and Fahim Ullah. Automated computer vision-based construction progress monitoring: a systematic review. *Buildings*, 12(7):1037, 2022. doi:10.3390/buildings12071037.

[3] Guangan Chen, Michiel Vlaminck, Wilfried Philips, and Hiep Luong. Multi-view 3D reconstruction for construction site monitoring. In *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 469–476. SCITEPRESS, 2024. doi:10.5220/0012308300003660.

[4] Alexander Braun, Sebastian Tuttas, Uwe Stilla, and André Borrmann. BIM-based progress monitoring. In *Building Information Modeling: Technology Foundations and Industry Practice*, pages 463–476. Springer, 2018. doi:10.1007/978-3-319-92862-3_28.

[5] Miguel Arturo Vega Torres, Anna Ribic, Borja García de Soto, and André Borrmann. BIMCaP: BIM-based AI-supported LiDAR-camera pose refinement. *arXiv preprint arXiv:2412.03434*, 2024. doi:10.48550/arXiv.2412.03434.

[6] Slávek Zbirovskỳ and Václav Nežerka. Open-source automatic pipeline for efficient conversion of large-scale point clouds to IFC format. *Automation in Construction*, 177:106303, 2025. doi:10.1016/j.autcon.2025.106303.

[7] Yasutaka Furukawa, Carlos Hernández, et al. Multi-view stereo: a tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1-2):1–148, 2015. doi:10.1561/0600000052.

[8] Mani Golparvar-Fard, Feniosky Peña-Mora, Carlos A Arboleda, and SangHyun Lee. Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs. *Journal of computing in civil engineering*, 23(6):391–404, 2009. doi:10.1061/(ASCE)0887-3801(2009)23:6(391).

[9] Mani Golparvar-Fard, Feniosky Peña-Mora, and Silvio Savarese. Integrated sequential as-built and as-planned representation with D^{4}AR tools in support of decision-making tasks in the AEC/FM industry. *Journal of Construction Engineering and Management*, 137(12):1099–1116, 2011. doi:10.1061/(ASCE)CO.1943-7862.0000371.

[10] Mani Golparvar-Fard, Feniosky Pena-Mora, and Silvio Savarese. Monitoring changes of 3D building elements from unordered photo collections. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 249–256. IEEE, 2011. doi:10.1109/ICCVW.2011.6130250.

[11] Mani Golparvar-Fard, Feniosky Pena-Mora, and Silvio Savarese. Automated progress monitoring using unordered daily construction photographs and IFC-based building information models. *Journal of Computing in Civil Engineering*, 29(1):04014025, 2015. doi:10.1061/(ASCE)CP.1943-5487.0000205.

[12] Changchang Wu. VisualSFM: A Visual Structure from Motion System. `http://ccwu.me/vsfm/index.html`, 2011. Software, accessed 2025-12-24.

[13] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. doi:10.1109/CVPR.2016.445.

[14] Agisoft LLC. Agisoft photogrammetry software. `https://www.agisoft.com`, 2025. Commercial software, accessed 2025-12-24.

[15] Hadi Mahami, Farnad Nasirzadeh, Ali Hosseininaveh Ahmadabadian, and Saeid Nahavandi. Automated progress controlling and monitoring using daily site images and building information modelling. *Buildings*, 9(3):70, 2019. doi:10.3390/buildings9030070.

[16] Hany Omar, Lamine Mahdjoubi, and Gamal Kheder. Towards an automated photogrammetry-based approach for monitoring and controlling construction site activities. *Computers in Industry*, 98:172–182, 2018. doi:10.1016/j.compind.2018.03.012.

[17] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[18] Autodesk, Inc. Autodesk revit. `https://www.autodesk.com/products/revit/`, 2025. BIM authoring software, accessed 2025-12-24.

[19] Sebastian Tuttas, Alexander Braun, André Borrmann, and Uwe Stilla. Acquisition and consecutive registration of photogrammetric point clouds for construction progress monitoring using a 4D BIM. *PFG–journal of photogrammetry, remote sensing and geoinformation science*, 85(1):3–15, 2017. doi:10.1007/s41064-016-0002-z.

[20] Mustafa K Masood, Antti Aikala, Olli Seppänen, and Vishal Singh. Multi-building extraction and alignment for as-built point clouds: a case study with crane cameras. *Frontiers in Built Environment*, 6:581295, 2020. doi:10.3389/fbuil.2020.581295.

[21] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139–1, 2023. doi:10.1145/3592433.

[22] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010. doi:10.1002/wics.101.

[23] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth Anything: unleashing the power of large-scale unlabeled data. In *CVPR*, 2024. doi:10.1109/CVPR52733.2024.00987.

[24] Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, et al. SAM 3: segment anything with concepts. 2025. doi:10.48550/arXiv.2511.16719.

[25] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi:10.1109/TIP.2003.819861.

[26] Berthold KP Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the optical society of America A*, 4(4):629–642, 1987. doi:10.1364/JOSAA.4.000629.

Real-Time Risk Assessment for Underground Space Using IoT, Online Learning, and Bayesian Networks

Milad Mousavi[1], Xuesong Shen[1], Zhigang Zhang[2], Khalegh Barati[1], and Binghao Li[3]

[1] School of Civil and Environmental Engineering, UNSW Sydney, Australia
[2] Chongqing Research Institute of China Coal Technology & Engineering Group, Chongqing, China
[3] School of Mineral and Energy Resources Engineering, UNSW Sydney, Australia
milad.mousavi@unsw.edu.au, x.shen@unsw.edu.au, zzg-2@163.com, khalegh.barati@unsw.edu.au, binghao.li@unsw.edu.au

Abstract –

Underground working environments are inherently hazardous, with explosion risks posing a persistent threat. Traditional static risk assessments struggle with dynamic conditions. This paper presents a novel real-time explosion risk prediction framework integrating Internet of Things (IoT) sensor data, Long Short-Term Memory (LSTM) networks, and Bayesian Networks (BNs). The system monitors critical indicators (e.g., methane, wind speed) via IoT. The LSTM captures temporal patterns via online learning, and its outputs feed into a BN, which combines predictions with expert knowledge for interpretable risk levels. A case study on a real underground coal mine demonstrated accurate methane concentration forecasting (R^2 = 0.93), with errors decreasing over time. The BN assessment dynamically reflected increased explosion risk upon introduction of an ignition source. This integrated approach offers real-time, interpretable support for explosion risk mitigation.

Keywords –

Explosion Risk Prediction; Dynamic Risk Assessment; Process Safety; Internet of Things; Long Short-Term Memory; Bayesian Networks

1 Introduction

Developing and operating underground spaces poses unique hazards characterized by spatial opacity, geological uncertainty, and dynamic environmental fluctuations such as gas emissions and seismic activity [1]. These risks, exemplified by fatal gas explosions, endanger workers and infrastructure, causing substantial economic losses [2]. Traditional static risk analysis methods, assuming stable environments, fail to incorporate time-dependent fluctuations [3]. Consequently, there is a pressing need for dynamic, real-time risk monitoring and prediction systems that can adapt to the evolving conditions characteristic of underground industrial settings.

Recent technological advancements offer promising avenues to address these challenges. The proliferation of the Internet of Things (IoT) enables continuous, real-time data collection on critical environmental indicators through wireless sensor networks. However, current monitoring systems primarily visualize data, leaving it untapped for deeper risk insights [4]. Furthermore, compatibility issues among data sources and the complexity of analyzing multivariate sensor streams hinder generalization [5]. Simultaneously, data-driven models, particularly deep learning architectures like Long Short-Term Memory (LSTM) networks, have demonstrated superior capability in capturing complex temporal patterns in time-series data. However, these models often operate as "black boxes," providing limited interpretability and failing to integrate crucial domain expertise [6]. On the other hand, knowledge-driven probabilistic models like Bayesian Networks (BNs) excel in representing expert knowledge and facilitating transparent, causal reasoning but may lack the adaptive learning capacity required for highly dynamic environments [7].

This paper addresses these limitations by introducing a novel dynamic risk prediction system that integrates IoT sensors, LSTM networks, and BNs into a unified, real-time framework. The technical novelty of this framework lies in an automated, real-time pipeline coupling continuous online LSTM learning (updating weights per time-step) with instantaneous Bayesian probabilistic inference via an Application Programming Interface (API), providing dynamic adaptability and interpretability for high-risk mining environments.

2 Literature Review

This section surveys the current state of research and technological applications in underground risk management.

2.1 Risk Management in Underground Settings

Traditional risk analysis in underground domain employs both qualitative techniques, such as safety checklists and the Delphi method, and quantitative techniques, including event tree analysis, fault tree analysis, and neural networks [8]. Typically, these methods rely on expert knowledge, historical data, and static models that assume stable conditions, thereby failing to capture the rapid environmental shifts common underground.

This limitation has prompted a shift towards dynamic modeling. For instance, Wu, et al. [9] proposed a quantitative risk assessment method for coal mine gas explosions based on BNs and Computational Fluid Dynamics (CFD), emphasizing that a dynamic model integrating real-time conditions could enhance production safety. Furthermore, the interconnected nature of underground systems leads to cascading hazards. Bai, et al. [10] developed an energy-based coupling risk assessment model for utility tunnels, identifying multiple cascading accident scenarios and stressing the need for stronger preventive actions. Hai, et al. [11] noted that existing methods often perform static analysis on individual risks while overlooking dynamic interactions, recommending the development of better risk indicator systems and the incorporation of interdependencies among risk factors.

2.2 IoT Applications in Monitoring Underground Hazards

The adoption of IoT technology has significantly advanced the monitoring of factors such as gas, dust, temperature, humidity, airflow, and seismic activity in underground mining. IoT devices enable swift data transmission, facilitating a dynamic approach to risk assessment. Onifade, et al. [12] highlighted how IoT enables smart ventilation systems that deliver precise amounts of fresh air to specific areas, using sensor networks to transmit real-time data to control rooms for immediate alerts and automated adjustments. Wu, et al. [13] developed a dynamic information platform for underground coal mine safety using IoT, integrating data acquisition, transmission, analysis, and application layers to collect and analyze vast amounts of operational data. Despite these advancements, a significant challenge remains. Ślęzak, et al. [4] pinpointed that the big data collected is typically used for raw data visualization and monitoring, with deeper analytical potential largely untapped. Applying intelligent forecasting models and data analytics to this data could predict near-future conditions and substantially enhance decision support systems.

2.3 Bayesian Risk Assessment in Underground Environments

BNs have become a crucial tool in underground risk management due to their probabilistic nature and ability to handle complex, uncertain relationships. Their applications range from accident cause analysis to dynamic risk modeling. Xuecai, et al. [14] developed a data-driven BN model for analyzing coal and gas outburst accidents, learning its structure and parameters from a comprehensive dataset of real-world incidents. In scenarios with insufficient data, expert knowledge is integrated. Tong, et al. [15] constructed a BN for mine gas explosions using expert knowledge and the Delphi method to establish conditional probabilities, featuring multi-scale nodes with dynamic probability updates. Mousavi, et al. [16] introduced an IoT and BN framework for online safety risk management, emphasizing the need for accurate BN structures and comprehensive sensor data. However, large and diverse data streams can challenge BN accuracy and computational efficiency. Moradi, et al. [17] noted that relying solely on BNs for complex assessments with numerous sensors and heterogeneous data becomes impractical, underscoring the importance of integrating BNs with data-driven prediction engines for proactive, dynamic risk monitoring.

2.4 Data-Driven Risk Prediction in Underground Spaces

Data-driven models, particularly LSTM networks, have emerged as powerful tools for uncovering complex patterns in time-series data from underground environments. Their application focuses on predicting critical parameters to provide early warnings. Kumari, et al. [18] proposed a UMAP and LSTM model for predicting fire and explosibility risks in sealed-off areas of coal mines, using historical and real-time gas sensor data to forecast concentrations and calculate fire ratios. Dey, et al. [19] implemented a hybrid CNN-LSTM model with IoT sensors to predict miner health quality and gas concentration levels, achieving high prediction accuracy.

Despite these strides, significant concerns remain. A primary issue is the black-box nature of these deep learning models, which fail to deliver context-specific insights or support interactive what-if scenario analysis, limiting customized decision-making guidance [20]. This lack of interpretability challenges trust in the algorithms, necessitating careful integration with engineering knowledge [21]. Furthermore, a notable gap in the literature is the absence of online learning frameworks specifically tailored for underground spaces. Online learning, which involves continuously training a model on newly arriving data streams [22], is essential for real-time adaptability in dynamic environments but has not

been extensively explored. This gap highlights the need for methodologies that seamlessly integrate deep learning with online learning to ensure predictive models can evolve continuously in response to changing underground conditions [23].

3 Research Methodology

The proposed IoT-LSTM-Bayesian system, as illustrated in Figure 1, operates through three modules, with static segments occurring pre-operation and dynamic segments occurring periodically during operation. The first module involves IoT deployment for real-time data acquisition and preprocessing. The second module updates an LSTM model with real-time data and predicts future values. The third module evaluates target risks using a pre-developed BN and the LSTM predictions, reporting risk evolution to users.

3.1 Data Specifications and Preparation

Data is derived from IoT sensors (e.g., for methane, wind speed, temperature, dust) deployed in critical locations like work fronts and ventilation corridors. Data is recorded at intervals from minutes to hours and transmitted via technologies like Wi-Fi. The real-time nature is crucial for adapting to dynamic conditions but requires robustness against sensor failures.

Data preparation involves three key tasks. First, missing values are imputed using forward fill to maintain temporal order. Second, anomalies, such as test-induced sensor readings, are identified via domain knowledge and replaced with the last known correct value. Third, the sliding window technique segments input data into overlapping fixed-length windows to capture temporal dependencies for the LSTM.

3.2 Online LSTM Learning and Prediction

LSTM networks are chosen for their ability to model long-term temporal dependencies in sequential data, such as methane concentration patterns. This memory mechanism is particularly critical in underground environments where gas accumulation and dispersion exhibit complex, time-lagged trends that simpler models fail to capture. An LSTM cell contains forget, input, and output gates that regulate information flow via sigmoid and hyperbolic tangent functions [24].

The model undergoes online learning, updating itself with each new data sample at every time step. This enables adaptation to evolving trends. Model performance is evaluated at each step using Mean Squared Error (MSE), Mean Absolute Error (MAE), and

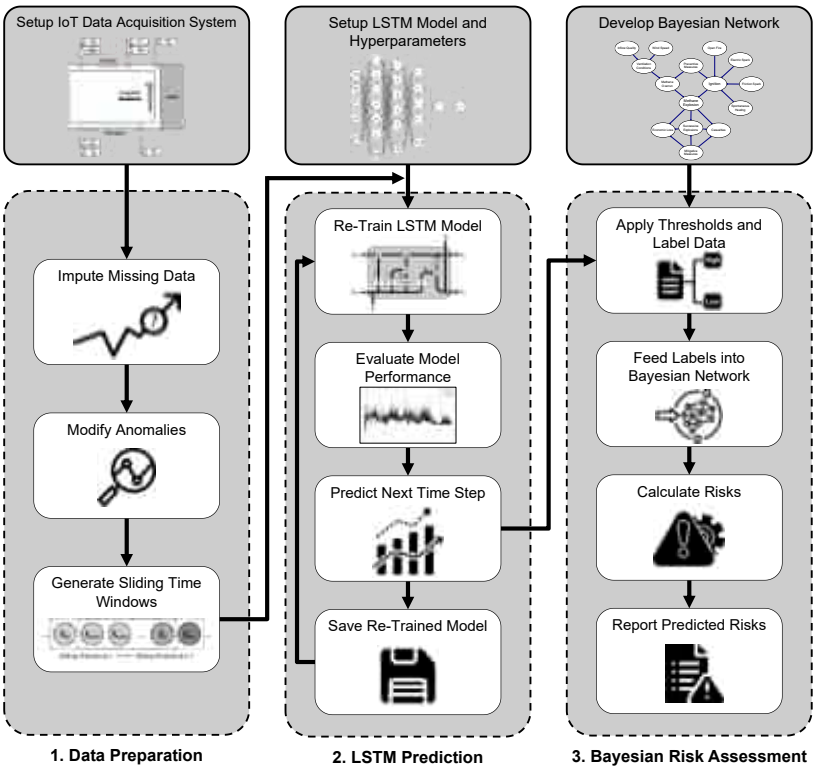


Figure 1. Overview of the proposed IoT-LSTM-Bayesian risk prediction system.

the Coefficient of Determination (R^2). Lower MSE/MAE and an R^2 closer to 1 indicate better performance. The trained model then predicts the next step's values. This model is saved and reused for subsequent retraining.

3.3 Bayesian Risk Assessment

BNs are chosen for their ability to integrate data-driven insights with expert knowledge into a transparent, interpretable probabilistic framework. A BN is a directed acyclic graph where nodes represent variables and edges represent conditional dependencies, quantified via Conditional Probability Tables (CPTs) [25]. The joint probability distribution is calculated using Eq. (1). Inference, updating probabilities given new evidence, is performed using Bayes' theorem (Eq. 2).

$$P(U) = P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | Parent(X_i)) \quad (1)$$

$$P(Parent(X_i)|X_i) = \frac{P(Parent(X_i))P(X_i|Parent(X_i))}{P(X_i)} \quad (2)$$

In the dynamic system, predicted environmental indicators from the LSTM are discretized using predefined thresholds (e.g., methane concentration labeled as negligible or serious) and fed as evidence into specific BN nodes. The BN then updates probabilities throughout the network, providing real-time, interpretable risk insights.

4 Case Study on Underground Coal Mine

Methane explosions, resulting from methane concentrations reaching 5-15% in the presence of an ignition source, are among the most severe hazards in underground coal mining [26]. Concentrations exceeding 1% are typically considered hazardous, triggering safety protocols that halt production and incur significant financial losses [13]. This case study evaluates the proposed system using environmental monitoring data from a real longwall underground coal mine in China, focusing on methane concentration prediction and explosion risk assessment.

4.1 Data Characteristics and Preparation

Data was collected from eight sensors deployed at the mine's working face: five methane concentration sensors (T0-T4), one temperature sensor (TP1), one wind speed sensor (WS1), and one dust concentration sensor (D1). Figure 2 shows the spatial distribution of the sensors used in this study. Data spanned 3 January 2022 to 12 May 2023 (494 days) at one-minute intervals, yielding 710,367 samples. Table 1 presents the data characteristics. The raw dataset contained missing values and outliers. Missing values were imputed using forward fill, replacing them with the last recorded value to preserve temporal order. Outliers, primarily caused by controlled sensor calibration tests that artificially raised methane levels above 1%, were identified using domain knowledge from mine experts and replaced with the last known valid reading. Data from all eight sensors formed a six-step sliding window to capture multivariate spatial-temporal dynamics, targeting the critically located T0 sensor for prediction.

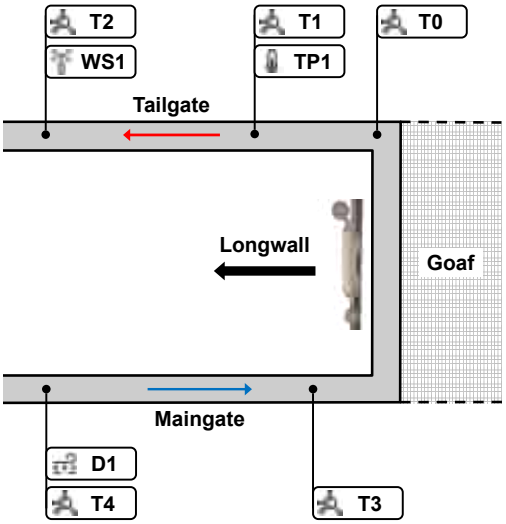


Figure 2. Deployment layout of the environmental sensors within the coal mine.

4.2 Online LSTM Learning and Prediction

A predictive LSTM model, whose architecture is shown in Figure 3, was constructed with two LSTM layers (256 and 128 units, respectively) followed by two dense layers (32 and 1 unit). The Rectified Linear Unit (ReLU) activation function was used, and the model was compiled with the Adam optimizer having a learning rate of 0.001. The hyperparameter selection, including the six-step window length, was based on domain knowledge and empirical testing related to gas dynamics in mines. The model was trained using online learning; at each one-minute time step, a single data sample (the current sliding window) was used to update the model weights via the *train_on_batch* function in *Keras*. Subsequently, the *predict_on_batch* function was used to forecast the T0 methane concentration for the next immediate time step.

Table 1. An overview of the characteristics of the sensor readings.

Variable	Unit	Sensor Code	Min	Max	Mean	SD
Methane Concentration	% CH4	T0	0.000	0.985	0.121	0.083
		T1	0.000	0.995	0.146	0.084
		T2	0.000	0.995	0.184	0.093
		T3	0.000	1.000	0.064	0.042
		T4	0.000	0.947	0.052	0.032
Wind Speed	m/s	WS1	0.003	6.345	1.193	0.525
Temperature	°C	TP1	9.400	20.700	15.315	2.600
Dust Concentration	ppm	D1	0.000	238.067	1.250	5.796

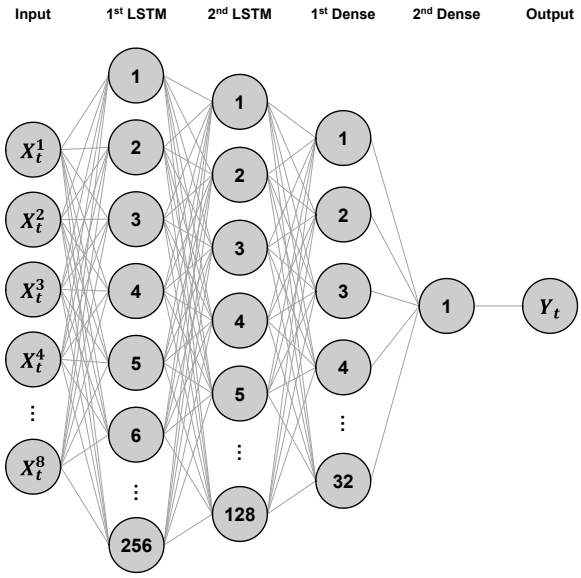


Figure 3. Architecture of the developed LSTM network.

Figure 4 compares the actual methane concentrations measured by sensor T0 and those predicted by the developed LSTM model. The model's performance evolved significantly over time as it engaged in continuous online learning, updating its weights incrementally with each new minute of data using the *train_on_batch* function. Initially, with scant training data, predictions displayed considerable deviation from actual readings; this early phase underscores the importance of expert oversight before deploying model outputs for operational decisions. However, as sequential processing continued, the model rapidly adapted to fluctuations and, aided by spatially distributed multivariate inputs, demonstrated strong responsiveness to sudden spikes. The prediction error, evaluated cumulatively using MAE and MSE, decreased sharply within the first 250 time steps and then plateaued at a low, stable level, indicating successful convergence. Over the entire dataset, the final MAE reached 0.01, MSE reached 0.0005, and the R^2 achieved 0.93. This high R^2 value reflects a strong explanatory power and indicates a high level of predictive accuracy for methane concentration, validating the LSTM's suitability for real-time hazard forecasting in dynamic underground environments.

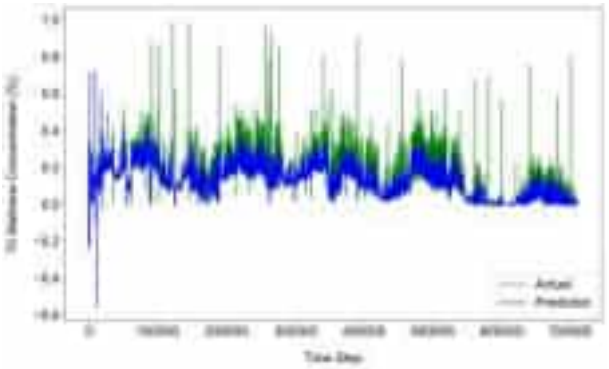


Figure 4. Comparison of the predicted and actual values over time.

4.3 Methane Explosion Bayesian Network

Due to a lack of comprehensive historical data on methane explosions, the BN was constructed using a knowledge-driven methodology. An initial network structure was drafted from literature, identifying causal and consequential variables related to ignition sources, gas accumulation, ventilation, safety interventions, and explosion consequences. This structure builds upon the authors' previous work [27]. This structure was refined through three rounds of expert consultation involving academic specialists and a senior mine ventilation officer. Based on expert feedback, variables like atmospheric pressure and oxygen concentration were removed for being redundant or constant, while practical proxies like wind speed (for airflow) and inflow methane quality were

added. Various methane sources were aggregated into a single "methane overrun" node, as sensors measure cumulative concentration.

The final BN, as shown in Figure 5, comprised 15 discrete nodes organized into five groups: (1) Causes: ignition sources like friction spark, open fire; (2) Environmental Conditions: methane overrun, wind speed, inflow quality, ventilation conditions; (3) Accident Trigger: methane explosion; (4) Consequences: casualties, economic loss, successive explosions; and (5) Control Actions: preventive and mitigative measures. Nodes were defined as binary or ordinal states (e.g., negligible/serious, none/slight/severe). CPTs were populated via a structured expert elicitation process, which is fully detailed and validated in [16]. The structure allowed certain nodes (e.g., methane overrun) to receive direct sensor input, while others (e.g., ignition sources) could be set manually for scenario analysis.



Figure 5. Refined BN for methane explosion in underground coal mines.

4.4 LSTM-Bayesian Risk Assessment

Figure 6 displays the architecture of the LSTM-Bayesian fusion system. The system integrated the prediction and assessment layers. At each time step, the continuous methane concentration predicted by the LSTM for sensor T0 was discretized using a 1.0% threshold; values above were labeled serious, others negligible. This step converted the numerical prediction into a format compatible with the BN's categorical reasoning framework. This labeled evidence was then input into the "methane overrun" node of the BN via *SMILE Engine* [28], the API of *GeNIe Modeler*. The BN performed automatic probabilistic inference to update the risk levels of all dependent nodes, including the "methane explosion" node.

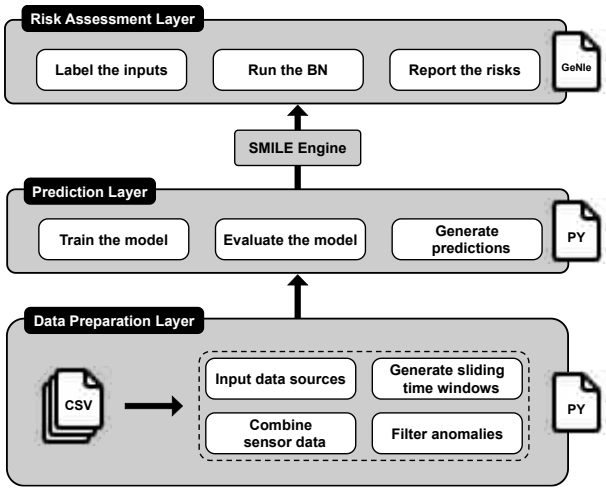


Figure 6. Architecture of the LSTM-Bayesian risk assessment system.

The system's performance was examined under two evidential scenarios. In Scenario 1, representing normal operations, only the predicted methane concentration was used as evidence, while ignition source nodes retained their prior low probabilities. The BN calculated a stable explosion risk probability of approximately 3.5% throughout the period. In Scenario 2, simulating a localized open flame near T0 for three hours, the "open fire" node was set to 1 during this interval. This nearly doubled the explosion risk to 6.8%. Given methane consistently remained below 1% during the study, low absolute probabilities are expected. In practice, operators utilize these relative risk spikes to trigger tiered advisory interventions before deterministic alarms mandate shutdowns.

4.5 Discussion and Implications

The case study validates the framework's core functionalities. The LSTM component demonstrated effective online learning, with predictive accuracy for methane concentration improving to a high level (R^2=0.93) as it adapted to the mine's dynamic environment. The integration with the BN provided a crucial interpretative layer. It translates predicted gas levels into contextualized probabilistic risks, enabling automated safety protocols (e.g., machinery power-offs) within the 1-minute forecast window. Unlike reactive 1% alarms that mandate costly complete operational shutdowns, the system detects ascending trends early. This grants operators crucial lead time to proactively intensify ventilation or slow production, preventing catastrophic financial losses and fatal explosive states. This hybrid approach bridges data-driven prediction and explainable reasoning to support critical decision-making.

5 Conclusions

An innovative system for dynamic risk management in underground environments was introduced, integrating IoT, LSTM, and BNs. A case study in a Chinese coal mine demonstrated its capability to accurately predict methane concentrations and assess explosion risks in real time. The LSTM model achieved high accuracy (R^2=0.93), and the BN provided interpretable, scenario-sensitive risk updates. This integration addresses limitations of purely data-driven or knowledge-driven methods, enhancing safety and operational efficiency.

While the extensive 494-day dataset establishes a robust proof-of-concept, relying on a single mine limits generalizability. Ongoing future work will conduct cross-site validation, benchmark against baseline models, develop 5-10 minute multi-step forecasts for human intervention, apply cost-sensitive learning, and explore digital twin integration.

Acknowledgments

This research is supported by funding from the Australian Research Council (ARC) Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), and the ARC Research Hub for Human-Robot Teaming for Sustainable and Resilient Construction (IH240100016). The authors thank BayesFusion, LLC for software access.

References

[1] M. Q. Huang, J. Ninić, and Q. B. Zhang, "BIM, machine learning and computer vision techniques in underground construction: Current status and future perspectives," *Tunnelling and Underground Space Technology,* vol. 108, p. 103677, 2021/02/01/ 2021, doi: https://doi.org/10.1016/j.tust.2020.103677.

[2] Y. Liang and Q. Liu, "Early warning and real-time control of construction safety risk of underground engineering based on building information modeling and internet of things," *Neural Computing and Applications,* vol. 34, no. 5, pp. 3433–3442, 2022/03/01 2021, doi: https://doi.org/10.1007/s00521-021-05755-8.

[3] G. Zhang, E. Wang, X. Liu, and Z. Li, "Research on risk assessment of coal and gas outburst during continuous excavation cycle of coal mine with dynamic probabilistic inference," *Process Safety and Environmental Protection,* vol. 190, pp. 405–419, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.psep.2024.08.054.

[4] D. Ślęzak *et al.*, "A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines," *Information Sciences,* vol. 451-452, pp. 112–133, 2018/07/01/ 2018, doi: https://doi.org/10.1016/j.ins.2018.04.026.

[5] M. Mousavi, X. Shen, Z. Zhang, K. Barati, and B. Li, "Online Deep Transfer Learning and Multi-Sensor Analysis for Enhanced Underground Monitoring," in *2025 10th International Conference on Machine Learning Technologies (ICMLT)*, 23–25 May 2025 2025, pp. 125–129, doi: 10.1109/ICMLT65785.2025.11193369.

[6] L. Cheded and R. Doraiswami, "A novel integrated framework for fault diagnosis with application to process safety," *Process Safety and Environmental Protection,* vol. 154, pp. 168–188, 2021/10/01/ 2021, doi: https://doi.org/10.1016/j.psep.2021.08.008.

[7] G. Yu, Y. Wang, Z. Mao, M. Hu, V. Sugumaran, and Y. K. Wang, "A digital twin-based decision analysis framework for operation and maintenance of tunnels," *Tunnelling and Underground Space Technology,* vol. 116, p. 104125, 2021/10/01/ 2021, doi: https://doi.org/10.1016/j.tust.2021.104125.

[8] M. Li, H. Wang, D. Wang, Z. Shao, and S. He, "Risk assessment of gas explosion in coal mines based on fuzzy AHP and bayesian network," *Process Safety and Environmental Protection,* vol. 135, pp. 207–218, 2020/03/01/ 2020, doi: https://doi.org/10.1016/j.psep.2020.01.003.

[9] Z. Wu, Y. Li, and Q. Jing, "Quantitative risk assessment of coal mine gas explosion based on a Bayesian network and computational fluid dynamics," *Process Safety and Environmental Protection,* vol. 190, pp. 780–793, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.psep.2024.08.080.

[10] Y. Bai *et al.*, "Energy-based coupling risk assessment (CRA) model for urban underground utility tunnels," *Reliability Engineering & System Safety,* vol. 250, p. 110255, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.ress.2024.110255.

[11] N. Hai, D. Gong, S. Liu, and Z. Dai, "Dynamic coupling risk assessment model of utility tunnels based on multimethod fusion," *Reliability Engineering & System Safety,* vol. 228, p. 108773, 2022/12/01/ 2022, doi: https://doi.org/10.1016/j.ress.2022.108773.

[12] M. Onifade, K. O. Said, and A. P. Shivute, "Safe mining operations through technological advancement," *Process Safety and Environmental Protection,* vol. 175, pp. 251–258, 2023/07/01/ 2023, doi: https://doi.org/10.1016/j.psep.2023.05.052.

[13] Y. Wu, M. Chen, K. Wang, and G. Fu, "A dynamic information platform for underground coal mine safety based on internet of things," *Safety Science,* vol. 113, pp. 9–18, 2019/03/01/ 2019, doi: https://doi.org/10.1016/j.ssci.2018.11.003.

[14] X. Xuecai *et al.*, "Accident causes data-driven coal and gas outburst accidents prevention: Application of data mining and machine learning in accident path mining and accident case-based deduction," *Process Safety and Environmental Protection,* vol. 162, pp. 891–913, 2022/06/01/ 2022, doi: https://doi.org/10.1016/j.psep.2022.04.059.

[15] X. Tong, W. Fang, S. Yuan, J. Ma, and Y. Bai, "Application of Bayesian approach to the assessment of mine gas explosion," *Journal of Loss Prevention in the Process Industries,* vol. 54, pp. 238–245, 2018/07/01/ 2018, doi: https://doi.org/10.1016/j.jlp.2018.04.003.

[16] M. Mousavi, X. Shen, Z. Zhang, K. Barati, and B. Li, "IoT-Bayes fusion: Advancing real-time environmental safety risk monitoring in underground mining and construction," *Reliability Engineering & System Safety,* vol. 256, p. 110760, 2025/04/01/ 2025, doi: https://doi.org/10.1016/j.ress.2024.110760.

[17] R. Moradi, S. Cofre-Martel, E. Lopez Droguett, M. Modarres, and K. M. Groth, "Integration of deep learning and Bayesian networks for condition and operation risk monitoring of complex engineering systems," *Reliability Engineering & System Safety,* vol. 222, p. 108433, 2022/06/01/ 2022, doi: https://doi.org/10.1016/j.ress.2022.108433.

[18] K. Kumari *et al.*, "UMAP and LSTM based fire status and explosibility prediction for sealed-off area in underground coal mine," *Process Safety and Environmental Protection,* vol. 146, pp. 837–852, 2021/02/01/ 2021, doi: https://doi.org/10.1016/j.psep.2020.12.019.

[19] P. Dey, S. K. Chaulya, and S. Kumar, "Hybrid CNN-LSTM and IoT-based coal mine hazards monitoring and prediction system," *Process Safety and Environmental Protection,* vol. 152, pp. 249–263, 2021/08/01/ 2021, doi: https://doi.org/10.1016/j.psep.2021.06.005.

[20] H. Chen *et al.*, "Augmented reality, deep learning and vision-language query system for construction worker safety," *Automation in Construction,* vol. 157, p. 105158, 2024/01/01/ 2024, doi: https://doi.org/10.1016/j.autcon.2023.105158.

[21] C. Zhou, Y. Gao, E. J. Chen, L. Ding, and W. Qin, "Deep learning technologies for shield tunneling: Challenges and opportunities," *Automation in Construction,* vol. 154, p. 104982, 2023/10/01/ 2023, doi: https://doi.org/10.1016/j.autcon.2023.104982.

[22] Z. Fan, C. Liu, D. Cai, and S. Yue, "Research on black spot identification of safety in urban traffic accidents based on machine learning method," *Safety Science,* vol. 118, pp. 607–616, 2019/10/01/ 2019, doi: https://doi.org/10.1016/j.ssci.2019.05.039.

[23] Y. Pan, M. Wu, L. Zhang, and J. Chen, "Time series clustering-enabled geological condition perception in tunnel boring machine excavation," *Automation in Construction,* vol. 153, p. 104954, 2023/09/01/ 2023, doi: https://doi.org/10.1016/j.autcon.2023.104954.

[24] P. Lyu, N. Chen, S. Mao, and M. Li, "LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion," *Process Safety and Environmental Protection,* vol. 137, pp. 93–105, 2020/05/01/ 2020, doi: https://doi.org/10.1016/j.psep.2020.02.021.

[25] A. TohidiFar, M. Mousavi, and A. Alvanchi, "A hybrid BIM and BN-based model to improve the resiliency of hospitals' utility systems in disasters," *International Journal of Disaster Risk Reduction,* vol. 57, p. 102176, 2021/04/15/ 2021, doi: https://doi.org/10.1016/j.ijdrr.2021.102176.

[26] J. Brodny, D. Felka, and M. Tutak, "The use of the neuro-fuzzy model to predict the methane hazard during the underground coal mining production process," *Journal of Cleaner Production,* vol. 368, p. 133258, 2022/09/25/ 2022, doi: https://doi.org/10.1016/j.jclepro.2022.133258.

[27] M. Mousavi, X. Shen, and B. Li, "Online Safety Risk Management for Underground Mining and Construction Based on IoT and Bayesian Networks," presented at the Proceedings of the 40th International Symposium on Automation and Robotics in Construction, Chennai, India, 2023/07/07, 2023.

[28] BayesFusion. "SMILE: Structural Modeling, Inference, and Learning Engine." https://www.bayesfusion.com/smile/ (accessed 24 October 2025.

Interpretable Real-Time Construction Safety Monitoring via Video-Text Retrieval

Sejun Park[1] Jaehoo Kim[2] Dongmin Lee[3]

[123]School of Civil, Environmental, and Architectural Engineering, Korea University, Republic of Korea
sejun0909@korea.ac.kr, wogn0505@korea.ac.kr, leedm@korea.ac.kr

Abstract

Recent construction sites increasingly employ real-time safety monitoring systems that apply rule-based logic to frame-level object detections. However, because such approaches rely on isolated detections and manually defined rules, they provide limited evidence for how workers, PPE, actions, locations, and hazards jointly constitute safety risk, and their predictions often fluctuate over time, leading to false alarms and missed detections. To address these limitations, we propose an interpretable framework for streaming construction safety monitoring based on video–text retrieval, which captures spatiotemporal context and provides explicit textual explanations. The proposed method evaluates short video segments by retrieving relevant evidence from a structured safety text bank and estimating risk from the retrieved results, while also presenting representative evidence sentences for operator verification. Experiments on representative scaffold safety scenarios demonstrated strong risk classification performance, with accuracy of up to 97.6%, while the retrieved evidence consistently outperformed captioning baselines in explanatory quality, achieving top-1 BERTScore-F1 values of 0.925–0.943 and top-1 SPICE values of 0.618–0.727. The system also achieved a mean end-to-end latency of 210.8 ms on a single GPU, supporting real-time operation. These findings indicate that video–text retrieval offers a practical and interpretable approach to spatiotemporally informed construction safety monitoring without site-specific retraining.

Keywords

construction safety, real-time monitoring, interpretability, video–text retrieval

1 Introduction

Computer-vision (CV)-based real-time safety monitoring has become a central approach for detecting hazardous situations in construction sites, where visual data are continuously analysed to identify workers, equipment, and site conditions. However, construction sites remain highly dynamic environments: construction phases, spatial layouts, and equipment placements change frequently, while viewpoint and illumination changes, occlusions, and motion blur caused by worker movements continuously alter the visual conditions under which CV models operate.

Consequently, the visibility and appearance of target entities can vary substantially from frame to frame, causing detection outputs to fluctuate even within the same ongoing situation. This temporal instability in detections often results in intermittent misses and spurious detections over time, i.e., temporal inconsistency. Because frame-wise decision-making treats each frame independently, it cannot effectively use adjacent temporal evidence to stabilize predictions, which in turn increases false positives and false negatives and undermines the reliability of hazard detection[1].

Moreover, construction-site hazards are not determined solely by the presence of a particular object in an isolated frame, but by how multiple risk cues, such as workers, personal protective equipment (PPE) compliance, actions, objects, and locations, interact and evolve in spatiotemporal context[2]. Consequently, reliance on frame-level detections alone can either miss hazardous events or incorrectly classify normal operations as dangerous[3]. These observations motivate moving beyond frame-wise detection toward video-based safety monitoring that incorporates temporal context and supports more reliable interpretation of how risk cues jointly indicate hazardous situations.

In response to these challenges, prior studies have explored several video-based approaches for safety monitoring. Temporal modelling methods, such as RNN/LSTM-based approaches, aggregate information across consecutive frames to reduce prediction fluctuations and improve temporal consistency [4], while video anomaly detection methods learn normal patterns from videos and assign anomaly scores to clips that deviate from them[5]. However, both approaches mainly produce scores, labels, or detections, providing limited explicit evidence about why a situation is risky or how specific risk cues contribute to the assessment. Video

captioning methods improve interpretability by describing scenes in natural language[6]. However, generated descriptions can vary across similar scenes and may omit or overstate critical evidence, while repeated clip-wise decoding also increases computational cost and end-to-end latency in real-time deployment.

Thus, existing approaches remain limited in two practical respects for construction safety monitoring: explicit interpretability and real-time suitability. To address these limitations, we propose a framework for real-time construction safety monitoring that constructs a structured text bank encoding relationships among risk cues from safety regulations and real-world site data, and applies video–text retrieval to provide interpretable evidence for streaming inputs.

2 Related Work

2.1 Computer Vision-based construction safety monitoring

Construction safety monitoring has evolved toward detecting workers, equipment, and site conditions from vision data (e.g., CCTV/UAV videos) and issuing alarms by rule-based assessment of PPE compliance or hazardous states. Recently, pipelines have been proposed that first detect objects using real-time single-stage detectors and then integrate knowledge-based (ontology-driven) reasoning as post-processing to determine violations or hazardous situations[7]. In addition, PPE monitoring has progressed beyond a binary "worn/not worn" decision, with attempts to combine instance segmentation and tracking to associate PPE with individual workers and to refine compliance states at a finer granularity[8].

Nevertheless, frame-level, detection-based risk judgment is highly susceptible to occlusions, viewpoint/illumination changes, and motion blur, causing predictions to fluctuate and making errors accumulate as false positives and false negatives over time. To mitigate this, prior work has proposed aggregating decisions over a temporal window to reduce false alarms, or enhancing spatial reasoning with depth information [1][9]. However, such temporal and spatial augmentations primarily aim to stabilize alarm outputs, and they remain limited in explaining why a scene is judged as hazardous—i.e., which relationships among PPE, actions, locations, and hazard factors constitute the basis for the decision.

2.2 Unsupervised/Self-supervised Video Anomaly Detection (VAD)

To overcome the limitations of frame-based decision-making, a growing body of work has explored applying unsupervised/self-supervised video anomaly detection (VAD) to safety monitoring. In general, VAD learns normal patterns from videos and then assigns anomaly scores or detects abnormal segments by identifying sequences that deviate from the learned normal distribution. In construction safety monitoring, a study combined unsupervised GAN-based learning with pseudo-anomaly synthesis and introduced the WeTeam22 dataset to detect unsafe worker behaviors[10].

However, because VAD typically outputs anomaly scores or abnormal intervals, it is difficult to directly translate the results into concrete risk cues that operators can readily understand and act upon. Recently, benchmarks have been proposed that encourage models to go beyond mere anomaly detection and explain what happened, why it occurred, and what evidence supports the risk judgment[11]. This trend suggests that producing such explanations generally requires additional annotations and/or a separate language reasoning module. Moreover, construction sites exhibit substantial distribution shifts due to changes in phases, spatial configurations, and camera conditions, which induces drift in the normal-data distribution and imposes considerable burdens in collecting, curating, maintaining, and re-adapting normal data for unsupervised learning.

2.3 Video Captioning-based construction safety monitoring

Video captioning aims to complement situation understanding by describing video scenes in natural language. Vision-language models (VLMs), such as Qwen2.5, SmolVLM, and InternVL, process video or multi-frame inputs to perform clip-level scene understanding and generate textual descriptions accordingly[12][13][14]. In another high-risk domain, transportation, research has continued toward explainable video understanding—for example, domain-specific captioning approaches such as TrafficVLM, which describe incidents in a stepwise manner and incorporate output control mechanisms[15].

In contrast, within the construction domain, captioning has been used primarily for recording and documentation, while it has received limited attention as a real-time interpretation mechanism for safety monitoring[16]. This is partly because scene distributions vary substantially across sites, making it difficult to build training data, and because accurately reflecting regulatory semantics—such as PPE compliance or proximity to hazardous zones—often requires vocabulary standardization. Moreover, when deployed in streaming settings, clip-wise text decoding must be repeatedly executed, which can increase computational cost and end-to-end latency.

Overall, prior approaches have advanced either

toward stabilizing alarms or toward providing natural-language descriptions, but mechanisms that consistently present explanations of streaming scenes while accounting for temporal context remain limited. To address this gap, we propose a low-latency framework that combines a structured text bank encoding relationships among key risk factors in construction site with video–text retrieval, thereby achieving both temporal consistency and interpretability in real-time monitoring.

3 Method

This study proposes a real-time monitoring framework that interprets construction-site videos via similarity-based retrieval against a text bank and simultaneously provides a risk score and evidence-grounded explanations. After aligning text and video in a shared embedding space, the framework retrieves text entries most similar to each segment embedding to form an evidence set, and then aggregates unsafe evidence to compute a risk score. Figure 1 illustrates the overall workflow of the proposed framework. The workflow begins with constructing a structured text bank from safety knowledge and indexing its text embeddings in the joint embedding space. During real-time inference, the input video stream is segmented, each segment is encoded into a video embedding, and top-K similar text entries are retrieved by similarity search. The retrieved texts are then used to compute the risk score RS(t), and the results are rendered as real-time overlay outputs.

3.1 Text Bank Construction in a Joint Embedding Space

To provide evidence for risk assessment in terms of relationships among key risk cues, we construct a text bank that describes the work subject, PPE state, action, location, and hazard factor. Safety risk is not determined solely by the presence of a single object; rather, it emerges from combinations of who (SUBJECT) performs what action (ACTION) under which protective condition (SUBJECT_STATE), at which location (LOCATION), and with which associated hazard factor (RISK). This formulation is aligned with semantic scene representations used in computer vision and multimodal scene understanding, while explicitly extending them to construction-safety attributes such as PPE condition and hazard relevance[17]. In construction safety, PPE compliance and work location are particularly decisive determinants of risk, and prior studies have emphasized the importance of entity interactions and contextual information[18]. Accordingly, texts structured around subject, PPE state, action, location, and risk provide an appropriate basis for composing evidence cues.

Each text entry c_i is encoded by the text encoder $f_T(\cdot)$ of a pretrained video–text dual encoder, producing a text embedding $e_T(i) \in \mathbb{R}^d$, which is stored in the bank. The collection of these embeddings forms a joint embedding space $\mathbb{R}^d$ that enables similarity-based interpretation of streaming scenes. To enable low-latency retrieval, all text embeddings are pre-indexed in an offline stage, allowing efficient text search under real-time monitoring constraints.

3.2 Streaming Video Segmentation

In our framework, the input stream is processed by grouping frames into fixed-length segments. Because a segment comprises consecutive frames over a short time interval rather than a single frame, it is less sensitive to instantaneous noise and can capture both motion dynamics and scene cues[1].

Concretely, we partition the input stream into segments V_t of duration t seconds and apply a sliding window with stride S seconds to generate temporally

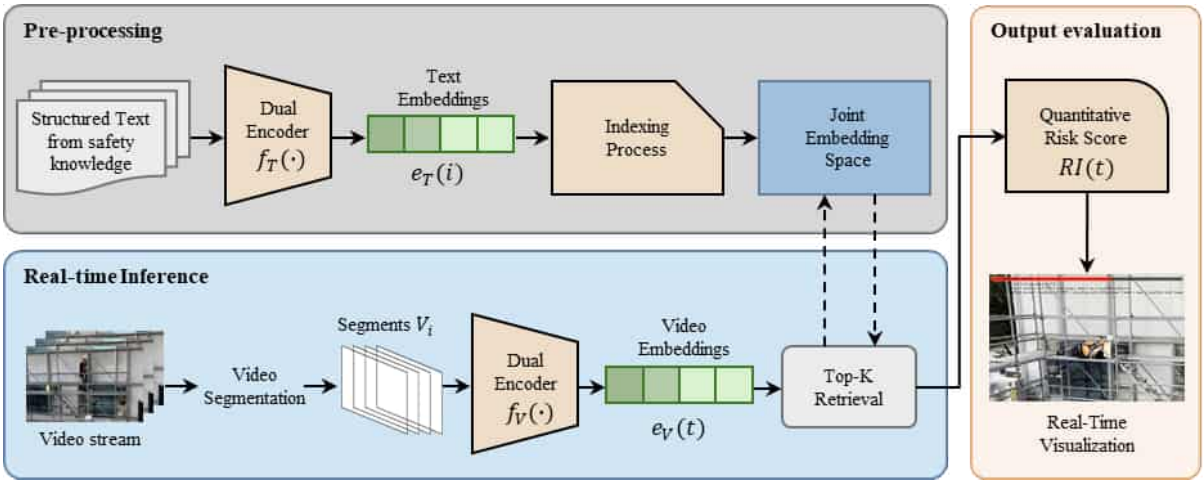


Figure 1. Research framework

overlapping segments along the timeline. All subsequent steps—embedding extraction, evidence retrieval, risk score computation, and visualization—are performed at the segment level.

3.3 Video Embedding Extraction

Each segment V_t is encoded into a video embedding $e_V(t) \in \mathbb{R}^d$ using the video encoder $f_V(\cdot)$ of the same dual-encoder model. Given the consecutive frames that constitute the segment, the video encoder summarizes the scene information over the corresponding time interval into a single fixed-length vector. The resulting video embedding is represented in the same joint embedding space $\mathbb{R}^d$ as the text embeddings and is subsequently used to select evidence texts via similarity-based retrieval.

3.4 Similarity-Based Retrieval of Evidence Texts

Given the video embedding $e_V(t)$ at time t, we retrieve the top-K texts from the text-bank embedding set $\{e_T(i)\}$ with the highest cosine similarity to form an evidence set $E_K(t)$. The video–text matching score is defined as the cosine similarity $\mathrm{sim}(\cdot)$, computed as the inner product after applying ℓ_2-normalization to all embeddings:

$$E_K(t) = \mathrm{TopK}_i(\mathrm{sim}(e_V(t), e_T(i))). \quad (1)$$

Here, $E_K(t)$ denotes the candidate set of evidence texts that describe the scene observed at time t.

3.5 Evidence-Grounded Risk Score

Based on the retrieved evidence set $E_K(t)$, we compute a segment-level Risk Score $RS(t)$. Our formulation relies on two assumptions: (i) in hazardous scenes, the proportion of unsafe evidence among the retrieved texts increases, and (ii) evidence with higher retrieval scores (similarities) provides more direct explanatory power for risk assessment.

Each text c is assigned a binary label $u(c) \in \{0,1\}$ indicating whether it constitutes unsafe evidence. In this study, texts that contain explicit worker-centered risk cues—such as missing PPE and proximity to missing guardrails/openings—are labeled as $u(c) = 1$ (unsafe), whereas texts describing normal working conditions are labeled as $u(c) = 0$ (safe). These labels are assigned consistently using a rule-based procedure grounded in the OSHA guide A Guide to Scaffold Use in the Construction Industry and the AI-Hub real-time video dataset for work-at-height scenarios [19]. For each evidence text $c \in E_K(t)$, its weight $w(c)$ is set to the cosine similarity returned by the retrieval step.

With these definitions, the risk score is computed as the weighted proportion of unsafe evidence among the retrieved top-K texts and normalized to the range $[0, 100]$:

$$RS(t) = 100 \cdot \frac{\sum_{c \in E_K(t)} w(c)\, u(c)}{\sum_{c \in E_K(t)} w(c)} \quad (2)$$

Accordingly, $RS(t)$ increases when a larger number of highly similar evidence texts are labeled as unsafe.

3.6 Real-Time Visualization with Evidence Captions

The proposed framework outputs a segment-level risk score $RS(t)$ sequentially over time for streaming inputs, while simultaneously presenting a subset of the retrieved evidence texts (top-N) as evidence captions. This enables operators to inspect both the current risk level and the textual evidence underlying the risk assessment. For real-time monitoring, the final outputs are continuously rendered as an overlay on the video frames.

4 Experiments and Results

4.1 Safety Text Bank Construction and Indexing

In this study, we built a text bank to interpret real-time monitoring scenes in terms of relationships among key risk factors. Each text is structured into five fields—SUBJECT / SUBJECT_STATE / ACTION / LOCATION / RISK—and the vocabulary for each field was defined by extracting and organizing terms from the OSHA guide and the AI-Hub dataset. The numbers of tokens per field are 3 for SUBJECT, 6 for SUBJECT_STATE, 44 for ACTION, 37 for LOCATION, and 19 for RISK, yielding a total of 151,296 texts through combinatorial generation. In addition, each text entry c_i is assigned a binary label $u(c_i) \in \{0,1\}$ indicating whether it constitutes unsafe evidence.

To compare the text bank and video segments within the same embedding space, we employ a pretrained video–text dual encoder, Perception Encoder (PE-Core). PE-Core maps video and text representations into a shared embedding space via contrastive vision–language alignment learned from large-scale data. In our experiments, the embedding dimension is $d = 1024$[20]. For real-time top-K retrieval over the large-scale text bank, we pre-index the text embeddings offline using a FAISS IVF-PQ (IndexIVFPQ) approximate nearest neighbor (ANN) index.

4.2 Streaming Inference Settings

We partition the real-time streaming video into segments of length $t = 2.0$s and apply a sliding window

with stride $S = 0.5$ s to continuously update the risk signal. For each segment, we uniformly sample 16 frames as input to the video encoder to produce a segment embedding $e_V(t)$. In the online stage, we retrieve $K = 10$ texts to construct $E_K(t)$, and in the visualization stage, we select the top $N = 2$ entries as evidence captions and render them as an overlay on the video stream. All streaming experiments were conducted on a single NVIDIA GeForce RTX 5070 Ti (16GB) GPU with Driver 591.44 and CUDA 13.1.

4.3 Dataset and Ground Truth Annotation

To assess field deployability, we recorded three staged video clips in a scaffold environment at Building 208, Chung-Ang University, Seoul, Republic of Korea. The clips correspond to three hazard scenarios: Vertical (1 min 50 s), referring to vertically overlapping work; Mobile (1 min 11 s), referring to mobile-scaffold movement with a worker riding on the scaffold; and Fall (48 s), referring to falling events. For evaluation, we manually assigned segment-level binary ground-truth labels (unsafe = 1, safe = 0) for all clips. Notably, the unsafe label was defined to cover not only the hazardous moment itself, but also the precursor interval immediately before the event and the recovery interval afterward when applicable. In addition, we manually wrote detailed descriptive captions for each segment, which were used to evaluate the interpretability of the retrieved evidence texts.

For baseline comparison, the same segmented clips were additionally evaluated using three open-source vision-language models: Qwen2.5-VL, SmolVLM2-2.2B, and InternVL2.5-2B. Each model was prompted to generate a one-sentence neutral description of the visible construction scene and to classify the scene as safe or unsafe based only on visible evidence. To ensure a consistent output format across models, we used a fixed instruction requiring the response to contain exactly two fields, caption: and scene_safety:.

4.4 Interpretability of Evidence Captions

Table 1 presents a scenario-wise comparison (Vertical, Mobile, and Fall) of how well the descriptions generated or retrieved by each model support the ground-truth (GT) explanatory captions for each segment. The evaluation was conducted using BERTScore-F_1, which measures sentence-level semantic similarity, and SPICE, which reflects structural semantic alignment in terms of objects, actions, and relations. The baseline models (Intern, Qwen, and Smol) generate a single final caption for each segment; therefore, their performance was evaluated using the top-1 F_1 score of that single output. In contrast, the proposed method (Ours) retrieves a set of evidence texts ranked by similarity, and thus reports not only top-1 performance but also the maximum score within the top- K retrieved set (max F_1@5 and max F_1@10).

Table 1. BERTScore Between Evidence Captions and Ground-Truth Text

Scenario	Model	Metric	Top-1 F_1	Max F_1@5	Max F_1@10
Vertical	Intern	BERT	0.888	-	-
Vertical	Qwen	BERT	0.886	-	-
Vertical	Smol	BERT	0.876	-	-
Vertical	Ours	BERT	0.925	0.938	0.941
Mobile	Intern	BERT	0.888	-	-
Mobile	Qwen	BERT	0.891	-	-
Mobile	Smol	BERT	0.879	-	-
Mobile	Ours	BERT	0.925	0.933	0.934
Fall	Intern	BERT	0.893	-	-
Fall	Qwen	BERT	0.891	-	-
Fall	Smol	BERT	0.890	-	-
Fall	Ours	BERT	0.943	0.948	0.950

Overall, the proposed method achieved the highest caption alignment across all scenarios. In terms of BERTScore-F_1, the top-1 F_1 of Ours was 0.925 for Vertical, 0.925 for Mobile, and 0.943 for Fall, while the corresponding max F_1@5 scores increased to 0.938, 0.933, and 0.948, and the max F_1@10 scores further increased to 0.941, 0.934, and 0.950. In contrast, the top-1 BERTScore-F_1 values of the baseline models remained within 0.876-0.888 for Vertical, 0.879-0.891 for Mobile, and 0.890-0.893 for Fall. This indicates that the evidence retrieved by the proposed method maintains a higher degree of semantic consistency with the GT descriptions.

In Table 2 the difference becomes even more pronounced in SPICE. The top-1 SPICE of Ours was 0.618 for Vertical, 0.647 for Mobile, and 0.727 for Fall, while the corresponding max F_1@5 scores increased to 0.679, 0.720, and 0.767, and the max F_1@10 scores further improved to 0.715, 0.734, and 0.779. In contrast, the top-1 SPICE scores of the baseline models were limited to 0.119-0.193 for Vertical, 0.183-0.225 for Mobile, and 0.144-0.190 for Fall. These results suggest that the baseline models fail to sufficiently capture the objects, actions, and spatial relations that constitute the actual scene.

Table 2. SPICE score Between Evidence Captions and Ground-Truth Text

Scenario	Model	Metric	Top-1 F_1	Max F_1@5	Max F_1@10
Vertical	Intern	SPICE	0.193		
Vertical	Qwen	SPICE	0.128		
Vertical	Smol	SPICE	0.119		
Vertical	Ours	SPICE	0.618	0.679	0.715
Mobile	Intern	SPICE	0.225		
Mobile	Qwen	SPICE	0.183		
Mobile	Smol	SPICE	0.208		
Mobile	Ours	SPICE	0.647	0.720	0.734
Fall	Intern	SPICE	0.190		
Fall	Qwen	SPICE	0.144		
Fall	Smol	SPICE	0.178		
Fall	Ours	SPICE	0.727	0.767	0.779

In other words, although the captions produced by existing generative models may appear linguistically natural, they may still fail to accurately describe the key evidence underlying hazardous situations. Therefore, such single generated captions are difficult to rely on directly as interpretable outputs or as a basis for safety judgment. In contrast, the proposed method not only shows high semantic similarity to the GT descriptions but also achieves a substantial advantage in SPICE, confirming that it provides evidence that more accurately reflects the structural meaning of the scene.

In addition, the proposed method consistently maintained strong performance across all three scenarios: Vertical, Mobile, and Fall. This suggests that its descriptive capability is not limited to a particular scenario type, but can robustly reflect both semantic and structural information across diverse work situations and risk contexts. The substantial margin observed in SPICE across all scenarios further indicates that the proposed method more robustly captures scene-level objects, actions, and spatial relations, and therefore can provide highly interpretable evidence under a wider range of site conditions.

4.5 Risk Score Performance

Risk-score performance was evaluated by applying multiple thresholds (0.1-0.9) to the proposed Risk Score, classifying each segment as unsafe if its score exceeded the given threshold and as safe otherwise, and then comparing the resulting predictions with the segment-level ground-truth labels. Table 3 summarizes the segment-level classification performance under each threshold. As the threshold increased from 0.1 to 0.7, Accuracy steadily improved from 88.5% to 97.6%, while Recall remained at 1.00 throughout this range. F1 also increased consistently from 0.936 to 0.985. In contrast, when the threshold was further increased to 0.8 and 0.9, Accuracy remained high at 97.5% and 96.1%, respectively, but Recall decreased to 0.983 and 0.957, and F_1 was 0.985 and 0.977.

Table 3. Risk Score Performance Across Thresholds (Acc. = Accuracy)

Threshold	Acc(%)	Recall	F_1
0.1	88.5	1	0.936
0.2	88.7	1	0.937
0.3	88.7	1	0.937
0.4	90.0	1	0.944
0.5	92.8	1	0.959
0.6	94.9	1	0.970
0.7	97.6	1	0.985
0.8	97.5	0.983	0.985
0.9	96.1	0.957	0.977

These results indicate that gradually increasing the threshold improves classification reliability, but excessively high thresholds may cause some hazardous segments to be missed. In particular, a threshold of 0.7 can be regarded as the most balanced operating point overall, as it achieved the highest Accuracy (97.6%) while simultaneously maintaining Recall of 1.00 and the highest F_1 score (0.985). Therefore, in this study, 0.7 was selected as the representative threshold for the subsequent analysis of temporal variation patterns.

Figure 2 shows the temporal evolution of $RS(t)$ under this representative threshold in the streaming setting ($t = 2.0$s $S = 0.5$s) across three hazard scenarios. In the graph, the risk score rises distinctly during the annotated hazardous intervals and is generally well aligned with the shaded GT-unsafe regions. This suggests that the proposed Risk Score appropriately reflects the timing of hazardous events across different hazard situations, including Vertical, Mobile, and Fall.

Overall, the proposed Risk Score enabled the identification of a stable operating point through threshold analysis, and under this threshold, it showed high segment-level classification performance and favorable temporal alignment across all three hazard scenarios. These results support that, even in a streaming-based safety monitoring environment, the proposed method can achieve both strong detection performance and stable operational behavior.

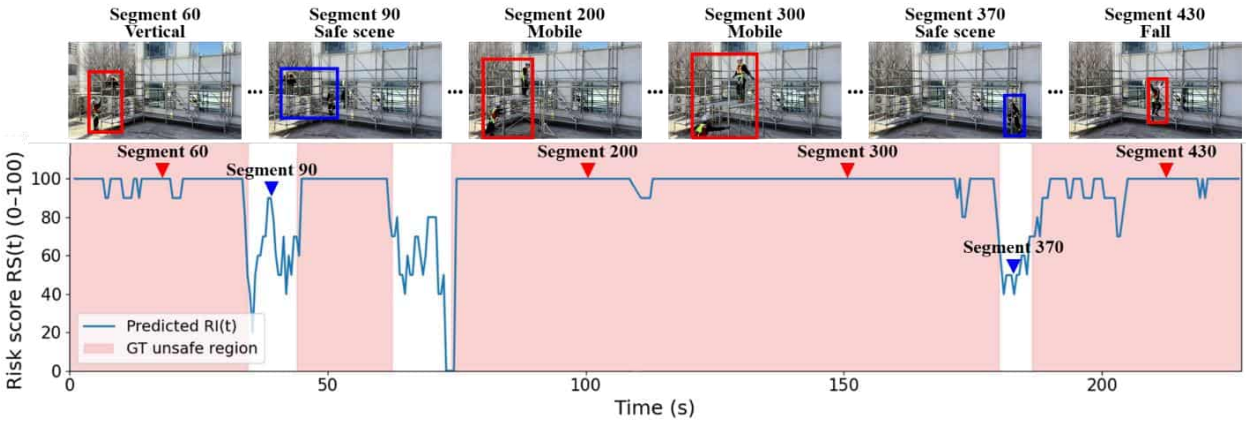


Figure 2. Risk Score Align with Unsafe Events in Streaming Video

4.6 End-to-End Streaming Feasibility

Table 4 reports the end-to-end per-segment processing latency of each model under the streaming stride setting ($S = 0.5$s) to verify whether the system satisfies the real-time condition $L \leq S$.

Table 4. Comparison for Real-Time Operation

Model	Stride	Latency	Realtime
Intern	500.0 ms	1274.3ms	✗
Qwen	500.0 ms	1663.8ms	✗
Smol	500.0 ms	3748.6ms	✗
Ours	500.0 ms	210.8 ms	✓

In this study, real-time capability is defined in terms of timeliness, that is, whether the system can update its output without delay as an event unfolds. Because the streaming pipeline updates its output every 0.5s, real-time operation is considered feasible only when the mean per-segment processing latency L does not exceed the stride S. As shown in Table 3, the proposed method achieved a mean latency of 210.8 ms, which is substantially smaller than the stride of 500.0 ms and therefore satisfies the real-time condition. In contrast, the baseline models, InternVL2.5-2B, Qwen2.5-VL, and SmolVLM2-2.2B, showed mean latencies of 1274.3 ms, 1663.8 ms, and 3748.6 ms, respectively, all of which exceed the stride. This indicates that real-time online operation is difficult for these models under the same setting. These results show that only the proposed method can provide timely segment-level updates within the required streaming interval. Therefore, the proposed framework not only demonstrates effective hazard interpretation performance, but also secures a processing speed that is practically applicable to real-time safety monitoring even in a single-GPU environment.

5 Conclusion

This study proposed a real-time construction safety monitoring framework that interprets streaming site videos through text-bank-based retrieval and provides both a segment-level risk score $RS(t)$, and supporting evidence texts. The framework constructs a structured text bank, aligns video and text in a shared embedding space, and retrieves top-K evidence texts for each segment to compute and visualize risk without site-specific retraining.

Experimental results showed that the proposed Risk Score achieved stable classification performance across threshold settings, with 0.7 identified as the most balanced operating point (Accuracy = 97.6%, Recall = 1.00, and F_1 = 0.985). Under this threshold, the temporal response of the risk score was generally well aligned with the GT-unsafe intervals across the three hazard scenarios: Vertical, Mobile, and Fall. In addition, the proposed method consistently outperformed the generative baseline models in interpretability, showing stronger semantic and structural alignment with the GT explanatory captions.

The proposed framework also demonstrated practical real-time capability. Under the streaming setting of $S = 0.5$ s, the mean end-to-end latency was 210.8 ms, satisfying the condition $L \leq S$, whereas the baseline VLMs exceeded the allowable interval. These results suggest that the proposed method can support low-latency safety monitoring in streaming environments.

This study has several limitations. Validation was conducted on a limited set of hazard scenarios and a single experimental setting, so generalization under more diverse conditions remains to be verified. Future work will extend the framework through broader evaluation, expansion of the hazard taxonomy and text bank, and adaptive alarm strategies based on temporal accumulation, smoothing, and dynamic thresholds.

Acknowledgement

This research was supported by a grant(RS-2022-00143493) from Digital-Based Building Construction and Safety Supervision Technology Research Program funded by Ministry of Land, Infrastructure and Transport of Korean Government.

This research was supported by the Young Researcher Program (RS-2026-25529211) through the Korea Agency for Infrastructure Technology Advancement (KAIA), funded by the Ministry of Land, Infrastructure and Transport of the Republic of Korea.

References

[1] S. F. A. Zaidi, J. Yang, M. S. Abbas, R. Hussain, D. Lee, and C. Park. Vision-based construction safety monitoring utilizing temporal analysis to reduce false alarms. Buildings, 14(6):1878, 2024.

[2] Y. Li, H. Wei, Z. Han, N. Jiang, W. Wang, and J. Huang. Computer vision-based hazard identification of construction site using visual relationship detection and ontology. Buildings, 12(6):857, 2022.

[3] W. Fang, L. Ding, P. E. Love, H. Luo, H. Li, F. Peña-Mora, et al. Computer vision applications in construction safety assurance. Automation in Construction, 110:103013, 2020.

[4] X. Li, T. Hao, F. Li, L. Zhao, and Z. Wang. Faster R-CNN-LSTM construction site unsafe behavior recognition model. Applied Sciences, 13(19):10700, 2023.

[5] W. Sultani, C. Chen, and M. Shah. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE CVPR, pages 6479–6488, 2018.

[6] X. Zhou, A. Arnab, S. Buch, S. Yan, A. Myers, X. Xiong, et al. Streaming dense video captioning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 18243–18252, 2024.

[7] S. K. Lee and J. H. Yu. Ontological inference process using AI-based object recognition for hazard awareness in construction sites. Automation in Construction, 153:104961, 2023.

[8] Y. R. Lee, S. H. Jung, K. S. Kang, H. C. Ryu, and H. G. Ryu. Deep learning-based framework for monitoring wearing personal protective equipment on construction sites. Journal of Computational Design and Engineering, 10(2):905–917, 2023.

[9] M. S. Abbas, R. Hussain, S. F. A. Zaidi, D. Lee, and C. Park. Computer vision-based safety monitoring of mobile scaffolding integrating depth sensors. Buildings, 15(13):2147, 2025.

[10] C. Ding, Q. Liu, X. Guo, T. Xue, and Z. Wang. Identifying unsafe behaviors of construction workers through an unsupervised multi-anomaly GAN approach. Automation in Construction, 165:105509, 2024.

[11] H. Du, S. Zhang, B. Xie, G. Nan, J. Zhang, J. Xu, et al. Uncovering what why and how: A comprehensive benchmark for causation understanding of video anomaly. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18793–18803, 2024.

[12] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge, S. Song, et al. Qwen2.5-VL technical report. arXiv preprint arXiv:2502.13923, 2025.

[13] A. Marafioti, O. Zohar, M. Farré, M. Noyan, E. Bakouch, P. Cuenca, et al. SmolVLM: Redefining small and efficient multimodal models. In Conference on Language Modeling (COLM), 2025.

[14] Z. Chen, J. Wu, W. Wang, W. Su, G. Chen, S. Xing, et al. InternVL: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 24185–24198, 2024.

[15] Q. M. Dinh, M. K. Ho, A. Q. Dang, and H. P. Tran. Trafficvlm: A controllable visual language model for traffic video captioning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7134–7143, 2024.

[16] B. Xiao, Y. Wang, Y. Zhang, C. Chen, and A. Darko. Automated daily report generation from construction videos using ChatGPT and computer vision. Automation in Construction, 168:105874, 2024.

[17] D. Yang, M. Kim, S. Mac Kim, B. W. Kwak, M. Park, J. Hong, et al. Llm meets scene graph: Can large language models understand and generate scene graphs? A benchmark and empirical study. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), Volume 1: Long Papers, pages 21335–21360, 2025.

[18] S. M. H. Abidi, S. M. Raza, and S. Y. Shin. SafeVision: Vision-language reasoning for context-aware safety monitoring. Neurocomputing, 669:132479, 2026.

[19] Occupational Safety and Health Administration (OSHA). A Guide to Scaffold Use in the Construction Industry. OSHA Publication 3150 (Revised 2002), 2002.

[20] D. Bolya, P.-Y. Huang, P. Sun, J. H. Cho, A. Madotto, C. Wei, T. Ma, J. Zhi, J. Rajasegaran, H. A. Rasheed, J. Wang, M. Monteiro, H. Xu, S. Dong, N. Ravi, S.-W. Li, P. Dollár, and C. Feichtenhofer. Perception Encoder: The best visual embeddings are not at the output of the network. NeurIPS, 2025

NUS3D: A BIM-Synthetic, Mobile-Scan and TLS Point Cloud Dataset for Robotic 3D Scene Understanding in Built Environment

Vincent J.L. Gan[1], Mingkai Li[1], Jingxuan Li[1], Chao Yin[2], Boyu Wang[3]

[1] National University of Singapore, Singapore
[2] Hong Kong University of Science and Technology, HK
[3] New York University (Abu Dhabi), UAE

Abstract –
Representation learning for 3D scenes in the architecture, engineering and construction domains is becoming important, as it facilitates Scan-to-BIM, 3D perception and robotic deployment in built environments. This study presents NUS3D, a dataset that consolidates BIM-synthetic, mobile-scan and terrestrial laser scan (TLS) point clouds for building interiors. Originated in 2022, we attempted to integrate heterogeneous data sources to capture the spectrum of 3D data characteristics including structural and M&E (mechanical and electrical) components. BIM-synthetic point clouds provide noise-free geometry that serves as a design-intent reference, mobile-scan contributes dynamic observations collected in real-world conditions, TLS offers high-precision and dense measurements of as-built scenes. Such mixed-source composition enables training and evaluation of algorithms for representation learning, allowing models to acquire generalizable spatial and geometric features that are transferable across synthetic and real, static and dynamic, as well as design and as-built domains. This paper includes a semantic segmentation test using several BIM and TLS M&E point clouds. Experiment demonstrates the potential of NUS3D for advancing future research on 3D scene understanding, Scan-to-BIM, and robotic perception in the built environment.

Keywords –
Point Cloud; 3D Scene Understanding; Scan-to-BIM; Robotic Perception.

1 Introduction

Robots are increasingly integrated into built environments [13], yet autonomous perception in these settings remains highly challenging. Such environments range from unstructured construction sites, where layouts change over time, to semi-structured spaces with known geometry but movable elements like furniture, equipment and human activity present. Unlike controlled conditions, these spaces are in constant flux such as clutter, moving obstacles, and human presence alter scene geometry, complicating environment perception.

Point cloud semantic segmentation serves as a key enabler for 3D scene understanding. In practice, 3D point clouds can be obtained through passive reconstruction or active sensing. Passive reconstruction typically relies on cameras to infer geometric information. This includes monocular approaches that estimate depth from a single image, often supported by deep learning techniques [7]. Stereo or multi-view approaches exploit disparity or multi-view geometry (structure-from-motion, multi-view stereo) to generate point clouds [25]. In contrast, active sensing relies on external sensors to directly acquire geometry of the object. For ground-based measurements, LiDAR (Light Detection and Ranging) is more capable of producing dense point clouds in large-scale environments. Among these systems, terrestrial laser scanners (TLS) are particularly effective, offering superior accuracy and measurement range. However, TLS does not support real-time point loud registration that are usually required in scanning complex, occluded environments where the sensor should move around the built environment [13].

As such, researchers are looking into mobile scanning (MLS) in which LiDAR are leveraged with robot to improve the mobility of 3D scanning [13,28]. In this regard, Simultaneous Localization and Mapping (SLAM) has gained increasing attention. SLAM enables robots to determine their position within an environment while simultaneously generating a 3D map of that environment. It allows real-time tracking of the scanning system while reconstructing the 3D scene [10,14]. This approach enhances the efficiency of mobile LiDAR mapping and point cloud acquisition, particularly in indoor building and construction environments where cluttered and occluded spaces are commonly encountered.

In recent years, several 3D point cloud datasets have been released (see **Table 1**). Notably, ModelNet [26] is one of the earliest collections, comprising CAD models primarily intended for 3D object classification,

recognition and shape analysis. Semantic3D [12] provides a benchmark for outdoor semantic segmentation, featuring extensive terrestrial laser scans of streets, buildings, vegetation, and other urban structures. For autonomous driving and navigation, KITTI [11] remains a widely used dataset, containing road scenes captured with LiDAR and stereo cameras for benchmarking 3D detection and SLAM. Extending this effort, nuScenes [4] incorporates a rich set of LiDAR, camera, and radar recordings across urban traffic environments.

The TUM RGB-D dataset [23] focuses on visual SLAM evaluation, offering indoor RGB-D sequences with ground truth trajectories, and has become a de facto standard for benchmarking SLAM. Meanwhile, ScanNet [6] introduces richly annotated RGB-D indoor scans, including reconstructed meshes and semantic labels, which have since become a foundation for indoor scene parsing tasks. Similarly, the Stanford Large-Scale 3D Indoor Spaces (S3DIS) dataset [2] offers labeled point clouds of complex building interiors. Both the ScanNet and S3DIS were employed in Scan-to-BIM studies.

Table 1. Representative 3D datasets used in computer vision, robotics and built environment research

Dataset		Description	Reference
ModelNet	CAD (indoor objects)	CAD models for object classification, recognition, and shape analysis.	[26]
Semantic 3D	Outdoor urban scenes	Large-scale outdoor laser point cloud dataset including streets, buildings, churches, and vegetation.	[12]
KITTI	Urban scenes	Classic dataset for autonomous driving with LiDAR, GPS/IMU for 3D detection.	[11]
nuScenes	Urban scenes	Dataset covering roads, including LiDAR, cameras, and radars.	[4]
TUM RGB-D SLAM	Indoor SLAM	RGB-D sequences with ground-truth trajectories.	[23]
ScanNet	Indoor scenes (RGB-D scans)	Real indoor scene with RGB-D sequences, reconstructed meshes, and annotations.	[6]
S3DIS	Indoor large building spaces	Point clouds of large-scale indoor buildings with semantic labels.	[2]

Despite the breadth of existing resources, several limitations remain when applying these datasets. Built environments, particularly indoor spaces and building-street interfaces with indoor-outdoor transitions, are characterized by clutter, occlusions, narrow passages, mixed lighting, and weakly textured surfaces, posing substantial challenges that differ from existing datasets. Particularly, such environments often contain dense mechanical, electrical and plumbing components and spatial relationships with structural components.



Figure 1. Prospect for built environment 3D dataset

Another limitation lies in the method used to generate these datasets. The type and quality of point cloud data available from passive and active sensing, as well as static or dynamic mobile approaches involve different trade-offs in terms of accuracy, completeness, noise and operational efficiency. Looking ahead, built environment datasets should include multiple data sources such as BIM-synthetic point clouds, which represent noise-free design intent; mobile scans that capture dynamic observations under real-world conditions; and TLS data, which offer dense, high-precision ground truth. This mixed-source composition supports robust training and evaluation of 3D representation learning models with improved generalization. Specifically, there is growing momentum toward robotic data acquisition (see **Figure 1**). Integrating mobile robotic LiDAR mapping with

BIM-synthetic and TLS data enables a holistic dataset for BIM-to-Scan and Scan-to-BIM verification, supporting reliable structural semantic segmentation, surface-level feature extraction, and enhances 3D scene understanding.

This paper presents our ongoing efforts to develop NUS3D, a hybrid dataset that consolidates BIM-synthetic, mobile-scan, and high-precision TLS point clouds to capture the diverse characteristics of 3D data in built environments, particularly with indoor and indoor-outdoor transitional spaces that are characterized by clutter, occlusions, semi-structured layouts and weakly textured surfaces. This composition supports the development and evaluation of geometric representation learning pipelines, enabling them to acquire generalizable geometric features transferable across synthetic and real data as well as design-intent and as-built domains. Such capabilities are expected to advance scene understanding for environment perception, robotic deployment and Scan-to-BIM research.

2 Related Work

2.1 Data Acquisition for Indoor Environments

In cluttered changing environment, conventional static terrestrial laser scanning remains accurate but it is often time-consuming and limited in flexibility for large or continuously evolving spaces. There is a growing deployment of mobile robots in clutter, unstructured environments, such as scanning building geometry with UAV-UGV [14] and automated mobile robot navigation to produce survey-quality 3D point clouds of construction sites.

Ibrahim, Golparvar-Fard and El-Rayes [17] introduced metrics and methods for assessing the quality of reality capture, particularly in scenarios where unmanned ground rovers are deployed for construction monitoring and inspection. Cheng et al. [5] proposed an improved coverage planning that optimizes waypoint ordering and sequences for robotic data acquisition. Such integration opens the possibility of quality-assurance-driven scanning, where robots adapt their movement trajectories to optimize coverage and improve point cloud quality.

2.2 Point Cloud Quality Assessment

Research has been undertaken to explore how errors affect scene understanding, including those arising from SLAM drift, LiDAR occlusions, and inaccuracies in BIM models. Esfahani et al. [9] provide groundwork by conducting a quantitative investigation into how levels of automation and modeler training affect Scan-to-BIM accuracy, offering predictive models of geometric uncertainty under various workflows.

The establishment of point cloud quality metrics, such as density, coverage and spatial distribution, remains important for defining minimum data requirements for Scan-to-BIM outcomes. Liu et al. [20] contributes toward this goal by introducing BIMNet, an IFC-based large-scale dataset and corresponding benchmark focused on as-built BIM reconstruction from. Furthermore, the deviation analysis method proposed by Anil et al. [1] demonstrates how geometric deviations between point clouds and BIMs can be systematically analysed to locate, characterize, and diagnose errors in as-built models.

2.3 Representation Learning

Another research avenue lies in the evaluation of representation learning and scene understanding, including transformer-based architecture, contrastive learning, and recent 3D Gaussian splatting approaches. For instance, Perez-Perez et al. [22] proposed a learning-based pipeline combining semantic labels and geometric descriptors, leveraging classifiers (SVM, AdaBoost) and Markov Random Field optimization, to segment point clouds for automated 3D model fitting. Semantic-geometric pipeline was demonstrated in industrial contexts enriching RGB-point clouds with instance segmentation for geometry fitting to yield structural and M&E models [21]. Another study [18] proposes an instance-level clustering method for road markings from mobile LiDAR, robust to occlusion and complex geometric conditions across 28 datasets. Complementing these, Hu and Brilakis [16] introduced instance segmentation anchored in design-intent via matching IFC models to as-built scenes using PROSAC-based detection and DBSCAN clustering, which achieves high precision in cluttered, occluded scans.

Another promising direction is the development of geometry-aware AI, in which BIM-derived priors, such as the verticality of walls or the planarity of slabs, are incorporated to regularize predictions in real-world scans. New representation learning pipelines can leverage BIM-synthetic point clouds as training data, combined with domain adaptation to bridge discrepancies between synthetic and real scans [15]. These approaches can be extended to zero- or few-shot scenarios by applying large vision-language models (e.g., OpenMaskDINO3D, SegPoint, Reason3D).

Beyond data capture and representation learning, geometric synchronization is another direction where point cloud data streams are continuously used for geometric updates. Bosché et al. [3] have combined Scan-to-BIM and Scan-vs-BIM techniques in a unified approach for automated comparison of as-built and as-planned M&E works. Future work seeks to explore automatic as-built model updating, where deviations between reality capture and the design model are detected

and translated into structured updates, enabling self-correcting digital representations. Erişen et al. [8] proposed a multi-task domain-adapted deep learning framework capable of generating semantically annotated 3D reconstructions directly from a single RGB image, delivering mesh accuracy and semantic fidelity.

These developments indicate a research trajectory that integrates sensor-agnostic reconstruction (especially mobile scans) with 4D spatiotemporal geometric updates. Together, the research paves the way toward enriched and adaptable 3D scene understanding that evolves with real-world environments.

3 Developing NUS3D Dataset

The proposed NUS3D dataset aims to integrate multiple sources of 3D point clouds to support the testing and verification of scene understanding and environment perception. It comprises three subsets shown in **Table 2**.

BIM-synthetic point clouds are derived from BIM by sampling geometric primitives and surfaces directly from building elements, producing noise-free and complete synthetic representations. Mobile-scan point clouds are captured using active or passive 3D scene reconstruction systems supported by SLAM algorithms. TLS point clouds are obtained through high-resolution time-of-flight laser scanning from ground-based stations, providing dense and accurate data. It includes structural and mechanical and electrical (M&E) components.

Table 2. Structure of NUS3D dataset

Subset	Method	Description
BIM-Synthetic Point Clouds	BIM 3D model	This subset comprises synthetic 3D point clouds generated from BIM providing complete and noise-free geometric representations of building elements.
Mobile-Scan Point Clouds	Handheld or robotic scans with SLAM	This subset comprises real-world 3D point clouds using handheld or robotic scan (e.g., RealSense L515 or SHARE SLAM S20), capturing practical variations in density and noise.
TLS Point Clouds	Terrestrial laser scanner	This subset comprises high-density 3D point clouds using terrestrial laser scanners (e.g., FARO®), providing accurate as-built geometry while reflecting challenges such as occlusion in real environments.

Note: As different sensing devices exhibit varying levels of accuracy, density, and noise characteristics, this paper does not differentiate such sensor-specific variations at a finer level.

3.1 BIM-Synthetic Point Clouds

BIM-synthetic point clouds are synthetic datasets produced by sampling geometric primitives and surfaces directly from building elements within a BIM model. This process ensures that representations are both noise-free and comprehensive, providing ideal ground truth for downstream learning. A notable application of this subset is its use in BIM-to-scan semantic segmentation, where synthetic point clouds serve as training data to overcome the domain gap between as-designed and as-built data. For example, a domain adaptation network, DawNet, augments segmentation performance by combining hybrid attention mechanisms with feature whitening, significantly enhancing accuracy when adapting to noisy real-world scans [15].

3.2 Mobile-Scan Point Clouds

Figure 2 illustrates mobile-scan point clouds conducted since 2022. The environments present dense furniture, narrow corridors, reflective surfaces and severe occlusions, combined with semi-structured layouts and transitions between indoor and outdoor domains.

Handheld active or passive 3D scene reconstruction systems are deployed for data acquisition with the aid of robots for some scenes. The 3D scene reconstruction are undertaken with SLAM to support real-time mobile 3D mapping under real-world conditions. The workflow begins with data capture, as investigated in previous studies [14,24]. LiDAR measures depth by emitting laser pulses and calculating distances based on time-of-flight, generating depth maps with spatial geometry. Then, the data captured are matched via SLAM to incrementally reconstruct dense colorized 3D point clouds along the robot or handheld sensor trajectory.

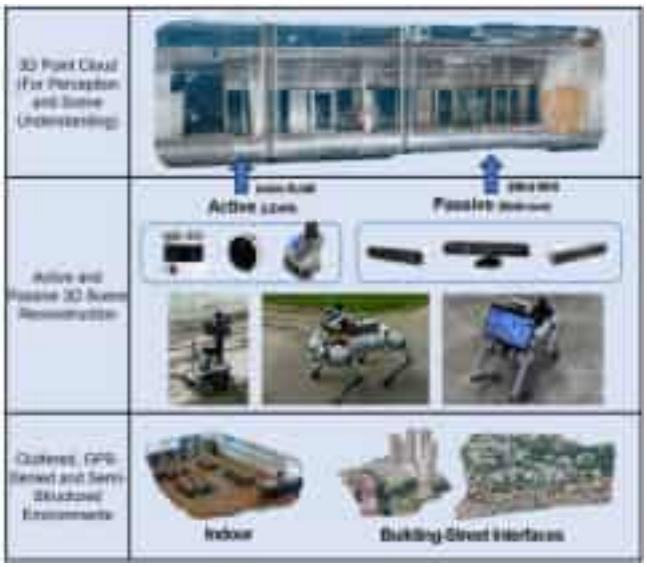


Figure 2. Mobile-scan point cloud acquisition

3.3 TLS Point Clouds

TLS point clouds are generated using high-resolution time-of-flight laser ranging, which captures dense and highly accurate 3D point clouds from ground-based scanning stations. Each point encodes a precise spatial coordinate, enabling faithful reconstructions of architectural and structural elements such as walls, floors, ceilings, columns, and detailed furnishings, including desks, chairs, and equipment. This results in a comprehensive and true-to-life digital representation that supports tasks ranging from spatial measurement and geometric analysis to the built facilities.

Compared with mobile-scan or BIM-synthetic point clouds, TLS exhibits higher density and accuracy, but their quality can still be influenced by environmental conditions. Variations in lighting, weakly textured surfaces and reflective materials across indoor and outdoor environments can introduce challenges in point cloud registration.

4 Cross-Domain Point Cloud Semantic Segmentation and Evaluation

4.1 Semantic Segmentation

In this paper, a segmentation experiment is undertaken as an illustrative example to demonstrate the potential benefit of NUS3D dataset. The evaluation focuses on augmenting TLS data with BIM-synthetic point clouds to explore cross-domain learning. The comparison evaluates TLS-only as well as TLS and BIM-synthetic training, using M&E components. In this study, ResPointNet++ [27] is adopted as a representative backbone for controlled evaluation, while broader benchmarking across additional architectures and tasks will be investigated in future work.

Semantic segmentation aims to assign each point a meaningful interpretation, either in the form of a semantic category or an object-level instance label. Classical geometry-based approaches typically utilize local surface descriptors, such as normals, curvature, and smoothness, to group points with similar differential properties, while model-fitting techniques identify structural primitives (e.g., planes, cylinders) through parametric estimation and consensus-based optimization [3]. In contrast, learning-based methods leverage neural architectures designed for irregular point sets to learn hierarchical feature representations and perform dense, point-wise classification under supervised or weakly supervised settings. The network is trained for 200 epochs on an NVIDIA RTX 4090 GPU.

The experiments are conducted on four real TLS scan M&E scenes (P1-P4) [19] together with four BIM-synthetic scenes (V1-V4). The synthetic point clouds are generated by exporting BIM models to OBJ format and importing them into CloudCompare, where virtual scanner positions are placed to mimic realistic acquisition effects such as occlusion and view-dependent incompleteness. The training and testing scenes are shown in **Figure 3**. To assess the impact of synthetic data, two training regimes are examined under identical hyperparameter settings: (i) a model trained solely on real TLS scans, and (ii) a model trained on the combined set of real TLS scans and BIM-synthetic point clouds as shown in **Figure 4**. This experimental design allows for a controlled evaluation of how synthetic augmentation contributes to segmentation performance across different M&E component types.

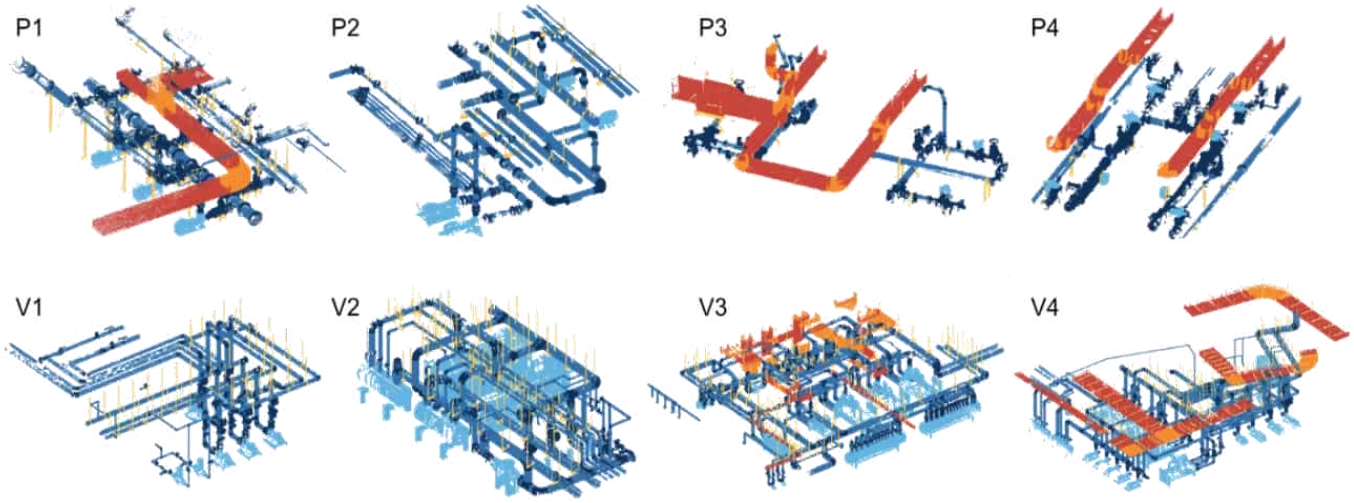


Figure 3. Dataset visualization: training – P1, P2; testing (real) – P3, P4; testing (hybrid) – P3, P4, V1-V4

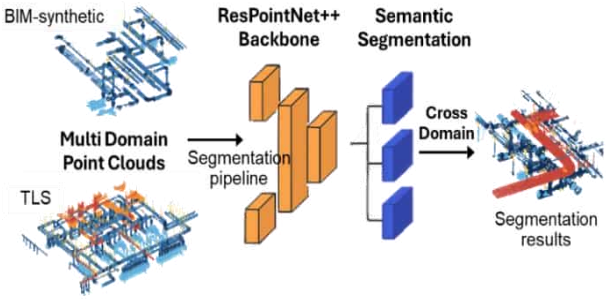


Figure 4. Cross-domain semantic segmentation method

The semantic segmentation output can be refined through a series of post-processing operations, including noise suppression, structural connectivity analysis, and primitive-based geometric fitting. These steps enforce topological consistency, preserve object boundaries, and improve robustness in dense or cluttered environments.

4.2 Evaluation of Segmentation Performance

Evaluation of point cloud segmentation quantitatively assesses the agreement between predicted labels and ground truth. Class-wise performance is typically measured using Intersection over Union (IoU), which evaluates the overlap between the predicted point set and the ground-truth point set of each semantic category. For a given class c, IoU is defined as:

$$\mathrm{IoU}_c = \frac{|\mathcal{P}_c \cap \mathcal{G}_c|}{|\mathcal{P}_c \cup \mathcal{G}_c|} \tag{1}$$

where $\mathcal{P}_c$ and $\mathcal{G}_c$ denote the sets of points predicted and annotated as class c, respectively. The mean Intersection over Union (mIoU), obtained by averaging IoU over all C classes, provides a class-balanced indicator:

$$\mathrm{mIoU} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{IoU}_c \tag{2}$$

Additional metrics include Overall Accuracy (OA):

$$\mathrm{OA} = \frac{\sum_{c=1}^{C} |\mathcal{P}_c \cap \mathcal{G}_c|}{N} \tag{3}$$

where N is the total number of points. OA measures the proportion of correctly classified points across all categories. Furthermore, per-class accuracy is reported to reveal the model's behavior on individual component types, which is particularly important under class imbalance or when different M&E components exhibit varying geometric complexity.

5 Results and Discussion

Table 3 presents the segmentation performance. The segmentation results are also visualized in **Figure 5**. Incorporating synthetic scenes into the training set yields substantial performance gains across nearly all M&E categories. The model trained solely on real scans exhibits limited generalization, particularly for components that are sparse or absent in the real dataset. This leads to extremely low IoU values for Duct fittings (8.93%), Hangers (32.70%), and Machinery (0.17%), while even common classes such as Ducts and Pipes remain suboptimal. In contrast, the hybrid training regime significantly boosts per-class performance: Machinery (+34.27%), Hangers (+27.32%), and Duct fittings (+24.37%) show the largest increases, and even major categories such as Ducts and Pipes improve by 15.00% and 8.27%, respectively. Overall, mIoU increases from 35.37% to 54.86% (+19.49%), and OA improves by more than 11%.

These results highlight the value of supplementing TLS real scans with BIM-synthetic point clouds. The BIM-synthetic data broadens the geometric and structural variability of component instances, enabling the network to better recognize rare, small, or partially occluded elements in real evaluations. The experimental results further demonstrate the potential of the NUS3D hybrid dataset to support the training and improvement of 3D

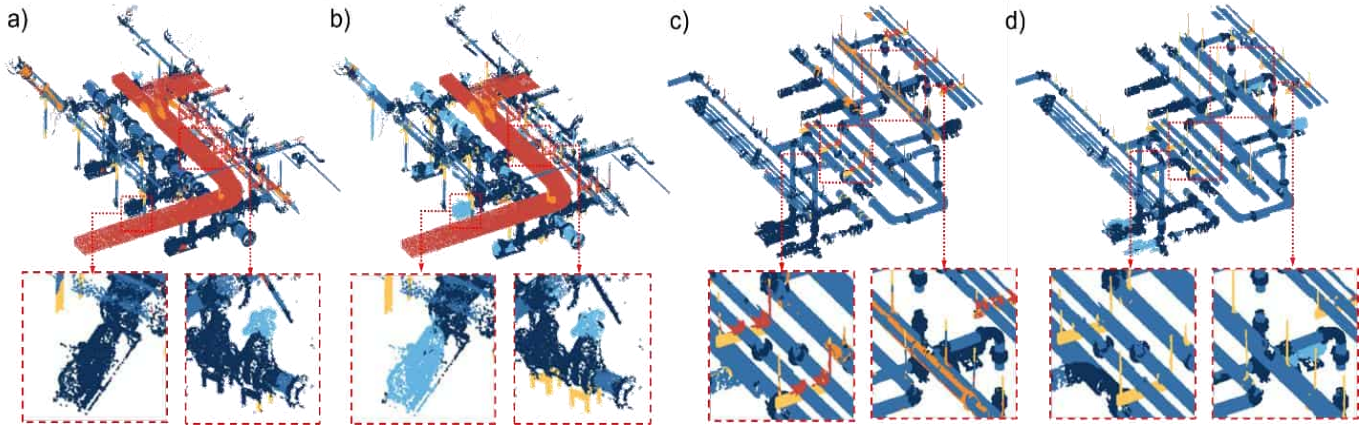


Figure 5. Semantic segmentation results: a) Real – P1; b) Hybrid – P1; c) Real – P2; d) Hybrid – P2

representation learning models. By jointly leveraging BIM-synthetic and TLS data, the dataset improves generalization across design and as-built domains.

Table 3. Segmentation accuracy

Category	TLS	TLS + BIM	Improve-ment
Duct	56.38	71.37	+15.00
Duct fitting	8.93	33.31	+24.37
Hanger	32.70	60.02	+27.32
Machinery	0.17	34.44	+34.27
Pipe	69.21	77.48	+8.27
Pipe fitting	44.85	52.55	+7.70
mIoU	35.37	54.86	+19.49
OA	66.84	77.96	+11.12

6 Conclusions

This study presents ongoing efforts for developing a hybrid dataset that integrates BIM-synthetic, mobile-scan and TLS point clouds to capture the diverse geometric characteristics of the built environment. NUS3D bridges the gap between design-intent and as-built representations, providing a unified foundation for 3D representation learning and scene understanding for scan-to-BIM and robotic perception. The dataset enables evaluation of algorithms for semantic segmentation, geometric reconstruction, and cross-domain transfer between synthetic and real, static and dynamic, as well as design and as-built domains. Demonstrations on BIM-synthetic and TLS-acquired M&E scenes further illustrate its capability.

Looking ahead, our research will be extended to incorporate mobile-scan point clouds, further enriching the dataset diversity. A cross-domain representation learning pipeline focused on mobile 3D scanning with BIM and TLS point clouds will be investigated with applications to robotic perception, construction progress monitoring and scan-to-BIM and BIM-to-scan verification. These capabilities will support research on point cloud representation learning and generalizable 3D scene understanding.

Acknowledgements

This research is supported by Singapore MOE AcRF Tier-1 Grant (Ref: A-8004304-00-00). Any opinions, findings and conclusions or recommendations expressed in the material are those of the author(s) and do not reflect the views of the grantors.

The dataset presented in this paper integrates point clouds collected by the research group since 2022. The authors would like to express their appreciation to researchers affiliated with the lab and Centre for Digital Building Technology. Users are encouraged to cite this paper as the overarching reference, and, where applicable, cite the corresponding subset publications when using the data.

References

[1] E.B. Anil, P. Tang, B. Akinci, D. Huber, Deviation analysis method for the assessment of the quality of the as-is Building Information Models generated from point cloud data, Automation in Construction 35 (2013) 507-516.

[2] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, S. Savarese,, 3D Semantic Parsing of Large-Scale Indoor Spaces, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 1534–1543.

[3] F. Bosché, M. Ahmed, Y. Turkan, C.T. Haas, R. Haas, The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components, Automation in Construction 49 (2015) 201-213.

[4] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom,, nuScenes: A multimodal dataset for autonomous driving, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 11621–11631.

[5] Z. Chen, H. Wang, K. Chen, C. Song, X. Zhang, B. Wang, J.C.P. Cheng, Improved coverage path planning for indoor robots based on BIM and robotic configurations, Automation in Construction 158 (2024) 105160.

[6] A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, M. Nießner,, ScanNet: Richly-annotated 3D reconstructions of indoor scenes, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 2432–2443.

[7] D. Eigen, C. Puhrsch, R. Fergus,, Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS) (2014) 2366 - 2374.

[8] S. Erişen, M. Mehranfar, A. Borrmann, Single Image to Semantic BIM: Domain-Adapted 3D Reconstruction and Annotations via Multi-Task Deep Learning, Remote Sensing 17 (16) (2025) 2910.

[9] M.E. Esfahani, C. Rausch, M.M. Sharif, Q. Chen, C. Haas, B.T. Adey, Quantitative investigation on the accuracy and precision of Scan-to-BIM under different modelling scenarios, Automation in Construction 126 (2021) 103686.

[10] V.J. Gan, D. Hu, Y. Wang, R. Zhai, Automated indoor 3D scene reconstruction with decoupled mapping using quadruped robot and LiDAR sensor, Computer‐Aided Civil and Infrastructure Engineering 40 (15) (2025) 2209-2226.

[11] A. Geiger, P. Lenz, C. Stiller, R. Urtasun,, Vision meets robotics: The KITTI dataset, International Journal of Robotics Research (2013) 1231–1237.

[12] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, M. Pollefeys,, Semantic3D.net: A new large-scale point cloud classification benchmark, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2017) 91–98.

[13] D. Hu, V.J. Gan, C. Yin, Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors, Automation in Construction 152 (2023) 104949.

[14] D. Hu, V.J.L. Gan, T. Wang, L. Ma, Multi-agent robotic system (MARS) for UAV-UGV path planning and automatic sensory data collection in cluttered environments, Building and Environment 221 (2022) 109349.

[15] D. Hu, V.J.L. Gan, R. Zhai, Automated BIM-to-scan point cloud semantic segmentation using a domain adaptation network with hybrid attention and whitening (DawNet), Automation in Construction 164 (2024) 105473.

[16] Z. Hu, I. Brilakis, Matching design-intent planar, curved, and linear structural instances in point clouds, Automation in Construction 158 (2024) 105219.

[17] A. Ibrahim, M. Golparvar‐Fard, K. El‐Rayes, Metrics and methods for evaluating model‐driven reality capture plans, Computer‐Aided Civil and Infrastructure Engineering 37 (1) (2022) 55-72.

[18] J. Jung, E. Che, M.J. Olsen, C.E. Parrish, Y. Turkan, S. Yoo, Instance-Based Clustering of Road Markings with Wear and Occlusion from Mobile Lidar Data, Journal of Computing in Civil Engineering 38 (4) (2024) 04024021.

[19] M. Li, V.J.L. Gan, B. Wang, Integrating hierarchical segmentation and vision-language reasoning for spatially complex and occluded MEP point clouds, Automation in Construction 179 (2025) 106455.

[20] Y. Liu, H. Huang, G. Gao, Z. Ke, S. Li, M. Gu, Dataset and benchmark for as-built BIM reconstruction from real-world point cloud, Automation in Construction 173 (2025) 106096.

[21] F. Noichl, Y. Pan, M.S. Mafipour, A. Braun, I. Brilakis, A. Borrmann, From enriched point cloud to structural and MEP models: An automated approach to create semantic-geometric models for industrial facilities, Computing in Civil Engineering 2023, 2023, pp. 92-99.

[22] Y. Perez-Perez, M. Golparvar-Fard, K. El-Rayes, Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment, Automation in Construction 125 (2021) 103584.

[23] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers,, A benchmark for the evaluation of RGB-D SLAM systems, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2012) 573–580.

[24] T. Wang, V.J.L. Gan, Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning, Automation in Construction 149 (2023) 104810.

[25] T. Wang, V.J.L. Gan, Enhancing 3D reconstruction of textureless indoor scenes with IndoReal multi-view stereo (MVS), Automation in Construction 166 (2024) 105600.

[26] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao,, 3D ShapeNets: A deep representation for volumetric shapes, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015) 1912–1920.

[27] C. Yin, B. Wang, V.J.L. Gan, M. Wang, J.C.P. Cheng, Automated semantic segmentation of industrial point clouds using ResPointNet++, Automation in Construction 130 (2021) 103874.

[28] R. Zhai, J. Zou, V.J. Gan, X. Han, Y. Wang, Y. Zhao, Semantic enrichment of BIM with IndoorGML for quadruped robot navigation and automated 3D scanning, Automation in Construction 166 (2024) 105605.

Discrete-Event Simulation–Based Optimization of Time Buffers for Space-Constrained Modular Production Lines

Nima Moghimi[1*], Sahar H. Shamaee[2], Leonard Irerih[3], Qipei Mei[4], Vicente A. Gonzalez[5], Farook Hamzeh[6]

[1*]Corresponding Author, Ph.D. Student, Department of Civil and Environmental Engineering, University of Alberta, Canada
[2]Graduate Research Assistant, Department of Civil and Environmental Engineering, University of Alberta, Canada
[3]MSc Student, Department of Civil and Environmental Engineering, University of Alberta, Canada
[4]Assistant Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
[5]Hal Kvisle Professor and Tier 1 Canada Research Chair in Digital Lean Construction, Infrastructure Human Tech (IHT) Lab, Strategic Projects Insight Centre in Engineering (SPICE), Department of Civil and Environmental Engineering, University of Alberta, Canada
[6]Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
nmoghimi@ualberta.ca, hamedsha@ualberta.ca, irerih@ualberta.ca, qipei@ualberta.ca, vagonzal@ualberta.ca, hamzeh@ualberta.ca

Abstract

Modular construction offers significant efficiency benefits over traditional methods, yet its production systems remain vulnerable to workflow variability caused by product customization and manual operations. While buffering is a standard strategy to mitigate such variability, the unique constraints of modular factories (e.g., specifically the large physical footprint of modules and high labor costs) often render traditional inventory and capacity buffers impractical. This study proposes a novel time buffer allocation methodology to absorb process variability and protect throughput in space-constrained modular production lines. Adopting a simulation-optimization approach, the study integrates Discrete-Event Simulation (DES) with the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm to determine optimal project-specific time slacks. The framework is evaluated through a case study of a prefabricated timber wall assembly line in Edmonton, Canada. The results demonstrate that the optimized time buffer strategy reduces the workload at the bottleneck loading station by approximately 20% compared to a continuous flow scenario, effectively preventing starvation. Crucially, the optimized approach avoids the upstream blocking and excessive Work-in-Progress (WIP) accumulation observed under uniform buffering policies. The findings provide factory managers with a practical, data-driven method for scheduling releases that harmonizes upstream production efficiency with downstream logistics constraints.

Keywords –
Modular Construction; Simulation-Optimization; Buffer Allocation; Discrete-Event Simulation; Time Buffers

1 Introduction

Modular construction offers significant advantages over traditional methods, including reduced project schedules and improved quality control [1]. However, the operational reality of modular factories is characterized by substantial variability. The make-to-order nature of the industry necessitates extensive customization in component geometry, structural details, and finishing requirements, which introduces heterogeneous processing times at individual workstations [2,3]. This inherent variability, compounded by manufacturing tolerances and material handling uncertainties, frequently leads to production imbalances, bottleneck formation, and reduced overall efficiency [4,5].

To mitigate such variability, manufacturing systems typically employ inventory, capacity, or time buffers [6]. While effective in high-volume manufacturing, standard buffering strategies face unique hurdles in modular construction. Inventory buffers are often impractical because the large physical footprint of wall panels and volumetric modules severely restricts the available floor space between workstations [7–9]. Similarly, capacity buffering, maintaining excess labor or equipment, is constrained by the high costs of skilled labor, union restrictions, and crew specialization [10–12].

Given physical and economic constraints, time

buffers emerge as the most operationally feasible mechanism for absorbing process variability, utilizing temporal slack to protect critical workstations without requiring the additional space or redundant resources associated with inventory or capacity buffers [13,14]. This approach is well supported by the literature. Building on the foundational concept of strategic temporal protection for release-based scheduling introduced by Gao [15], subsequent work has shown that time buffers can be used to approximate the Buffer Allocation Problem by reformulating traditional integer programming models into more tractable linear formulations [16]. Furthermore, this approach was validated in the construction domain through the development of a stochastic optimization model integrating Monte Carlo simulation with critical chain techniques, which achieved a 37% reduction in project delays and a substantial improvement in schedule reliability [17].

However, existing research on time buffering remains predominantly grounded in Critical Chain Project Management (CCPM), prioritizing project-level schedule protection over the station-to-station flow control required for manufacturing contexts [18,19]. While CCPM methodologies have introduced advanced sizing techniques, including network decomposition [14] and entropy-based methods [20], they do not adequately address the dynamics of flow lines, an area where buffer allocation research has traditionally focused on inventory rather than time [21,22]. Consequently, there is a distinct lack of methodologies tailored to the stochastic dynamics of modular production lines. The unique combination of high process variability, space constraints, and make-to-order production in modular construction creates specific buffering challenges that necessitate station-level stabilization—a gap currently unaddressed by existing optimization frameworks.

This study addresses this gap through a Design-Science approach by developing and evaluating a prescriptive simulation-optimization decision-support artifact for time-buffer allocation in modular prefabrication. By integrating discrete-event simulation [23] with evolutionary optimization techniques [24], the proposed framework identifies project-specific buffer times that mitigate downstream starvation while avoiding excessive upstream blocking, thereby improving overall production performance under space constraints. The artifact is demonstrated and evaluated through an industrial case study at a prefabricated timber construction factory in Edmonton, Canada, providing actionable, data-driven guidance for planners in space-constrained manufacturing environments [25].

2 Methodology

This study adopts a Design Science Research (DSR) methodology [26], which emphasizes the development and evaluation of prescriptive solutions to real-world operational problems. The DSR artifact developed in this paper is a simulation-optimization decision-support framework for project-specific time-buffer allocation in space-constrained modular construction factories. The framework is implemented through a simulation-based optimization engine that recommends release slacks for downstream loading activities. Consistent with DSR, the study is structured around three stages: (i) problem identification and motivation, where spatial limitations and stochastic variability restrict the use of conventional inventory and capacity buffers; (ii) framework design and implementation, where the discrete-event simulation model and the optimization mechanism are developed; and (iii) evaluation, where the proposed framework is assessed in an industrial case study and benchmarked against baseline buffering policies.

2.1 Framework Design: Simulation-Optimization Approach

The proposed framework is designed using a simulation-optimization approach in which the Discrete-Event Simulation (DES) model functions as a black-box objective function [27]. In this framework, the optimizer treats the simulation as an evaluative function f(x) that accepts candidate time buffer allocations as inputs and returns performance metrics as outputs, without requiring explicit mathematical formulation of system dynamics [28]. This black-box approach is particularly suitable for modular construction environments where stochastic variability, customization-induced heterogeneity, and spatial constraints create complexities that resist traditional mathematical optimization methods [6]. The optimizer iteratively explores the buffer allocation space, executing the simulation for each candidate configuration and refining the solution toward optimal configurations based on performance feedback [29].

2.1.1 Discrete-Event Simulation Model Development

The discrete-event simulation model captures the stochastic dynamics of modular construction production lines by representing the system as a sequence of discrete events, such as job arrivals, station completions, and transfers between workstations, each causing a change in system state [30]. Jobs are characterized by distinct types reflecting different product configurations and customization requirements. Processing times at each workstation are sampled from job-specific probability distributions to represent production variability [31]. Blocking and starving are explicitly modelled through

limited intermediate buffers, such that a station may be starved by insufficient upstream supply or blocked by saturated downstream capacity. The primary performance metric is Workload Time, defined as the total active processing time required for a segment to complete its assigned jobs.

2.1.2 Optimization Framework

The optimization framework addresses a constrained single-objective minimization problem. The decision variables are the time-buffer parameters assigned to entities (jobs, tasks, or production batches) within a defined two-stage production line segment.

Let $\mathbf{s} = (s_1, s_2, \dots, s_n)$ denote the vector of time-buffer parameters, where each $s_i \in [s_{\min}, s_{\max}]$ represents the time slack assigned to entity i.

The optimizer treats the simulation as a black-box stochastic function [32]. For each stochastic simulation replication k with random inputs ω_k, the simulation returns:

$W_{\text{up}}(\mathbf{s}, \omega_k)$: The total active processing time accumulated by the upstream stage.

$W_{\text{down}}(\mathbf{s}, \omega_k)$: The total active processing time accumulated by the downstream stage.

The optimization objective is to minimize the expected total workload of the coupled two-stage segment:

$$\min_{\mathbf{s}\in[s_{\min},s_{\max}]^n} \mathbb{E}_\omega\left[W_{\text{up}}(\mathbf{s}, \omega) + W_{\text{down}}(\mathbf{s}, \omega)\right] \quad (1)$$

Since the true expected value is intractable, a Monte Carlo approximation is employed over K simulation replications with independent random seeds $\{\omega_k\}_{k=1}^K$:

$$\hat{J}(\mathbf{s}) = \frac{1}{K}\sum_{k=1}^{K}\left[W_{\text{up}}(\mathbf{s}, \omega_k) + W_{\text{down}}(\mathbf{s}, \omega_k)\right] \quad (2)$$

The problem is subject to simple box constraints:

$$s_{\min} \le s_i \le s_{\max}, i = 1, \dots, n. \quad (3)$$

Operational constraints (resource limits, blocking/starving rules) are enforced internally within the simulation logic.

2.1.3 Optimization Algorithm

Because the objective is evaluated through a discrete-event simulation treated as a black-box stochastic function, and its expectation is approximated via Monte Carlo replications, analytic gradients are unavailable and finite-difference gradients are unreliable due to simulation noise and non-smooth event logic. Therefore, a derivative-free stochastic optimization algorithm is used to iteratively propose candidate buffer vectors *S* and refine the search. In this study, the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is employed due to its effectiveness in handling continuous bounded domains and stochastic objective functions [33].

CMA-ES adapts its search distribution based on the history of evaluated solutions, making it well-suited for expensive black-box optimization. Depending on problem dimensionality and computational budget, alternative metaheuristic approaches such as particle swarm optimization, simulated annealing, or surrogate-assisted methods may also be suitable [34].

2.2 Evaluation Design

The framework is evaluated using an industrial case study of a modular wall-panel production line under space-constrained conditions. Performance is assessed by comparing the optimized time-buffer policy against baseline buffering strategies (e.g., zero buffering and a uniform buffering rule) under identical production scenarios. All candidate policies are evaluated through multiple Monte Carlo replications to account for stochastic variability and to ensure fair comparison of expected workload outcomes for the coupled upstream–downstream segment. The case-study implementation and benchmarking results are presented in the section 3, which operationalizes the framework in an industrial setting and reports the comparative evaluation outcomes.

3 Framework Evaluation

3.1 Case Study Overview

This section presents the implementation of the proposed framework in the wall assembly line of a prefabricated timber factory in Edmonton, Alberta, producing components for residential modular buildings. The factory operates two adjacent plants manufacturing walls, roofs, floors, stairs, decks, and finishing elements. The wall line was selected because it reflects the main characteristics of the target production environment: hybrid automation-manual operations, severely constrained intermediate buffers, and high product variety driven by customization. Together, these features create heterogeneous workstations and specification-dependent processing-time variability. Figure 1 shows the wall assembly line and overall production flow.

The line consists of six interconnected sections: Framing, Sheeting, CNC Nailing, Packaging, Doors and Windows (DW) Installation, and Loading. Framing uses one CNC machine and one crew to assemble multiwall (MW) panel frames, which move through limited-capacity buffers to Sheeting, CNC Nailing, and Packaging. After Packaging, interior panels are routed to storage, while the remaining panels proceed to four parallel DW lanes for finishing before being transferred

by cart to the Loading buffer. At Loading, two crews use a loading cart to place panels onto delivery trailers.

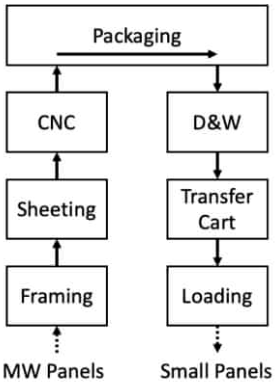


Figure 1: Schematic Layout of the Wall Assembly Line.

3.2 Simulation Model Development

Processing time data were collected through field observation and sampled production records, including panel durations and job specifications (e.g., interior/exterior classification and installation requirements). This provided a practical basis for model parameterization in a data-limited industrial setting, although the input assembly process remained partially manual. A sample of panel records per station was obtained, capturing both quantitative processing times and contextual factors influencing variability. Triangular distributions were selected for modeling processing times due to practical advantages in data-limited environments [35,36]. Parameters were derived from empirical data and validated with shopfloor experts.

The simulation model was developed in Python using SimPy to capture the stochastic dynamics of the wall assembly line. The model represents process flow, resource contention, and buffer limitations using capacity-limited SimPy Store objects enforcing first-in, first-out (FIFO) discipline. Single-capacity stores represent transfer elements; multi-capacity stores model buffers (e.g., DW: 4 lanes × 4 panels = 16 panels capacity; Loading: 6 lanes × 4 panels = 24 panels capacity). Blocking and starving arise naturally from resource constraints. Routing logic dynamically assigns panels at Packaging based on panel type and queue lengths, replicating observed load-balancing behavior.

Model validation against real production data (84 MW panels, 135 small panels) demonstrated acceptable accuracy: Framing ($R^2 = 0.95$), DW ($R^2 = 0.89$), Loading ($R^2 = 0.88$), with mean absolute percentage errors of 16–18% (acceptable for complex construction simulations) [36].

3.3 Optimization Specification

This study focuses on optimizing the time buffer between the upstream production segment (Framing through DW Installation) and the Loading station. Loading, constrained by external trailer logistics, can load a maximum of three projects per day, regardless of project size or upstream production rate. This external constraint positions Loading as a system bottleneck regarding project throughput. Furthermore, the transfer cart section, identified as the internal bottleneck based on panels' expected processing times, feeds directly into the constrained Loading buffer.

Currently, plant management decouples Loading from upstream stations using an average one-day time buffer. However, this uniform buffer policy does not account for project-specific variability or system dynamics. Given the severely limited inventory buffer capacity (24 panels total across six lanes), excessive time buffers cause WIP accumulation that blocks or slows upstream production, while insufficient buffers lead to Loading starving and underutilization [37]. This segment therefore represents an excellent case for testing the effectiveness of the proposed time buffer optimization technique, where the goal is to determine project-specific loading slacks that balance upstream production flow with downstream logistics constraints.

Projects flow through the upstream production segment (Framing through DW Installation), generating production workload W_{prod}. The time buffer s decouples this from the constrained Loading station (workload W_{load}), as illustrated in Figure 2. The optimization objective is to determine per-project buffer allocations that minimize the combined workload $W_{\text{prod}} + W_{\text{load}}$.

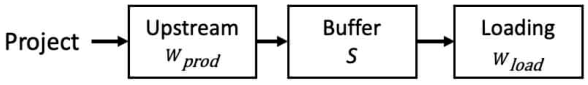


Figure 2: Schematic representation of the proposed time buffer allocation approach.

The selected test schedule comprises 14 projects of varying sizes, mimicking a one-week workload of the production line. For each project i, the decision variable s_i represents the time slack (loading buffer in hours) between when project i reaches the Loading buffer from upstream stations and when it is released to the Loading station. The bounds on s_i are determined by project-specific delivery deadlines: $s_i \in [0, d_i]$, where d_i is the maximum allowable slack for project i based on its deadline. For the plant's operations, typical deadlines following the loading date are within 2–3 days; thus, bounds reflect this practical constraint.

The production workload $W_{\text{prod}}(\mathbf{s}, \omega)$ represents the

total active processing time required by the upstream (i.e., Framing, Sheeting, CNC Nailing, Packaging, DW Installation, and transfer cart stages) to complete all 14 projects, while the loading workload $W_{\text{load}}(\mathbf{s}, \omega)$ measures the Loading station's active processing time for the same project set. The optimization objective is to minimize the expected total workload of the coupled production-Loading system:

$$\min_{\mathbf{s}\in[0,d_1]\times\cdots\times[0,d_{14}]} \mathbb{E}_{\omega}\left[W_{\text{prod}}(\mathbf{s}, \omega) + W_{\text{load}}(\mathbf{s}, \omega)\right] \quad (4)$$

With an effective buffering strategy, the total makespan for finishing the 14-project set is expected to remain relatively constant; however, strategic time buffer allocation changes the distribution of workload and utilization patterns between the production segment and Loading. By decoupling the upstream production from the constrained Loading station through optimized project-specific slacks, the buffer creates a more consistent arrival flow at Loading, smoothing demand variability and reducing starving events. This workload redistribution improves Loading station utilization while preventing upstream blocking, thereby enhancing overall system efficiency [37].

The CMA-ES algorithm was configured with population size proportional to the 14-dimensional decision space and standard initial step size. Monte Carlo estimation used $K = 50$ simulation replications during candidate evaluation and final validation of the best solution, ensuring reliable stochastic objective function approximation [38].

3.4 Results

Three scenarios were simulated to evaluate the proposed time-buffer optimization framework: (1) continuous production with no buffer slack, representing uninterrupted flow; (2) uniform buffering with an 8-hour slack for all projects, reflecting current plant practice; and (3) optimized buffering with project-specific slacks determined by the optimization model. The test schedule included 14 projects of varying sizes released over a four-day horizon. Figure 3 compares the execution timelines of the upstream production segment and the Loading station under the three scenarios.

In the continuous production scenario, the upstream segment completed its work in 31.77 hours, consistent with the four-day production window, but the Loading station required 21.92 hours and extended to a fifth day because of the three-project-per-day trailer logistics constraint. This indicates poor decoupling, with the Loading station frequently starved despite uninterrupted upstream flow.

Under the uniform 8-hour buffer policy, Loading workload decreased to 16.71 hours, indicating fewer starving events and more stable downstream operation. However, upstream workload increased to 37.12 hours, exceeding the four-day window and requiring substantial overtime. This occurred because the uniform slack caused excessive WIP accumulation in the space-constrained buffer, which increased upstream blocking and reduced throughput.

The optimized buffer scenario achieved a better balance between upstream and downstream performance. Loading workload was reduced to 17.11 hours, close to the uniform-buffer case, while upstream workload remained at 31.83 hours, comparable to the continuous-flow scenario and still within the four-day production window. This shows that project-specific slacks can effectively decouple Loading from upstream variability, improving downstream utilization without causing upstream blocking or significant overtime. Although Loading still required five days because of the external logistics constraint, its workload was about 20% lower than in the continuous scenario, creating opportunities to reassign the Loading crew to other value-adding tasks.

The results highlight the critical importance of strategic time buffer allocation in production systems with coupled internal and external bottlenecks. While a uniform buffer policy improves downstream efficiency, it negatively impacts upstream production due to physical space constraints. The proposed optimization framework effectively addresses this challenge by determining project-specific slacks that absorb variability without overwhelming buffer capacity. The optimized solution achieves the best of both scenarios: it maintains upstream production efficiency comparable to a continuous flow model while achieving downstream workload reductions similar to a large uniform buffer, thereby maximizing overall system efficiency and labor productivity. These findings provide a validated foundation for implementing dynamic, data-driven buffering strategies in modular construction factories.

Overall, the results demonstrate that strategic time-buffer allocation is critical in systems with coupled internal and external bottlenecks. A uniform buffer improves downstream stability but degrades upstream performance under space constraints. In contrast, the proposed framework identifies project-specific slacks that absorb variability without overwhelming buffer capacity. As a result, it preserves upstream efficiency while achieving downstream workload reductions similar to the uniform-buffer case. From an implementation perspective, the optimized slacks can be directly incorporated into production schedules as project-specific release delays from upstream production to the Loading buffer, providing managers with simple and actionable scheduling guidance.

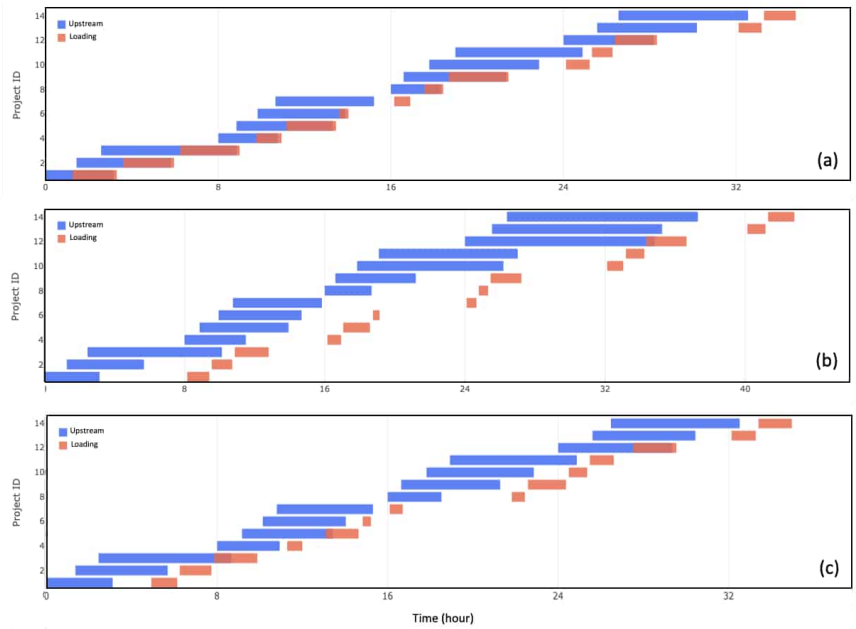


Figure 3: Project execution timelines for upstream segment and loading station under three buffering scenarios: (a) no buffer, (b) uniform 8-hour buffer, and (c) optimized project-specific buffers.

3.5 Practical Implementation

The proposed framework is intended to operate as a decision-support layer on top of the factory's existing planning process rather than as a stand-alone tool that production planners must model manually. In practice, the required inputs are already generated during routine weekly or monthly planning, including the project list, project sequence, target release dates, and station-level production assumptions. These data can be transferred from the plant's scheduling spreadsheets or production management system into the simulation-optimization engine with limited preprocessing.

From an implementation perspective, the main technical requirement is not that every planner or manager possess simulation expertise, but that the organization have access to an industrial engineer, analyst, or digital-production specialist capable of maintaining the simulation model and periodically updating its input parameters. Once this model is established, end users interact only with the results, which can be delivered as a lookup table or simple scheduling rule: for each project in the weekly schedule, apply the recommended loading slack before releasing it to the Loading buffer. In this way, the complex stochastic behavior captured by the DES-CMA-ES framework is translated into concrete release times that are directly usable in operational planning.

This implementation logic minimizes disruption to current manual or semi-automated scheduling practices while still enabling data-driven decision-making. The operational value of the approach lies in converting simulation outputs into actionable project-level dispatch adjustments that reduce Loading crew overtime, improve labor utilization, and maintain upstream production within the expected four-day window. Nevertheless, wider deployment at industrial scale would benefit from a user-friendly interface and automated data exchange between the scheduling platform and the optimization engine, which remains an important direction for future development.

4 Conclusion

This study addressed production variability in modular construction factories where space constraints limit the use of inventory buffers and labor costs restrict capacity buffering. Following a Design Science Research (DSR) methodology, it developed a simulation-optimization framework that combines discrete-event simulation with CMA-ES to allocate project-specific time buffers without requiring additional space or redundant resources. The framework was evaluated on a prefabricated timber wall production line in Edmonton, Canada, as a real-world demonstration case under space-constrained modular production conditions. The optimized time-buffer policy outperformed both zero-buffer (continuous) operation and the plant's uniform buffering practice, reducing bottleneck Loading workload by approximately 20% while preventing downstream starvation. Unlike uniform buffering—which stabilizes the downstream station by blocking upstream work and extending lead times—the optimized

solution maintained upstream efficiency within the targeted four-day window. Overall, the results indicate that time buffers can provide a practical alternative to inventory buffers in space-constrained off-site construction. From a managerial perspective, the framework translates complex stochastic system behavior into simple project-specific scheduling rules that can improve labor utilization and throughput reliability.

While the findings are promising, the present study should be interpreted as a preliminary industrial evaluation with several limitations. First, the framework was assessed using a single demonstration case on one prefabricated timber wall production line, which limits immediate generalizability to other factory layouts, product mixes, and operating conditions. Second, as with any simulation-based study, the analysis depends on modelling assumptions regarding workflow logic, variability representation, and system constraints that may differ across industrial settings. Third, the DES input data in this study were assembled from field observations and sampled production records, which was practical for the case-study context but still required partial manual preprocessing.

Future research should therefore validate the framework across additional production lines, broader project portfolios, and alternative buffering scenarios. It should also extend the method to more complex factory environments, such as multi-line and multi-product systems with interacting bottlenecks. In addition, more systematic data acquisition should be explored through integration with digital production reporting systems, shop-floor sensing, and digital-twin-oriented data pipelines, enabling more automated model updating and more scalable industrial implementation.

5 Acknowledgement

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance grant ALLRP 581074–22, Discovery Grant RGPIN-2020-04420, and NSERC Alliance—Alberta Innovates Advance Program 232403286. All findings and conclusions expressed in this paper are those of the authors and do not reflect those of the contributors.

6 References

[1] Egege CO. Off-site modular construction as a method of improving construction quality and safety. International Journal of Structural and Civil Engineering Research 2018;7:259–68.

[2] Alsakka F, Yu H, El-Chami I, Hamzeh F, Al-Hussein M. Digital twin for production estimation, scheduling and real-time monitoring in offsite construction. Comput Ind Eng 2024;191:110173. https://doi.org/10.1016/j.cie.2024.110173.

[3] Bhatia A, Moselhi O, Han SH. Optimized production scheduling for modular construction manufacturing. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 40, IAARC Publications; 2023, p. 270–7.

[4] Shahtaheri Y, Rausch C, West J, Haas C, Nahangi M. Managing risk in modular construction using dimensional and geometric tolerance strategies. Autom Constr 2017;83:303–15.

[5] Panahi R, Louis J, Podder A, Pless S, Swanson C, Jafari M. Automated progress monitoring in modular construction factories using computer vision and building information modeling. 40th International Symposium on Automation and Robotics in Construction. Retrieved from: the US Department of Energy Office of Scientific and Technical Information: https://www. osti. gov/biblio/2203915, 2023.

[6] Zhang M, Pastore E, Alfieri A, Matta A. Buffer allocation problem in production flow lines: A new Benders-decomposition-based exact solution approach. IISE Trans 2022;54:421–34.

[7] Sadeghi H, Cheung CM, Yunusa-Kaltungo A, Manu P. A systematic review of occupational safety and health in modular integrated construction. Saf Sci 2025;189:106897.

[8] Salhab D, Alsakka F, Hamzeh F. A narrative review of workspace planning in construction: Challenges and insights. Proc., 31st Annual Conf. Int. Group for Lean Construction (IGLC31), 2023, p. 1208–19.

[9] Moghimi N, Arshang O, Hamzeh F. Digital twins in offsite construction: Current implementations, challenges, and future pathways. Autom Constr 2026;183:106820.

[10] Salama T, Salah A, Moselhi O. Integration of linear scheduling method and the critical chain project management. Canadian Journal of Civil Engineering 2018;45:30–40.

[11] Srisuwanrat C, Ioannou PG. The investigation of lead-time buffering under uncertainty using simulation and cost optimization. Proceedings of the 15th Annual Conference of the International Group for Lean Construction, 2007, p. 580–9.

[12] Barkokebas B, Al-Hussein M, Hamzeh F. Assessment of digital twins to reassign multiskilled workers in offsite construction based on lean thinking. J Constr Eng Manag 2023;149:04022143.

[13] Nie X, Zheng Y, Gu X, Su B, Wang B. Calculation of buffer size on critical chain based on duration distribution, multiresource constraints, and relay potential. Sci Program 2022;2022:6591223.
[14] She B, Chen B, Hall NG. Buffer sizing in critical chain project management by network decomposition. Omega (Westport) 2021;102:102382.
[15] Gao H. Building robust schedules using temporal protection: An empirical study of constraint-based scheduling under machine failure uncertainty. 1997.
[16] Alfieri A, Matta A, Pastore E. The time buffer approximated Buffer Allocation Problem: A row–column generation approach. Comput Oper Res 2020;115:104835.
[17] Abuhussain M. Advanced Time Buffer Optimization Model for Construction Schedule Management Using Stochastic Analysis. vol. 2025. 2024.
[18] Zhang J, Song X, Díaz E. Buffer sizing of critical chain based on attribute optimization. Concurrent Engineering 2014;22:253–64.
[19] Zohrehvandia S, Soltanib R. Journal of Project Management. Journal of Project Management 2022;7:121–32.
[20] Sun W, Abdullah LN, Khalid FB, Suhaiza Binti Sulaiman P. EAR-CCPM-Net: A Cross-Modal Collaborative Perception Network for Early Accident Risk Prediction. Applied Sciences 2025;15:9299.
[21] Weiss S. Optimization of buffer allocations in stochastic flow lines 2016.
[22] Kassoul K, Cheikhrouhou N, Zufferey N. Simultaneous allocation of buffer capacities and service times in unreliable production lines. Int J Prod Res 2024;62:644–64.
[23] Afifi M, Al-Hussein M, Abourizk S, Fotouh A, Bouferguene A. Discrete and continuous simulation approach to optimize the productivity of modular construction element. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 33, IAARC Publications; 2016, p. 1.
[24] Rahman M, Han SH. Performance Evaluation of Genetic Algorithm and Particle Swarm Optimization in Off-Site Construction Scheduling. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 41, IAARC Publications; 2024, p. 275–82.
[25] Barkokebas B, Martinez P, Bouferguene A, Hamzeh F, Al-Hussein M. Digitalization-based process improvement and decision-making in offsite construction. Autom Constr 2023;155:105052.
[26] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Quarterly 2004:75–105.
[27] Moghimi N, Kashani H. A probabilistic framework to design the financing scheme for resilience enhancement of infrastructure systems. Reliab Eng Syst Saf 2025;262:111179.
[28] Yeomans JS, Imanirad R, Yang X-S. A stochastic biologically-inspired metaheuristic for modelling-to-generate-alternatives. Lecture Notes in Management Science 2013;5:1–9.
[29] Juan ÁA, Keenan P, Martí R, McGarraghy S, Panadero J, Carroll P, et al. A review of the role of heuristics in stochastic optimisation 2023.
[30] Alsakka F, Darwish M, Yu H, Hamzeh F, Al-Hussein M. The impacts of lean implementation revealed in the course of building a digital twin of a construction manufacturing facility. Proc. 30th Annual Conference of the International Group for Lean Construction (IGLC), Emonton, Canada, 2022, p. 890–901.
[31] Bhatia APS, Han S, Moselhi O. A simulation-based statistical method for planning modular construction manufacturing. Journal of Information Technology in Construction 2022;27.
[32] Moghimi N, Kashani H. Probabilistic Modeling to Develop the Optimal Seismic Resilience Enhancement Strategy for Industrial Facilities. Journal of Infrastructure Systems 2026;32:04025041. https://doi.org/10.1061/JITSE4.ISENG-2754.
[33] Hansen N, Ostermeier A. Completely derandomized self-adaptation in evolution strategies. Evol Comput 2001;9:159–95.
[34] Bianchi L, Dorigo M, Gambardella LM, Gutjahr WJ. Metaheuristics in stochastic combinatorial optimization: a survey. TechReport: Dalle Molle Institute for Artificial Intelligence 2006.
[35] AbouRizk SM, Halpin DW, Sawhney A. Modelling Uncertainty in Construction Simulation. MACHINE INTELLIGENCE AND PATTERN RECOGNITION 1994;17:287.
[36] Law AM, Kelton WD, Kelton WD. Simulation modeling and analysis. vol. 3. Mcgraw-hill New York; 2007.
[37] Hopp WJ, Spearman ML. Factory physics. Waveland Press; 2011.
[38] Tekin E, Sabuncuoglu I. Simulation optimization: A comprehensive review on theory and applications. IIE Transactions 2004;36:1067–81.

Automatic Semantic Parsing of Structural Drawings and Rebar Schedules Based on a Hybrid Deep Learning Framework

Rong-Lu Hong[1], Jin-Bin Im[1], Lijing Xu[1], Wooshin Shim[1], Seunghyeon Wang[1] and Ju-Hyung Kim[1]

[1] Department of Architectural Engineering, Hanyang University, Seoul, Republic of Korea
abraham0814@hanyang.ac.kr, verylyzeng@hanyang.ac.kr, jbim0202@hanyang.ac.kr, js_math@naver.com, wsh1019@hanyang.ac.kr , kcr97jhk@hanyang.ac.kr

Abstract

Reinforcement detailing quality in reinforced concrete structures is directly associated with structural safety and project costs. However, manually cross-checking structural construction drawings against rebar schedules remains inefficient and highly susceptible to human error. Although demand for digital transformation in construction has become increasingly urgent, achieving high-precision automated information extraction continues to be challenging owing to the ultra-high-resolution of engineering drawings and the semantic complexity of domain-specific annotations. To address these issues, this paper proposes a staged visual semantic parsing framework tailored to structural construction drawings and rebar schedules. The proposed approach incorporates an overlapping tiling strategy to preserve local contextual information within large-scale drawings and constructs an end-to-end pipeline that integrates YOLOv11-based text detection with TrOCR-based sequence recognition, enabling accurate dense rebar annotation and tabular data extraction. Experimental results from real-world engineering datasets demonstrate that the proposed method exhibits strong robustness under high-density text conditions and complex background interference, achieving a Character Accuracy of 90.93% (CER = 0.0907) on the held-out test set, outperforming the best baseline by approximately 37.2 percentage points. These findings validate the feasibility and accuracy of leveraging deep-learning techniques for automatic structured information reconstruction from two-dimensional engineering documents, establishing a technical foundation for advancing automated quantity take off, intelligent code-compliance assessment, and digital quality management in the construction industry.

Keywords –

Structural Drawings; Rebar Schedule; Optical Character Recognition; Text Recognition

1 Introduction

The construction quality of reinforced concrete (RC) structures is a fundamental determinant of the lifecycle safety performance and service durability of infrastructure systems [1]. As a core element in ensuring structural load-bearing capacity and long-term durability, reinforcement detailing accuracy not only forms the cornerstone of quality control but also represents a critical variable in project cost management and quantity surveying [2]. Previous statistical studies have indicated that reinforcement costs account for approximately 30% of the total project cost in conventional RC structures [1]. Reinforcement deviations and the resulting rework pose substantial risks to effective cost control [3]. Consequently, in modern engineering projects characterized by a large number of components and highly complex specification systems, project managers are required to rigorously verify and document key parameters, such as reinforcement quantity, diameter, spacing, and layout patterns, as presented in structural construction drawings and rebar schedules, to ensure compliance between construction deliverables and design specifications [2,4].

However, in the digital transformation of the construction industry, automated structural data acquisition poses fundamental challenges. Although building information modeling (BIM) and computer-aided design (CAD) technologies have been widely adopted, 2D rasterized drawings, such as PDF files or scanned documents, remain the legally authoritative single source of truth in critical processes, including multi-party coordination, construction approval, and contractual delivery [4]. While CAD-to-BIM conversion theoretically provides an alternative pathway for data acquisition, its applicability is severely constrained by its strong dependence on strict drafting standards and layer conventions [5]. In complex engineering practice, heterogeneity in design habits, software version

incompatibility, and non-standardized layer management frequently result in significant semantic ambiguity when directly parsing geometric primitives [4,5]. By contrast, computer-vision-based approaches aim to emulate the cognitive processes of human engineers by interpreting information through visual features rather than file attributes [6]. Therefore, developing a generic parsing method that does not rely on specific design software or layer standards and that can deconstruct drawing content directly from the visual domain is essential for achieving true automation and robustness [7].

Although computer-vision techniques have made notable progress in general domains such as document analysis and legal contract review [8,9], their application to structural engineering drawings remains hindered by unresolved technical barriers. Existing general-purpose optical character recognition (OCR) systems are primarily designed for documents with well-aligned layouts and homogeneous backgrounds [7], whereas structural drawings exhibit distinctive characteristics, including high information density, large spatial scale, and weak semantic explicitness [10]. First, structural drawings typically possess extremely high resolutions, often exceeding 4K, and direct input into generic detection models frequently leads to graphic memory overflow or loss of small object features [11]. Second, reinforcement annotations commonly employ highly compressed domain specific syntax, such as "Φ8@200," and are often subject to severe visual overlap with dimension lines, grid lines, and background noise [1,10]. This complex spatial topology and specialized symbolic semantics combination renders generic document-processing techniques insufficient for practical accuracy in engineering scenarios [7].

To address the challenges posed by high-resolution, dense annotations, and compressed semantics in structural construction drawings, this paper proposes a staged visual semantic parsing approach that decomposes the task into text region detection and semantic recognition. In the detection stage, an adaptive overlapping tiling strategy is introduced to preserve local contextual information and reduce misdetections caused by small-scale text fragmentation [12]. Based on this strategy, a YOLOv11-based detection network was employed to achieve stable and accurate text region localization. For semantic recognition, a Transformer-based sequence recognition model was adopted to enhance intercharacter dependency modeling and improve robustness under complex background conditions [13]. By integrating YOLOv11 and TrOCR, an end-to-end parsing framework was constructed, enabling coordinated detection and recognition optimization. Experimental results confirmed the robustness and generalization capability of the proposed method in complex structural drawing scenarios.

2 Related Work

Early studies on text parsing in engineering drawings primarily relied on traditional image-processing techniques and handcrafted feature design methods such as morphological operations and connected component analysis[14]. Using this class of approaches, Ahmed et al. [15] achieved text and symbol extraction from engineering drawings, demonstrating relatively high accuracy under conditions in which drawing formats were fixed and background noise was minimal. Moreno-García et al. [14] combined rule-based reasoning with clustering algorithms to classify symbols and connection relationships in piping and instrumentation diagrams, thereby validating the feasibility of such methods in controlled environments [14,15]. However, these approaches are highly sensitive to environmental conditions, as variations in the scanning resolution, noise interference, symbol overlap, and background clutter can lead to substantial performance degradation. In practical engineering archives, in which drawing quality varies significantly, traditional methods relying on handcrafted features struggle to maintain a stable and robust parsing performance [7].

In recent years, deep-learning approaches, particularly those based on convolutional neural networks (CNNs), have been widely recognized and validated for visual analysis tasks in construction engineering [16]. Compared with traditional image-processing techniques, these methods are more effective in handling complex scenes while simultaneously improving accuracy and robustness, thereby providing strong technical support for construction informatization and intelligent decision-making [2,4]. In the context of the automatic engineering drawing parsing, previous studies commonly modeled text extraction as a two-stage pipeline consisting of text detection followed by text recognition. For example, Ravagli et al. [17] reformulated drawing text extraction as a two-stage task and employed the CTPN along with Tesseract to extract information from architectural floor plans. However, this approach exhibits limited effectiveness when dealing with non-horizontal text. To improve practical performance, Saba et al. [18] integrated an EAST detector with a CRNN-based EasyOCR module and achieved improved performance in terms of handling rotated and inclined text. Similarly, Schönfelder et al. [7] developed a recognition pipeline combining YOLOv7 and PARSeq, covering multiple types of drawing annotations and providing valuable methodological references for subsequent research.

To achieve deeper feature representation and stronger global context modeling, Zhang and He [8] introduced a method based on YOLOv10 and LORE for parsing complex engineering tables, demonstrating that deep-learning models significantly outperform traditional

approaches in terms of handling structured data. Khallouli et al. [9] further demonstrated the advantages of Transformer-based architectures, specifically TrOCR, in processing engineering drawings containing multi-lingual content and complex fonts, with a performance clearly surpassing that of conventional CNN-based baselines. However, within the specific context of RC structural drawings, previous studies have largely been limited to relatively simple tasks, such as text classification or geometric element extraction. For instance, Ma et al. [1] attempted to separate geometric and non-geometric content to extract reinforcement shapes; however, but approach remains insufficient for handling highly compressed composite semantic annotations such as "Φ8@200." Furthermore, existing general-purpose models often overlook the information loss caused by tiling large-format drawings during pre-processing and lack systematic investigation tailored to the joint interpretation of reinforcement schedules and structural drawings.

3 Material and Methods

3.1 Data collection

The dataset used in this study was collected from 31 real engineering projects, all involving the aboveground structural systems of apartment buildings across multiple provinces and cities in China. Each project comprises one structural construction drawing and one corresponding rebar schedule, yielding a total of 62 high-resolution engineering images (31 matched pairs). These projects originate from multiple design firms, encompassing a range of drafting styles and drawing conventions, thereby ensuring a degree of diversity in page layout, font style, and annotation practice within the dataset. The structural component types covered include beams, columns, walls, and foundations. The image resolution ranged from 7560 × 5174 to 15000 × 5914 pixels, reflecting the large-scale and high-information-density characteristics of structural engineering drawings.

As illustrated in Fig. 1(a), structural construction drawings are primarily used to represent the spatial arrangement of structural components and their corresponding identification labels, such as component names denoted by codes such as GBZ1. Fig. 1(b) shows the corresponding rebar schedules, which systematically describe the internal rebar configuration of each component, including the component name, elevation information, quantity and specification of longitudinal reinforcement (e.g., number of bars and diameter), type and spacing of stirrups or ties, and, in certain cases, cross-sectional details illustrating the spatial arrangement of reinforcement.

In addition to rebar-related content, the original structural drawings contained auxiliary textual information, including contractor details, client information, and general project notes. To protect project privacy and prevent irrelevant information from affecting model learning, the auxiliary regions were anonymized using mosaic processing prior to annotation. Text regions were manually annotated using the PPOCRLabel tool. A category-based annotation strategy was adopted: rebar-related semantic elements (component name, elevation, longitudinal reinforcement quantity and specification, stirrup or tie type and spacing) were annotated using green bounding boxes and served as exclusive targets for model training; auxiliary textual information was annotated using blue bounding boxes and selectively retained or excluded according to task requirements.

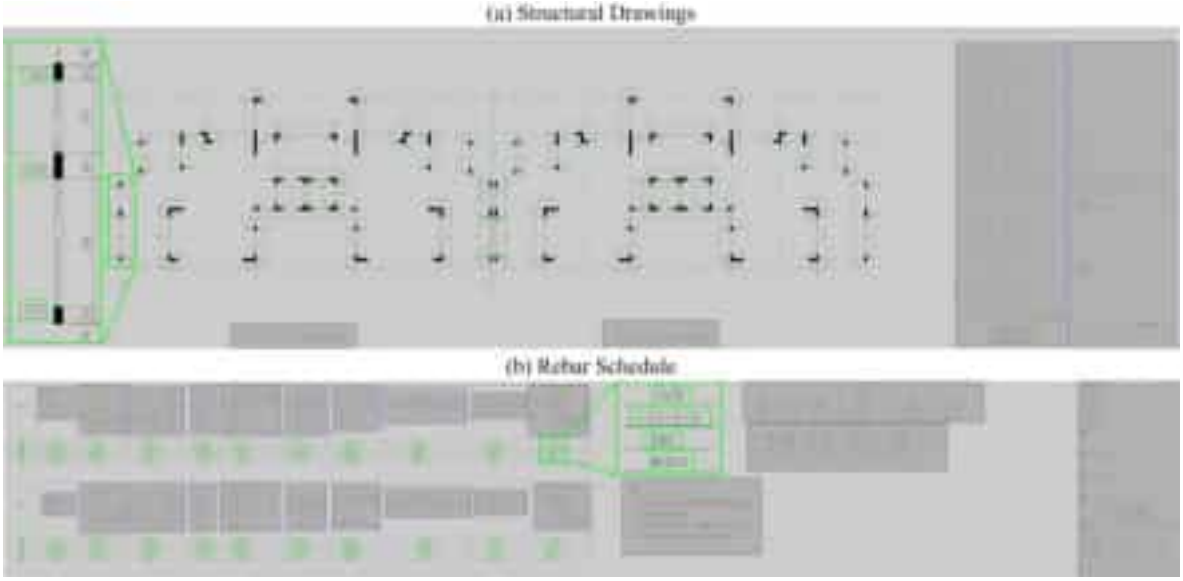


Figure 1. Example of structural engineering drawings: (a) structural construction drawing; (b) rebar detail schedule

Although cross-sectional detail drawings are graphical rather than textual in nature, their positions were also annotated and retained separately, given their potential value for downstream structural information association. To ensure consistency between the annotation data and recognition output, symbols representing different reinforcement grades were normalized to Φ during annotation, thereby reducing recognition noise caused by symbol variation and scanning inconsistencies.

3.2 Data Preprocessing Strategy

Owing to the ultra-high-resolution characteristics of the original structural construction drawings, directly feeding the entire image into a deep-learning model often results in excessive GPU memory consumption and degrades the recognition performance for dense small-scale text targets. To address this issue, as illustrated in Fig. 2, an overlap tiling strategy was introduced during the data pre-processing stage to decompose large-scale engineering drawings into manageable units [12].

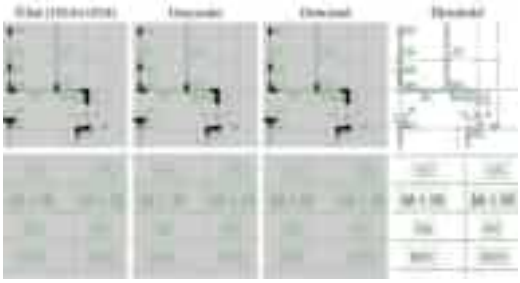


Figure 2. Example of data preprocessing.

Specifically, each original image was cropped into sub-images with a fixed size of 1024 × 1024 pixels. An overlap region of 200–300 pixels was applied between adjacent tiles to prevent text elements from being fragmented, while simultaneously reducing the risk of disrupting tabular structures or splitting component identifiers across tiles. For regions located at image boundaries where the cropped area was smaller than 1024 × 1024 pixels, white background padding was applied to achieve a uniform input size, thereby ensuring consistency during the model inference stage. Through this tiling procedure, 3510 valid cropped samples were generated from original 62 drawings.

Following spatial tiling, additional pre-processing operations tailored for text recognition tasks were applied to the cropped images, including grayscale conversion, noise suppression, and threshold-based segmentation. These steps enhance the contrast of the text regions while attenuating background interference. The described pre-processing pipeline preserves the integrity of the rebar-related semantic information and provides stable and standardized input data for subsequent text detection and Transformer-based sequence recognition models.

3.3 Hybrid Pipeline Architecture

This study proposes a staged hybrid OCR parsing architecture for structural construction drawings in which reinforcement semantic parsing is decomposed into two sequential stages: text-region detection and character-sequence recognition. This separation reduces learning complexity and enables independent optimization of each stage [17].

As shown in Fig. 3, after tiling and pre-processing, the drawings were first processed during the detection stage to localize reinforcement-related text regions and output bounding boxes. The detected regions were then cropped and forwarded to the recognition stage to generate the corresponding character sequences. This staged design effectively addresses the coexistence of dense text distributions and highly compressed semantic expressions typical of structural construction drawings.

3.3.1 Detection Stage

The detection stage employs the YOLOv11 Small model to localize text regions in structural drawings [19]. Model parameters were initialized using yolo11s.pt weights pre-trained on the COCO dataset. As a single-stage detection architecture, YOLOv11 maintains

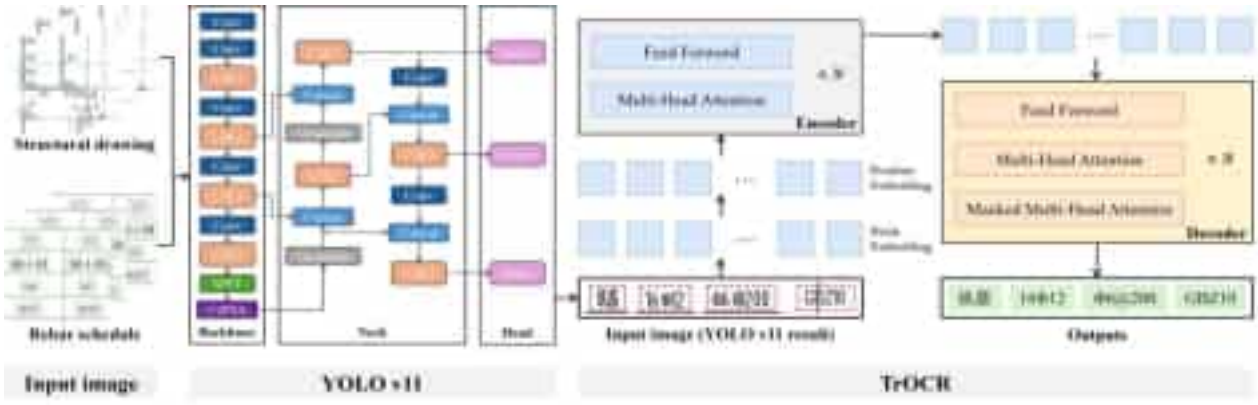


Figure 3. Hybrid Pipeline Architecture for Drawing OCR Analysis

reliable localization performance for small-scale and densely distributed text targets, making it suitable for engineering drawing scenarios. Text detection was formulated as a single-class object-detection task, allowing training to focus on spatial localization accuracy without category-level ambiguity.

3.3.2 Recognition Stage

In the recognition stage, a Transformer-based TrOCR model is adopted to perform character sequence modeling of the text regions produced by the detection stage. By constraining the recognition task to pre-detected text regions, the proposed framework substantially reduces the interference of background structures on character modelling.

The TrOCR model is initialized using a hybrid strategy [13]. The visual encoder is loaded from pre-trained Microsoft/trocr-base-printed weights, providing visual feature representations with a strong generalization capability. By contrast, the text decoder is initialized from scratch using a lightweight transformer architecture specifically adapted to the highly regularized and syntactically compact reinforcement annotations in structural drawings. This design leverages pre-trained visual representations while avoiding the modelling redundancy of general-purpose language model decoders in domain-specific scenarios.

3.4 Experimental Strategy

All experiments were conducted on a Windows 11 Pro platform using Python 3.9 and PyTorch 2.8.0. The hardware configuration consisted of an NVIDIA GeForce RTX 5080 GPU and an AMD Ryzen 9 9950X processor. Separate datasets and training strategies were designed for the detection and recognition stages to reflect their distinct task characteristics.

3.4.1 Detection Model Training

The text-detection model was trained using a dataset generated through the tiling procedure, which contains 3510 cropped engineering drawing images. The dataset was divided into training, validation, and test sets in a ratio of 8:1:1 to ensure sufficient training data while enabling independent performance monitoring and final evaluation. The model was trained for up to 100 epochs using an early stopping strategy based on the validation performance. Training was terminated if no improvement was observed for 20 consecutive epochs to mitigate overfitting under limited-sample conditions. Owing to the GPU memory constraints associated with the 1024 × 1024-pixel inputs, the batch size was set to eight. The default YOLO optimization configuration was adopted, including built-in learning-rate warm-up and decay mechanisms, to facilitate stable transfer-learning adaptation to the text-localization task [19].

The detection performance was evaluated using the mean average precision (mAP), including mAP50 and mAP50:95. [6]mAP50 corresponds to an intersection over union (IoU) threshold of 0.50, whereas mAP50:95 represents the average precision computed over IoU thresholds ranging from 0.50 to 0.95 with a step size of 0.05 [10]. IoU is defined as

$$IoU = \frac{Area\left(B_{pred} \cap B_{gt}\right)}{Area\left(B_{pred} \cup B_{gt}\right)} \tag{1}$$

where B_{pred} and B_{gt} denote the predicted and ground-truth bounding boxes, respectively. The detection was considered correct when the IoU exceeded a given threshold. AP is defined as the area under the precision–recall curve, as follows:

$$AP = \int_0^1 p(r)dr \tag{2}$$

As text detection was formulated as a single-class detection task in this study, the mAP was equivalent to the AP of this class under the corresponding IoU threshold.

3.4.2 Recognition Model Training

The training data for the text-recognition model were generated by cropping the text region output from the detection stage, yielding 11427 character-level annotated samples. The dataset was divided into training, validation, and test sets in the same 8:1:1 ratio. The recognition model was trained for 50 epochs by using a staged training strategy. During the first 10 epochs, the visual encoder was frozen and only the decoder was trained. Subsequently, the higher encoder layers were partially unfrozen, and the learning rate was reduced to one-fifth of its initial value to enhance domain adaptation while preserving the pre-trained visual features.

The training employed the AdamW optimizer with an initial learning rate of 1×10^{-4} and a batch size of 32. Weight decay and label smoothing were applied, and early stopping based on the validation set character error rate (CER) was used, with training terminated after 10 consecutive epochs without improvement.

Recognition performance was evaluated using the CER and character accuracy (CA), respectively defined as

$$CER = \frac{S + D + I}{N} \tag{3}$$

$$CA = 1 - \frac{S + D}{N} \tag{4}$$

where *S*, *D*, and *I* denote the number of substitutions, deletions, and insertions, respectively, and *N* is the total number of characters in the ground-truth text.

4 Results

4.1 Text Detection Results

The training process was systematically analyzed to evaluate the convergence behavior and localization performance of the text-detection model. Training was terminated early at epoch 61 on the basis of validation performance. As illustrated in Fig. 4, the box loss and DFL loss for both the training and validation sets decreased steadily throughout the optimization process, with no observable divergence between the two curves, confirming the stability of convergence. The detection model surpassed an mAP of 0.95 at IoU = 0.50 within the early training epochs, and the stricter mAP50:95 metric reached approximately 0.636 toward the final stage, indicating that the model attains reliable performance in both coarse localization and precise boundary regression.

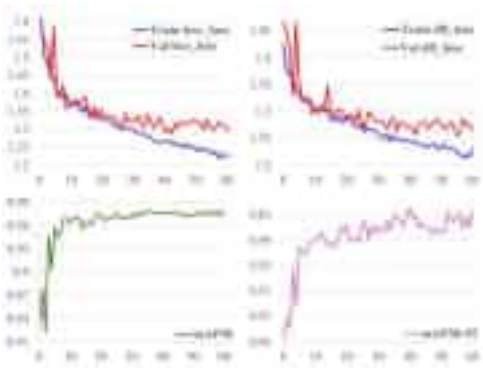


Figure 4. Training curves of the detection model

Fig. 5 presents representative detection results on the test set. The predicted bounding boxes exhibit a high degree of spatial overlap with the ground truth annotations, with no notable false positives or missed detections observed, confirming the effectiveness of the adopted detection strategy in the context of engineering drawings and providing a reliable input foundation for the subsequent text recognition stage.

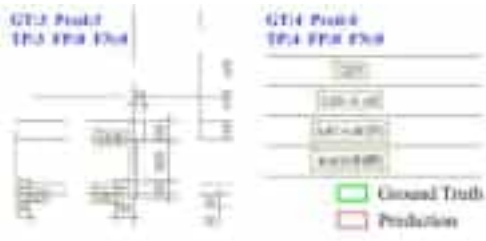


Figure 5. Typical examples of detection test results

4.2 Text Recognition Results

The performance of the text-recognition module is analyzed following the introduction of enhanced regularization strategies. As shown in Fig. 6, the training process converged stably, with the training loss and validation loss decreasing in parallel while maintaining a narrow gap. The validation CER reached its optimal value of 0.0281 (CA = 97.19%) at epoch 7. After unfreezing the top two encoder layers at epoch 11, no appreciable performance degradation was observed. Training was terminated early at epoch 17 based on validation performance, indicating that the model achieves favorable convergence and generalization under the combined regularization and staged unfreezing strategy.

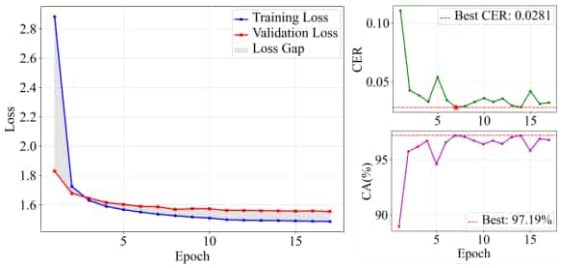


Figure 6. Training curves of the text recognition

To evaluate the effectiveness of the proposed pipeline, three representative open-source OCR methods (docTR, EasyOCR, and Tesseract) were employed as baselines for comparative experiments. None of the baseline methods incorporated a dedicated object detection module; instead, they applied the same sliding window strategy to segment the original engineering images into patches for direct recognition. The results are summarized in Table 1. Among the three baselines, EasyOCR achieved the best performance, yet its CER remained as high as 0.4622 (CA = 53.78%), with a precision of only 0.196, indicating that patch-based recognition using fixed windows can improve text recall but still introduces substantial background noise and redundant detections in the absence of semantic-level text localization, falling short of the high-accuracy recognition demands of engineering drawings. The proposed YOLOv11 + TrOCR pipeline, through a staged detect-then-recognize strategy, achieved a CA of 90.93% (CER = 0.0907) on the held-out test set, representing an improvement of approximately 37.2 percentage points over the best baseline EasyOCR (CA = 53.78%), fully demonstrating the necessity and effectiveness of incorporating a semantic-level dedicated detection module for structural drawing recognition.

Table 1. Comparison with Baseline OCR Model

Model	Tesseract	docTR	EasyOCR	**Ours**
CER	0.9341	0.6791	0.4622	**0.0907**
CA	6.59	32.09	53.78	**90.93**

Table 2 presents representative recognition results on the test set along with their corresponding error types. For standard structural annotation texts that appear frequently in engineering drawings, such as rebar diameter notations 18Φ14, component identifiers GBZ4, and stirrup spacing expressions Φ8@200, the model achieved zero-character error rate, indicating stable and reliable recognition capability for conventional annotation text.

Table 2. Typical examples of text recognition test results

Image	GT	Pred	CER	Error Type
	18Φ14	18Φ14	0	-
	编号	编号	0	-
	GBZ4	GBZ4	0	-
	Φ8@200	Φ8@200	0	-
	8.700~ 52.200	8.700~ 2.200	0.154	Digit Missing
	2Φ12+ 6 Φ14 (实心)	2Φ12+ 6 Φ14 (实必)	0.188	Char Confusion
	-2.520	-25.20	0.333	Decimal Shift

To further identify directions for model optimization, a character-level error decomposition was conducted on the 131 non-exact-match samples. It should be noted that the majority of these samples contained only one-to-two-character deviations, and the overall CER remained at a low level. Among a total of 304 edit operations, Digit Missing accounted for the highest proportion (148 occurrences, 48.7%), primarily manifesting as the omission of individual digits by the model and affecting 15.9% of all test samples. Decimal Shift ranked second (52 occurrences, 17.1%, affecting 8.7%), mainly attributable to the high visual similarity between elevation values and dimensional annotations, which makes it difficult for the model to distinguish between the two based solely on local image features. Char Confusion and special symbol errors together accounted for only 10.5% of all edit operations, occurring infrequently yet carrying relatively notable impact at the semantic level. Other errors (72 occurrences, 23.7%) were predominantly caused by missing or extraneous spaces and did not affect practical semantic parsing. The above analysis indicates that the residual recognition deviations of the current model are primarily concentrated in numerically dense short texts, and performance can be further improved through the introduction of structured numerical constraints or domain-adaptive post-processing strategies.

5 Conclusions

This study proposed a staged visual semantic parsing framework for 2D rasterized structural construction drawings, targeting the automated extraction of reinforcement information from structural construction drawings and rebar schedules. A high-resolution engineering drawing dataset comprising 62 drawings, was constructed, and an overlap tiling strategy was employed to process large-format images, effectively alleviating GPU memory constraints and text fragmentation.

From a methodological perspective, a hybrid OCR pipeline integrating YOLOv11 and TrOCR was developed in which reinforcement semantic parsing was decomposed into text-detection and -recognition stages. Experimental results demonstrate that the text-detection model can stably localize dense text regions under complex structural drawing backgrounds, thereby providing reliable inputs for the recognition stage. The text recognition model achieves a CA of 90.93% (CER = 0.0907) on the held-out test set, representing an improvement of approximately 37.2 percentage points over the best baseline method EasyOCR (CA = 53.78%) under the same sliding window strategy. Analysis of representative samples showed that the model accurately recognized high-frequency structural annotations, including reinforcement diameters, component identifiers, and spacing expressions, whereas recognition errors were primarily concentrated in complex text instances characterized by dense numerical content or highly compressed semantics.

Despite achieving stable recognition performance, several limitations remain. This study normalized different reinforcement grades using the symbol Φ to ensure recognition consistency, without performing fine-grained differentiation among different grade symbols. Furthermore, the current scope is confined to character-level text detection and recognition, and does not yet encompass the reconstruction of structured engineering semantics or the cross-drawing information association between structural construction drawings and rebar schedules. To bridge this gap, future research will proceed along four stages: (1) structured information reconstruction, parsing raw OCR output into structured fields (component name, rebar diameter, spacing, and quantity) based on domain-specific grammatical rules (e.g., the KS D 3504 reinforcement notation standard) and regular expressions; (2) component-level entity matching, using recognized component identifiers (e.g., GBZ1) as indices to establish one-to-one or one-to-many cross-drawing mapping relationships with corresponding rebar schedule entries; (3) compliance verification, automatically comparing matched structural information against design codes to verify mandatory requirements such as minimum rebar spacing and concrete cover

thickness; and (4) automated quantity extraction, calculating the reinforcement quantity for each component to support quantity surveying and cost management. Future work will also expand the fine-grained recognition of reinforcement-grade symbols and validate the generalization capability of the model across a broader range of engineering project types and drafting standards.

Acknowledgement

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (grant number RS202400356697).

References

[1] Z. Ma, Q. Zhao, Y. Zhu, T. Cang, and X. Hei. An Automatic Extraction Method of Rebar Processing Information Based on Digital Image. Mathematics, 10:2974, 2022.

[2] U. Woo, M. Lee, T. Lee, H. Choi, S.-M. Kang, and K.-K. Choi. Real-time rebar spacing measurement system for quality control in construction. Autom. Constr., 171:106012, 2025.

[3] X. Yuan, A. Smith, F. Moreu, R. Sarlo, C. D. Lippitt, M. Hojati, S. Alampalli, and S. Zhang. Automatic evaluation of rebar spacing and quality using LiDAR data: Field application for bridge structural assessment. *Autom. Constr.*, 146:104708, 2023.

[4] Y. Xiang, A.-M. Mahamadu, and L. Florez-Perez. Engineering information format utilisation across building design stages: An exploration of BIM applicability in China. *J. Build. Eng.*, 95:110030, 2024.

[5] C. Zhang, A.-X. Zhu, L. Zhou, M. Che, and T. Qiu. Constraints for Improving Information Integrity in Information Conversion From CAD Building Drawings to BIM Model. *IEEE Access*, 8:81190–81208, 2020.

[6] S. Wang, S. Park, J. Kim, and J. Kim. Safety Helmet Monitoring on Construction Sites Using YOLOv10 and Advanced Transformer Architectures with Surveillance and Body-Worn Cameras. *J. Constr. Eng. Manag.*, 151:04025186, 2025.

[7] P. Schönfelder, F. Stebel, N. Andreou, and M. König. Deep learning-based text detection and recognition on architectural floor plans. *Autom. Constr.*, 157:105156, 2024.

[8] Z. Zhang and Y. He. Research on Intelligent Verification of Equipment Information in Engineering Drawings Based on Deep Learning. *Electronics*, 14:814, 2025.

[9] W. Khallouli, M. S. Uddin, A. Sousa-Poza, J. Li, and S. Kovacic. Leveraging Transformer-Based OCR Model with Generative Data Augmentation for Engineering Document Recognition. *Electronics*, 14:5, 2025.

[10] Y. Man, Q. Zhang, Y. Cai, and Z. Chang. A Target Detection Algorithm for Marks Layout in Detail Drawings of Steel Structure. *Int. J. Steel Struct.*, 25:557–570, 2025.

[11] L. Jamieson, C. F. Moreno-Garcia, and E. Elyan. Towards fully automated processing and analysis of construction diagrams: AI-powered symbol detection. *Int. J. Doc. Anal. Recognit. IJDAR*, 28:71–84, 2025.

[12] A. Rezvanifar, M. Cote, and A. B. Albu. Symbol Spotting on Digital Architectural Floor Plans Using a Deep Learning-based Framework. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 2419–2428, Seattle, WA, USA, 2020.

[13] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei. TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37:13094–13102, Washington, DC, USA, 2023.

[14] C. F. Moreno-García, E. Elyan, and C. Jayne. New trends on digitisation of complex engineering drawings. *Neural Comput. Appl.*, 31:1695–1712, 2019.

[15] S. Ahmed, M. Liwicki, M. Weber, A. Dengel, Automatic Room Detection and Room Labeling from Architectural Floor Plans, in: 2012 10th IAPR Int. Workshop Doc. Anal. Syst., 2012: pp. 339–343.

[16] S. Wang, M. Kim, H. Hae, M. Cao, and J. Kim. The Development of a Rebar-Counting Model for Reinforced Concrete Columns: Using an Unmanned Aerial Vehicle and Deep-Learning Approach. *J. Constr. Eng. Manag.*, 149:04023111, 2023.

[17] J. Ravagli, Z. Ziran, and S. Marinai. Text Recognition and Classification in Floor Plan Images. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, pages 1–6, Sydney, Australia, 2019.

[18] A. Saba, R. Hantach, and M. Benslimane. Text Detection and Recognition from Piping and Instrumentation Diagrams. In *Proceedings of the 2023 8th International Conference on Image, Vision and Computing (ICIVC)*, pages 711–716, Dalian, China, 2023.

[19] R. Khanam and M. Hussain. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv preprint arXiv*:2410.17725, 2024.

Towards Predicting Blockages in Urban Stormwater Drainage Systems with Machine Learning

Ashen Krishantha[1], Palaneeswaran Ekambaram[1], Morshed Alam[1], Iqbal Hossain[1], Mengqi Huang[1], Abdul Aziz[2]

[1]School of Engineering, Swinburne University of Technology, Australia
[2]Frankston City Council, Victoria, Australia
abatagalleranhawadig@swin.edu.au

Abstract –

Occurrences of blockages in stormwater drainages could be a significant issue, especially in urban areas, due to various challenges for infrastructure operation and maintenance. Predicting blockage occurrences is a complex problem that involves numerous parameters and an adequate understanding of their impacts, interactions and the nonlinear behaviours of certain parameters. This research study aims to develop a data-driven predictive model-based framework for potential urban stormwater drainage blockages. The framework would be using selective explainable machine learning methods with a focus on improving model transparency. This paper presents an interim summary of employing some machine learning models, such as K-nearest neighbours, logistic regression, random forest, artificial neural networks, and XGBoost. The prototype framework discussed in this paper has been developed and evaluated using a set of practical datasets obtained from historical maintenance records of a local city council in Victoria state of Australia, along with some climatic, environmental, socio-economic and asset-related datasets compiled from various sources. In the pilot testing of models of prototype framework, XGBoost demonstrated the best predictive performance, showing a strong ability to capture non-linear relationships in the data. The SHAP global feature importance analysis was used to explore the importance of chosen model parameters, which revealed that specific rainfall and vegetation cover in the catchment zones are the key influencing parameters for predicting potential occurrences of blockages in stormwater drainage systems. Hopefully, the study outcomes at the outset would demonstrate that explainable machine learning can support the development of a smart stormwater drainage asset management digital twin infrastructure.

Keywords – Stormwater infrastructure, Blockage Prediction, Machine Learning, Explainable AI, Predictive Maintenance, Asset Management

1 Introduction

Urban stormwater drainage systems are essential components of modern cities, providing protection against flooding, safeguarding public health, and supporting other operational urban activities. However, these systems face significant challenges due to ageing infrastructure, climate change, growing urbanisation, and expanding cities[1]. Under these conditions, drainage blockages have become a major operational challenge, which reduces the drainage capacity and significantly increases the likelihood of urban flooding[2]. Stormwater blockages are caused by tree root intrusion, sediment deposition, debris and litter accumulation[3, 4]. Therefore, managing these challenges requires effective maintenance strategies that can support informed decision-making in stormwater asset management.

Current stormwater maintenance, such as manual inspections and CCTV inspections, is mostly reactive or schedule-based, where inspections and cleaning activities are driven by public complaints, costly and time-consuming[5, 6]. Consequently, researchers and local councils have increasingly focused on developing proactive approaches that can identify potential blockage issues before failures occur. Such approaches can help enhance system efficiency, prioritise high-risk assets, and reduce the likelihood of unexpected blockage-related failures.

Recent advances in machine learning have led to the development of data-driven prediction models for urban drainage systems. Using historical maintenance data, these models can predict drainage issues at the asset level, enabling proactive maintenance planning [7]. Machine learning techniques are well-suited to capture complex and non-linear relationships between input and output variables, which helps in understanding how key

influencing factors contribute to drainage blockage occurrence[8]. According to [9], Predictive models can achieve substantial cost savings, including reductions of up to 60% in preventive maintenance costs through a 39% decrease in preventive maintenance frequency. Despite their strong predictive performance, one of the main challenges of these models is the lack of transparency, as many high-performing machine learning models are often treated as "black boxes"[10]. Explainable artificial intelligence (XAI) helps transform black-box machine learning models into transparent models by explaining how factors and their contributions affect the final output. This transparency enables decision makers to trust predictions, justify maintenance priorities and allocate resources effectively across infrastructure like drainage networks [11].

Therefore, this study uses classical machine learning models, together with explainable AI methods, to predict and interpret the occurrence of drainage blockages. Historical maintenance data and environmental and climatic variables are used to develop an explainable prediction framework. The study aims to assess model performance, improve understanding of model behaviour, and show how the results can support proactive and risk-based asset management. In future applications, the developed models can be integrated into proactive drainage management frameworks, including digital twin systems, where explainable predictions can support monitoring, maintenance prioritisation, and informed decision-making.

2 Background of Research

Blockage formation in urban drainage systems is a complex process governed by the interaction of physical, environmental, climatic, operational, and socio-economic factors. This complexity limits the effectiveness of traditional analytics approaches, such as statistical approaches to capture nonlinear relations between factors and blockage formation. For example, Statistical models for predicting blockages in drainage systems have several limitations, particularly related to data availability and model assumptions. These models are highly dependent on the quality and resolution of input data, and inaccurate or incomplete records can significantly reduce prediction reliability. For example, Markov chain models require large and consistent datasets describing pipe condition and blockage history, which are often unavailable in municipal records, leading to reduced prediction accuracy [12]. While Log Gaussian Cox models are capable of representing spatio-temporal patterns, they are computationally intensive and require a long simulation time [13]. Similarly, Cox proportional hazards models rely on assumptions that may not adequately represent the nonlinear behaviour. As a result, the applicability of these models remains limited in practical drainage management contexts.

A wide range of machine learning (ML) and deep learning (DL) models have been applied in previous studies to diagnose and predict failures within urban sewer systems. Classical ML approaches, including decision trees, random forests, logistic regression, and support vector machines, have demonstrated strong predictive capability for blockage prediction trained on historical asset and inspection data [14-16]. More recently, deep learning models, such as neural networks, graph-based neural networks, and autoencoder-based frameworks, have enabled the integration of spatiotemporal data and large feature sets to predict deposit accumulation, siltation severity, and cleaning requirements with improved accuracy [17-20]. Despite these advances, most existing predictive models focus on wastewater or combined sewer systems, with limited attention given to stormwater drainage systems. Tran [21], developed a neural network model to assess serviceability conditions of stormwater drainage, but predictive modelling for stormwater-specific blockage behaviour remains underdeveloped.

Many existing studies use a limited set of input variables, mainly focusing on pipe attributes such as diameter, length, age, material, depth, and soil type, as well as population and property density. Although these inputs are commonly used, environmental and catchment-related factors such as land use, vegetation cover (e.g., NDVI), impervious area, and the number of dry days preceding a blockage event are employed in only a limited number of studies. These factors strongly influence runoff, debris movement, and sediment build-up in stormwater systems. Their effects on stormwater blockages are still not well studied, highlighting a clear research gap in stormwater blockage prediction. Moreover, some of these studies try to understand important features from feature importance, but systematic use of model explainability to interpret predictions is also limited in existing literature. This study incorporates stormwater-specific variables into blockage prediction, which have been rarely considered in previous studies. In addition, it introduces the use of explainable machine learning in drainage infrastructure blockage prediction to improve model interpretability. SHAP-based analysis is applied to provide transparent and interpretable insights into how different variables influence blockage formation, supporting more informed maintenance decision-making.

3 Methodology

This section describes the methodology adopted to develop an explainable machine learning framework for predicting stormwater drainage blockages. The overall workflow includes dataset preparation, model development and the use of explainable machine learning for interpreting model outputs. Each component of the methodology is described as shown in Figure 1.

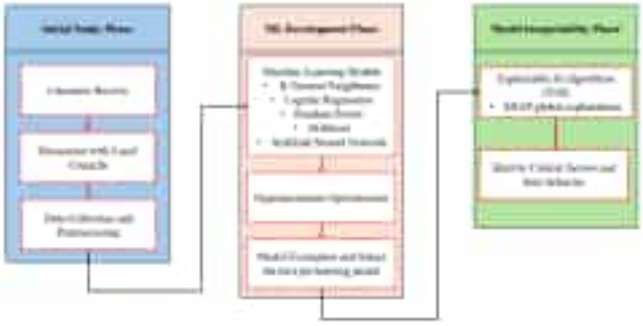


Figure 1. Research Methodology

3.1 Dataset and Variables

The dataset used in this study was provided by the Frankston council in Victoria, Australia. It comprises historical stormwater maintenance records collected between 2019 and 2025, totalling 4407 records. The stormwater drainage network includes 40,083 pipes, totalling about 1,059 km, culverts, and open drains, connected by around 41,003 stormwater pits. Pipe diameters range from 150 mm to 3,300 mm, representing both local collection pipes and larger downstream conveyance pipes. A key limitation of the dataset is that blockage information is primarily recorded at the pit level, which limits the direct use of detailed pipe attributes, such as diameter, length, and slope, as input parameters.

The factors considered in this study were identified through a comprehensive review of existing literature and experts' discussions with the local council based on their operational experience and practical knowledge of stormwater drainage management. The input factors used in this study, along with their corresponding data sources, are summarised in Table 1.

Table 1. Factors Considered in this Study

Input factor	Unit	Description
Rainfall	mm	Rainfall amount before the blockage event
Number of dry days	Days	Days without rainfall before the blockage event
Land Use	%	Impervious land use in the catchment
Number of Connections	-	Number of pipes connected to pits
NDVI	-	Vegetation cover across the area
Network density	(km/km^2)	Drainage network length per area
Population density	(Persons/km^2)	Population per unit area

Rainfall is the cumulative rainfall before a blockage event as an indicator of stormwater runoff. The number of dry days was defined as the consecutive days with rainfall less than 1 mm before a blockage, representing dry-weather periods that influence debris accumulation [22]. Some variables, such as land use and population density, were assumed to remain constant throughout the year due to the lack of high-frequency temporal data. Vegetation cover (NDVI) was derived from Sentinel-2 satellite imagery, providing spatial information on vegetation conditions across the study area.

3.2 Model Development

3.2.1 Types of Models

Classical machine learning models, including K-Nearest Neighbours (KNN), Logistic Regression, Random Forest (RF), XGBoost, and an Artificial Neural Network (ANN), were developed to predict the occurrence of stormwater blockages using the prepared dataset. These models were selected because they have been widely applied in previous studies for prediction and classification problems [23, 24]. Thus, their predictive accuracy and reliability have been well-suited for drainage infrastructure and asset management problems.

3.2.2 Training and Optimization

Model training was performed using historical data, with the dataset divided into training and testing subsets. The dataset consists of 1669 blockage records (37.9% of total records) and 2738 non-blockage records (62.1% of total records). Given this relatively moderate class distribution, no specific class imbalance handling

techniques were applied in this study. Hyperparameter tuning was conducted using a grid search (GS) approach to optimise model performance and prevent overfitting. Five-fold cross-validation (5-fold CV) was applied during training to ensure model robustness and generalisation across different data samples. The optimal model configuration was selected based on consistent performance across validation folds. Additionally, a simple baseline model (logistic regression) was trained without hyperparameters to assess the performance gains of hyper parameterized models. Table 2 shows the optimized hyperparameters used for each model.

Table 2. Optimized hyperparameters of each model

Model	Hyperparameters	Optimum
XGBoost	Max depth	6
	Estimators	400
	Learning rate	0.1
RF	Estimators	400
	Max depth	20
	Minimum sample split	5
KNN	Weights	Distance
	Neighbours	15
	Metric	Manhattan
LR	C	0.7
	Penalty	L1
ANN	Number of hidden layers	2
	Number of neurons per layer	(64,32)

3.2.3 Model Evaluation

The performance of the trained models was evaluated using multiple evaluation metrics obtained from testing data split to assess prediction accuracy (see equations 1-5).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

$$F1score = \frac{(2 \times Precision \times Recall)}{(Precision+Recall)} \quad (4)$$

$$ROC - AUC = Area\ under\ the\ ROC\ curve \quad (5)$$

Where TP is true positives, and FP are false positives. TN is true negatives, and FN are false negatives. The evaluation metrics were obtained using a 70/30 training and testing data split for each model to identify the most accurate model performance. A comparative analysis was conducted across all models, and the best-performing model was subsequently selected for further analysis.

3.3 Explainable AI

To address the black-box nature of machine learning models, explainable artificial intelligence (XAI) techniques were incorporated. SHAP-based global interpretation analyses were used to provide how environmental, climatic, and asset-related factors influence stormwater blockage occurrence. This interpretability enhances confidence in the model predictions and supports transparent, data-driven decision-making for drainage asset management.

4 Results

4.1 Model Performance Evaluation

The trained machine learning models were systematically evaluated to assess their predictive capability. This evaluation is critical to ensure (a) the ability of the proposed machine learning framework to capture complex and non-linear relationships among climatic, environmental, and related other factors under limited data availability, and (b) the robustness of the trained models in accurately predicting blockage occurrence for previously unseen (test) data. Following model training and hyperparameter optimization, performance was assessed using an independent test dataset, as shown in the table 3 below. Table 3 presents a comparative evaluation of the five classification models using Accuracy, F1-score, and ROC–AUC on the test dataset.

Table 3. Results and Model Comparison

Model	Accuracy	F_1 Score	ROC-AUC
XGBoost	0.81	0.74	0.88
RF	0.80	0.74	0.87
KNN	0.77	0.71	0.85
ANN	0.72	0.56	0.76
LR (With hyperparameters)	0.61	0.55	0.66
LR (Baseline model)	0.66	0.35	0.65

According to Table 3, among all models, XGBoost achieved the best overall performance, with an accuracy of 0.81, an F1-score of 0.74, and the highest ROC–AUC value of 0.88. Random Forest showed comparable performance (accuracy = 0.80, ROC–AUC = 0.87), indicating that ensemble tree-based models are also

effective for capturing the complex and non-linear relationships governing stormwater drainage blockages. KNN demonstrated moderate predictive capability (accuracy = 0.77, F_1 = 0.71, ROC–AUC = 0.85), while Logistic Regression (F_1 = 0.55, ROC–AUC = 0.66) and ANN (F_1 = 0.56, ROC–AUC = 0.76) showed comparatively lower performance, especially in terms of F1-score and ROC–AUC. ANN models typically require larger datasets to effectively capture complex patterns and avoid overfitting. In contrast, classical machine learning algorithms such as Random Forest and boosting models (e.g., XGBoost) often perform better with less training data and require less extensive hyperparameter tuning [25]. Moreover, a baseline Logistic Regression model without hyperparameter tuning was implemented to provide a simple benchmark for comparison of optimized models over traditional ML models. The baseline model produced performance (accuracy = 0.66, F1 = 0.35, ROC–AUC = 0.65). In terms of accuracy, it has a slight accuracy improvement over the Logistic Regression with hyperparameters, but overall performances of optimized models outperforming traditional ML models. Figure 2 shows ROC–AUC curves for the tested models, illustrating their classification performance on the test dataset.

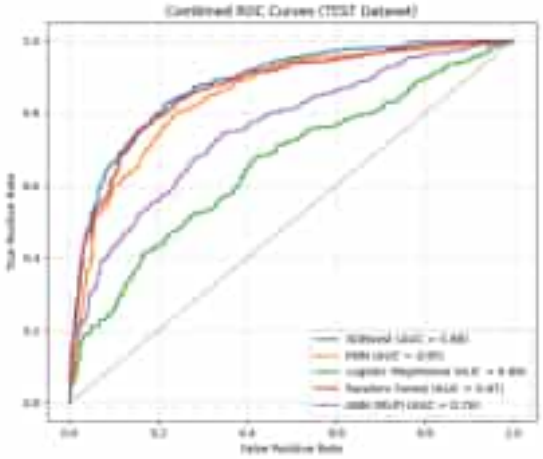


Figure 2. ROC-AUC Curves for each model

In addition to standard evaluation metrics, the confusion matrix for the best-performing model was analysed to better understand the model's misclassification patterns, as shown in Figure 3. The presence of false positives indicates that some non-blockage cases were predicted as blockages, which may lead to unnecessary inspections or preventive maintenance actions. Conversely, false negatives are cases in which actual blockages were predicted as non-blockages, potentially leading to undetected issues in the drainage network.

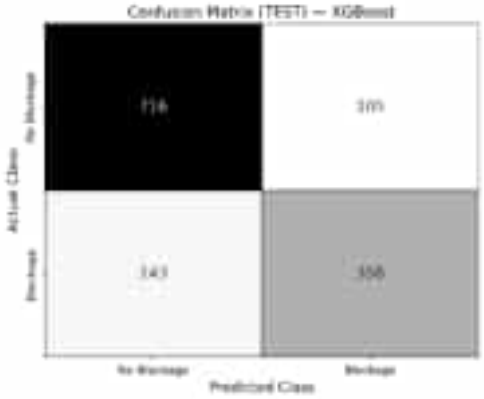


Figure 3. Confusion Matrix for test data set for XGBoost model

According to the figure, most samples are correctly classified compared to misclassified cases, indicating that the model performs well in distinguishing blockage and non-blockage events.

4.2 SHAP Explanation

To interpret the best-performing model, SHAP-based explainable AI analysis was applied to the XGBoost classifier. The explanation is shown in Figure 3.

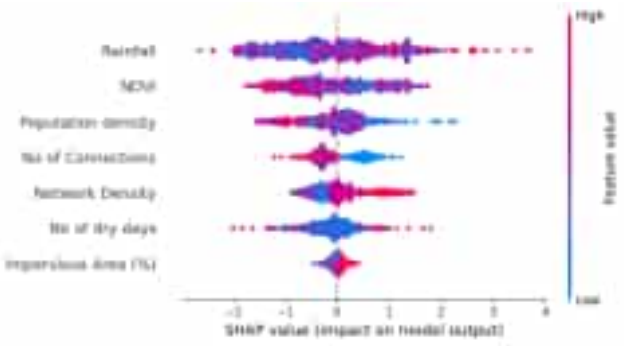


Figure 4. SHAP Summary plot

The SHAP summary plot reveals that rainfall is the most influential predictor of blockage occurrence, followed by NDVI, population density, and the number of connections. Higher rainfall values generally increase the likelihood of blockages because rainfall generates higher runoff that carries debris and sediment into pits and pipes. The SHAP results for NDVI indicate that higher and mean NDVI values are associated with an increased likelihood of blockage, likely due to greater vegetation coverage, which contributes organic debris to the drainage system. Although higher population density generally increases blockage likelihood, the SHAP plot

shows that low population density can also contribute to higher blockage risk in certain cases. This occurs when low-density areas may interact with other dominant factors, such as high rainfall, NDVI, etc. Network characteristics (e.g., network density and number of connections) also exhibit significant contributions, highlighting the influence of infrastructure complexity on blockage formation. Higher values of antecedent dry days and impervious area percentage increase blockage risk, but their influence is relatively smaller. These results not only enhance model transparency but also provide physically and operationally meaningful insights to support decision-making.

5 Conclusions and Future Works

This paper developed and evaluated a data-driven predictive framework for the occurrence of stormwater drainage blockages, focusing on model performance and interpretability. Among the machine learning approaches, the XGBoost model achieved the best predictive performance, capturing complex, non-linear relationships between drainage blockages and external influencing factors. The use of explainable artificial intelligence helped address the black-box nature of the model by clearly identifying the key factors influencing blockage predictions. The results show that rainfall and vegetation indicators (NDVI) are the dominant factors, and other environmental factors have a strong influence on blockage occurrence. The outputs of this study can be integrated with decision-support systems such as multi-criteria decision-making (MCDM) methods to support inspection prioritization, maintenance scheduling, and resource allocation. For instance, explainable blockage risk factors can be translated into criteria weights or constraint rules within MCDM frameworks, enabling transparent ranking of assets that require priority maintenance or inspection. Moreover, the direct effect of pipe attributes is limited in this study; they are still useful when combined with environmental and climatic factors. The main reason for this is that maintenance data are recorded at the pit level rather than for individual pipes. Therefore, methods such as pit–pipe mapping link each pit to its connected pipe segments based on the network topology, allowing pipe attributes such as diameter, length, slope, and material to be incorporated indirectly into the model and could improve predictive performance up to a certain level in future versions. Overall, these findings demonstrate both the strengths and limitations of the proposed approach and support its use as a data-driven tool for risk-based stormwater drainage asset management.

Future work focuses on extending the proposed predictive framework to improve interpretability, robustness, and practical applicability. This includes incorporating local explanations and feature dependency analysis using explainable AI to better understand how blockages vary across individual cases and operating conditions. The model can be integrated with additional input variables, such as traffic intensity, soil type, humidity, and temperature, which influence debris accumulation, sediment transport, and moisture conditions within the drainage network. In addition, the current occurrence-based prediction will be extended towards a risk-based formulation by incorporating blockage frequency and severity to predict the blockage risk for a better understanding of critical drainage assets. Also, future studies could extend the current model by incorporating graph-based or hybrid deep learning approaches, such as Graph Neural Networks (GNNs), which can explicitly capture spatial dependencies, network connectivity, and upstream–downstream interactions within urban drainage systems, since the input variables in the current model are treated independently and do not explicitly represent network relationships.

The framework will also be extended to incorporate real-time sensor data, including water level and flow measurements, to enable early blockage detection and warning alerts. In addition, coupling the predictive model with hydraulic modelling will enable scenario-based assessment of system performance under different rainfall and blockage conditions. By integrating real-time sensor data with physics-based hydraulic models, this model can be improved as a data analytics layer of a digital twin environment of stormwater drainage systems. Furthermore, 3D BIM visualisation will be used for asset condition monitoring to represent critical drainage assets under various predicted blockage risks to enhance visualisation capabilities.

6 References

[1] A. Ferdowsi and K. Behzadian, "Urban Drainage Infrastructures Toward a Sustainable Future," Springer International Publishing, 2024, pp. 111-119.

[2] M. Van Bijnen, H. Korving, and F. Clemens, "Impact of sewer condition on urban flooding: an uncertainty analysis based on field observations and Monte Carlo simulations on full hydrodynamic models," *Water science and technology,* vol. 65, no. 12, pp. 2219-2227, 2012.

[3] D. DeSilva, D. Marlow, D. Beale, and D. Marney, "Sewer Blockage Management: Australian Perspective," *Journal of pipeline systems engineering and practice.,* vol. 2, no. 4, pp. 139-145, 2011, doi: 10.1061/(ASCE)PS.1949-1204.0000084.

[4] M. Shivashankar, M. Pandey, and A. K. Shukla, "Numerical Investigation on the Evaluation of the Sediment Retention Efficiency of Invert Traps in an Open Rectangular Combined Sewer Channel," *Journal of Hazardous, Toxic, and Radioactive Waste,* vol. 27, no. 3, pp. 139-148, 2023.

[5] S. Ma, T. Zayed, J. Xing, and Y. Shao, "A state-of-the-art review for the prediction of overflow in urban sewer systems," *Journal of cleaner production.,* vol. 434, p. 139923, 2024, doi: 10.1016/j.jclepro.2023.139923.

[6] S. R. Tatiparthi *et al.*, "Real-Time detection of sewer water levels and blockages using UHF-RFID sensors," *Water Research,* Article vol. 278, 2025, Art no. 123380, doi: 10.1016/j.watres.2025.123380.

[7] O. Aounali, W. Shepherd, and S. Tait, "Integrating dynamic features into machine learning models for predicting sewer network failures: a Random Forest approach," *Urban Water Journal,* pp. 1-16, 2025, doi: 10.1080/1573062x.2025.2589081.

[8] S. Ray, "A quick review of machine learning algorithms," in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, 2019: IEEE, pp. 35-39.

[9] C. I. Ossai, "Data-driven approaches for smart water systems," (in English), *Water e-Journal,* vol. 10, no. 2, 2025 2025.

[10] A. Garzón, Z. Kapelan, J. Langeveld, and R. Taormina, "Machine Learning - Based Surrogate Modeling for Urban Water Networks: Review and Future Research Directions," *Water Resources Research,* vol. 58, no. 5, 2022, doi: 10.1029/2021wr031808.

[11] D. Thakker, B. K. Mishra, A. Abdullatif, S. Mazumdar, and S. Simpson, "Explainable artificial intelligence for developing smart cities solutions," *Smart Cities,* vol. 3, no. 4, pp. 1353-1382, 2020.

[12] A. Altarabsheh, M. Ventresca, A. Kandil, and D. Abraham, "Markov chain modulated Poisson process to stimulate the number of blockages in sewer networks," *Canadian journal of civil engineering = Revue canadienne de génie civil.,* vol. 46, no. 12, pp. 1174-1186, 2019, doi: 10.1139/cjce-2018-0104.

[13] C. Spychala, C. Dombry, and C. Goga, "Variable selection methods for Log-Gaussian Cox processes: A case-study on accident data," *Spatial statistics.,* vol. 61, p. 100831, 2024, doi: 10.1016/j.spasta.2024.100831.

[14] J. Bailey, E. Harris, E. Keedwell, S. Djordjevic, and Z. Kapelan, "Developing Decision Tree Models to Create a Predictive Blockage Likelihood Model for Real-World Wastewater Networks," *Procedia engineering.,* vol. 154, pp. 1209-1216, 2016, doi: 10.1016/j.proeng.2016.07.433.

[15] M. Hassouna, M. Reis, M. Al Fairuz, and A. Tarhini, "Data-Driven Models for Sewer Blockage Prediction," 2019 2019: IEEE, doi: 10.1109/iccece46942.2019.8941848.

[16] E. Okwori, M. Viklander, and A. Hedström, "Spatial heterogeneity assessment of factors affecting sewer pipe blockages and predictions," *Water research.,* vol. 194, p. 116934, 2021, doi: 10.1016/j.watres.2021.116934.

[17] Y. Q. Jiang, C. L. Li, Y. T. Zhang, R. B. Zhao, K. F. Yan, and W. H. Wang, "Data-driven method based on deep learning algorithm for detecting fat, oil, and grease (FOG) of sewer networks in urban commercial areas," *Water Research,* vol. 207, Dec 2021, Art no. 117797, doi: 10.1016/j.watres.2021.117797.

[18] M. Ribalta *et al.*, "Sediment Level Prediction of a Combined Sewer System Using Spatial Features," *Sustainability,* vol. 13, no. 7, p. 4013, 2021, doi: 10.3390/su13074013.

[19] M. Ribalta, R. Béjar, C. Mateu, L. Corominas, O. Esbrí, and E. Rubión, "Sewer sediment deposition prediction using a two-stage machine learning solution," *Journal of Hydroinformatics,* vol. 26, no. 4, pp. 727-743, 2024, doi: 10.2166/hydro.2024.144.

[20] J. H. Wu, L. Ma, X. Chen, and T. Broyd, "Research on interpretable graph neural network agent model for siltation diagnosis in Urban-Scale sewer systems," *Tunnelling and Underground Space Technology,* vol. 162, Aug 2025, Art no. 106666, doi: 10.1016/j.tust.2025.106666.

[21] D. H. Tran, A. W. M. Ng, and B. J. C. Perera, "Neural networks deterioration models for serviceability condition of buried stormwater pipes," *Engineering Applications of Artificial Intelligence,* vol. 20, no. 8, pp. 1144-1151, Dec 2007, doi: 10.1016/j.engappai.2007.02.005.

[22] S. Draude, E. Keedwell, R. Hiscock, and Z. Kapelan, "A statistical analysis on the effect of preceding dry weather on sewer blockages in South Wales," *Water Science & Technology,* vol. 80, no. 12, pp. 2381-2389, 2019. [Online]. Available: https://doi.org/10.2166/wst.2019.453.

[23] E. Palaneeswaran, P. E. D. Love, M. M. Kumaraswamy, and T. S. T. Ng, "Mapping rework causes and effects using artificial neural networks," *Building Research & Information,* vol. 36, no. 5, pp. 450-465, 2008/10/01 2008, doi: 10.1080/09613210802128269.

[24] A. Alharthi, M. Alaryani, and S. Kaddoura, "A comparative study of machine learning and deep learning models in binary and multiclass classification for intrusion detection systems," *Array.,* vol. 26, p. 100406, 2025, doi: 10.1016/j.array.2025.100406.

[25] "NeurIPS 2022 Datasets and Benchmarks Paper," United States, 2022. [Online]. Available: https://papers.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf.

Automating Electrical Outlet and Circuit Assignment in BIM through AI-Assisted CAD–BIM Interoperability

Luis A. Cristancho [1], Kevin Torres [1], Omar Sánchez [2], Rodrigo F. Herrera [1] and Karen Castañeda [1]

[1] School of Civil Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2374631, Chile
[2] Department of Civil Engineering, Pontificia Universidad Javeriana, Bogotá D.C., 111711, Colombia
lcristancho@utb.edu.co, torreskd@javeriana.edu.co, omar_sanchezr@javeriana.edu.co, rodrigo.herrera@pucv.cl, karen.castaneda@pucv.cl

Abstract – Electrical system modeling in BIM environments still relies heavily on manual interpretation of 2D CAD drawings, which can propagate circuit and panelboard assignment errors and increase rework. This paper presents an AI-driven, a parametric CAD–BIM workflow that automates outlet-to-circuit assignment and circuit-to-panelboard connections while maintaining semantic traceability. CAD plan views are captured as images and processed with a multimodal large-language model (Microsoft Copilot). Through iterative prompt engineering, the system extracts outlet identifiers, associated distribution board, and element counts, and exports a structured Excel dataset for downstream automation. Autodesk Dynamo and Python then ingest the dataset, match it to Revit parametric families, and execute rule-based logic to generate circuits and connect them to the target panelboard. The approach was validated on five real case studies from a data-center project by comparing a conventional manual Revit procedure against the automated workflow using execution time and review-stage errors as performance metrics. For 42 outlets, the manual procedure took 31.5 minutes and produced six misassignments, whereas the automated routine completed in 0.033 seconds with zero errors; for 24 outlets, the manual procedure took 12 minutes with two errors, while the automated routine took 0.025 seconds with zero errors.

Keywords – Building Information Modeling (BIM); CAD–BIM Interoperability; Artificial Intelligence; Electrical System Automation; Parametric Modeling; Dynamo Scripting

1 Introduction

The increasing scale and technical complexity of contemporary construction projects have intensified the demand for accurate, consistent, and interoperable information management across the design and execution stages [1]. In this context, electrical system planning remains heavily dependent on two-dimensional CAD documentation, in which circuit layouts, outlet distributions, and panel assignments are commonly interpreted manually, introducing coordination risks and propagating errors into downstream BIM processes [2]. Although Building Information Modeling has enhanced multidisciplinary coordination and centralized data, its application to electrical design workflows is often constrained by fragmented information transfer between CAD drawings and BIM environments [3]. These limitations become particularly critical when repetitive components, standardized electrical conventions, and regulatory requirements must be maintained consistently across multiple project iterations [4]. Deficiencies in early-stage electrical data integration can lead to misaligned circuit assignments, inconsistent schedules, and increased rework during construction, directly affecting project cost, time, and model reliability. Current digital practices, including spreadsheet-based coordination and rule-based verification, provide partial support but continue to rely heavily on manual intervention. As a result, recent research has increasingly explored artificial intelligence and parametric programming as enabling technologies for interpreting structured graphical information and implementing automated logic within BIM workflows [5]. By combining AI-assisted extraction of graphical inputs with parametric scripts that support semantic matching and rule-driven classification, it becomes feasible to bridge the persistent gap between CAD-based electrical plans and BIM models [6]. This integration supports automated outlet identification, circuit correspondence, and panel allocation while preserving consistency between design documentation and digital representations. In this setting, the convergence of CAD inputs, AI-driven data structuring, and parametric automation within BIM platforms emerges as a promising pathway to enhance reliability, reduce human error, and improve efficiency in electrical system modeling [7].

The accelerated adoption of artificial intelligence across industrial domains has reshaped how technical information is processed, interpreted, and operationalized, particularly in environments with large volumes of graphical data and repetitive decision rules [8]. In construction-related workflows, AI-driven solutions have progressively enabled automated drawing interpretation, data structuring, and rule-based decision support, offering new opportunities to reduce manual

effort and improve consistency in digital models [9]. Within BIM-centered processes, these technologies have been applied to scheduling, coordination, and analytical tasks, providing near-real-time feedback that was previously unattainable with conventional tools. At the same time, advances in machine learning and generative systems have enabled the transformation of unstructured visual inputs into structured datasets that parametric environments can directly consume. This capability is especially relevant in electrical design processes, where information embedded in CAD drawings must be accurately translated into BIM elements while preserving naming conventions, circuit logic, and panel relationships. Despite these advances, adoption remains uneven due to organizational inertia, limited standardization, and the technical complexity of integrating AI outputs into interoperable BIM workflows [10]. Lightweight, task-oriented AI tools have recently emerged as a practical entry point for embedding intelligence into traditional manual processes, enabling incremental experimentation with lower implementation barriers [11]. However, a clear methodological gap persists in integrating AI-based graphical extraction, intermediate data processing, and parametric automation into a coherent end-to-end workflow. Addressing this gap is essential to move from isolated AI applications toward reliable automation strategies that support data-driven decision-making and consistent model generation in BIM-based electrical system design [12].

Although prior research has extensively examined the application of artificial intelligence and automation within BIM-based environments, most contributions have focused on isolated tasks or domain-specific use cases, leaving unresolved questions about their systematic integration into end-to-end design workflows [6]. Existing studies have explored AI-driven methods for graphical information extraction [13], parametric modeling strategies, and interoperability solutions between CAD and BIM platforms [14], demonstrating measurable improvements in efficiency and data consistency. However, these approaches are often implemented independently, without a unified methodological structure that integrates graphical interpretation, intermediate data processing, and automated model generation within a single workflow. Limited attention has been paid to electrical design processes, in which circuit logic, outlet distribution, and panel assignment rely heavily on manual interpretation of CAD drawings and on repetitive decision rules. The lack of integrated solutions to automate these tasks contributes to inconsistencies between design documentation and BIM models, increasing coordination effort and the propagation of errors. In response to this gap, this research aims to propose an AI-driven, parametric workflow that systematically links CAD-based electrical drawings, automated graphical information extraction, structured data processing, and BIM model generation through programming-based automation. The proposed approach enables the automated identification and assignment of electrical outlets and circuits while preserving semantic coherence [illegible]

parametric programming, and CAD–BIM interoperability within a unified framework, this study advances a reproducible, scalable methodology that addresses critical limitations identified in the current literature and supports more reliable digital modeling of electrical systems.

2 Literature Background

The growing complexity of construction projects and the increasing reliance on digital workflows have reinforced Building Information Modeling (BIM) as a central platform for managing geometric and semantic information throughout the project lifecycle [15,16]. Although BIM has demonstrated significant benefits in coordination and information integration, its effectiveness is often constrained by continued reliance on two-dimensional CAD-based documentation, which requires manual interpretation and limits automation potential [2]. Recent research has therefore shifted toward incorporating computational and data-driven approaches to enhance the transformation of graphical information into structured BIM environments [17]. Advances in artificial intelligence, parametric programming, and interoperability strategies have enabled new methods for extracting, processing, and reusing design data with reduced human intervention [18]. Studies indicate that integrating image-based analysis, rule-based logic, and visual programming can improve consistency, traceability, and scalability in BIM workflows, especially in disciplines characterized by repetitive elements and strict design conventions [19]. Despite these developments, the literature remains fragmented regarding how these approaches are jointly implemented within coherent methodological frameworks. This fragmentation highlights the need for a structured review of existing contributions addressing AI-based graphical information extraction, parametric automation in BIM, and CAD–BIM interoperability, as foundational components for advancing automated and reliable digital modeling processes.

Recent academic contributions reveal a growing focus on integrating artificial intelligence, parametric automation, and interoperability within BIM-based workflows, driven by the need to reduce manual modeling and improve semantic consistency across digital construction processes. Montas-Laracuente et al. [20] presented a systematic examination of AI-driven Scan-to-BIM strategies, highlighting the dominance of deep learning approaches for semantic segmentation of point clouds and images, while exposing limitations in end-to-end automation and BIM interoperability. Kutá et al. [21] offered a comprehensive overview of AI adoption throughout BIM-enabled project lifecycles, demonstrating advances in automation, generative processes, and digital twins. However, their analysis underscored the dependence on structured data and expert validation. Araya-Aliaga et al. [22] introduced an integrated AI–BIM framework to automate the generation and validation of geometric and semantic information, reporting higher model coherence, reduced manual effort, and improved data traceability. Harle [23]

extracting visual and geometric data within BIM environments, reporting enhanced element recognition alongside unresolved challenges related to model generalization and platform interoperability. Taken together, these studies reflect steady progress in AI-assisted BIM workflows, yet they consistently point to fragmented implementations across disciplines and project stages.

Building on these developments, further research has advanced AI–BIM integration to support safety monitoring, parametric reconstruction, and performance-driven analysis. Wang et al. [24] demonstrated the feasibility of computer vision–based systems for identifying construction activities and unsafe behaviors from site imagery, achieving high accuracy in real-time safety monitoring. Salmi et al. [25] proposed a BIM-based parametric modeling strategy for generating Mine Information Models from heterogeneous spatial datasets, showing that algorithmic rules and parametric objects enable accurate reconstruction with limited human intervention. Semjén et al. [26] synthesized existing knowledge on parametric and generative design within BIM workflows, revealing optimization gains during early design phases while noting persistent issues related to standardization and lifecycle integration. Kamel et al. [3] analyzed BIM–BEM interoperability processes, identifying semantic inconsistencies and data loss during information exchange. Su et al. [27] developed a BIM–machine learning framework for automated property valuation that combines IFC data with optimized regression models, achieving higher accuracy than traditional methods. Zhang et al. [28] integrated BIM for building energy efficiency, highlighting its role in energy prediction and lifecycle assessment, but noted that interoperability issues and costs limit adoption. Collectively, the literature depicts a landscape in which AI, parametric automation, and interoperability are treated as partially isolated components. In response to this gap, the present paper advances an integrated methodology that connects CAD-based electrical drawings, AI-assisted graphical information extraction, and parametric automation in Dynamo to automate the assignment and connection of electrical outlets in BIM models. This contribution introduces a structured, end-to-end workflow that enhances consistency, reduces errors, and improves efficiency in electrical modeling processes, directly addressing limitations identified in prior studies.

3 Research Method

3.1 Image-Based Data Extraction Process

The methodological workflow began by formalizing a procedure for extracting information from CAD-derived images of electrical layouts, with a focus on outlet tags and distribution boards, as shown in plan views. A multimodal large language interface (Microsoft Copilot) was then utilized to identify and parse the visual attributes embedded in the CAD imagery. To ensure reliable extraction, a structured prompt engineering process was conducted through iterative refinement, ambiguity in label detection, color-based classification, and record structuring. After approximately 20 iterations, a stable prompt specification was established, enabling the automated identification of three key variables for downstream automation: the outlet identifier, the associated panelboard, and the count of outlets in the drawing segment.

The extracted information was exported into a structured Microsoft Excel file, which served as the standardized input dataset for the subsequent parametric assignment of outlets to panels and circuits within Autodesk Revit. Prompt calibration was conducted using three representative CAD images. Simultaneously, two professionals with extensive experience in interpreting CAD electrical drawings independently reviewed the same images and generated reference reports, which served as a baseline for comparison. The final prompt achieved complete agreement with the expert-derived outputs on the calibration set, confirming its reliability as a preprocessing mechanism for transforming unstructured plan-image content into machine-readable data for BIM-based automation. **Error! Reference source not found.** shows the instructions for generating the prompt for data extraction.

Table 1. Prompt used in data extraction.

Id	Actions	Instructions
1	Label identification	- Extract all visible labels in the image. - Each label follows the format BaseText-#Number. - Read and process the image from top to bottom. - If the first element is blue, classify it as Secondary (S) and the next as Primary (P), alternating thereafter. - If the first element is red, classify it as Primary (P) and the next as Secondary (S), alternating thereafter.
2	Base-text cleaning	- Remove the suffix -#Number and keep only the base text (e.g., RPPBB112-#1 → RPPBB112).
3	Table generation	- The table must contain three columns: a) Type: P or S b) Base label: the base text without the suffix c) Count: a consecutive index starting at 1 and ending at the total number of elements.

Although the prompt reached full agreement with the expert-generated outputs for the calibration images, the AI extraction stage should be interpreted as a calibrated preprocessing component rather than as a universally deterministic procedure. Because multimodal large language models may exhibit non-deterministic behavior, the reproducibility of the extraction depends on maintaining the same prompt structure, image preparation criteria, and review protocol. The extraction is also sensitive to variations in drawing legibility, annotation density, color conventions, and screenshot quality, which may affect robustness when different CAD standards are used. For this reason, the workflow incorporates expert validation during calibration and fixes the final prompt before evaluation. In addition, the present study assessed the extraction process only with Microsoft Copilot; therefore, the transferability of the results to other AI tools remains a relevant limitation that should be examined in future research.

3.2 Development of an Automated Tool in a BIM Environment

The BIM tool that automates the assignment of electrical outlets to distribution boards, leveraging the structured dataset produced by the Microsoft Copilot–based analysis, was implemented using Autodesk Dynamo as the primary development environment. As a prerequisite, a dedicated library of parametric families was created in Autodesk Revit to represent outlets and distribution board, encoding the key attributes required for identification, matching, and subsequent parameter assignment. These families were designed to conform to the labeling logic and graphical conventions observed in the CAD drawings analyzed by the AI component, thereby ensuring semantic alignment among the source documentation, the extracted Excel dataset, and the target BIM model. The Dynamo algorithm was structured into three main modules: (1) an input-processing module responsible for reading, parsing, and validating the Excel file generated by the AI pipeline; (2) a matching module that compares outlet identifiers obtained from the spreadsheet with the corresponding identifiers embedded in the Revit family instances, incorporating basic normalization rules to reduce mismatches due to formatting differences; and (3) an assignment module that programmatically links each matched outlet instance to its corresponding panelboard within the BIM model by writing the appropriate parameter values. In this third module, once a valid name-based correspondence is detected, the algorithm automatically populates the outlet instance with the associated board information, enabling consistent model-wide traceability of outlet-to-board relationships. An initial prototype was subjected to desktop-based functional tests to verify correct behavior under the expected input conditions and with the developed parametric families. Following calibration of the matching and assignment logic, a first compiled version of the Dynamo workflow was produced as the baseline implementation for subsequent evaluation and refinement.

3.3 Validation and Calibration

Building on the procedures outlined in Sections 3.1 and 3.2, the validation and calibration processes were designed to test the proposed workflow under realistic modeling conditions and to evaluate its operational performance against a conventional manual baseline. This assessment compared the standard Revit-based circuit assignment process, which relied on direct interpretation of CAD electrical documentation, with the automated implementations using Dynamo and Python. The primary performance indicators for this comparison were execution time and the number of misassignments identified during the review stage. A unified set of five real-world case studies from a data center project was selected, featuring outlet rows with varying densities and configurations. This selection aims to assess the workflow's stability across different input conditions. For each case study, a reference assignment was generated from the CAD documentation and used as a benchmark circuit-to-panelboard relationships.

Calibration was conducted before validation by performing desktop-based functional tests on an initial prototype. This was done to verify that the developed parametric families operated correctly with the expected input structure. Refinements focused on consolidating the matching and assignment logic to create a stable baseline workflow. Subsequently, the parameters were fixed to prevent any adjustments during the evaluation runs. To ensure comparability, manual timing excluded file preparation overhead and captured only the assignment activity, while automated timing measured the end-to-end script runtime. Accuracy was assessed by counting discrepancies identified during the review stage and comparing them with the reference assignment list. This approach enabled consistent reporting across cases and facilitated replication of the evaluation protocol.

4 Results

4.1 Proposed framework

Figure 1 presents the proposed framework for automating outlet-to-circuit assignment and circuit-to-panel connection in a BIM environment, based on information extracted from CAD drawings with AI assistance. The workflow starts by compiling the project’s technical documentation, generating screenshots from the AutoCAD plan views, selecting rows or groups of outlets, and verifying the legibility of electrical symbols and annotations. This step is repeated until a consolidated image is obtained with sufficient resolution and the largest possible number of clearly identifiable elements. The consolidated image is then processed in Microsoft Copilot using structured prompts that guide the interpretation of annotations and symbols, enabling the identification of outlets and the extraction of the required attributes, outlet identifier, circuit, and panel, which are organized into a dataset exported in Excel format. In the automation stage, the Excel file is imported into Dynamo to locate the corresponding outlets in the Revit model, retrieve their spatial coordinates (X, Y, and Z), establish spatial ordering, and execute Python scripts that generate circuits and connect each outlet to its corresponding distribution board. The outcome is an updated BIM model in which the electrical assignment consistently reflects the information specified in the CAD drawings, reducing manual effort during circuiting and panel configuration.

4.2 Evaluation framework and algorithmic workflow

To rigorously evaluate the performance of the Python-based automated workflow for assigning electrical outlets to circuits in BIM environments, five real-world case studies from a data center project were defined. The evaluation compared the traditional manual procedure in Revit, which relies on direct interpretation of electrical CAD drawings, with the proposed automated solution. Execution time and the number of errors detected during the review stage were selected as

in the literature as critical factors influencing BIM model reliability and efficiency in MEP disciplines. Figure 2 shows a representative row of outlets requiring circuit assignment, highlighting the high level of geometric and functional repetition that characterizes this configuration and that, under manual workflows, increases the likelihood of inconsistencies. This operational context is particularly relevant, as previous studies on BIM automation emphasize that the greatest efficiency gains occur in repetitive, rule-governed tasks driven by structured graphical information.

Algorithmic architecture comprises four clearly defined stages, as shown in Figure 2**Error! Reference source not found.**. In the first stage, the system automatically identifies outlet families within the BIM model and extracts their spatial coordinates (X, Y, Z). The second stage processes the Y-coordinate exclusively, which is critical in this case study, given the vertical arrangement of the outlets. Sorting these values in descending order establishes a spatial hierarchy consistent with the logic embedded in the original electrical drawings. Subsequently, a Python-based node filters secondary outlets, recognizing that they appear in paired configurations with primary outlets and that the study focuses only on assigning primary elements. The third stage automatically generates electrical circuits in Revit, associating each primary outlet with its corresponding circuit. In contrast, the fourth stage performs the bulk connection of all generated circuits to the selected panelboard, completing the workflow without manual intervention (see Figure 2). This sequential structure ensures traceability between extracted geometric data and electrical assignment decisions, directly addressing limitations reported in prior CAD–BIM interoperability solutions.

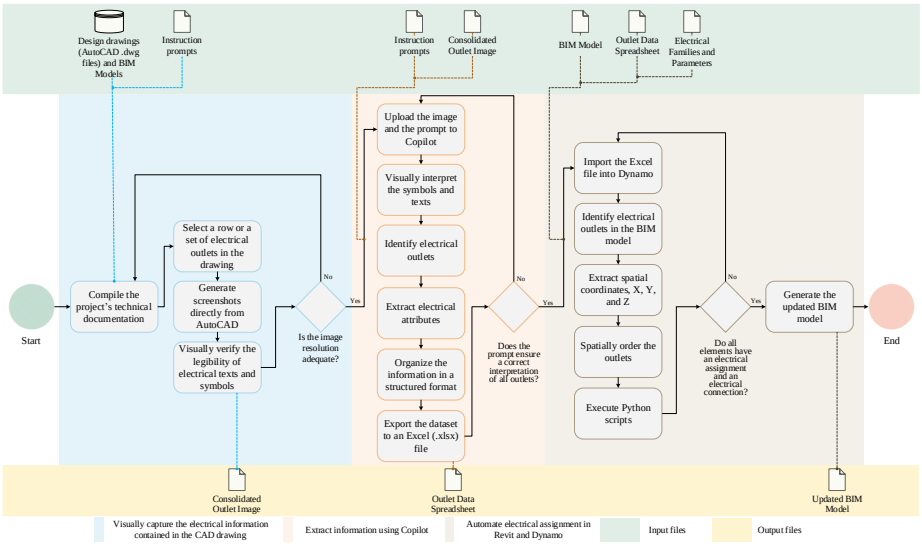


Figure 1. Proposed framework.

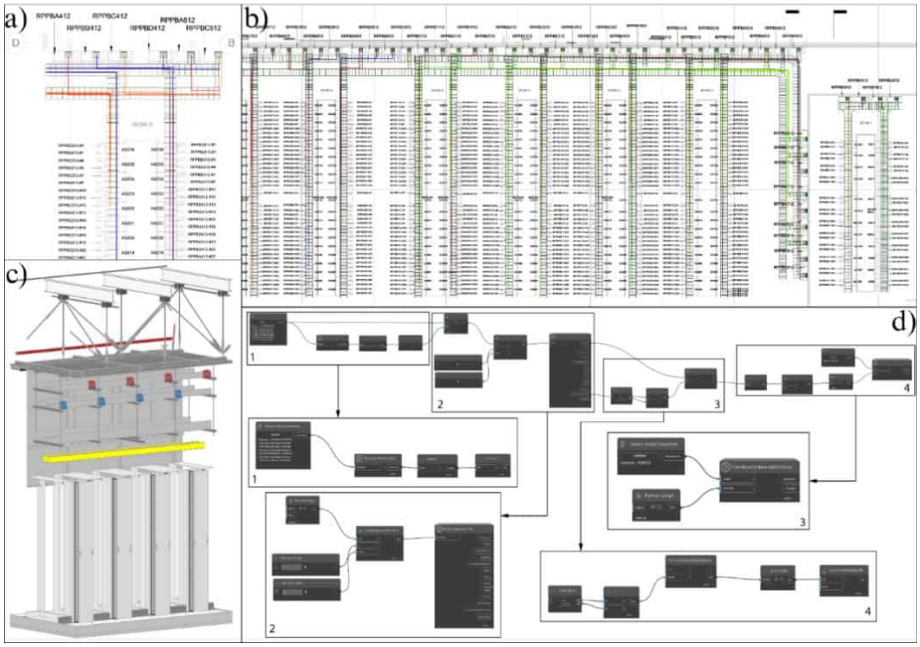


Figure 2. a) Row of outlets that must be assigned to a circuit. b) Number of sockets to be assigned to a circuit in the model. c) Dynamo Algorithm. d) Outlets are identified in blue and red.

4.3 Case study analysis and critical performance assessment.

The validation of automated BIM workflows requires empirical evidence that goes beyond proof-of-concept demonstrations and explicitly examines their behavior under realistic modeling conditions. In this context, the proposed AI-driven parametric workflow was evaluated through a unified set of five real case studies derived from a data center project, each representing rows of electrical outlets with varying densities, configurations, and operational constraints. The assessment compares the conventional manual circuit-assignment process in Revit, guided by CAD-based electrical documentation, with the automated solution implemented through Python scripts and Dynamo integration. Execution time and assignment accuracy were adopted as the primary performance indicators, since both metrics directly affect modeling efficiency, coordination reliability, and the quality of subsequent construction documentation (Table 22. Impact of AI-driven parametric automation on efficiency and accuracy in CAD–BIM electrical workflows.). The reported time corresponds exclusively to the circuit-assignment stage: manual assignment time in Revit versus the algorithm execution time once the input is available. The preparation of that input file, generated by the AI component after the extraction process, was not included in the measurement because prompt refinement represents a one-time, non-recurring calibration cost, while CAD image capture was also excluded because it is independent of the assignment process under analysis. The results reveal a consistent and substantial reduction in execution time when automation is applied, with gains increasing proportionally to the number of outlets involved. This behavior suggests that the algorithm exhibits favorable scalability characteristics, particularly in environments dominated by repetitive elements and a strict sequencing logic. In large outlet rows with more than 40 elements, the automated workflow effectively mitigates the cumulative delays inherent to manual selection, verification, and circuit creation, a limitation frequently reported in BIM-based electrical coordination studies.

In terms of accuracy and model reliability, the automated workflow performed consistently across all evaluated cases, although a slight discrepancy in Case 3 (48 outlets) indicates that minor variations may still occur under specific conditions. By contrast, the manual approach showed increasing inconsistencies as task repetition and dataset size grew, mainly due to incorrect circuit–panel associations, misinterpretation of CAD references, and omissions in repetitive assignments. Supported by geometric filtering, rule-based ordering, and parametric control, the deterministic algorithm produced consistent and reproducible outcomes that reduced human-induced variability. Although time savings were evident even in small-scale scenarios, the advantage of automation became more pronounced in larger datasets, where manual execution led to disproportionate increases in time and error risk. Unlike prior AI-assisted BIM approaches focused on isolated tasks, the proposed workflow integrates graphical

interpretation, data processing, and circuit assignment into a unified pipeline, strengthening model consistency and supporting scalable electrical system modeling in complex BIM environments. Table 2 summarizes the case studies, comparing manual and automated methods by execution time and errors.

Table 22. Impact of AI-driven parametric automation on efficiency and accuracy in CAD–BIM electrical workflows.

Case Studies	# Outlets	Manual method		Automated method	
		Time (min)	Errors	Time (min)	Errors
Case 1	42	31.50	6	0.03	0
Case 2	24	12.20	2	0.02	0
Case 3	48	72.50	10	0.01	1
Case 4	12	8.32	1	0.02	0
Case 5	6	4.33	0	0.02	0

4.4 Comparative impact on efficiency, accuracy, and model reliability

Beyond improvements in execution time and error reduction, the comparative evaluation reveals structural limitations in prevailing electrical modeling automation approaches. Prior studies often focus on partial automation or decision-support systems that accelerate isolated tasks while maintaining manual interpretation during critical CAD–BIM translation stages. In contrast, the proposed workflow demonstrates that substantial performance gains arise when rule definition, geometric filtering, and assignment logic are fully formalized within a unified algorithmic process. This approach effectively decouples modeling effort from the number of elements, overcoming the efficiency decline commonly observed in manual and semi-automated BIM practices as task repetition increases.

From an accuracy perspective, the findings indicate an epistemic shift rather than a minor technical improvement. Manual workflows depend on continuous cognitive validation, introducing variability, fatigue-related errors, and inconsistencies that scale with repetition. While previous studies mitigate these issues through visual checks or post-processing validation, the proposed method embeds decision rules directly into the execution logic, eliminating interpretative ambiguity at its source. The results therefore challenge the assumption that human oversight is essential for correctness in complex electrical assignments, showing that deterministic, geometry-driven logic can outperform manual reasoning in highly repetitive and regulated design contexts.

The workflow also improves BIM model reliability compared with fragmented automation strategies reported in prior research. Whereas many studies automate extraction or classification without resolving downstream connectivity, the proposed approach ensures consistent outlet-to-circuit and circuit-to-panel relationships, producing a more stable, auditable, and reusable model for coordination, simulation, and operation. Overall, the results support viewing AI-driven and parametric CAD–BIM automation not merely as a productivity enhancement, but as a structural mechanism for enforcing design logic and reducing systemic uncertainty.

Conclusions

This study provides three main theoretical contributions. First, it proposes an end-to-end CAD-to-BIM automation workflow for electrical outlet assignment by integrating graphical capture, AI-assisted information extraction, structured data processing, and BIM-based parametric execution. Second, it defines the role of AI as a mechanism for converting unstructured electrical CAD information into structured datasets that support rule-based BIM automation while reducing manual interpretation. Finally, it introduces a reproducible rule-driven logic for outlet-to-circuit and circuit-to-panelboard assignment, demonstrating how repetitive MEP tasks can be formalized into deterministic BIM workflows that improve semantic consistency in model authoring.

The theoretical contributions of this study have practical applications in industry workflows. The proposed method integrates AI-based data structuring with Dynamo and Python scripting in Revit, significantly reducing the time for outlet and circuit assignments while enhancing consistency during model reviews. Validation through five case studies highlighted two representative tests: in the first, manual assignment of 42 outlets took 31.5 minutes with six errors, while the automated workflow completed it in 0.033 seconds without mistakes. In the second, the manual process for 24 outlets took 12 minutes with two errors, and the automated method finished in 0.025 seconds error-free. These results demonstrate a shift from expert-dependent processes prone to human error to a rule-based, structured input workflow, reducing rework and improving model reliability, especially in projects with repetitive electrical configurations.

Despite the contributions of this study, several limitations remain. The empirical assessment was based on two data center case studies with high geometric repetition and a spatial layout that enabled classification mainly along a single coordinate, which may restrict the generalizability of the findings to projects with more heterogeneous designs, multiple distribution logics, or different outlet-matching patterns. In addition, the

proposed workflow depends on consistent outlet family identification and naming across CAD documentation, the AI-assisted structured dataset, and BIM parameters, so variations in drafting standards, screenshot quality, or prompt sensitivity may affect its robustness. The system also does not explicitly verify compliance with electrical engineering regulations, including circuit load capacity, load balancing, or conformity with applicable codes, as these were treated as predefined input conditions. Future research should therefore validate the approach across a wider range of building typologies and electrical design conventions, explore direct data extraction from CAD sources instead of image captures, incorporate automated checks for electrical rules and model consistency, and examine adoption-related aspects, including setup effort, script maintainability, and input governance at the organizational scale.

5 Acknowledgements

Kevin Torres and Luis Cristancho gratefully acknowledge the Doctorate in Intelligent Industry program at the Pontificia Universidad Católica de Valparaíso for the academic training and institutional support provided during their doctoral studies. Karen Castañeda gratefully acknowledges the financial support provided by the Dirección de Investigación of the Vicerrectoría de Investigación, Creación e Innovación, VINCI-DI, at the Pontificia Universidad Católica de Valparaíso, through the DI Regular PUCV 2026 project, grant number 044.783/2026.

6 References

[1] T. Wang, H.M. Chen, Integration of building information modeling and project management in construction project life cycle, Autom. Constr. 150 (2023) 104832. https://doi.org/10.1016/J.AUTCON.2023.104832.

[2] Y. Kim, S. Chin, S. Choo, BIM data requirements for 2D deliverables in construction documentation, Autom. Constr. 140 (2022) 104340. https://doi.org/10.1016/J.AUTCON.2022.104340.

[3] E. Kamel, A.M. Memari, Review of BIM's application in energy simulation: Tools, issues, and solutions, Autom. Constr. 97 (2019) 164–180. https://doi.org/10.1016/J.AUTCON.2018.11.008.

[4] J. Miller, L. Schneiderbauer, M. Hauer, A. Jäger, G. Fröch, R. Pfluger, S. Moser, Enhancing Interoperability Between Building Information Modeling and Building Energy Modeling: Alphanumerical Information Exchange for Energy Optimization in Early Design Stages, Applied Sciences 2025, Vol. 15, Page 5789 15 (2025) 5789. https://doi.org/10.3390/APP15105789.

[5] M.E. Agri, A. Le, Q. Phung, AI integration in architectural design and management: professionals' perspectives, Architectural Engineering and Design Management (2025). https://doi.org/10.1080/17452007.2025.2548911;WGROUP:STRING:PUBLICATION.

[6] A.A. Khan, A.O. Bello, M. Arqam, F. Ullah, Integrating Building Information Modelling and Artificial Intelligence in Construction Projects: A Review of Challenges and Mitigation Strategies, Technologies 2024, Vol. 12, Page 185 12 (2024) 185. https://doi.org/10.3390/TECHNOLOGIES12100185.

[7] J.L. Alves, R.P. Palha, A.T. de Almeida Filho, Towards an integrative framework for BIM and artificial intelligence capabilities in smart architecture, engineering, construction, and operations projects, Autom. Constr. 174 (2025) 106168. https://doi.org/10.1016/J.AUTCON.2025.106168.

[8] G. Culot, M. Podrecca, G. Nassimbeni, Artificial intelligence in supply chain management: A systematic literature review of empirical studies and research directions, Comput. Ind. 162 (2024) 104132. https://doi.org/10.1016/J.COMPIND.2024.104132.

[9] A.S. Mosallam, Q. Amarkhil, M. Hegab, A. Alroomi, AI-Driven Semantic Framework for Automated Construction Planning and Scheduling with BIM and Digital Twin Integration, Engineering Proceedings 2025, Vol. 112, Page 3 112 (2025) 3. https://doi.org/10.3390/ENGPROC2025112003.

[10] Y. Arayici, E. Balodimou, M. Saki, C. Eduardo Gomes de Souza, C. Kowal Chinelli, C. Alberto Pereira Soares, O. Celso Longo, BIM as a Tool for Developing Smart Buildings in Smart Cities: Potentialities and Challenges, Architecture 2025, Vol. 5, Page 103 5 (2025) 103. https://doi.org/10.3390/ARCHITECTURE5040103.

[11] T. Grønsund, M. Aanestad, Augmenting the algorithm: Emerging human-in-the-loop work configurations, The Journal of Strategic Information Systems 29 (2020) 101614. https://doi.org/10.1016/J.JSIS.2020.101614.

[12] Y. Yan, C. Wang, F. Zhang, L. Wu, J. Zang, T. Liu, S. Tang, Z. Ye, Development and Implementation of a System for Electrical Engineering BIM Detailed Design in Construction Projects, Buildings 2025, Vol. 15, Page 2960 15 (2025) 2960. https://doi.org/10.3390/BUILDINGS15162960.

[13] T. Ye, M. He, Y. Wang, J. Wang, L. Zhu, J. Zhang, An Optimized Intelligent Segmentation Algorithm for Concrete Cracks Based on Transformer, Electronics 2025, Vol. 14, Page 1720 14 (2025) 1720. https://doi.org/10.3390/ELECTRONICS14091720.

[14] A.S. Allam, M. Nik-Bakht, Integrating industry foundation classes and knowledge graphs for automated deconstruction planning, Journal of Building Engineering 106 (2025) 112564. https://doi.org/10.1016/J.JOBE.2025.112564.

[15] N.S.A. Rahim, S. Ismail, C. Subramaniam, S.N.H. Abdullah Habib, S. Durdyev, Building Information Modelling Strategies in Sustainable Housing Construction Projects in Malaysia, Sustainability 2023, Vol. 15, Page 2313 15 (2023) 2313. https://doi.org/10.3390/SU15032313.

[16] S.D. Datta, M.H.R. Sobuz, N.J. Mim, A.D. Nath, Investigation on the effectiveness of using building information modeling (BIM) tools in project management: a case study, Journal of Construction 22 (2023) 306–320. https://doi.org/10.7764/RDLC.22.2.306.

[17] L. Garlet, M.K. Bracht, R. Lamberts, A.P. Melo, J. O'Donnell, Artificial intelligence to enhance BIM-BEPS integration via IFC: Challenges, solutions, and future directions, Advanced Engineering Informatics 69 (2026) 103824. https://doi.org/10.1016/J.AEI.2025.103824.

[18] M. Ehtsham, G. Parisi, F. Pedone, F. Rossi, M. Zincani, E. Congiu, C. Marchionni, AI-Powered Advanced Technologies for a Sustainable Built Environment: A Systematic Review on Emerging Challenges, Sustainability (Switzerland) 17 (2025) 8005. https://doi.org/10.3390/SU17178005/S1.

[19] R. Valdebenito, E. Forcael, Integrating Artificial Intelligence and BIM in Construction: Systematic Review and Quantitative Comparative Analysis, Applied Sciences 2025, Vol. 15, Page 12470 15 (2025) 12470. https://doi.org/10.3390/APP152312470.

[20] N. Montás-Laracuente, E. Delgado-Martos, C. Pesqueira-Calvo, G. Intra Sidola, A.M. Maitín, Á.J. García-Tejedor, A. Nogales, A systematic review of artificial intelligence for capturing real-world structures into building information modelling, Journal of Building Engineering 113 (2025) 114093. https://doi.org/10.1016/J.JOBE.2025.114093.

[21] D. Kutá, M. Faltejsek, The Role of Artificial Intelligence in the Transformation of the BIM Environment: Current State and Future Trends, Applied Sciences 2025, Vol. 15, Page 9956 15 (2025) 9956. https://doi.org/10.3390/APP15189956.

[22] E. Araya-Aliaga, E. Atencio, F. Lozano, J. Lozano-Galant, Automating Dataset Generation for Object Detection in the Construction Industry with AI and Robotic Process Automation (RPA), Buildings 2025, Vol. 15, Page 410 15 (2025) 410. https://doi.org/10.3390/BUILDINGS15030410.

[23] S.M. Harle, Advancements and challenges in the application of artificial intelligence in civil engineering: a comprehensive review, Asian Journal of Civil Engineering 25 (2024) 1061–1078. https://doi.org/10.1007/S42107-023-00760-9/FIGURES/5.

[24] B. Wang, C. Yin, H. Luo, J.C.P. Cheng, Q. Wang, Fully automated generation of parametric BIM for MEP scenes based on terrestrial laser scanning data, Autom. Constr. 125 (2021) 103615. https://doi.org/10.1016/J.AUTCON.2021.103615.

[25] J. Salmi, Z. Ye, J. Ninic, R. Heikkilä, BIM for mining - Automated generation of information models using a parametric modelling concept, International Journal of Rock Mechanics and Mining Sciences 186 (2025) 106032. https://doi.org/10.1016/J.IJRMMS.2025.106032.

[26] Á. Semjén, J. Szép, Integrating generative and parametric design with BIM: A literature review of challenges and research gaps in construction design, Applications in Engineering Science 23 (2025) 100253. https://doi.org/10.1016/J.APPLES.2025.100253.

[27] T. Su, H. Li, Y. An, A BIM and machine learning integration framework for automated property valuation, Journal of Building Engineering 44 (2021) 102636. https://doi.org/10.1016/J.JOBE.2021.102636.

[28] T. Zhang, X. Yang, Z. Wu, G. Zhai, D.T. Doan, Q. Sun, H. Gao, Application of Building Information Modeling for Energy Efficiency: A Systematic Review, Buildings 2025, Vol. 15, Page 3722 15 (2025) 3722. https://doi.org/10.3390/BUILDINGS15203722.

Industrial3D: A Large-Scale Industrial MEP Point Cloud Dataset and Benchmark for Complex Industrial Scene Understanding

Chao Yin[1,2], Bing Sun[1], Difeng Hu[3], Boyu Wang[4], Mingkai Li[5], Jack C.P. Cheng[1]

[1]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[2]Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou, China
[3]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China
[4]New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[5]Department of the Built Environment, National University of Singapore, Singapore

Abstract -

Automated point cloud understanding is critical for modern digital construction workflows, yet existing datasets focus predominantly on architectural elements, leaving complex mechanical, electrical, and plumbing (MEP) systems severely underrepresented. This paper introduces Industrial3D, the first large-scale, high-fidelity dataset specifically designed for industrial MEP scene understanding. Industrial3D comprises over 610 million expertly labeled points across 13 water treatment facilities, captured at 6mm terrestrial LiDAR resolution with 12 specialized MEP and structural classes, representing 4× the scale of existing construction datasets. The dataset exhibits extreme class imbalance (215:1 ratio) and geometric ambiguity where over 86% of tail-class points share cylindrical primitives with the dominant pipe class, creating systematic challenges absent in architectural benchmarks. We evaluate representative methods across learning paradigms, including fully-supervised approaches (KPConv, ResPointNet++, PosPool, and a class-balanced variant) and unsupervised methods (GrowSP). Our evaluation reveals that current state-of-the-art methods face significant limitations in specialized industrial domains, particularly for small, functionally critical components such as valves, strainers, and reducers. Fully-supervised methods achieve approximately 53–54% mIoU, while unsupervised GrowSP reaches only 11.7% mIoU and fails completely on most tail classes. These results point to research opportunities in long-tailed learning and geometry-aware segmentation for industrial scenes. Industrial3D provides a resource for advancing automated construction verification and Scan-to-BIM workflows. Dataset and code will be publicly available at: `https://github.com/pointcloudyc/Industrial3D`.

Keywords -

Point Cloud Dataset; MEP Systems; Semantic Segmentation; Benchmark; Terrestrial Laser Scanning; Industrial Scene Understanding

1 Introduction

Automated point cloud understanding is essential for modern digital construction workflows, including as-built verification, progress monitoring, and automated Building Information Model (BIM) generation [1, 2]. Terrestrial laser scanning (TLS) has become the standard for rapid, high-fidelity 3D data acquisition, capturing billions of geometric measurements of complex mechanical, electrical, and plumbing (MEP) systems. However, a critical bottleneck persists: transforming raw point clouds into actionable semantic understanding remains challenging, particularly for complex industrial environments where MEP systems exhibit intricate geometric relationships, extreme scale variations, and severe occlusion.

Consider a typical water treatment facility: pipes with identical cylindrical geometry serve different functions, small components like valves are embedded within larger structural systems, and dense equipment arrangements create challenging visibility conditions. Current point cloud understanding methods achieve strong results on architectural datasets like S3DIS [3] and ScanNet [4], but struggle with the specialized demands of industrial MEP environments. Existing datasets exhibit three fundamental limitations: (1) *domain mismatch*, focusing on furniture and architectural elements rather than MEP systems; (2) *insufficient scale*, with inadequate point density and scene complexity for fine-grained industrial analysis; and (3) *limited class diversity*, lacking specialized MEP components critical for construction applications.

To address these limitations, we introduce Industrial3D, the first large-scale, high-fidelity dataset specifically designed for industrial MEP scene understanding. Industrial3D comprises over 610 million expertly labeled points across 13 water treatment facilities, captured at 6mm TLS resolution with 12 specialized MEP and structural classes. We also evaluate representative methods across learning paradigms, including fully-supervised approaches (KPConv [5], ResPointNet++ [6], PosPool [7], and a class-

balanced variant using CB loss [8]) and the unsupervised method GrowSP [9], providing baseline performance metrics and analysis of domain-specific challenges.

2 Related Work

2.1 Point Cloud Datasets

Large-scale annotated datasets have driven progress in point cloud understanding. S3DIS [3] provides 273 million points across 271 rooms with 13 classes focused on architectural elements and furniture. ScanNet [4] expands to 1,500 indoor scans with 20 classes but maintains similar architectural bias. Semantic3D [10] offers outdoor scenes with 4 billion points but limited class diversity. Construction-specific datasets remain scarce: CLOI [11] includes industrial objects but has limited public availability, while PSNet5 [6] provides a medium-scale MEP-specific publicly available dataset with 80 million points across 5 classes, yet remains limited in scale and class diversity. Table 1 summarizes the comparison.

These datasets share critical limitations for MEP understanding: domain mismatch (architectural focus versus MEP complexity), insufficient resolution for fine-grained components, and absence of specialized classes like valves, flanges, and pumps. Industrial3D addresses these gaps as the first large-scale public dataset targeting MEP understanding, with 610M+ points at 6mm resolution across 12 specialized classes—4× the scale of existing construction datasets.

Table 1. Comparison of Industrial3D with existing point cloud datasets. M = million, B = billion. * indicates limited public availability.

Dataset	Type	Points	Classes	Imbal.
S3DIS	Indoor	273M	13	62:1
ScanNet	Indoor	242M	20	~25:1
Semantic3D	Outdoor	4B	8	75:1
CLOI*	Industrial	140M	10	–
PSNet5	Industrial	80M	5	~50:1
Industrial3D	**Industrial**	**610M**	**12**	**215:1**

2.2 Point Cloud Semantic Segmentation

Deep learning has driven significant progress in point cloud semantic segmentation [12]. PointNet [13] pioneered neural architectures for 3D point clouds, while PointNet++ [14] extended this with hierarchical feature learning. Recent advances include kernel-based convolutions (KPConv [5]), efficient random sampling strategies (RandLA-Net [15]), and local aggregation operators (PosPool [7]).

Construction-specific applications have further explored these architectures. Our prior work ResPointNet++ [6] integrated residual learning into PointNet++ for industrial MEP segmentation with 5 classes on the PSNet5 dataset, and SE-PseudoGrid [16] introduced squeeze-and-excitation attention for piping component classification. Beyond supervised learning, unsupervised methods like GrowSP [9] learn semantic groupings without labels through primitive clustering and superpoint growing. Industrial3D enables systematic evaluation of these representative methods on real-world industrial MEP scenes with 12 classes.

3 Industrial3D Dataset

3.1 Data Acquisition and Annotation

The Industrial3D dataset was captured from 13 large-scale rooms within operational water treatment facilities in Hong Kong. These scenes were selected for their diverse and complex arrangements of MEP and structural components. The total scanned area covers approximately 20,000 m^2, encompassing over 2.3 billion raw 3D points. Table 2 provides statistics of seven representative scenes.

Table 2. Statistics of seven representative scenes in Industrial3D, showing the scale and diversity of industrial environments captured.

Scene	Length (m)	Width (m)	Area (m^2)
Chiller	45.1	19.2	865.92
OGB	86.2	42.5	3,663.50
OSCG	15.3	38.2	584.46
PBF2_1	58.0	22.0	1,276.00
PBF2_2	68.4	38.2	2,612.88
SPH	36.1	26.9	971.09
WRT2	53.3	23.7	1,263.21
Total	-	-	**11,237.06**

High-fidelity point cloud data was acquired using a Leica BLK360 terrestrial laser scanner [17], as shown in Figure 1. This instrument captures 360,000 points per second with 6mm accuracy at 10-meter range, which is critical for capturing small MEP components. Multiple scans were performed within each room from different locations to ensure comprehensive coverage. Raw scan data was processed using Autodesk ReCap software for registration and format conversion.

The annotation process required approximately 754 person-hours to complete, with five domain experts labeling over 610 million points. We defined 12 semantic classes encompassing MEP and structural elements, as detailed in Table 3. Using CloudCompare [18], annotators followed a systematic pipeline: tiling large scenes into manageable chunks, extracting point clusters for each class, merging results, and refining boundaries with outlier removal. Figure 2 illustrates representative instances of each semantic category. Notably, the geometric similarity among pipe-related classes (pipe, elbow, tee, reducer) creates significant geometric ambiguity, where over 86%



Figure 1. Data acquisition setup for Industrial3D (source from [6]). (a) Partial view of the WRT2 scene showing complex industrial piping systems. (b) Leica BLK360 terrestrial laser scanner during data acquisition with iPad controller for scan registration and preview.



Figure 2. Representative instances of each semantic category in Industrial3D, visualized as synthetic point clouds sampled from BIM models. Note: these are idealized representations; actual LiDAR scans exhibit significant occlusion, noise, and incomplete coverage due to terrestrial scanning geometry.

of tail-class points share cylindrical primitives with the dominant pipe class.

Table 3. The 12 semantic classes in Industrial3D, grouped by frequency. Head classes account for 77% of total points; tail classes each represent less than 3%.

Class	Description
Head classes (77% of points)	
Pipe	Cylindrical conduits for fluid transport
RectangularBeam	Rectangular structural support members
Duct	Rectangular air handling conduits
Common classes	
Ibeam	I-shaped structural steel beams
Tank	Large storage vessels
Tail classes (<3% each)	
Flange	Circular connecting components
Valve	Flow control devices
Pump	Mechanical fluid movement devices
Strainer	Filtration devices in pipe systems
Elbow	Angular pipe connectors
Tee	T-shaped pipe junction fittings
Reducer	Pipe diameter transition fittings

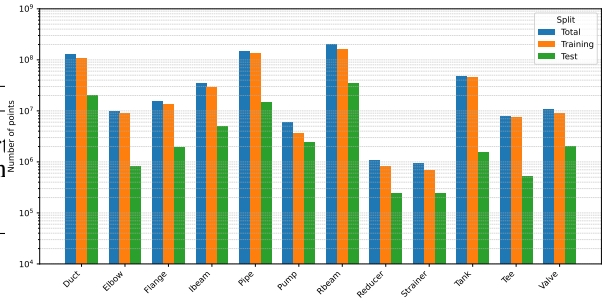


Figure 3. Category-wise point distribution in Industrial3D (logarithmic scale). The severe class imbalance (215:1 ratio) between head classes (Pipe, RectangularBeam, Duct) and tail classes (Reducer, Strainer, Tee) is characteristic of real-world industrial environments.

3.2 Dataset Statistics and Visualization

The final dataset comprises over 610 million expertly labeled points. Each point includes 3D coordinates (X, Y, Z), intensity, and RGB color values. The dataset exhibits significant class imbalance with a 215:1 ratio between the most and least frequent classes—3.5× more severe than S3DIS (62:1). Figure 3 shows the category-wise point distribution on a logarithmic scale. Figure 4 presents the distribution of semantic categories across different industrial areas. Figure 5 provides qualitative visualizations of two representative scenes.

3.3 Dataset Characteristics

Industrial3D differs from existing datasets in several ways: (1) **Scale**, 4× larger than existing industrial datasets with 610M+ labeled points; (2) **Public availability**, as the first large-scale publicly available industrial MEP dataset; (3) **Class diversity**, with 12 specialized MEP and structural classes absent from architectural benchmarks; and (4) **Challenge severity**, the most extreme class imbalance (215:1) among comparable benchmarks, combined with geometric ambiguity where functionally different components share similar cylindrical shapes.

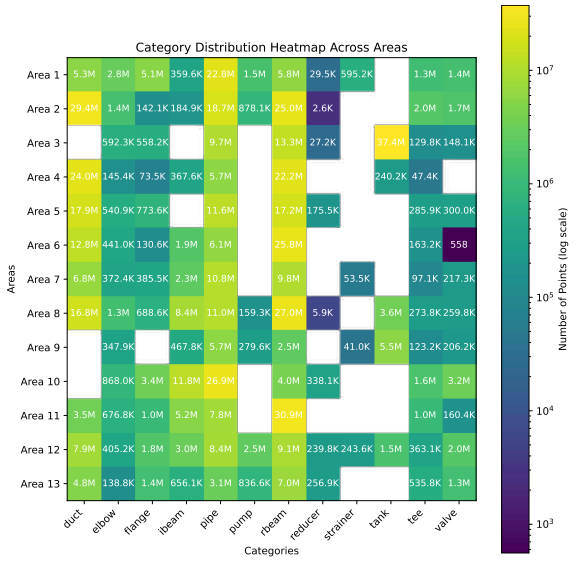


Figure 4. Heatmap showing distribution of semantic categories across industrial areas in Industrial3D. Values are displayed in millions (M) or thousands (K) of points. Structural elements are distributed across all areas, while specialized functional components are concentrated in operation-specific zones.

4 Benchmark Experiments

4.1 Experimental Setup

We partition Industrial3D into a test set (Area 6 and Area 12) and a training set (remaining areas), following the S3DIS area-based evaluation protocol [3]. Area 6 (PBF2_2, Primary Biological Filter) and Area 12 (SPH, Sludge Press House) were selected as test sets because they represent spatially separated facilities with distinct operational characteristics: Area 6 features dense piping arrangements typical of biological filtration systems, while Area 12 contains all diverse mechanical equipment including presses and pumps. This split ensures evaluation on unseen facility types, testing generalization capability across different industrial environments. We evaluate representative methods across two learning paradigms: (1) fully-supervised approaches including ResPointNet++ [6], KPConv [5], PosPool [7], and ResPointNet++ with class-balanced (CB) loss [8]; and (2) the unsupervised method GrowSP [9]. We report overall accuracy (OA), mean Intersection-over-Union (mIoU), and per-class IoU as evaluation metrics.

4.2 Fully-Supervised Results

Table 4 presents benchmark results for fully-supervised methods. All methods achieve similar overall performance (mIoU $\approx$ 53–54%) but exhibit substantial gaps on tail classes. ResPointNet++ achieves 53.31% mIoU

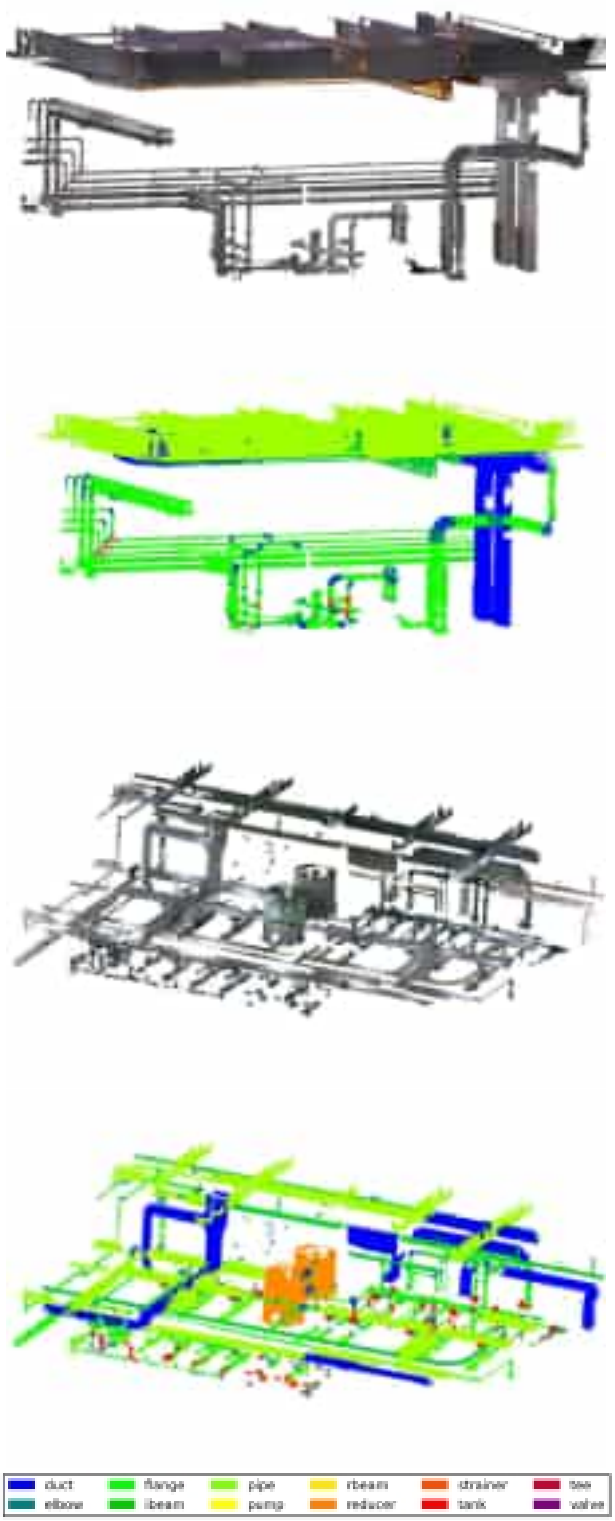


Figure 5. Qualitative visualization of two representative scenes from Industrial3D. Each scene shows: (1) raw point cloud data with RGB color; (2) semantic annotations with class-specific coloring. Rows from top to bottom: Area 6-1 (dense piping region) raw data and annotated data, and Area 12 (Sludge Press House - SPH scene) raw data and annotated

with strong performance on head classes (Duct: 91.36%, Ibeam: 99.41%, RectBeam: 95.19%) but struggles with rare components (Reducer: 0.00%, Strainer: 0.00%). KPConv shows similar patterns with 53.65% mIoU. PosPool achieves non-zero performance on Reducer (34.48%) but comparable overall mIoU at 53.18%. The class-balanced variant (ResPointNet++ w. CB) improves Reducer performance to 14.92% and achieves the highest mIoU at 54.05%, demonstrating that addressing class imbalance provides modest gains.

The results reveal a "dual crisis" in industrial point cloud segmentation: statistical rarity combined with geometric ambiguity. Critical functional components (valves, pumps, strainers) achieve significantly lower IoU than structural elements. The complete failure on Strainer (0.00% for all methods) results from multiple factors: (1) insufficient training samples (Strainer represents only 0.6% of training points), (2) geometric similarity to pipes (both are cylindrical), and (3) small physical size making these components difficult to distinguish from surrounding clutter. For Reducer and Tee, geometric ambiguity is the primary issue, these components share cylindrical primitives with pipes, causing systematic confusion. The class-balanced variant shows modest improvement on Reducer (14.92% IoU versus 0.00%), confirming that frequency-based re-weighting helps but cannot fully resolve geometric ambiguity. For an in-depth treatment including proposed solutions, we refer readers to our long-tailed learning methods paper [19].

4.3 Unsupervised Results

Table 5 presents results for GrowSP [9], which learns semantic groupings through unsupervised primitive clustering and superpoint growing. We evaluate three stages of the GrowSP pipeline: initial primitive extraction, superpoint formation, and final primitive merging to understand performance at each processing level.

GrowSP achieves 84.57% OA and 42.06% mIoU at the superpoint stage, demonstrating reasonable performance on geometrically distinct classes such as Tank (90.61%), RectBeam (86.23%), and Pipe (73.46%). However, performance degrades substantially for fine-grained MEP components: Elbow (4.19%), Reducer (0.00%), and Strainer (0.00%) are nearly or completely unrecognized. The primitive merging stage leads to severe performance collapse (mIoU drops to 11.73%), indicating that the merging heuristics fail to preserve semantic boundaries in complex industrial scenes.

These results demonstrate that while GrowSP can capture coarse geometric structure without supervision, it fundamentally struggles with the fine-grained distinctions required for MEP components. The 11-point mIoU gap between GrowSP (42.06%) and supervised methods (∼54%) indicates that domain-specific supervision remains essential for accurate industrial scene understanding.

5 Conclusion and Future Work

We introduced Industrial3D, the first large-scale, high-fidelity point cloud dataset for industrial MEP scene understanding. With over 610 million labeled points across 12 specialized classes from 13 water treatment facilities, Industrial3D provides 4× the scale of existing construction datasets while exhibiting authentic industrial challenges: severe class imbalance (215:1), geometric ambiguity where 86% of tail-class points share primitives with head classes, and complex occlusion patterns. Our benchmark evaluation reveals that fully-supervised methods achieve approximately 53–54% mIoU but struggle significantly with tail classes. The unsupervised GrowSP method achieves only 11.7% mIoU and fails completely on multiple tail classes (Reducer, Strainer, Tee at the merged stage). These results point to fundamental research gaps in handling geometrically ambiguous and statistically rare components.

Addressing these challenges requires multi-faceted approaches. For extreme class imbalance, techniques such as class-balanced loss functions [8], re-sampling strategies, and two-stage training can help models learn from tail classes. However, as our class-balanced experiments show (Reducer improving from 0.00% to 14.92% IoU), re-weighting alone is insufficient due to geometric ambiguity. Effectively distinguishing functionally different but geometrically similar components (e.g., pipes versus reducers) requires geometry-aware segmentation that leverages topological relationships, cylindrical primitive fitting, and multi-modal feature fusion combining geometric (XYZ), appearance (RGB), and intensity information. Industrial3D provides the testbed for developing such specialized approaches.

Future work will expand benchmarking to include weakly-supervised methods (SQN [20], CPCM [21], and WSSS-ST [22]), self-supervised approaches (PointContrast [23], OcCo [24]), and foundation models (Point-SAM [25], Reason3D [26]) to evaluate their effectiveness under sparse supervision and zero-shot settings. Beyond semantic segmentation, Industrial3D enables research in instance segmentation, geometric primitive fitting, and topological relationship inference, critical steps toward fully automated Scan-to-BIM workflows. We will release Industrial3D publicly to support research in automated construction verification and digital twin generation.

Acknowledgments

This work was supported by the China Postdoctoral Science Foundation (No. 2023M740761).

Table 4. Benchmark results of fully-supervised methods on the Industrial3D test set. OA = overall accuracy (%), mIoU = mean IoU (%), per-class IoU (%). CB indicates class-balanced weighting loss. Best results per column are in bold.

Method	mIoU	Duct	Elbow	Flange	Ibeam	Pipe	Pump	RectBeam	Reducer	Strainer	Tank	Tee	Valve
ResPointNet++	53.31	91.36	46.05	50.92	**99.41**	76.23	35.95	95.19	0.00	0.00	97.06	12.03	35.56
KPConv	53.65	91.35	**45.44**	50.95	98.35	75.96	42.60	95.04	0.00	0.00	97.72	7.72	**38.69**
PosPool	53.18	90.25	32.93	42.88	98.93	**76.35**	31.27	94.18	**34.48**	0.00	95.59	9.17	32.14
ResPointNet++ w. CB	**54.05**	**91.30**	41.88	**50.71**	99.08	75.69	**43.81**	**95.07**	14.92	0.00	**99.04**	4.68	32.76

Table 5. Benchmark results of the unsupervised method GrowSP on the Industrial3D test set. OA = overall accuracy (%), mIoU = mean IoU (%), per-class IoU (%). The three rows represent different stages of the GrowSP pipeline.

Stage	OA	mIoU	Duct	Elbow	Flange	Ibeam	Pipe	Pump	RectBeam	Reducer	Strainer	Tank	Tee	Valve
Superpoints	84.57	42.06	68.07	4.19	18.64	66.15	73.46	59.14	86.23	0.00	0.00	90.61	12.98	25.18
Primitives	84.00	40.29	67.58	4.19	15.35	66.15	71.70	49.20	85.99	0.00	0.00	90.61	12.98	19.75
Merged Primitives	41.81	11.73	16.37	6.66	6.18	17.56	55.02	0.00	39.00	0.00	0.00	0.00	0.00	0.00

References

[1] Pingbo Tang, Daniel Huber, Burcu Akinci, Robert Lipman, and Alan Lytle. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction*, 19(7):829–843, 2010.

[2] Frédéric Bosché, Mahmoud Ahmed, Yelda Turkan, Carl T Haas, and Ralph Haas. The value of integrating scan-to-bim and scan-vs-bim techniques for construction monitoring using laser scanning and bim: The case of cylindrical mep components. *Automation in Construction*, 49:201–213, 2015.

[3] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1534–1543, 2016.

[4] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5828–5839, 2017.

[5] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019.

[6] Chao Yin, Boyu Wang, Vincent JL Gan, Mingzhu Wang, and Jack CP Cheng. Automated semantic segmentation of industrial point clouds using respointnet++. *Automation in Construction*, 130:103874, 2021.

[7] Ze Liu, Han Hu, Yue Cao, Zheng Zhang, and Xin Tong. A closer look at local aggregation operators in point cloud analysis. In *Computer Vision–ECCV 2020*, pages 326–342. Springer, 2020.

[8] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[9] Zihui Zhang, Bo Yang, Bing Wang, and Bo Li. Growsp: Unsupervised semantic segmentation of 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17619–17629, 2023.

[10] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan D Wegner, Konrad Schindler, and Marc Pollefeys. Semantic3d.net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*, 2017.

[11] Eva Agapaki and Ioannis Brilakis. Cloi-net: Class segmentation of industrial facilities' point cloud datasets. *Advanced Engineering Informatics*, 45: 101121, 2020.

[12] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12): 4338–4364, 2020.

[13] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017.

[14] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on

point sets in a metric space. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[15] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11108–11117, 2020.

[16] Chao Yin, Jack CP Cheng, Boyu Wang, and Vincent JL Gan. Automated classification of piping components from 3d lidar point clouds using se-pseudogrid. *Automation in Construction*, 139: 104300, 2022.

[17] Leica Geosystems. Leica blk360 imaging laser scanner. On-line: `https://leica-geosystems.com/products/laser-scanners/scanners/blk360`, 2018.

[18] Daniel Girardeau-Montaut. Cloudcompare: 3d point cloud and mesh processing software. On-line: `https://www.cloudcompare.org/`, 2023.

[19] Chao Yin, Qing Han, Zhiwei Hou, Yue Liu, Anjin Dai, Hongda Hu, Ji Yang, and Wei Yao. Resolving primitive-sharing ambiguity in long-tailed industrial point cloud segmentation via spatial context constraints. *arXiv preprint arXiv:2601.19128*, 2026.

[20] Qingyong Hu, Bo Yang, Guangchi Fang, Yulan Guo, Ales Leonardis, Niki Trigoni, and Andrew Markham. Sqn: Weakly-supervised semantic segmentation of large-scale 3d point clouds. In *Computer Vision–ECCV 2022*, pages 600–619. Springer, 2022.

[21] Lizhao Liu, Zhuangwei Zhuang, Shangxin Huang, Xunlong Xiao, Tianhang Xiang, Cen Chen, Jingdong Wang, and Mingkui Tan. Cpcm: Contextual point cloud modeling for weakly-supervised point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[22] Chao Yin, Bo Yang, Jack C.P. Cheng, Vincent J.L. Gan, Boyu Wang, and Ji Yang. Label-efficient semantic segmentation of large-scale industrial point clouds using weakly supervised learning. *Automation in Construction*, 148:104757, 2023. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.104757. URL `https://www.sciencedirect.com/science/article/pii/S0926580523000171`.

[23] Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas J Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *Computer Vision–ECCV 2020*, pages 574–591. Springer, 2020.

[24] Hanchen Wang, Qi Liu, Xiangyu Yue, Joan Lasenby, and Walterio W Mayol-Cuevas. Unsupervised point cloud pre-training via occlusion completion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9782–9792, 2021.

[25] Yuchen Zhou, Jiayuan Gu, Tung Yen Chiang, Fanbo Xiang, and Hao Su. Point-SAM: Promptable 3d segmentation model for point clouds. In *The Thirteenth International Conference on Learning Representations*, 2025.

[26] Kuan-Chih Huang, Xiangtai Li, Lu Qi, Shuicheng Yan, and Ming-Hsuan Yang. Reason3d: Searching and reasoning 3d segmentation via large language model. In *International Conference on 3D Vision (3DV)*, 2025.

Attention-Based Deep Reinforcement Learning for Task Sequence Optimization of Robotic Connection Operations

Yizhe Wang[1], Cong Zhang[2], Yihai Fang[1]

[1]Department of Civil and Environmental Engineering, Monash University, Australia
[2]College of Computing and Data Science, Nanyang Technological University, Singapore
yizhe.wang@monash.edu, cong030@e.ntu.edu.sg, yihai.fang@monash.edu

Abstract

Construction robotics has significant potential to improve the efficiency, quality, and safety of structural assembly, yet its practical deployment is constrained by the limited availability of online task sequence optimization methods. This study focuses on gantry robots performing connection tasks in floor frame assembly under limited nail magazine capacity. To improve operational efficiency, the sequencing problem is formulated as a split-delivery vehicle routing problem (SDVRP). An attention-based deep reinforcement learning (DRL) model is employed to learn routing policies, and its performance is compared against an exact solver (Gurobi), a genetic algorithm (GA), and a nearest-neighbor (NN) heuristic. Experiment results show that the DRL policy achieves solutions within 2.1% of the optimal benchmark while requiring only 0.1s per instance. Additional tests across different nail capacities demonstrate robustness to distribution shifts. Overall, the results highlight the potential of attention-based DRL for fast and high-quality task sequence optimization, representing the first application of this approach to robotic structural assembly connection operations.

Keywords –

Deep Reinforcement Learning; Optimization; Construction Robot; Task Sequence; Structural Assembly Connection

1 Introduction

The construction sector continues to face sustained challenges from skilled labour shortages, worker health and safety concerns, and modest growth in productivity [1]. Within this context, structural assembly constitutes a fundamental construction activity in which discrete components are connected to form larger systems such as walls, floors, and roofs [2,3]. These operations are particularly well suited to partial or full automation using robotic systems owing to their repetitive and physically demanding nature [4].

With recent advances in robotics, gantry robots have been adopted across diverse applications such as timber prefabrication [5], rebar tying [6], furniture manufacturing [7], and shipbuilding [8], owing to their ability to transport materials efficiently across large workspaces. These robotic systems are increasingly regarded as promising platforms for automating structural assembly tasks. An example of an overhead gantry robot performing floor-frame connection operations is illustrated in Figure 1.

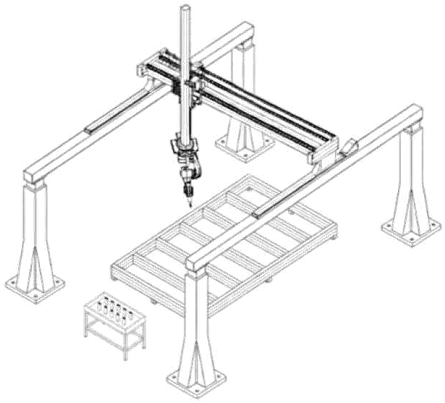


Figure 1. Example of a gantry robot performing connection tasks in timber floor-frame assembly

In practical construction settings, human workers naturally plan and adapt their own task sequences, whereas robots rely on explicitly specified task sequences to operate efficiently [9]. When gantry robots are used for connection tasks in structural assembly, their productivity depends critically on the quality of task sequence planning.

This study addresses this challenge by formulating task sequence optimization for robotic connection operations as a split-delivery vehicle routing problem (SDVRP) and solving it using an attention-based deep reinforcement learning (DRL) model. The performance of the trained policy is compared against heuristic methods and an exact solver, and its generalization capability is evaluated across varying problem settings. This proposed method achieves solutions within 2.1% of

the optimal benchmark while requiring only 0.1s per instance, and maintains competitive performance when nail capacities remain close to the training dataset range.

2 Literature Review

2.1 Gantry Robot Applications in Structural Assembly

Gantry systems have long been used in the construction industry, with early applications appearing in the 1980s. During this period, gantry platforms were installed in temporary site factories on top of buildings under construction and were progressively raised as each floor was completed [10]. With continued advancements, gantry robots have been increasingly adopted for assembly operations. In industry practice, IronBot (for rebar lifting and placement) and TyBot (for rebar tying) developed by Advanced Construction Robotics are notable examples. They exhibit a high level of autonomy, including automated rebar grasping, placement, intersection detection, and tying execution [5].

In the research domain, Chai et al. presented an integrated timber construction platform incorporating a gantry-based multi-robot system, sensing, control, and a timber machining center to enable multifunctional prefabrication of timber components [6]. Wenna et al. developed a digital-twin framework for 3D path planning of a large-span curved-arm gantry robot for automated wooden-furniture production, enabling the generation of collision-free and efficient trajectories [7]. Other studies have focused on gantry robots for connection operations. For example, Saleem et al. designed a six-degree-of-freedom gantry welding robot through mathematical modelling and mechanical design [11]. Similarly, an improved RRT* algorithm was proposed for gantry-mounted welding robots to enhance planning efficiency and reliability in complex environments [8].

The majority of prior research focuses on path planning and control methods to ensure safe and efficient operation. However, few studies address task sequencing optimization, despite its importance for assembly efficiency. This gap motivates the present study on gantry robotic connection operations in structural assembly.

2.2 Task Sequencing in Robotic Structural Assembly

Although task sequencing has been studied only rarely for gantry robotic structural assembly, research on a broader range of robotic systems, including aerial robots and robotic arms, provides useful insights.

In multi-robot structural assembly, Muñoz et al. integrate a constraint-based vehicle routing problem (VRP) solver with a hierarchical task network (HTN) planner [12]. Assembly tasks are abstracted into a VRP formulation, where the VRP solver provides an initial task allocation and coarse sequencing. The HTN planner subsequently expands these assignments into executable schedules. The two planners operate in a closed feedback loop, and the method is validated through simulations of aerial robots assembling a truss-like structure. The VRP solver part relies on heuristic search strategies rather than exact optimization.

Yang et al. address assembly sequence planning for timber wall framing using a robotic arm [13]. Their method jointly considers stud placement and fastening operations while incorporating task precedence constraints. A modified A* algorithm is used to identify an executable assembly sequence, where the accumulated end-effector travel distance forms the actual cost, and the remaining cost is estimated using a minimum spanning tree (MST)-based heuristic. This approach highlights the role of domain-specific heuristics in navigating the assembly sequence planning.

You et al. explored the use of large language models (LLMs) for structural assembly task sequencing [14]. Their system, RoboGPT, leverages ChatGPT's reasoning capabilities to autonomously generate and adapt task sequences. The approach is validated using a single robotic arm executing assembly tasks, demonstrating the feasibility of LLM-driven task planning through direct application programming interface (API) integration.

To the best of the authors' knowledge, no existing studies on robotic structural assembly have employed DRL for task sequence planning. Among the various optimization methods, attention-based DRL has shown strong potential for tackling combinatorial optimization problems, in some cases outperforming traditional metaheuristic algorithms [15], and has already been explored in related manufacturing applications [16]. A further advantage of DRL is its fast inference: once trained, the policy produces near-instantaneous solutions without a repeated optimization process, enabling online adaptation to robotic structural assembly.

In summary, task sequencing research for gantry robotic structural assembly remains limited, and DRL-based optimization in structural assembly is still lacking. This study addresses these gaps by applying an attention-based DRL model [17] to an SDVRP formulation of gantry robot connection tasks in structural assembly.

3 Methodology

3.1 Problem Formulation and Abstraction

This study investigates a common application scenario of gantry robots in structural assembly, in which a robot equipped with a nailing end-effector executes

floor frame connections. The robot must periodically return to a storage location for refilling after a certain number of nailing operations because the nailing end-effector has a limited magazine capacity. Consequently, the execution sequence of the nailing tasks directly determines the robot's total travel distance and, in turn, the assembly efficiency.

During assembly, the gantry robot operates within a predefined workspace free of internal obstacles, enabling collision-free motion of the end-effector. In this study, the robot is assumed to first travel in the horizontal plane from the storage location to a point vertically above the target nailing position, and then perform a vertical approach to the connection plane to execute the nailing operation. Since this vertical motion contributes a constant and negligible portion of the overall travel effort, the travel cost between the storage and each nailing task is approximated by the two-dimensional Euclidean distance in the horizontal plane.

Based on the abstraction, the robotic nailing process is modeled as an SDVRP. The SDVRP is a variant of the classical VRP defined on a set of customer nodes and a depot, in which each customer has a demand, vehicles have limited capacity, and demands may be fulfilled through multiple vehicle visits. The goal is to determine a set of routes originating from and returning to the depot that minimizes the total travel cost while satisfying vehicle capacity constraints.

In the robotic structural assembly nailing setting, the storage location is treated as the depot, the capacity of the nailing end-effector as the vehicle capacity, and each connection task as a customer node with a corresponding demand. The objective is to determine a sequence of nailing operations that minimizes the total travel distance while satisfying the capacity constraint.

Among the available methods for solving the SDVRP, attention-based DRL offers fast inference, strong combinatorial optimization capability, and promising potential for complex construction scenarios [15,17–19]. Accordingly, this study adopts an attention-based DRL framework for the formulated SDVRP.

3.2 Optimization Framework

The Transformer is an encoder–decoder architecture built around the attention mechanism, enabling highly parallelizable sequence modeling through self-attention blocks [20]. Building on the Transformer architecture, attention-based models can effectively capture global dependencies and combinatorial structure, making them well-suited to sequential decision-making under constraints. Accordingly, routing problems such as the SDVRP can be formulated as sequence generation: instance features (e.g., locations, demands, and capacity) are embedded and encoded by an attention-based encoder [21], and a decoder then sequentially selects the next node to construct one or more feasible tours [17].

While Transformers in natural language processing are typically trained in a supervised, next-token-prediction manner using large text corpora, deep models for routing problems are often trained without access to optimal solutions. Instead, these models are commonly optimized using reinforcement learning, where the policy interacts with an environment, receives a scalar reward reflecting solution quality (e.g., total route length or cost), and is updated via policy-gradient methods [15,17,22].

This study employs the attention-based transformer model proposed by Kool et al. [17] to solve the SDVRP. Since its introduction in 2019, this model has become a milestone in attention-based DRL for combinatorial optimization and has been widely applied to routing problems, such as the traveling salesman problem (TSP), the orienteering problem (OP), the capacitated vehicle routing problem (CVRP), SDVRP and the prize-collecting TSP (PCTSP). In this framework, the routing problem is formulated as a sequential decision-making process, in which a parameterized stochastic policy is learned to construct feasible solutions. The policy is implemented using an attention-based encoder–decoder neural network and optimized with the REINFORCE algorithm [23] with a deterministic greedy rollout baseline.

The following subsections describe the proposed optimization framework in detail, including the Markov decision process (MDP) formulation, policy definition, neural network architecture, and DRL training procedure in the context of the SDVRP.

3.2.1 MDP Formulation and Policy Definition

The SDVRP is formulated as a finite-horizon Markov decision process (MDP), where the agent sequentially constructs a feasible routing solution. At each decision step t, the state s_t captures the problem instance and the current routing status, including the current location, remaining vehicle capacity, and the remaining demand of each task. An action a_t corresponds to selecting the next node to visit, subject to feasibility constraints enforced by action masking. An episode terminates when all tasks are fully served, and the solution quality is evaluated by a terminal cost corresponding to the total travel distance.

Based on this MDP formulation, a stochastic policy $p_\theta(\pi|x)$ is learned to construct a complete routing solution for a given problem instance x. A feasible solution is represented by an ordered sequence of node indices $\pi = (\pi_1, \dots, \pi_n)$, where π_t denotes the node selected at decision step t. The objective of the policy is to generate a complete sequence that satisfies all problem constraints while minimizing the overall travel cost.

The attention-based model defines this policy in an autoregressive manner. Specifically, the conditional

probability of generating a complete solution sequence π is factorized as (1).

$$p_\theta(\pi|x) = \prod_{t=1}^{n} p_\theta(\pi_t \mid x, \pi_{1:t-1}) \quad (1)$$

Here, $\pi_{1:t-1}$ denotes the partial sequence constructed up to step $t-1$, and θ represents the parameters of the neural network.

3.2.2 Neural Network Architecture

The neural network adopts the standard attention-based encoder-decoder architecture proposed by [17]. The encoder consists of three stacked attention layers, each comprising a multi-head self-attention sublayer followed by a node-wise feed-forward sublayer. Skip-connections and batch normalization are applied in each sublayer to stabilize training.

Each input node is represented by its spatial coordinates and associated attributes (e.g., demand), which are first projected into a shared embedding space. Through self-attention, the encoder aggregates information across all nodes, producing context-aware embeddings that capture the global structure of the routing instance.

The decoder constructs a solution in an autoregressive manner. At each decision step, it attends to the encoder embeddings and incorporates information from the partial solution constructed so far. A masking mechanism is applied to exclude infeasible actions, such as selecting already visited nodes or actions that would violate capacity feasibility. The decoder then outputs a probability distribution over all admissible nodes, from which the next node is selected.

3.2.3 DRL Training with REINFORCE

With the policy parameterized by the neural network architecture described above, the policy parameters θ are trained using the REINFORCE algorithm [23] by minimizing the expected cost of solutions sampled from the stochastic policy. For a given problem instance x, the training objective $\mathcal{L}(\theta|x)$ is defined as (2).

$$\mathcal{L}(\theta|x) = E_{p_\theta(\pi|x)}[L(\pi)] \quad (2)$$

$L(\pi)$ is a scalar cost function that evaluates the quality of a solution π. The training objective is optimized using gradient descent, which is calculated with the REINFORCE gradient estimator with a baseline in (3).

$$\nabla\mathcal{L}(\theta|x) = E_{p_\theta(\pi|x)}\big[\big(L(\pi) - b(x)\big)\nabla \log p_\theta(\pi|x)\big] \quad (3)$$

$b(x)$ is a baseline that depends only on the instance x and is used to reduce the variance of the gradient estimate. In this study, $b(x)$ is given by the cost of a deterministic greedy rollout of the current policy.

When the sampled solution π achieves a lower cost than the baseline, the term $L(\pi) - b(x)$ becomes negative, and the gradient update increases the log-probability $\log p_\theta(\pi|x)$ of the selected actions. Conversely, if the sampled solution is worse than the baseline, the update decreases this probability. This formulation supports end-to-end learning without optimal solutions or stepwise supervision.

All experiments were run on a workstation with an Intel Core Ultra 9 185H CPU and an NVIDIA GeForce RTX 4070 Laptop GPU. The DRL model was trained using a fixed set of hyperparameters throughout the study, and the main training settings are summarized in Table 1.

Table 1. Training hyperparameters of the attention-based DRL model

Hyperparameter	Value
Dataset size	1 million instances per epoch
Number of epochs	100
Learning rate	$1e^{-5}$
Batch size	512
Optimizer	Adam
Training duration	Approximately 100 minutes

4 Case Study

4.1 Application Scenario

This study applies an attention-based DRL approach to a representative floor-frame nailing task in structural assembly, with the aim of demonstrating both its effectiveness and practical advantages in structural assembly settings.

Figure 1 in Section 1 illustrates the overall gantry robot system considered in this study. The gantry structure provides overhead motion and defines the robot's operational workspace. A nailing end-effector, integrating both a nail gun and a nail magazine, is mounted on the moving carriage to enable autonomous connection operations. The target floor frame structure to be assembled is on the floor, while the nail storage area, where nail magazines are refilled, is near the boundary of the gantry system. During operation, the robot travels within the gantry workspace to access each connection location and periodically returns to the storage area when the magazine capacity is exhausted.

Figure 2 presents a plan view of the gantry robot workspace and the corresponding floor frame nailing tasks. The dashed boundary indicates the reachable workspace of the gantry system. The floor frame consists of two pairs of horizontal header joists at the top and bottom, connected by 9 vertical cross joists. Nailing demand is intentionally defined as non-uniform across

the joints to reflect realistic fastening requirements. Specifically, the four corner joints between the header joists and outermost cross joists, along with the two midspan joints on the top and bottom headers, are designated as strengthened connections requiring twice as many nails as standard joints.

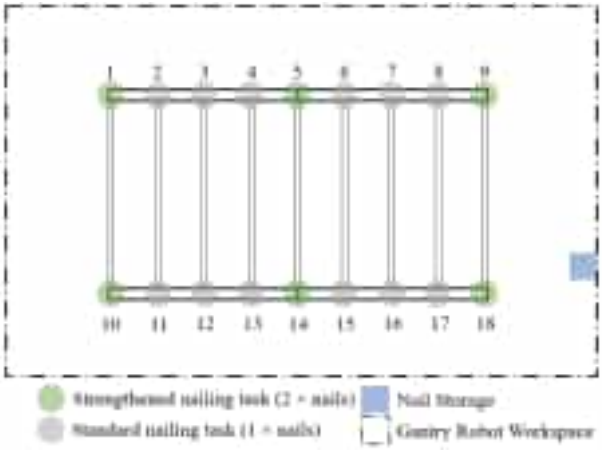


Figure 2. Plan view of the gantry robot workspace and floor-frame assembly

The training instances in this study are generated using a set of predefined rules with a fixed seed. The floor-frame to be assembled is first placed within a 4 m × 2 m rectangular region, with the length of the vertical cross joists sampled uniformly between 1 m and 2 m. The 9 vertical joists are then positioned with center-to-center spacings drawn from a uniform range of 0.3-0.5 m, yielding overall frame dimensions between 2.4 m × 1 m and 4 m × 2 m. The reachable workspace of the gantry system is defined by sampling its plan dimensions between 4 m × 2 m and 8 m × 6 m. The nail storage location is selected along the boundary of this workspace such that it always lies outside the range of the assembled floor frame. For each instance, the nail demand at a standard nailing task is drawn uniformly from 5-11 nails, while the demand at a strengthened nailing task is set to twice this value. The magazine capacity of the nailing end-effector is fixed at 100 nails for all experiments.

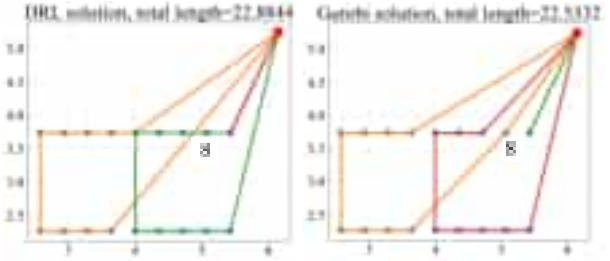
4.2 Results

4.2.1 Comparison of DRL with Exact and Heuristic Solvers

With all instances generated using a fixed random seed (seed = 123), 100 problem instances were created under this setting, and the solutions produced by the trained DRL model were compared against those obtained by Gurobi [24], Genetic Algorithm (GA) [25], and Nearest Neighbour (NN) baselines.

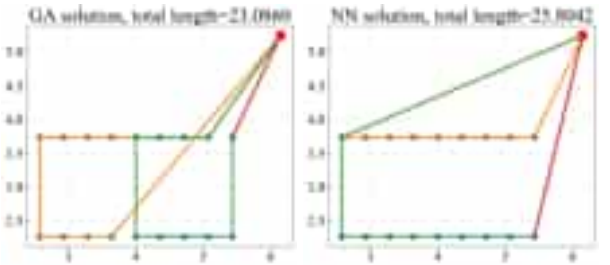
Gurobi was used as the optimal benchmark solver. It can obtain globally optimal solutions for small- to medium-scale routing problems within reasonable time, but its computational cost grows rapidly with problem size due to NP-hardness. For large-scale instances (e.g., over 100 connection tasks), solving can take hours or even days, making exact optimization impractical under limited time constraints [24,26].

Figure 3 illustrates the routing solutions obtained by different optimization methods for the first instance generated under seed=123. In this instance, each standard connection task requires 9 nails, while each strengthened connection task requires 18 nails. Overall, both the DRL and GA approaches produce solutions that are close to the optimal routes generated by Gurobi, whereas NN exhibits substantially greater deviation from optimality. Specifically, the primary discrepancy between the DRL result and the Gurobi solution occurs at Connection Task 8: in the Gurobi solution, this task is completed during the return segment of Route 1, thereby reducing the total travel distance.



DRL solution:
Route 1: 13,12,11,10,1,2,3,4,5
Route 2: 18,17,16,15,14,5,6,7,8,9
Route 3: 9

Gurobi solution:
Route 1: 4,3,2,1,10,11,12,13,8
Route 2: 9
Route 3: 7, 6, 5, 14, 15, 16, 17, 18



GA solution:
Route 1: 13,12,11,10,1,2,3,4,6,8
Route 2: 8,7,5,14,15,16,17,18,9
Route 3: 9

NN solution:
Route 1: 9,8,7,6,5,4,3,2,1
Route 2: 1,10,11,12,13,14,15,16,17,18
Route3: 18



Figure 3. Comparison of routing solutions obtained by DRL, Gurobi (optimal), GA and NN for a representative problem instance (1st instance with seed=123)

To mitigate the effects of randomness and robustly evaluate the effectiveness of the DRL approach, all methods were evaluated across the full set of 100 generated instances. The average total travel distance,

percentage gap to the optimal solution, and average computation time are summarized in Table 2. In addition to a deterministic greedy rollout, a stochastic DRL variant based on top-p (nucleus) Monte Carlo sampling was also evaluated, generating 128 candidate routes per instance and selecting the best-performing solution [27].

The results show that both DRL methods (greedy and sampling) achieve solutions within approximately 2.1% of the optimal benchmark, similar to the performance of GA. In contrast, NN exhibits the largest degradation from optimality, with an average gap of 9.6%. In terms of computational efficiency, the DRL inference time for both greedy and sampling remains low (around 0.1s), making it suitable for near-real-time applications. While NN provides the fastest solutions (approximately 0.001s), this advantage is negligible in practical construction planning contexts. GA requires significantly more computation time (over 9s), whereas Gurobi solves each instance in over 1s.

Table 2. Average solution and computation time over 100 generated instances (seed=123)

Solver	Distance (m)	Gap to Best	Time (s)
Gurobi	20.0429	0.00%	1.374
DRL (Greedy)	20.4682	2.12%	0.078
DRL (Sampling)	20.4518	2.04%	0.102
GA	20.4631	2.10%	9.100
NN	21.9632	9.58%	0.001

4.2.2 Generalization Performance under Varying Nail Capacities

The proposed DRL policy is trained on problem instances with a fixed nail capacity of 100. Ideally, such a policy should not only perform well on the training distribution, but also generalize to related problems with different nail capacities. To assess this generalization capability, an out-of-distribution evaluation is conducted on 100 instances generated with seed 123 by varying the nail capacity from 50 to 150 at intervals of 10.

Table 3 reports the average total travel distance obtained by Gurobi and DRL for each capacity level, along with the percentage gap between the DRL solution and the optimal Gurobi solution. When the nail capacity lies between 80 and 150, the DRL policy maintains a reasonably stable performance, with the gap to the optimum remaining below 7%. This indicates that the learned policy can adapt to moderate changes in capacity without substantial degradation in solution quality. However, as the capacity is reduced to 70 and below, the gap increases sharply, reaching 22.2% at capacity 70 and exceeding 40% at capacity 60, with the poorest performance observed at capacity 50 (66.9% gap).

Overall, these results suggest that the DRL policy exhibits satisfactory generalization for capacities within the vicinity of the training setting but deteriorates when applied to substantially different settings characterized by lower capacities. The underlying reasons are further discussed in Section 5.1.

Table 3. Generalization performance of the trained DRL policy under varying nail capacities, compared with Gurobi over 100 generated instances (seed = 123)

Nail Capacity	Gurobi Results (m)	DRL (Greedy) Results (m)	Gap to Best
150	15.6065	16.5800	6.24%
140	16.1545	16.9771	5.09%
130	16.9889	17.9866	5.87%
120	17.2348	18.4647	7.14%
110	18.9855	19.5970	3.22%
100(Trained)	20.0429	20.4682	2.12%
90	21.0676	21.8989	3.95%
80	22.8803	24.2583	6.02%
70	25.1728	30.7637	22.21%
60	27.2099	38.5350	41.62%
50	31.5817	52.7125	66.91%

5 Discussion

5.1 Interpretation of Results

Section 4.2.1 evaluates the trained attention-based DRL policy on randomly generated instances drawn from the same distribution as the training data and compares it against Gurobi, GA, and NN. The DRL policy attains solutions with an average gap of about 2.1% to the optimal benchmark, while requiring around 0.1s per instance. This runtime is compatible with online task sequence planning in structural assembly connection operations, yet still delivers high-quality routes. As problem complexity increases (e.g., to instances with 100 tasks), the constant-time inference property of the DRL policy suggests that sub-second runtimes can still be maintained, while the computational burden of Gurobi and GA is expected to grow substantially. For larger-scale or more constrained instances, Gurobi will experience a sharp increase in computational effort, with solution time extending to hours or longer. Taken together, these comparisons illustrate that the DRL approach occupies a favorable position in the speed-quality trade-off spectrum: sufficiently accurate for near-optimal planning and sufficiently fast for on-site deployment.

Section 4.2.2 examines the generalization of the trained policy to datasets with modified nail capacities. In the application scenario considered, the total nail demand ranges from 120 to 252 nails. When the capacity

is between 90 and 120, the gantry robot typically needs two to three tours to complete all connection tasks, which is consistent with the training instances. In this setting, the policy produces routes that remain close to optimal. For capacities between 130 and 150, some instances can be completed in a single tour, but most still require two tours; the policy continues to perform reasonably well when chaining more connection tasks into each tour. However, when the capacity drops to 80 or below, the robot must undertake many more tours (up to five tours at capacity 60 and six at capacity 50), a condition that never occurred during training. As a result, performance degrades noticeably.

To enhance robustness in such low-capacity settings, future training could expand the training distribution so that the policy encounters a wider range of demand-capacity ratios and more multi-tour scenarios. In the present study, the training instances were designed to reflect the practical application setting, where the nail magazine capacity is 100, and the nail demand per connection task ranges from 5 to 11. Although this setting is realistic for the target scenario, it limits the diversity of tour structures seen during training. Future work could therefore broaden the nail demand range during training, for example, from 2 to 20 nails per connection task, so that the model is exposed to more varied routing patterns and may generalize better to low-capacity cases.

5.2 Limitations and Future Research

Despite the encouraging results, this study has several limitations that point to important directions for future research.

First, the experimental scenarios considered in this study involve a relatively small number of connection tasks (18 tasks). Although the current results demonstrate the feasibility and promise of the proposed method, its computational advantage is expected to become more evident in larger and more complex problem instances. Future studies should therefore consider larger-scale cases and compare the proposed method with exact solvers such as Gurobi and heuristic methods in terms of inference time and solution quality under practical time constraints.

Second, the current formulation assumes a two-dimensional workspace without obstacles, and travel cost is approximated using Euclidean distance. The proposed approach therefore focuses on high-level task sequencing and routing, rather than low-level motion planning. In more realistic construction settings, the resulting task sequence could be combined with lower-level path planning methods, such as RRT* or Dijkstra-based algorithms, so that travel cost is estimated from feasible collision-free trajectories. Extending the framework in this direction remains an important future task. Third, the current framework assumes static task settings and does not account for replanning. In practical construction environments, robots may need to adapt their plans in response to unexpected events, such as delays, partial task completion, or human-robot interactions. Incorporating online replanning mechanisms and dynamic task updates constitutes an important extension of this work.

Finally, the present study evaluates the proposed approach at the algorithmic level only. Future work will focus on validating the method in laboratory-based robotic experiments, which is a necessary step towards real-world deployment in construction settings.

6 Conclusion

This study presented an attention-based DRL framework for task sequence optimization in gantry robotic connection tasks. By formulating the nailing operation as an SDVRP, the proposed method enables systematic optimization of both routing and refilling behavior under nail magazine capacity constraints. Benchmark comparisons with Gurobi, GA, and NN show that the trained DRL policy achieves a favorable balance between solution quality and computational efficiency: solutions remain within 2.1% of the optimal benchmark while requiring only 0.1s of inference time per instance. This constant-time inference property underscores the suitability of the DRL approach for online planning in construction robotics scenarios.

Generalization experiments show that the policy performs reliably for nail capacities close to the training setting but deteriorates when applied to substantially different conditions that require many additional tours. These findings suggest that robustness could be enhanced by training on a broader range of nail magazine capacities and by incorporating greater variability in nail demand across tasks.

Several limitations point to directions for future research. The experiments were conducted on relatively small problem instances in obstacle-free environments, so extending the method to larger assemblies is an important next step. The current framework also assumes static task settings and does not support replanning, which is crucial for handling disruptions in real construction operations. In addition, the approach has been evaluated only at the algorithmic level; laboratory-based robotic experiments are needed to assess practical feasibility. Overall, the results highlight the potential of attention-based DRL task sequence optimization to enable more efficient robotic connection operations.

References

[1] T. Bock, The future of construction automation: Technological disruption and the upcoming ubiquity of robotics, Autom Constr 59 (2015) 113–121

[2] S. Gao, R. Jin, W. Lu, Design for manufacture and assembly in construction: a review, Building Research and Information 48 (2020) 538–550.

[3] Y. Bai, Composites for Building Assembly, Springer Tracts in Civil Engineering, 2023.

[4] Y. Wang, Y. Fang, Y. Bai, Task planning for human-robot collaboration in structure assembly, Autom Constr 179 (2025) 106464.

[5] Advanced Construction Robotics Inc., Robotic Rebar Installation Solutions, (2025). https://www.constructionrobots.com/.

[6] H. Chai, L. Zhang, P.F. Yuan, Advanced Timber Construction Platform Multi-Robot System for Timber Structure Design and Prefabrication, in: Architectural Intelligence, Springer Nature Singapore, 2020: pp. 129–144.

[7] W. Wenna, D. Weili, H. Changchun, Z. Heng, H. Feng, Y. Yao, A digital twin for 3D path planning of large-span curved-arm gantry robot, Robot Comput Integr Manuf 76 (2022).

[8] X. Wang, J. Gao, X. Zhou, X. Gu, Path Planning for the Gantry Welding Robot System Based on Improved RRT*, Robot Comput Integr Manuf 85 (2024).

[9] J.G. Everett, A.H. Slocum, Automation and robotics opportunities: construction versus manufacturing, J Constr Eng Manag 120 (1994) 443–452.

[10] Y. Ikeda, T. Harada, Application of the Automated Building Construction System Using the Conventional Construction Method Together, in: Proceedings of the 23rd ISARC, Tokyo, Japan, 2006: pp. 722–727.

[11] M. Saleem, S. Jameel, A. Hassan, Z.H. Khan, Design and Fabrication of a Six DOF Gantry Robot for Welding Applications, in: ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, 2024: pp. 1940–1946.

[12] J. Muñoz-Morera, F. Alarcon, I. Maza, A. Ollero, Combining a hierarchical task network planner with a constraint satisfaction solver for assembly operations involving routing problems in a multi-robot context, Int J Adv Robot Syst 15 (2018). 10.1177/1729881418782088.

[13] C.-H. Yang, L.-T. Tsai, Y. Chen, S.-C. Kang, Assembly sequence planning method for robotic timber wall prefabrication, Autom Constr 181 (2026) 106669.

[14] H. You, Y. Ye, T. Zhou, Q. Zhu, J. Du, Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT, Buildings 13 (2023). 10.3390/buildings13071772.

[15] Y.-D. Kwon, J. Choo, B. Kim, I. Yoon, Y. Gwon, S. Min, POMO: Policy Optimization with Multiple Optima for Reinforcement Learning, (2020).

[16] Y. Chen, M. Chen, F. Yu, H. Lin, W. Yi, An Improved Ant Colony Algorithm with Deep Reinforcement Learning for the Robust Multiobjective AGV Routing Problem in Assembly Workshops, Applied Sciences (Switzerland) 14 (2024).

[17] W. Kool, H. Van Hoof, M. Welling, Attention, learn to solve routing problems!, 7th International Conference on Learning Representations, ICLR 2019 (2019) 1–25.

[18] R. Zhang, A. Prokhorchuk, J. Dauwels, Deep Reinforcement Learning for Traveling Salesman Problem with Time Windows and Rejections, Proceedings of the International Joint Conference on Neural Networks (2020).

[19] F. Berto, et al, RL4CO: An Extensive Reinforcement Learning for Combinatorial Optimization Benchmark, in: Association for Computing Machinery (ACM), 2025: pp. 5278–5289.

[20] A. Vaswani, et al, Attention Is All You Need, in: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, US, 2017: pp. 6000–6010.

[21] C. Zhang, Y. Wu, Y. Ma, W. Song, Z. Le, Z. Cao, J. Zhang, A review on learning to solve combinatorial optimisation problems in manufacturing, IET Collaborative Intelligent Manufacturing 5 (2022).

[22] I. Bello, H. Pham, Q. V. Le, M. Norouzi, S. Bengio, Neural combinatorial optimization with reinforcement learning, 5th International Conference on Learning Representations, ICLR 2017 - Workshop Track Proceedings (2017) 1–15.

[23] R.J. Williams, Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning, Mach Learn 8 (1992) 229–256.

[24] Gurobi Optimization LLC, Gurobi Optimizer Reference Manual, (2025). https://www.gurobi.com.

[25] J.H. Holland, Genetic Algorithms, Sci Am 267 (1992) 66–73.

[26] P. Toth, D. Vigo, The Vehicle Routing Problem, Society for Industrial and Applied Mathematics, 2002.

[27] A. Holtzman, J. Buys, L. Du, M. Forbes, Y. Choi, The Curious Case of Neural Text Degeneration, (2020).

XR-assisted semantic scene understanding of building interiors via AI segmentation and photogrammetric reconstruction

Bryan G. Pantoja-Rosero[1]

[1]Digital Intelligence for Structures and Construction Laboratory (DISC),
School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

bryan.pr@ntu.edu.sg

Abstract -

Semantic understanding of indoor building environments is essential for digitalization, facility management, robotics, and human–computer interaction. However, existing workflows typically decouple data acquisition, semantic interpretation, and 3D reconstruction into offline processes, limiting on-site verification and interaction. This paper presents an extended reality assisted framework for semantic understanding and digitalization of indoor scenes by integrating on-device artificial intelligence based semantic segmentation with photogrammetric reconstruction. Image frames captured using a mixed reality headset are processed in near real time using a lightweight YOLOv11-seg model to obtain object-level detections and segmentation masks. The same image data are subsequently used for Structure-from-Motion reconstruction, where 2D semantic information is spatially propagated into a 3D point cloud and mesh, yielding semantically enriched 3D scene models. These models are visualized and interactively validated by the user within the extended reality environment, enabling immediate assessment of digitalization completeness. The proposed framework demonstrates how extended reality can function as an active tool for near real-time semantic perception, 3D modeling, and user-guided scene digitalization.

Keywords -

Extended reality; Semantic segmentation; Structure-from-Motion; Building digitalization; 3D semantic modeling; Interior building modeling; Deep learning; Photogrammetry; Mixed Reality; Edge computing

1 Introduction

The semantic understanding of indoor building environments is a fundamental requirement for a wide range of applications, including building documentation, facility management, digital twins, robotics, and human–computer interaction [1, 2]. Accurately identifying and localizing building elements and objects within interior spaces enables more efficient digitalization workflows and supports downstream tasks such as asset inventory, spatial analysis, and interaction-aware planning [3]. Despite recent advances in computer vision, most existing semantic understanding approaches rely on offline image processing pipelines, where data are collected on site and analyzed afterward on external computing platforms [4]. This separation between data acquisition and semantic interpretation limits responsiveness, hinders on-site verification, and often requires repeated visits when additional information is needed or errors are discovered during post-processing.

Recent progress in artificial intelligence (AI), particularly deep learning–based semantic and instance segmentation, has significantly improved the automated recognition of objects and scene elements from images [5, 6]. These methods have demonstrated strong performance in both outdoor and indoor environments, enabling pixel-level delineation of semantic classes and supporting dense scene understanding [7]. However, most AI-driven semantic segmentation workflows remain inherently two-dimensional, operating on individual images or image sequences without explicit spatial grounding [8]. As a result, semantic information is often disconnected from the three-dimensional (3D) structure of the environment, limiting its interpretability and usefulness for spatial reasoning, digital modeling, and interactive applications [9].

In parallel, photogrammetric reconstruction techniques such as Structure-from-Motion (SfM) have become well established for generating accurate 3D representations of indoor scenes from image collections [10, 11]. While these methods provide detailed geometric models, they typically lack semantic richness, producing unstructured point clouds or meshes that require additional processing to identify meaningful objects or building elements [12, 13]. Bridging the gap between 2D semantic perception and 3D geometric reconstruction remains a key challenge, particularly in scenarios where rapid feedback, validation, and interaction with the reconstructed scene are required [14, 15].

Extended Reality (XR) technologies—including augmented, virtual, and mixed reality—offer a unique opportunity to address these challenges by tightly coupling perception, computation, and visualization within a single interactive platform [16, 17]. XR head-mounted devices

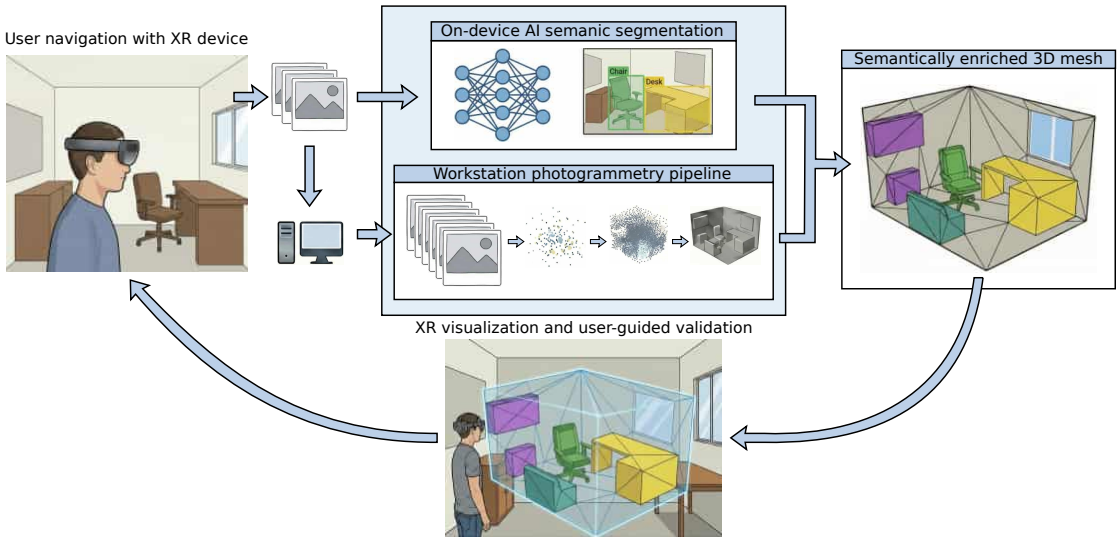


Figure 1. Closed-loop XR-assisted pipeline for semantic scene understanding. On-device semantic segmentation is performed on image frames acquired with a mixed reality headset, while photogrammetric reconstruction is executed on an external workstation. Semantic information is propagated into a 3D reconstruction and visualized in XR for interactive validation.

enable hands-free, spatially aware interaction with both physical and digital content, allowing users to observe, verify, and contextualize AI-generated information directly within the environment being analyzed [18]. Compared to conventional mobile devices or laptops, XR systems provide an immersive interface that supports near real-time understanding of spatial relationships, immediate validation of captured data, and intuitive exploration of semantic and geometric representations [19]. These capabilities are particularly beneficial for indoor scene digitalization, where occlusions, clutter, and complex layouts make purely image-based verification difficult.

Despite these advantages, current XR-based applications for semantic scene understanding remain limited. Many systems focus primarily on visualization, while AI-based semantic interpretation is performed offline or relies on cloud-based processing, introducing latency and restricting near real-time interaction [20, 21]. Furthermore, semantic outputs are often displayed in 2D overlays without enforcing spatial consistency across views or integrating semantic labels into a coherent 3D scene representation [22]. This limits the ability to verify semantic accuracy on site and reduces the usefulness of XR as an active tool for scene digitalization rather than a passive visualization interface.

To address these limitations, this paper proposes an XR-assisted framework for semantic understanding of indoor building scenes that integrates AI-based segmentation with photogrammetric reconstruction. In the proposed approach, images acquired using a mixed reality (MR) headset are processed using a deep learning–based segmentation model to identify semantic elements within the scene. The same image set is then used for SfM reconstruction to generate a 3D representation of the environment. By associating 2D semantic predictions with the reconstructed geometry, semantic information is spatially propagated into the 3D scene, enabling the creation of semantically enriched 3D models. These models are subsequently visualized and explored within the XR environment, allowing users to inspect, validate, and interact with semantic scene representations directly on site.

The contributions of this work are threefold:

- An XR-assisted pipeline for indoor semantic scene understanding, combining AI-based image segmentation and photogrammetric reconstruction within a unified workflow.
- A method for spatially consistent projection of 2D semantic information into 3D reconstructed scenes, enabling semantically enriched models of indoor environments.
- An interactive XR visualization framework that supports near real-time verification and exploration of semantic 3D scenes, facilitating immediate validation of digitalization results.

By shifting XR from a passive visualization tool to an active component of semantic scene understanding and digitalization, this work demonstrates the potential of XR-enabled AI systems to support more efficient, reliable, and interactive indoor environment modeling. Given the scope of this paper, the emphasis is placed on demonstrating the feasibility and end-to-end integration of the proposed workflow rather than on a comprehensive quantitative benchmarking of its components. The proposed frame-

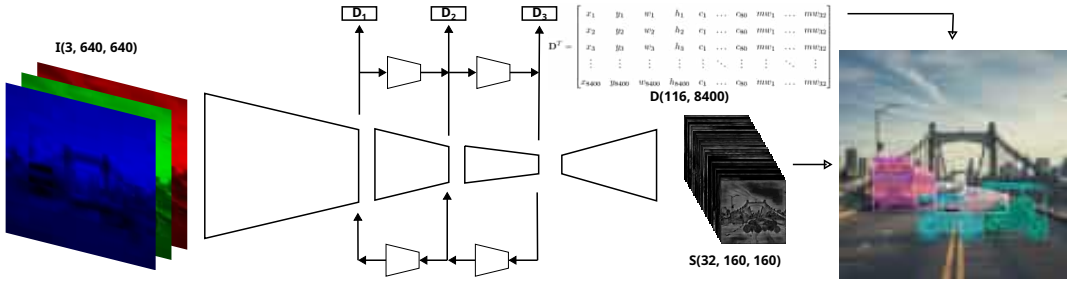


Figure 2. Overview of the YOLOv11-seg architecture used for on-device semantic perception, illustrating the processing pipeline from RGB image input to joint object detection and instance-level segmentation outputs.

work lays the groundwork for future extensions toward native 3D semantic segmentation, language-grounded scene queries, and tighter integration with building information models and digital twins.

2 Methodology

This work proposes an XR-assisted framework for semantic understanding and digitalization of indoor building scenes by integrating on-device AI-based semantic segmentation with photogrammetric reconstruction. The methodology is designed to enable near real-time semantic perception together with spatially consistent 3D modeling and on-site verification through XR. Building upon previous research integrating XR and AI for real-time infrastructure inspection [18], this work extends the approach from damage assessment to semantic scene understanding and 3D digitalization of indoor environments.

The overall pipeline is illustrated conceptually in Figure 1 and follows a closed-loop workflow. While the user navigates an indoor environment wearing an XR device (in this work, a Microsoft HoloLens 2 MR headset), image frames are continuously acquired and processed in near real time using an embedded deep learning model for semantic segmentation. In parallel, selected image frames are stored and transmitted to an external workstation, where a SfM pipeline is executed offline to reconstruct a three-dimensional representation of the scene. Semantic information extracted from the 2D images is subsequently associated with the reconstructed geometry, enabling the generation of a semantically enriched 3D model. This model is then transferred back to the XR device and visualized within the XR environment, allowing the user to interact with, inspect, and validate the reconstructed scene.

2.1 System overview

The proposed framework consists of three tightly coupled components:

1. XR-based semantic perception: A MR headset is used as both an image acquisition and visualization device. Image frames captured by the headset's RGB camera are processed on-device using a lightweight AI model based on the YOLOv11-seg architecture, pre-trained on the COCO dataset for semantic and instance-level object segmentation. During operation, the model performs near real-time inference, generating bounding boxes and segmentation masks for common indoor objects such as furniture and equipment. These predictions are visualized directly in the XR environment through spatially aligned 3D text labels, bounding boxes, and image-aligned planes that overlay segmented regions onto the physical scene.
2. Photogrammetric reconstruction and semantic mapping: In parallel with near real-time XR perception, image data are transmitted to an external computer, where a SfM pipeline is executed to reconstruct a 3D representation of the environment in the form of a point cloud or mesh. Point feature correspondences used during the SfM process are exploited to establish spatial relationships between 2D image pixels and 3D points. By leveraging the semantic segmentation masks generated during inference, semantic labels are propagated from the image domain into the reconstructed geometry, enabling object-level segmentation of the 3D scene.
3. XR-based visualization and validation of semantic 3D models: Once the reconstruction and semantic mapping are completed, the semantically enriched 3D model is transferred back to the XR device. The model is then visualized within the XR environment, where users can manipulate, inspect, and validate both geometry and semantic labels. This enables immediate assessment of digitalization completeness and semantic correctness. If inconsistencies or missing regions are identified, additional images can be acquired on site, closing the loop between perception, reconstruction, and validation.

A key feature of the proposed methodology is its iterative and user-in-the-loop nature. Rather than treating digitalization as an offline, post-processing task, the frame-

work is intended to allow users to continuously assess reconstruction quality and semantic coverage during acquisition. This reduces the risk of incomplete or erroneous models and minimizes the need for repeated site visits. Once a scene is considered satisfactorily digitazed, the workflow could be extended to adjacent rooms or spaces, supporting scalable indoor environment digitalization.

2.2 XR-based image acquisition and on-device semantic segmentation

Near real-time image acquisition and semantic perception are performed directly on a MR headset, which serves as both the sensing and visualization platform. During operation, the headset's RGB camera continuously captures image frames from the user's viewpoint as they navigate the indoor environment. These frames are processed on-device to enable near real-time semantic understanding of the scene without reliance on cloud-based computation. Performing inference directly on the device enables immediate feedback during acquisition, avoids dependency on network connectivity, and supports interactive, user-guided data collection within the XR environment.

Semantic perception is achieved using the YOLOv11-seg architecture (Figure 2), which performs joint object detection and instance-level segmentation in a single inference pass. In this work, a pre-trained YOLOv11-seg model trained on the COCO dataset is employed without additional fine-tuning. The COCO dataset provides a diverse set of common object categories typically found in indoor environments, such as chairs, desks, monitors, and other furniture or equipment, making it suitable for demonstrating semantic scene understanding in building interiors. By relying on a pre-trained model, the focus of this study is placed on system integration and spatial semantic mapping rather than model development or training.

For each processed image frame, the model outputs a set of object detections represented by bounding boxes, class labels, confidence scores, and corresponding pixel-wise segmentation masks. These outputs are generated in the 2D image domain but are immediately translated into spatially aligned visual elements within the XR environment. Bounding boxes are rendered as three-dimensional wireframe overlays aligned with the camera view, while object class labels are displayed as 3D text annotations positioned relative to the detected objects (Figure 3). This enables users to directly perceive semantic information in context, without interrupting their interaction with the physical environment.

In addition to bounding boxes and labels, segmentation results can be visualized through image-aligned 3D planes instantiated in the XR scene (Figure 4). Each plane is textured with the corresponding image frame, where segmented object regions are highlighted according to their predicted classes. These planes are spatially aligned with the real-world background, creating the perception that the segmentation results are embedded within the physical scene. This visualization strategy allows users to observe semantic boundaries while preserving spatial context.



Figure 3. YOLOv11-seg detection outputs visualized in the XR environment.



Figure 4. YOLOv11-seg segmentation results visualized as 3D textured planes in XR.

To avoid visual clutter and maintain usability, segmentation planes are not instantiated continuously for every processed frame. Instead, their creation is triggered explicitly by the user, enabling selective visualization of segmentation results when needed. Multiple planes corresponding to different viewpoints can be instantiated sequentially, supporting inspection of semantic predictions from multiple perspectives. This interaction-driven visualization approach leverages the strengths of XR by combining near real-time perception with user-controlled exploration of semantic information.

2.3 Photogrammetric reconstruction and semantic mapping

Photogrammetric reconstruction is employed to generate a three-dimensional representation of the indoor environment using the image data acquired during XR-based operation. Image frames captured by the MR device are transmitted to an external workstation, where, using Mesh-

room [11], a SfM pipeline executed to estimate camera poses and reconstruct the scene geometry in the form of a sparse point cloud, followed by a dense point cloud and surface mesh. The reconstructed geometry provides a spatial backbone onto which semantic information extracted from 2D images can be mapped.

A key aspect of the proposed framework is the association of 2D semantic segmentation results with the reconstructed 3D scene. During the SfM process, distinctive point features are detected and matched across multiple images to establish correspondences and triangulate 3D points. Because these features originate from known pixel locations in the input images, their projections can be directly related to the semantic segmentation masks generated during on-device inference. For each reconstructed 3D point, its originating 2D feature locations are checked against the corresponding segmentation masks to determine whether the point belongs to a specific semantic object class.

Using this association, semantic labels are propagated from the image domain into the 3D reconstruction. Points whose corresponding 2D features fall within segmented object regions are assigned the semantic class of that object. This results in a semantically labeled point cloud (Figure 5), where each point carries both geometric and semantic information. When multiple image observations are available for a given 3D point, semantic labels are aggregated to improve robustness and reduce the impact of occasional misclassifications.



Figure 5. Semantically segmented 3D point cloud generated from 2D YOLOv11-seg predictions.

To generate a semantically segmented surface representation, the labeled point cloud is further linked to the reconstructed mesh. For each mesh vertex, nearby reconstructed points are queried to determine its most likely semantic class. Mesh triangles whose vertices are consistently associated with the same semantic label are then identified and highlighted as belonging to a specific object or scene element (Figure 6). This process yields a semantic mesh in which object-level regions are explicitly delineated, enabling clear visualization and interaction at the surface level rather than only at sparse point locations.



Figure 6. Semantically segmented 3D mesh derived from the segmented point cloud.

The resulting semantically enriched 3D model preserves both, geometry and object-level semantic structure, providing a coherent representation suitable for visualization, interaction, and validation in XR environment. This representation is subsequently transferred back to the Hololens device, where it can be inspected and manipulated by the user, closing the loop between perception, reconstruction, and semantic scene understanding. Note that this semantic propagation from 2D observations to 3D geometry builds upon well-established approaches (e.g., SemanticFusion [8]), and is here adopted as a component within the proposed workflow rather than introduced as a novel contribution.

2.4 XR visualization and user-guided validation of semantic 3D models

After photogrammetric reconstruction and semantic mapping are completed, the semantically enriched 3D model is transferred back to the MR device. The model is then loaded and visualized within the XR environment, allowing the user to inspect the reconstructed scene directly in situ (Figure 7). This visualization step enables the seamless integration of geometry and semantic information within the physical context of the building interior.

Within the XR environment, the reconstructed model is rendered as an interactive 3D object. Semantic information is displayed through color-coded surfaces corresponding to different object classes, enabling intuitive identification of scene elements. Users can freely manipulate the model—such as rotating, translating, or scaling it—to examine its geometry and semantic segmented objects. This interaction facilitates rapid assessment of reconstruction completeness and semantic correctness.

A central feature of the proposed framework is its user-guided validation loop. By visualizing the semantic 3D



Figure 7. Semantically enriched 3D mesh visualized in the XR environment.

model immediately after reconstruction, users can identify missing regions, insufficient coverage, or incorrect semantic assignments while still on site. If deficiencies are observed, additional image data can be acquired using the XR device and incorporated into the reconstruction process, improving both geometric fidelity and semantic completeness. This iterative process reduces the likelihood of incomplete digitalization and minimizes the need for repeated site visits.

Once the user is satisfied with the reconstructed and semantically labeled scene, the workflow can be extended to additional rooms or adjacent spaces, supporting scalable indoor environment digitalization. By enabling immediate feedback and validation, the XR-based visualization closes the loop between data acquisition, AI-driven semantic perception, and 3D reconstruction, transforming digitalization from an offline post-processing task into an interactive, near real-time process.

3 Conclusions

This paper presented an XR-assisted framework for semantic understanding and digitalization of indoor building scenes by integrating on-device AI-based segmentation with photogrammetric reconstruction. By combining near real-time semantic perception, SfM–based 3D modeling, and immersive visualization within a single workflow, the proposed approach enables users to capture, verify, and interact with semantically enriched 3D scene representations directly on site.

Unlike conventional pipelines that decouple data acquisition, semantic interpretation, and reconstruction into separate offline stages, the proposed framework tightly couples these processes through XR. On-device semantic segmentation provides immediate contextual understanding of the environment, while photogrammetric reconstruction and semantic mapping allow this information to be propagated into a spatially consistent 3D representation. The XR-based visualization and user-guided validation loop further support immediate assessment of digitalization completeness and semantic correctness, reducing the risk of incomplete models and minimizing the need for repeated site visits.

Despite these advantages, several limitations remain. MR headsets offer limited computational resources compared to desktop or cloud-based platforms, making the deployment of large-scale spaces or large deep learning models challenging. As a result, lightweight segmentation models are required, which may limit achievable accuracy compared to offline processing. In addition, current XR hardware still presents ergonomic constraints, including device weight and comfort during prolonged use, as well as limited battery life that restricts continuous operation time. Environmental factors also affect usability; in particular, high ambient lighting conditions—common in construction sites—can reduce hologram visibility and impact the clarity of XR visualizations. Addressing these hardware limitations is largely dependent on future advances in XR device design and is beyond the scope of this work. This study focuses on demonstrating the feasibility of the proposed integrated workflow and therefore does not include a comprehensive quantitative evaluation, including an assessment of the effectiveness of the user-guided validation loop within the XR environment. Finally, photogrammetric reconstruction and semantic 3D mapping stages are performed offline on an external workstation and depend on the selected pipeline and available computing resources.

While the present study employs a pre-trained model based on the COCO dataset to demonstrate semantic scene understanding, the proposed framework is model-agnostic and can be readily adapted to other segmentation models trained on domain-specific datasets. This flexibility enables extension to a broad range of construction-related applications, including semantic segmentation of structural elements, condition assessment and monitoring of construction activities. Future work will explore native 3D semantic segmentation, tighter integration between reconstruction and inference pipelines, and language-grounded models as interaction mechanisms that allow users to query and manipulate semantic content directly within the XR environment. Additionally, it will include a systematic evaluation of the proposed framework. This will involve assessing 2D segmentation performance, evaluating 3D semantic consistency and reconstruction completeness, and measuring end-to-end latency of the pipeline. In addition, a user study will be conducted to quantify the impact of the XR-based validation loop on usability and data acquisition efficiency. Together, these directions highlight the potential of XR as an active tool for near real-time scene understanding, digitalization, and decision support in the built environment.

Acknowledgements

This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 1 (Project ID: RS52/25 — "Spatio-temporal visual perception and AI-agent-based scene representation for robotic support in construction").

References

[1] B G Pantoja-Rosero, A Rusnak, F Kaplan, and K Beyer. Generation of lod4 models for buildings towards the automated 3d modeling of bims and digital twins. *Automation in Construction*, 168:105822, 2024. ISSN 0926-5805. doi:10.1016/j.autcon.2024.105822. URL https://www.sciencedirect.com/science/article/pii/S0926580524005582.

[2] Bryan G Pantoja-Rosero, Radhakrishna Achanta, and Katrin Beyer. Damage-augmented digital twins towards the automated inspection of buildings. *Automation in Construction*, 150:104842, 2023. doi:10.1016/j.autcon.2023.104842. URL https://www.sciencedirect.com/science/article/pii/S0926580523001024.

[3] B G Pantoja-Rosero, R Achanta, M Kozinski, P Fua, F Perez-Cruz, and K Beyer. Generating lod3 building models from structure-from-motion and semantic segmentation. *Automation in Construction*, 141, 2022. doi:10.1016/j.autcon.2022.104430. URL https://www.sciencedirect.com/science/article/pii/S092658052200303X?via%3Dihub.

[4] A Chassignet and R Achanta. Partial annotations in active learning for semantic segmentation. *Automation in Construction*, 168:105828, 2024. doi:10.1016/j.autcon.2024.105828. URL https://www.sciencedirect.com/science/article/pii/S0926580524005648.

[5] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3D reconstructions of indoor scenes. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2432–2443, 2017. doi:10.1109/CVPR.2017.261.

[6] Bryan G Pantoja-Rosero, D Oner, Mateusz Kozinski, Radhakrishna Achanta, Pascal Fua, Fernando Pérez-Cruz, and Katrin Beyer. Topo-loss for continuity-preserving crack detection using deep learning. *Construction and Building Materials*, 344:128264, 2022. doi:10.1016/j.conbuildmat.2022.128264. URL https://www.sciencedirect.com/science/article/pii/S0950061822019250.

[7] Alexander Rusnak, Bryan G Pantoja-Rosero, Frédéric Kaplan, and Katrin Beyer. Automatic removal of non-architectural elements in 3d models of historic buildings with language embedded radiance fields. *Heritage*, 7(6):3332–3349, 2024. doi:10.3390/heritage7060157. URL https://www.mdpi.com/2571-9408/7/6/157.

[8] John McCormac, Ankur Handa, Andrew J. Davison, and Stefan Leutenegger. Semanticfusion: Dense 3D semantic mapping with convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4628–4635, 2017. doi:10.1109/ICRA.2017.7989538.

[9] Martin Runz, Maud Bufier, and Lourdes Agapito. Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 10–20, 2018. doi:10.1109/ISMAR.2018.00024.

[10] Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016. doi:10.1109/CVPR.2016.445.

[11] Carsten Griwodz, Simone Gasparini, Lilian Calvet, Pierre Gurdjos, Fabien Castan, Benoit Maujean, Gregoire De Lillo, and Yann Lanthony. Alicevision meshroom: An open-source 3d reconstruction pipeline. *Proceedings of the 12th ACM multimedia systems conference*, pages 241–247, 2021. doi:10.1145/3458305.3478443. URL https://dl.acm.org/doi/10.1145/3458305.3478443.

[12] Joseph Rowell. Semantic validation in structure from motion. *arXiv preprint*, 2023. doi:https://doi.org/10.48550/arXiv.2304.02420.

[13] Abhijit Kundu, Yin Li, Fuxin Li, and James M. Rehg. Joint semantic segmentation and 3D reconstruction from monocular video. In *Computer Vision – ECCV 2014*, pages 703–718. Springer, 2014. doi:10.1007/978-3-319-10599-4_45.

[14] Yunhan Yang, Xiaoyang Wu, Tong He, Hengshuang Zhao, and Xihui Liu. SAM3D: Segment anything in 3D scenes. *arXiv preprint*, 2023. doi:https://doi.org/10.48550/arXiv.2306.03908.

[15] Shuyuan Chen, Zhaoqing Wang, Jiepeng Wang, Yuexin Ma, Mingming Gong, Wenping Wang, and

Tongliang Liu. Panoslam: Panoptic 3D scene reconstruction via gaussian SLAM. *arXiv preprint*, 2024. doi:https://doi.org/10.48550/arXiv.2501.00352.

[16] J.I Fan and K Khoshelham. Augmented reality asset tracking using hololens. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-4-2021:121–127, 2021. doi:10.5194/isprs-annals-V-4-2021-121-2021.

[17] Juan Carlos Navares-Vázquez, Zhouyan Qiu, Pedro Arias, and Jesús Balado. Hololens 2 performance analysis for indoor/outdoor 3d mapping. *Journal of Building Engineering*, 108:112826, 2025. ISSN 2352-7102. doi:https://doi.org/10.1016/j.jobe.2025.112826.

[18] B.G. Pantoja-Rosero and S. Salamone. Integrating extended reality and ai-based damage segmentation for near real-time, traceable bridge inspections. *Automation in Construction*, 180:106567, December 2025. doi:10.1016/j.autcon.2025.106567. URL `https://doi.org/10.1016/j.autcon.2025.106567`.

[19] Boan Tao, Frédéric Bosché, and Jiajun Li. Mixed reality-based mep construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison. *Automation in Construction*, 168:105852, dec 2024. doi:10.1016/j.autcon.2024.105852.

[20] David DL Mascareñas, JoAnn P Ballor, Oscar L McClain, Miranda A Mellor, Chih-Yu Shen, Brian Bleck, John Morales, Li-Ming R Yeong, Benjamin Narushof, Philo Shelton, Eric Martinez, Yongchao Yang, Alessandro Cattaneo, Troy A Harden, and Fernando Moreu. Augmented reality for next generation infrastructure inspections. *Structural Health Monitoring*, 20(4):1957–1979, 2021. doi:10.1177/1475921720953846.

[21] Mahta Zakaria, Enes Karaaslan, and F Necati Catbas. Advanced bridge visual inspection using real-time machine learning in edge devices. *Advances in Bridge Engineering*, 3:27, 2022. ISSN 2662-5407. doi:10.1186/s43251-022-00073-y. URL `https://doi.org/10.1186/s43251-022-00073-y`.

[22] Kaveh Malek, Ali Mohammadkhorasani, and Fernando Moreu. Methodology to integrate augmented reality and pattern recognition for crack detection. *Computer-Aided Civil and Infrastructure Engineering*, 38(8):1000–1019, 2023. doi:10.1111/mice.12932.

Artificial Intelligence and Computer Vision for Construction Progress Monitoring: A Bibliometric and Thematic Review (2005–2026)

Kevin Torres [1], Luis A. Cristancho [1], Omar Sánchez [2], Rodrigo F. Herrera [1] and Karen Castañeda [1]

[1] School of Civil Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2374631, Chile
[2] Department of Civil Engineering, Pontificia Universidad Javeriana, Bogotá D.C., 111711, Colombia
torreskd@javeriana.edu.co, lcristancho@utb.edu.co, omar_sanchezr@javeriana.edu.co, rodrigo.herrera@pucv.cl, karen.castaneda@pucv.cl

Abstract - Reliable and timely construction progress monitoring remains challenging due to the dynamic, spatially complex, and visually cluttered nature of jobsite environments, where conventional inspection- and report-based approaches often yield subjective and intermittent assessments. This paper maps the adoption of artificial intelligence and computer vision for construction progress monitoring through a bibliometric and thematic review of 103 publications spanning 2005–2026. Using performance analysis and science mapping in Bibliometrix, the study characterizes scientific production, leading contributors, thematic structures, and topic evolution. Results indicate a marked growth and concentration of research output, led by the United States (61 publications) and China (43), followed by the United Kingdom (19), Germany (14), Australia (13), and South Korea (13). The thematic map reveals four interrelated research fronts: (1) computer vision as a transversal methodological core (e.g., machine learning, activity recognition, image processing, real-time monitoring), (2) construction progress monitoring as the primary application theme (strongly associated with deep learning, point clouds, object detection, and semantic segmentation), (3) photogrammetry and reality-capture pipelines as mature enabling technologies, and (4) BIM as an integration layer connecting field inference to planning and control. Trend analysis suggests a transition from early recognition and as-built modeling toward deep learning–driven, 3D-centric, and increasingly operational workflows. The findings provide an evidence-based agenda to strengthen benchmarking, generalize across sites, and tighten the coupling between vision outputs and project control systems.

Keywords – Artificial Intelligence; Computer Vision; Construction Progress Monitoring; Bibliometric Analysis; Trend Analysis

1 Introduction

The growing pressure to deliver construction projects within constrained schedules and budgets has heightened the need for reliable, objective, and continuous mechanisms to assess construction progress [1]. Conventional progress-monitoring practices, often reliant on manual inspections, periodic reports, and visual judgment, struggle to capture the dynamic and spatial complexity of construction activities [2]. As construction sites increasingly generate large volumes of visual data from fixed cameras, mobile devices, and aerial platforms, artificial intelligence and computer vision have gained prominence as technologies that transform visual information into structured indicators of project advancement [3]. Recent advances in deep learning and visual recognition enable automated detection of construction elements, activity states, and temporal changes directly from images and videos. These capabilities allow near-real-time progress evaluation, reduce subjectivity in reporting, and support more consistent performance tracking across project phases [4]. Furthermore, vision-based outputs can be aligned with planning and scheduling systems to provide data-driven insights into deviations between planned and actual progress. Despite the rapid expansion of technical solutions, research contributions remain dispersed across methods, application scenarios, and evaluation criteria, limiting the consolidation of knowledge in this field. A comprehensive understanding of how artificial intelligence and computer vision have been applied to construction progress monitoring, how research themes have evolved over time, and where methodological gaps persist is therefore essential. In this context, bibliometric and thematic analyses offer a systematic means to map scientific production, identify influential research streams, and highlight emerging directions in AI-enabled construction progress monitoring [5].

The rapid evolution of artificial intelligence and computer vision has significantly altered how construction processes are observed, measured, and analyzed, particularly in contexts marked by high visual complexity and continuous physical transformation. Vision-based technologies enable automated interpretation of site imagery, facilitating identification of construction elements, activity states,

and temporal changes that are difficult to capture through conventional monitoring approaches [6,7]. Within construction management research, these developments have supported the transition from subjective, report-based progress assessments to data-driven, repeatable evaluation mechanisms. Deep learning architectures applied to images and videos have demonstrated the ability to track work execution, detect deviations from planned schedules, and provide quantitative indicators of progress with increasing accuracy [8]. At the same time, integrating visual outputs with planning, scheduling, and digital modeling environments has expanded the analytical capabilities of progress-monitoring systems. The advancement of Artificial Intelligence (AI) applied to construction progress monitoring is closely connected to the development of automation and robotics. Specifically, robotic inspection systems and autonomous monitoring platforms are increasingly playing a vital role in boosting efficiency and accuracy at construction sites. In recent years, automation has evolved from emerging technology to a key component of progress monitoring, especially through robotic inspection systems. Progress in this field enables the integration of AI for hazard prediction, real-time surveillance, and autonomous decision-making, leading to substantial improvements in tracking construction activities and reducing risk-related delays. Despite these advances, research remains fragmented across algorithms, data sources, and application contexts, often emphasizing technical performance without systematically relating findings to broader research trends. Furthermore, differences in evaluation metrics, validation strategies, and reporting practices hinder comparative assessment across studies. As a result, the absence of a consolidated view of how artificial intelligence and computer vision have been applied to construction progress monitoring limits the identification of dominant research streams and emerging thematic directions [9]. Addressing this limitation requires a structured synthesis of scientific production that captures methodological evolution, thematic concentration, and research gaps over time. In this regard, bibliometric and thematic analyses provide a robust framework for organizing existing knowledge and supporting the strategic development of AI-enabled research on progress monitoring.

Despite the increasing number of studies on artificial intelligence and computer vision in construction, the literature remains fragmented across methodologies, application scales, and analytical objectives, limiting a comprehensive understanding of the field's evolution [10–13]. Previous studies have developed algorithms for activity recognition, object detection, and progress estimation using images, videos, and sensor-based data [14–16], often demonstrating strong technical performance under controlled conditions. However, most research prioritizes algorithm development without systematically examining the evolution of research themes, evaluation methods, and application contexts over time. Consequently, it remains unclear which technological paradigms have driven the field, how dominant research streams have emerged, and where persistent gaps hinder practical implementation. To address these limitations, this study conducts a bibliometric and thematic review of publications from 2005 to 2026 on artificial intelligence and computer vision for construction progress monitoring. By integrating performance analysis, scientific mapping, and thematic evolution assessment, the study identifies influential contributions, dominant research clusters, and emerging thematic trends shaping the field's intellectual structure. Beyond descriptive analysis, the review highlights conceptual imbalances, methodological gaps, and underexplored connections between algorithmic development and real-world deployment in construction environments, providing a foundation for future research and the development of scalable, reliable, and data-driven progress-monitoring solutions.

2 Literature Background

The growing demand for objective, continuous assessment has positioned progress monitoring as a key area in digital construction. Conventional practices based on site visits and subjective reporting offer limited capacity to capture evolving spatial and temporal conditions [14,15]. The availability of visual data from cameras, UAVs, and mobile devices has driven adoption of AI and computer vision for automated evaluation [11]. Recent studies show that deep learning enables systematic interpretation of site imagery, identifying elements, activity states, and execution stages with reduced reliance on manual judgment [16,17]. These advances shift monitoring toward data-oriented paradigms, where measurable indicators replace qualitative descriptions. Yet diversity in analytical approaches, acquisition strategies, and validation procedures reveals fragmented trajectories, complicating knowledge consolidation.

A substantial body of work has developed vision methods to detect activities and estimate progress from images and videos. Early studies focused on object recognition and feature-based techniques, while recent contributions employ deep neural networks to capture complex spatial–temporal relation [18]. Several works report strong performance in recognizing work states and quantifying progress under controlled conditions [19]. Others link visual outputs with scheduling to detect deviations between planned and executed work [20]. Persistent challenges include limited transferability, sensitivity to environmental variability, reliance on context-specific datasets, and inconsistent evaluation metrics, which constrain comparability and generalization.

Beyond algorithms, attention has turned to how AI and computer vision research has evolved as a domain. While individual studies provide technical insights,

fewer examine publication dynamics, thematic concentrations, and intellectual structures [10,21–23]. Addressing this gap requires approaches tracing evolution across methods and applications. Bibliometric and thematic analyses systematically identify influential publications, dominant streams, and emerging topics. By reviewing output from 2005–2026, this study offers a comprehensive view of the maturation of AI and computer vision in progress monitoring, highlighting enduring challenges, methodological shifts, and opportunities for scalable, reliable, data-driven solutions.

3 Research Method

3.1 Methodological rationale and scope definition

Research on AI and computer vision for construction progress monitoring has expanded rapidly, producing a heterogeneous and methodologically dispersed body of literature. Contributions differ in data types, analytical pipelines, validation strategies, and monitoring objectives, complicating the identification of dominant trajectories and consolidated paradigms. In this context, bibliometric analysis offers a reproducible framework to examine scientific production collectively, revealing structural patterns, thematic concentrations, and temporal dynamics beyond narrative or technology-centered reviews.

This study adopts a bibliometric and thematic review of peer-reviewed publications from 2005 to 2026, spanning early visual inspection and object recognition to recent advances in deep learning, 3D sensing, and data-intensive workflows. Three objectives guide the analysis: (1) evaluate the evolution and performance of scientific output, (2) characterize the intellectual and thematic structure, and (3) identify dominant and emerging trends shaping AI-enabled progress monitoring. These objectives provide a coherent foundation for interpreting methodological evolution and research priorities over time.

3.2 Selection of bibliometric techniques

The bibliometric methodology was designed to align directly with the research objectives, ensuring consistency between the analytical procedures and the evidence generated. Performance analysis was used to characterize the development of the field by examining publication growth, citation patterns, and the geographical distribution of scientific output, thereby identifying the countries, institutions, and research contributions that have shaped the evolution of artificial intelligence and computer vision applications for construction progress monitoring. To complement this assessment, citation-based indicators and collaboration networks were analyzed to evaluate scientific influence and knowledge exchange within the domain, revealing the central role of leading institutions, authors, and technological paradigms. In particular, the prominence of terms such as machine learning and deep learning among highly cited publications reflects the growing importance of data-driven approaches for automation and real-time monitoring in construction environments. The conceptual structure of the literature was examined through science-mapping techniques, including co-word analysis and thematic mapping, which enabled the identification of recurrent concepts, thematic relationships, and the relative maturity and relevance of different research streams. Finally, temporal trend analysis provided insight into the evolution of the field, highlighting a transition from early recognition-based and as-built modeling approaches toward deep learning-enabled, three-dimensional, and increasingly operational frameworks for construction progress monitoring.

Data collection and filtering process

Scientific data were collected from Scopus, chosen for its broad coverage of peer-reviewed journals in engineering, construction, and computer science, and its robust indexing standards. A structured search strategy was refined iteratively, combining preliminary inspection with team input. Keywords on artificial intelligence, computer vision, and progress monitoring were combined with Boolean operators to balance sensitivity and specificity.

The final search equation — [("construction progress" OR "progress monitoring" OR "progress tracking") AND ("computer vision" OR "image-based" OR "video-based" OR photogrammetry OR "visual inspection")] — targeted publications on visual data analysis, automated monitoring, and progress assessment. Records were refined through multi-stage filtering to ensure relevance and quality. Temporal limits (2005–2026) captured long-term evolution. Filters retained journal articles and reviews, excluding conference and editorial material, and focused on engineering, construction, and computer science. Titles, abstracts, and keywords were reviewed independently, with full-text checks for ambiguous cases. Inclusion required explicit use of AI or computer vision in construction progress monitoring and sufficient metadata for bibliometric analysis. In this study, conference proceedings were excluded from the bibliometric analysis mainly because they usually present preliminary results. While conferences offer valuable information and advancements, they tend to highlight initial findings that often lack the depth needed for a thorough analysis. By focusing on articles from scientific journals, the goal is to ensure a more consistent and comprehensive examination of the field's impact and emerging trends, leading to a more reliable assessment of the research landscape. This systematic process resulted in a curated corpus that supports reliable, reproducible bibliometric analysis.

3.3 Execution of bibliometric and thematic analysis

The bibliometric analysis was conducted in three stages: performance analysis, scientific mapping, and trend analysis. Performance indicators were computed using the Bibliometrix package in RStudio, enabling systematic assessment of annual publication output, citation trajectories, and country-level contributions. These metrics reveal how scientific activity has intensified and become more geographically distributed over time. The scientific mapping stage used author-provided keywords to construct co-occurrence networks that reflect the field's conceptual organization. To enhance consistency, synonymous terms and lexical variants were standardized using a controlled vocabulary developed collaboratively by the research team. Thematic maps were generated using centrality and density measures, allowing research clusters to be classified as motor, basic, emerging, or niche themes. Mathematical formulations for the equivalence index, centrality, and density were based on established bibliometric methodologies. The scientific mapping process was implemented using author keywords as the unit of analysis and co-word analysis as the relational technique. Prior to network construction, the keyword set was cleaned by normalizing spelling variants, singular and plural forms, and closely related expressions using a controlled vocabulary agreed upon by the research team. Only terms directly related to artificial intelligence, computer vision, and construction progress monitoring were retained for thematic interpretation. The thematic map was generated in Bibliometrix using Callon's centrality and density measures, which enabled classification of clusters as motor, basic, niche, or emerging themes. In parallel, the trend-topic analysis examined the temporal occurrence of keywords across the study period to identify shifts in conceptual emphasis and methodological development within the corpus. Finally, trend analysis examined temporal variations in keyword prominence, revealing shifts from early object recognition and as-built modeling toward deep learning–driven, three-dimensional, and operational monitoring workflows. This multi-layered analytical process supports a comprehensive interpretation of the maturation of AI and computer vision research for construction progress monitoring.

Evidence synthesis and analytical structure

The evidence synthesis was structured to directly reflect the three analytical dimensions defined by the research objectives. Performance analysis results describe growth patterns and the leading contributors shaping the field. Scientific mapping outcomes interpret the intellectual structure and thematic relationships among research fronts. Trend analysis findings highlight methodological transitions and emerging priorities in recent research. This organization ensures consistency between the analytical framework and the reported results, supporting transparent interpretation and replication. The methodological design emphasizes rigor, traceability, and reproducibility, providing a robust foundation for identifying future research opportunities in AI- and computer vision–based construction progress monitoring.

4 Results

4.1 Performance Analysis

The patterns of scientific production suggest both growth and concentration: from 2005 to 2026, the field has expanded rapidly as AI-driven monitoring capabilities have matured and diversified (**Error! Reference source not found.**).

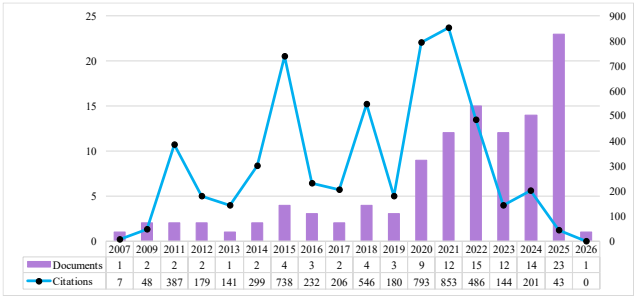


Figure 1. Scientific production by year.

The publication output indicates that the United States and China are the leading contributors in this area, with 61 and 43 publications, respectively, followed by the United Kingdom (19), Germany (14), Australia (13), and South Korea (13) (**Error! Reference source not found.**).



Figure 2. Scientific production by country.

Furthermore, the collaboration networks showed steady growth in international team interactions, indicating increased integration of global research in construction progress monitoring. The most frequently cited terms in the documents, such as 'machine learning' and 'deep learning,' were identified as the most influential in the field, highlighting the trend toward automating monitoring tasks with artificial intelligence. These results underscore the growing importance of AI and machine learning algorithms in

improving the accuracy and efficiency of progress tracking on construction projects.

4.2 Science Mapping

The thematic map suggests that the intellectual structure of research on artificial intelligence and computer vision for construction progress monitoring is not dispersed across isolated topics but rather concentrated around four interrelated fronts. Their relevance and developmental intensity can be interpreted in terms of their positions with respect to centrality and density (**Error! Reference source not found.**). This configuration indicates that the field has evolved through a combination of conceptual consolidation and methodological specialization, with certain themes gaining prominence by shaping the broader research agenda, while others advance along more focused, internally cohesive trajectories. Notably, the right side of the map, associated with greater centrality, highlights themes that exert a stronger influence on the domain's overall organization. Within this area, two clusters emerge as the main anchors of the literature, suggesting that the field is shaped by dual logic: one oriented toward methodological development and the other focused on assessing application-driven progress. Rather than representing disconnected lines of inquiry, these clusters reveal how the domain is increasingly structured around the interaction between enabling analytical techniques and the practical demands of construction monitoring. The first cluster is computer vision, marked by high centrality and dense associations with machine learning, activity recognition, image processing, and real-time monitoring, as well as newer links to digital twins and segmentation models. This highlights computer vision as a transversal methodological backbone supporting diverse monitoring tasks rather than an isolated application.

The second cluster corresponds to construction progress monitoring, a basic theme with high relevance but moderate density. It represents the main application driving the domain, though still consolidating. Emphasis lies on deep learning and 3D representations, oftenconnected with point cloud, object detection, semantic segmentation, and image recognition. These terms reflect a shift toward automated, data-driven interpretations of site conditions, moving beyond visual documentation to structured progress assessments.

In contrast, the upper-left quadrant groups photogrammetry as a niche theme, linked to UAV, laser scanning, 3D reconstruction, reality capture, and automated measurement. Its high density but low centrality shows methodological maturity in acquisition and reconstruction, yet isolation suggests it is treated more as an enabling technology than a driver of semantic interpretation.



Figure 3. Thematic map analysis.

The node interaction map (Figure 4) further supports these identified patterns, illustrating how construction progress monitoring acts as a central hub, intricately connected to point cloud data and deep learning techniques. Meanwhile, computer vision functions as a comprehensive aggregator, linking various recognition and inference tasks. This central role of progress monitoring within the network highlights its vital part in the ongoing development of AI-enabled monitoring systems. BIM (Building Information Modeling), in this context, serves as a key integration layer that connects technical innovations with management-focused objectives, helping to translate real-time data into actionable insights for decision-making. The structure shown in the map indicates an ongoing shift toward integrated monitoring pipelines, where 3D data capture, deep learning–based inference, and BIM-enabled environments work together. This shift suggests moving from isolated, independent technologies to a more operational, traceable, and decision-oriented approach to progress monitoring in construction projects. The implication is clear: the future of construction progress monitoring is evolving into a unified, adaptive system that tracks progress while continuously adjusting to the project's changing needs in real time.



Figure 4. Node interaction map.

Beyond identifying thematic clusters, analyzing node interactions reveals a layered structure in AI-enabled progress-monitoring systems. At the core, technologies like photogrammetry and three-dimensional reconstruction are essential for creating increasingly precise and reliable geometric models of

site conditions. These form the basis for higher-level analytical processes. Next, computer vision and deep learning conduct semantic inference, transforming raw visual data into recognizable construction components, activity states, and work progress assessments. This layer is crucial for converting large volumes of unstructured visual information into organized, usable data that supports progress evaluation. At the system-integration level, BIM functions as the backbone of a unified environment, connecting visual outputs with planning, scheduling, and control mechanisms. This link is vital to ensure that the insights produced by monitoring systems are not only accurate but also actionable within broader project management contexts. Connecting real-time data with decision-making tools in BIM creates a feedback loop that helps continuously assess and adjust project progress to meet overall goals. However, despite these advancements, the thematic structure also uncovers ongoing fragmentation in current research. Although the layers composing the monitoring pipeline are becoming more interconnected, they still tend to operate separately. While notable breakthroughs have been achieved in vision-based inference and geometric reconstruction, fewer studies have addressed interoperability and feedback mechanisms necessary to close the loop in progress monitoring systems. This fragmentation partly explains why many systems report high accuracy in controlled environments or pilot projects but face difficulties when scaled for real-world construction applications. The lack of a unified framework that integrates these components into operational systems may be a significant barrier to the widespread adoption of advanced monitoring solutions. Lastly, although individual parts of AI-driven construction progress monitoring have advanced markedly, further research into integrating these components into a seamless, closed-loop system remains vital. Moving beyond isolated technical innovations toward a fully operational, scalable system suitable for diverse construction environments presents the next challenge for research in this field.

4.3 Trend Analysis

The trend-topic analysis reveals a clear methodological and thematic transition in AI- and computer vision–based construction progress monitoring between 2009 and 2025 (**Error! Reference source not found.**). Early studies focused on as-built modeling and object recognition, emphasizing the alignment of site conditions with design intent through constrained recognition tasks. As the field evolved, topics such as laser scanning, daily construction images, image analysis, and imaging techniques became prominent, reflecting increased attention to reliable visual data acquisition and processing pipelines. From 2020 onward, the research focus shifted toward integrated progress assessment, with growing emphasis on BIM, 3D reconstruction, photogrammetry, and construction progress monitoring and evaluation. This evolution indicates a transition from isolated perception tasks to integrated monitoring workflows combining data capture, reconstruction, and interpretation. Recent trends further demonstrate the consolidation of contemporary AI paradigms.

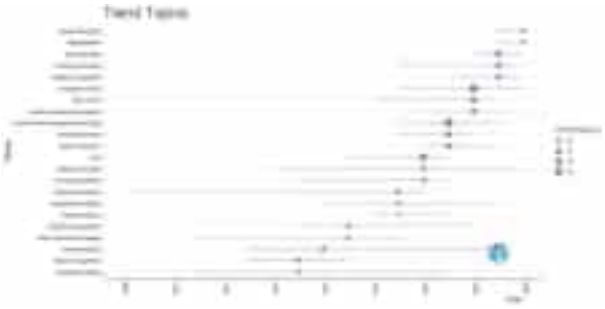


Figure 5. Trend analysis.

The increased frequency of computer vision and point cloud signals a maturation of 3D-driven monitoring. At the same time, the rise of deep learning, machine learning, and activity recognition reflects broader adoption of data-driven models capable of capturing complex spatial–temporal patterns in site environments. The emergence of digitalization and manual annotation near the end of the period suggests growing attention to dataset construction, labeling burden, and the governance of visual data at scale. Together, these trends depict a field moving beyond proof-of-concept to scalable, operational monitoring solutions that rely on robust 3D representations, learning-based inference, and tighter coupling with project information environments.

The trend analysis reveals not only technological advancements but also a progressive redefinition of construction progress monitoring. Early research approached progress assessment primarily as a visual correspondence problem, emphasizing object detection and the comparison between as-built conditions and design intent. As sensing and data acquisition technologies matured, the focus expanded toward automated spatial representation through photogrammetry and 3D reconstruction techniques, enabling more comprehensive and scalable monitoring workflows. More recently, the emergence of deep learning and activity recognition has shifted the field toward dynamic interpretations of construction progress, where monitoring is increasingly associated with evolving work states rather than static measures of completion. Despite these advances, the continued reliance on manual annotation and dataset preparation highlights a persistent challenge, as the rapid development of model architectures has not been matched by the availability of standardized and transferable training datasets, limiting model generalizability and large-scale implementation.

The corpus indicates that benchmark inconsistency is not merely a reporting issue, but a structural characteristic of the field. The analyzed studies address markedly different analytical targets, ranging from training image database development [6] and BIM-

linked indoor progress monitoring [8] to semantic information extraction [19], ontology-based image enrichment [20], and RTLS-integrated monitoring workflows [21]. This heterogeneity is also reflected in the thematic structure, where deep learning, point cloud, object detection, semantic segmentation, and BIM-related terms coexist within the same research domain. Trend analysis reinforces this interpretation, as the recent visibility of manual annotation and digitalization suggests continued dependence on project-specific data curation, while the relatively late prominence of BIM indicates that many studies still evaluate perception modules in isolation rather than within comparable end-to-end monitoring workflows. Consequently, reported performance is often tied to different tasks, datasets, validation settings, and deployment contexts, thereby weakening cross-study comparability and limiting the accumulation of transferable evidence.

5 Conclusions

This study contributes to the understanding of AI and computer vision applications in construction progress monitoring by consolidating research across methodologies, data acquisition strategies, applications, and evaluation approaches. The thematic analysis identifies four main research directions: construction progress monitoring, computer vision, photogrammetry and 3D reconstruction, and BIM integration with project information systems. The review also shows a methodological transition from conventional feature-based approaches to deep learning models capable of addressing the spatial and temporal complexity of construction environments. These findings reflect the evolution of monitoring systems from isolated analytical tools into integrated frameworks that support construction management and decision making.

The study also highlights practical implications for implementing automated monitoring systems in real construction projects. Effective monitoring frameworks require the integration of capture technologies such as fixed cameras, mobile devices, and UAVs with analytical models adapted to different monitoring levels, including activities, elements, operational states, and temporal changes. The integration of AI outputs with planning processes and BIM environments can improve performance tracking, reduce subjectivity, and support earlier detection of project deviations. However, some limitations remain. Bibliometric and thematic reviews depend on metadata quality, keyword consistency, and search strategies, which may restrict the inclusion of relevant studies. The absence of standardized datasets, evaluation metrics, and benchmarking practices also limits comparability across studies. Future research should focus on standardized evaluation protocols, open datasets, domain adaptation methods, and multimodal AI frameworks integrated with BIM-based management systems to improve robustness and applicability in real construction environments.

6 Acknowledgements

Kevin Torres and Luis Cristancho gratefully acknowledge the Doctorate in Intelligent Industry program at the Pontificia Universidad Católica de Valparaíso for the academic training and institutional support provided during their doctoral studies. Karen Castañeda gratefully acknowledges the financial support provided by the Dirección de Investigación of the Vicerrectoría de Investigación, Creación e Innovación, VINCI-DI, at the Pontificia Universidad Católica de Valparaíso, through the DI Regular PUCV 2026 project, grant number 044.783/2026.

References

[1] J.Y. Kim, G.H. Kim, Identifying Members of Common Structures Utilizing Three-Dimensional Detecting Information for 3D Point Cloud Model Application, Sustainability 2023, Vol. 15, Page 14073 15 (2023) 14073. https://doi.org/10.3390/SU151914073.

[2] L. Li, Z. Huang, J. Wang, B. Du, L. Dai, Automated construction monitoring based on computer vision: A comprehensive review, Developments in the Built Environment 25 (2026) 100832. https://doi.org/10.1016/J.DIBE.2025.100832.

[3] M. Sami Ur Rehman, M.T. Shafiq, F. Ullah, Automated Computer Vision-Based Construction Progress Monitoring: A Systematic Review, Buildings 2022, Vol. 12, Page 1037 12 (2022) 1037. https://doi.org/10.3390/BUILDINGS12071037.

[4] O. Moselhi, H. Bardareh, Z. Zhu, Automated Data Acquisition in Construction with Remote Sensing Technologies, Applied Sciences 2020, Vol. 10, Page 2846 10 (2020) 2846. https://doi.org/10.3390/APP10082846.

[5] H. Li, Y. Zhang, Y. Cao, J. Zhao, Z. Zhao, Applications of artificial intelligence in the AEC industry: a review and future outlook, Journal of Asian Architecture and Building Engineering 24 (2025) 1672–1688. https://doi.org/10.1080/13467581.2024.2343800;WGROUP:STRING:PUBLICATION.

[6] J. Hwang, J. Kim, S. Chi, J.O. Seo, Development of training image database using web crawling for vision-based site monitoring, Autom. Constr. 135 (2022) 104141. https://doi.org/10.1016/J.AUTCON.2022.104141.

[7] J. Oh, S. Hong, B. Choi, Y. Ham, H. Kim, Integrating text parsing and object detection for automated monitoring of finishing works in construction projects, Autom. Constr. 174 (2025) 106139. https://doi.org/10.1016/J.AUTCON.2025.106139.

[8] W. Wei, Y. Lu, T. Zhong, P. Li, B. Liu, Integrated vision-based automated progress monitoring of indoor construction using mask region-based convolutional neural networks and BIM, Autom. Constr. 140 (2022) 104327. https://doi.org/10.1016/J.AUTCON.2022.104327.

[9] N. Jacob-Loyola, F. Muñoz-La Rivera, R.F. Herrera, E. Atencio, Unmanned Aerial Vehicles (UAVs) for

Physical Progress Monitoring of Construction, Sensors 2021, Vol. 21, Page 4227 21 (2021) 4227. https://doi.org/10.3390/S21124227.
[10] T. Patel, B.H.W. Guo, Y. Zou, A scientometric review of construction progress monitoring studies, Engineering, Construction and Architectural Management 29 (2022) 3237–3266. https://doi.org/10.1108/ECAM-10-2020-0799.
[11] S.J. Badhan, R. Samsami, Artificial Intelligence (AI) in Construction Safety: A Systematic Literature Review, Buildings 15 (2025) 4034. https://doi.org/10.3390/BUILDINGS15224084/S1.
[12] M. Regona, T. Yigitcanlar, C.K.H. Hon, M. Teo, Mapping Two Decades of AI in Construction Research: A Scientometric Analysis from the Sustainability and Construction Phases Lenses, Buildings 2023, Vol. 13, Page 2346 13 (2023) 2346. https://doi.org/10.3390/BUILDINGS13092346.
[13] K. Chen, X. Zhou, Z. Bao, M.J. Skibniewski, W. Fang, Artificial intelligence in infrastructure construction: A critical review, Frontiers of Engineering Management 2024 12:1 12 (2024) 24–38. https://doi.org/10.1007/S42524-024-3128-5.
[14] K.M. Rashid, J. Louis, Times-series data augmentation and deep learning for construction equipment activity recognition, Advanced Engineering Informatics 42 (2019) 100944. https://doi.org/10.1016/J.AEI.2019.100944.
[15] M. Nur Sholeh, S. Fauziyah, R. Radian Khasani, Effect of Building Information Modeling (BIM) on reduced construction time-costs: A case study, E3S Web of Conferences 202 (2020). https://doi.org/10.1051/E3SCONF/202020202012.
[16] M. Zhang, S. Chen, X. Zhao, Z. Yang, Research on Construction Workers' Activity Recognition Based on Smartphone, Sensors 2018, Vol. 18, Page 2667 18 (2018) 2667. https://doi.org/10.3390/S18082667.
[17] M.A. Musarat, A.M. Khan, W.S. Alaloul, N. Blas, S. Ayub, Automated monitoring innovations for efficient and safe construction practices, Results in Engineering 22 (2024) 102057. https://doi.org/10.1016/J.RINENG.2024.102057.
[18] W.S. Alaloul, K.M. Alzubi, A.B. Malkawi, M. Al Salaheen, M.A. Musarat, Productivity monitoring in building construction projects: a systematic review, Engineering, Construction and Architectural Management 29 (2021) 2760–2785. https://doi.org/10.1108/ECAM-03-2021-0211.
[19] Y. Wang, B. Xiao, A. Bouferguene, M. Al-Hussein, H. Li, Vision-based method for semantic information extraction in construction by integrating deep learning object detection and image captioning, Advanced Engineering Informatics 53 (2022) 101699. https://doi.org/10.1016/J.AEI.2022.101699.
[20] C. Zeng, T. Hartmann, L. Ma, ConSE: An ontology for visual representation and semantic enrichment of digital images in construction sites, Advanced Engineering Informatics 60 (2024) 102446. https://doi.org/10.1016/J.AEI.2024.102446.
[21] D. Shamsollahi, O. Moselhi, K. Khorasani, Data integration using deep learning and real-time locating system (RTLS) for automated construction progress monitoring and reporting, Autom. Constr. 168 (2024) 105778. https://doi.org/10.1016/J.AUTCON.2024.105778.
[22] V.K. Bansal, M. Pal, Construction Projects Scheduling Using GIS Tools, International Journal of Construction Management 11 (2011) 1–18. https://doi.org/10.1080/15623599.2011.10773158.
[23] A.H. Dalir, Z. Pezeshki, M. Ravanshadnia, E. Krinitsky, I.A. Sultanguzin, Automatic Monitoring in Construction Projects: Scientometric Analysis and Visualization of Research Activities, Human-Centric Intelligent Systems 2025 5:1 5 (2025) 21–43. https://doi.org/10.1007/S44230-025-00089-3.
[24] G. Wang, Y. Zhou, D. Cao, Artificial intelligence in construction: Topic-based technology mapping based on patent data, Autom. Constr. 172 (2025) 106073. https://doi.org/10.1016/J.AUTCON.2025.106073.
[25] T.G. Mai, M. Nguyen, A. Ghobakhlou, W.Q. Yan, B. Chhun, H. Nguyen, Decoding a decade: The evolution of artificial intelligence in security, communication, and maintenance within the construction industry, Autom. Constr. 165 (2024) 105522. https://doi.org/10.1016/J.AUTCON.2024.105522.

Multi-fidelity Surrogate-integrated Modelling Framework: Application to Tall Energy Structures

Wuhua Xie[1] Ke Ke[1,*] Guohao Wang[3] and Honghu Chu[2,3]

[1]School of Civil Engineering, Chongqing University, China
[2]College of Civil Engineering and Architecture, Zhejiang University of Water Resources and Electric Power, China
[3]Bartlett School of Sustainable Construction, University College London, United Kingdom
wuhuaxie1995@gmail.com, ke.ke@cqu.edu.cn, ucbvanb@ucl.ac.uk, chuhh@zjweu.edu.cn

Abstract –

Tall renewable energy infrastructures are complex structural systems comprising numerous components. Under earthquakes, their designated energy-dissipating components may exhibit intricate material and geometric nonlinearities that require high-fidelity finite element (FE) modelling, whereas system-level seismic assessment typically demands extensive nonlinear time-history analyses, making uniformly refined modelling computationally prohibitive. This paper presents a multi-fidelity surrogate-integrated modelling framework that concentrates refined FE simulations on the critical component to construct a surrogate model, which is then integrated into an efficient system-level model with simplified representations for the remaining components. The surrogate was developed using a multilayer perceptron (MLP) trained to jointly predict a vector of hysteretic parameters, enabling shared representation learning to capture inter-parameter correlations and thereby improving predictive accuracy and output consistency. The framework was demonstrated on a representative 10-MW resilient jacket-type offshore wind turbine, in which an hourglass-like stiffened tower segment was characterised using a comprehensive hysteresis database – derived from 3,000 refined FE simulations – and integrated into an OpenSees-based system model. The results show that the joint-prediction MLP achieved higher predictive accuracy than independently trained single-output models, while enabling efficient large-scale nonlinear dynamic analyses.

Keywords –

Energy infrastructures; seismic assessment; surrogate; multi-fidelity modelling; hysteretic behaviour

1 Introduction

Over the past decade, renewable energy has expanded rapidly worldwide, driving the deployment of tall energy infrastructures such as wind turbine towers [1] and concentrated solar power towers [2]. The expansion of these installations into seismically active regions poses significant seismic risks to the infrastructures, owing to their highly slender and cantilever-dominated structural characteristics [3]. Such slender systems can be particularly sensitive to dynamic excitations, making it challenging to sustain functionality and structural integrity under seismic hazard demands [4].

To enhance the structural performance of tall energy infrastructures, several novel structural systems equipped with energy dissipation devices, such as tuned mass dampers [5, 6], tuned liquid dampers [7, 8], and ductile tower segments [9, 10], have been developed. Their seismic performance assessment is particularly challenging because it often requires extensive nonlinear time-history analyses (NLTHAs) to capture seismic response, damage evolution, and failure modes. However, modelling such systems inevitably involves a trade-off between computational accuracy and efficiency [11]. Applying uniformly high-fidelity modelling to the entire structural system becomes prohibitive when hundreds or thousands of NLTHAs are required for record-to-record variability or parametric studies [12]. For example, a prior study on the 10-MW jacket-type offshore wind turbine (JOWT) reported that a refined finite element (FE) model required 36 hours to compute a single case on a 24-core workstation, even without accounting for material nonlinearity [13]. In contrast, simplified beam-based models offer efficiency but may inadequately represent the critical components that exhibit strong geometric and material nonlinearities.

Recent advances in machine learning (ML) have provided a promising solution to address this trade-off by developing data-driven surrogate models that emulate high-fidelity nonlinear behaviour at a fraction of the

computational cost [14-18]. This paper proposes a multi-fidelity surrogate-integrated modelling framework to facilitate the seismic performance assessment of tall energy infrastructures. The key idea is to concentrate refined FE simulations on the critical component that dominates nonlinear deformation and energy dissipation, to construct a hysteresis database, and to train a surrogate model that maps structural design variables to a hysteretic representation. The trained surrogate model is then integrated into an efficient system-level model in which the remaining components are represented using computationally cheaper beam elements, without compromising the macroscopic fidelity of the nonlinear structural behaviour. This framework is demonstrated using a representative tall energy structure (a 10-MW resilient JOWT), where an hourglass-like stiffened tower segment (HSTS) is characterised using a hysteresis database derived from 3,000 refined FE simulations and integrated into an OpenSees system model.

A further emphasis of this study concerns the surrogate formulation. Instead of training multiple independent single-output regressors [15-18], this study adopts a single multilayer perceptron (MLP) trained to jointly predict a vector of hysteretic parameters. Multiple independently trained single-output models may yield outputs that are mutually incompatible and can violate the physical constraints implied by the actual structural behaviour. In contrast, the multi-output MLP leverages shared latent representations, allowing the model to learn cross-parameter dependencies, inherently constraining its outputs to a jointly feasible region, thereby promoting physical consistency among the outputs.

2 Methodology

This study develops a multi-fidelity surrogate-integrated modelling framework that combines refined FE simulations with data-driven surrogate modelling and an efficient system-level model. The framework targets the critical component (or component group) that dominates nonlinear deformation and energy dissipation, to which the accuracy of structural behaviour prediction is highly sensitive, while representing the remaining components using computationally cheaper modelling techniques suitable for system-level dynamics. Figure 1 shows the stepwise workflow of the proposed framework:

(1) Parameterised high-fidelity modelling of the critical component. A refined FE model is formulated and parameterised to support automated generation of component variants over a prescribed design space.

(2) Design-space sampling and database generation. A structural parameter matrix is constructed to span the practical design space, and cyclic/monotonic analyses are conducted on refined FE models to generate a hysteresis database representing the hysteretic behaviour.

(3) Hysteresis parameterisation. The simulated hysteretic curves are mapped to a phenomenological parameter set (e.g., stiffness, strength, ductility, hardening, and deterioration parameters) through calibration, yielding paired datasets of structural and hysteretic parameters.

(4) Surrogate model training with joint prediction. An MLP is trained to infer the hysteretic-parameter vector directly from the structural parameters by learning a shared latent representation and predicting all hysteretic parameters jointly.

(5) Reduced-order representation and system-level integration. The trained surrogate is assigned to a parameterised reduced-order representation of the critical component (e.g., concentrated plasticity formulation), which is then integrated into an efficient system-level model. All other components are modelled using simplified yet reliable elements that preserve global dynamics and essential interaction effects.

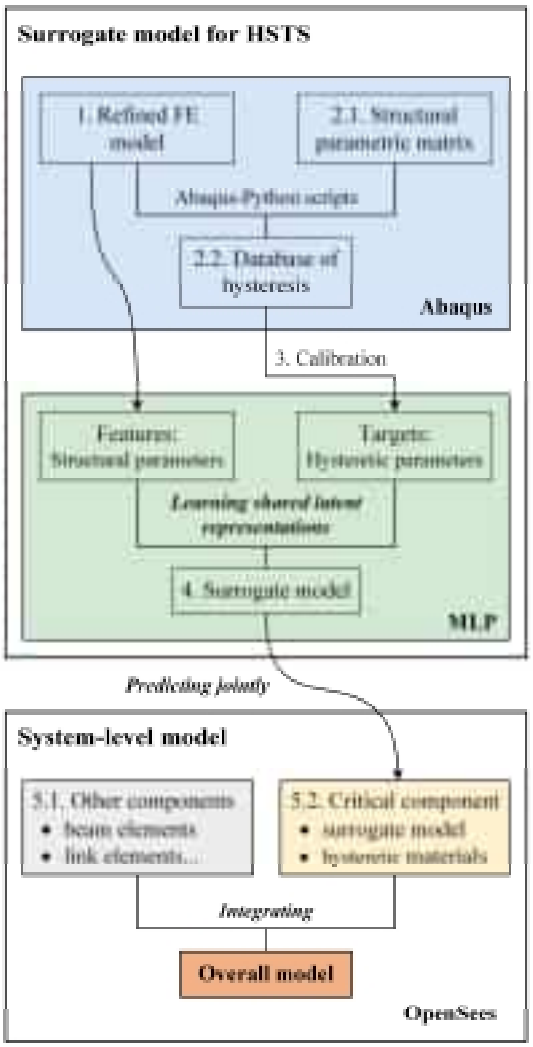


Figure 1. Multi-fidelity surrogate-integrated modelling framework

This multi-fidelity integration inevitably trades local micro-scale detail for system-level tractability. Nevertheless, by concentrating high-fidelity effort on the critical nonlinear component and ensuring that its macroscopic hysteretic behaviour is accurately reproduced via a joint-prediction surrogate, the framework provides an effective balance between fidelity and efficiency for seismic performance assessment of tall energy infrastructures.

3 Framework implementation on a 10-MW resilient JOWT

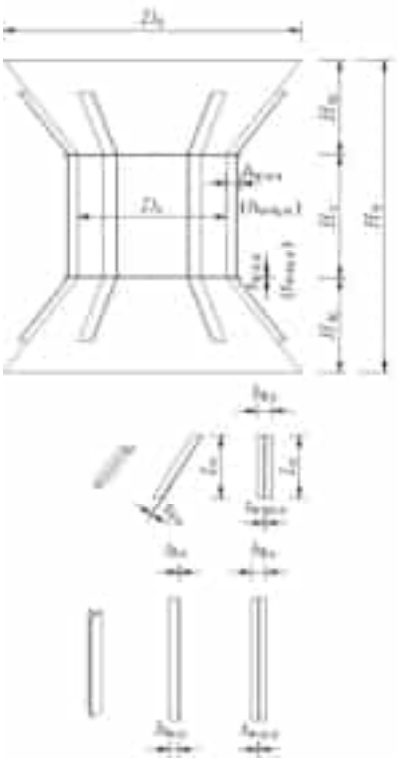
3.1 Reference structure

The multi-fidelity surrogate-integrated modelling framework is demonstrated using a representative 10-MW resilient JOWT proposed by Xie et al. [18]. This novel structural system introduces a HSTS – as the energy dissipation device – in lieu of the conventional cylindrical tower segment connecting the upper tower and transition piece of JOWTs. The HSTS is a novel component featuring complex thin-walled and stiffened configurations and necessitates a refined FE analysis to capture its initial imperfections, intricate material and geometric nonlinearities. In the following, a surrogate model is established for the HSTS and subsequently integrated into the system-level model, with the remaining components, such as the tower, rotor-nacelle assembly, and jacket foundation, modelled using computationally cheaper elements.

3.2 Automated FE modelling and hysteresis dataset generation

The refined FE models of the HSTS were established using Abaqus. Figure 2 provides further details regarding the basic geometry and structural parameters of the HSTS. This model adopted a simplified piecewise-linear material model, which is characterised by the following: (1) an elastic modulus of 210 GPa, (2) a yield stress of 355 MPa, and (3) a 1% kinematic linear strain hardening up to an ultimate stress of 470 MPa. Furthermore, the first-order 4-noded doubly curved rectangle (S4R) element was used.

Next, a comprehensive hysteresis database was generated. The database contains 3000 moment-rotation (M-θ) hysteretic curves obtained from refined FE simulations, covering a broad range of feasible HSTS design configurations. A structural parameter matrix was first constructed by Latin hypercube sampling within prescribed ranges selected to remain consistent with relevant design standards for stiffened-shell components. A set of Abaqus-Python scripts was then developed to enable automated processing, including:



	Symbol	Description	Value range for the design space
Shells	D_0	Diameter of the original tower segment	7465 (mm)
	H_0	Height of the original tower segment	8000 (mm)
	D_t	Reduced diameter	3000–5200 (mm)
	H_c	Height of the cylindrical shell	2400–4000 (mm)
	t_c	Thickness of the cylindrical shell	28–40 (mm)
	t_t	Thickness of the truncated conical shells	28–40 (mm)
	H_t	Height of the truncated conical shells	$(H_0 - H_c) / 2$ (mm)
Longitudinal stiffeners	N_s	Number of stiffeners	4, 5, 6, 7, 8
	$h_{w,s}$	Width of the web	150–300 (mm)
	$t_{w,s,1}$	Thickness of the web at the middle	20–40 (mm)
	$t_{w,s,2}$	Thickness of the web at the two ends	10–40 (mm)
	$b_{f,s}$	Width of the flange	100–200 (mm)
	$t_{f,s}$	Thickness of the flange	10–40 (mm)
	L_s	Height of the stiffener at the two ends	$(1/3 \sim 2/3) \times H_c$ (mm)
Ring stiffeners	$t_{w,r,1}$	Thickness of the web at the middle	20–40 (mm)
	$h_{w,r,2}$	Width of the web at the two ends	150–300 (mm)
	$t_{w,r,2}$	Thickness of the web at the two ends	20–40 (mm)
	$h_{w,r,1}$	Width of the web at the middle	$h_{w,r}$ + 100 (mm)

Figure 2. Basic geometry and structural parameters of the HSTS

(1) automatic generation of parameterised HSTS FE models for all sampled designs; (2) FE analyses under monotonic and cyclic loading histories consistent with FEMA 461 standard component-testing protocols; (3) post-processing of simulation results to extract the M-θ hysteretic curves; and (4) calibration of each hysteretic curve to a phenomenological representation using the modified Ibarra-Medina-Krawinkler (mIMK) model [19] parameters (Figure 3), including K_e, θ_p, θ_{pc}, M_y, M_c, and Λ.

Figure 4 summarises the calibrated mIMK parameters with their maximum, mean, minimum, and standard deviation (SD) values. Figure 5 presents representative hysteretic curves, comparing the simulations with their mIMK calibrations. The results demonstrate a high degree of accuracy.

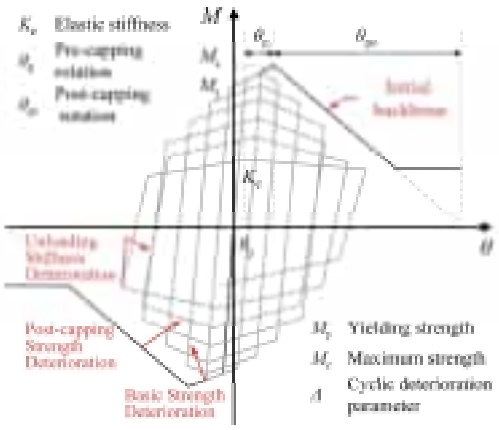


Figure 3. Illustration of the mIMK model

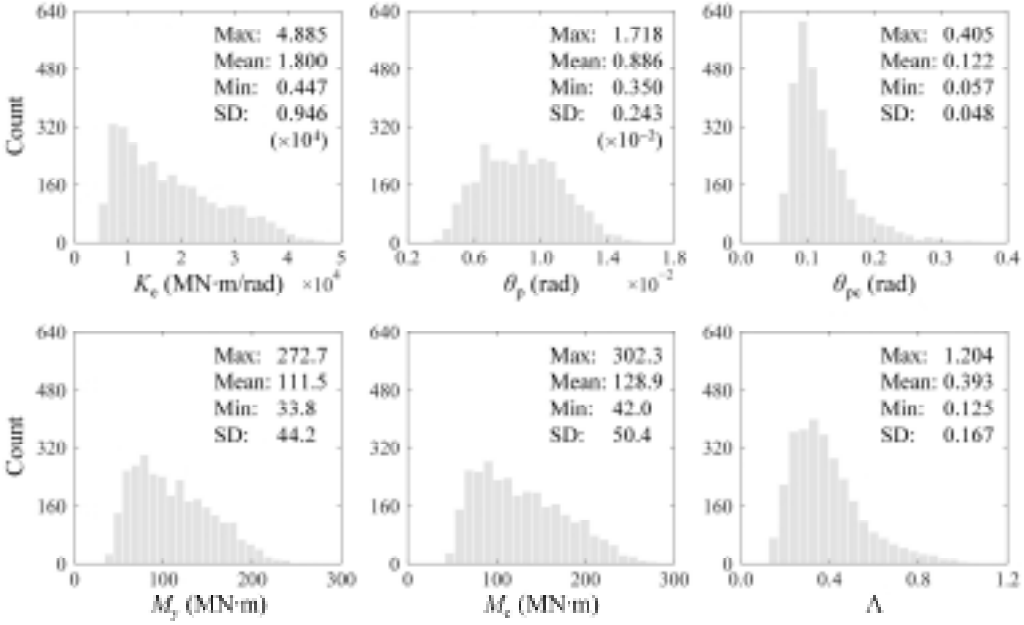


Figure 4. Calibrated mIMK parameters with the maximum, mean, minimum, and standard deviation (SD) values

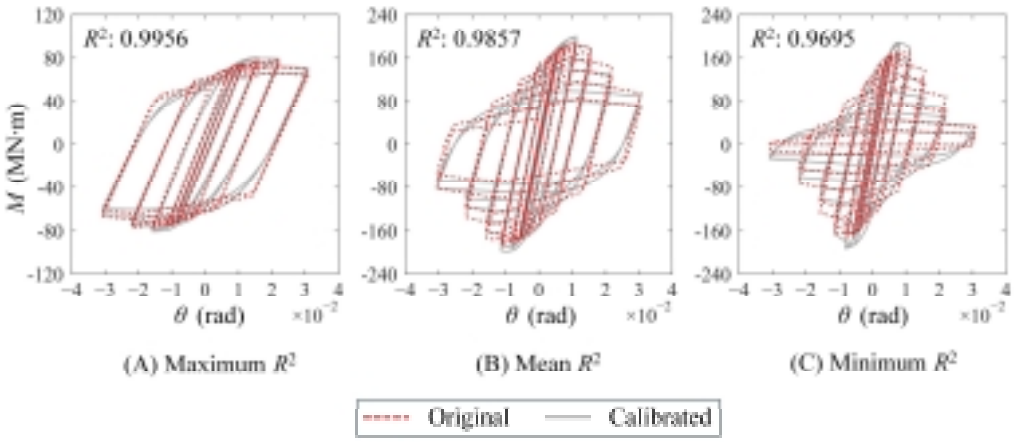


Figure 5. Representative results of hysteresis calibration

3.3 MLP surrogate model

An MLP is a feedforward neural network that learns a nonlinear mapping from input features to output targets through fully connected layers. In this study, the surrogate model is formulated as a single MLP with a vector-valued output, enabling the joint prediction of a vector of hysteretic parameters. The MLP model utilised the ReLU activation function and was trained using the Adam optimiser. The objective was to minimise the mean squared error (MSE) loss function. Training was terminated if the validation score did not improve by more than 10^{-4} for 50 consecutive iterations.

The dataset was first split into an 80/20 train/test partition (test set used only for the final evaluation). Within the training set, Optuna-based hyperparameter tuning used 5-fold cross-validation. For each fold, both inputs and outputs were standardised, and fitted on the fold training subset only to avoid data leakage. Table 1 summarises the hyperparameter search space for Optuna.

Table 1. MLP hyperparameter search space

Hyperparameter	Search space	Best
Number of hidden layers	[1, 3]	2
Units in layer *i*	[32, 320]	314/100
L2 regularisation	$[10^{-3}, 5\times10^{-2}]$	0.03634
Batch size	{32, 64, 128, 256}	32
Initial learning rate	$[10^{-3}, 5\times10^{-3}]$	0.00386

The Optuna objective maximised the mean cross-validated weighted coefficient of determination (R^2), computed as a weighted sum of per-output R^2 values. The weights for [K_e, θ_p, θ_{pc}, M_y, M_c, Λ] were set to [0.10, 0.25, 0.20, 0.10, 0.10, 0.25], following preliminary trials that indicated comparatively lower fit quality for θ_p, θ_{pc}, and Λ. The best hyperparameters were selected based on the cross-validated weighted R^2. Next, the final MLP model was trained on the full training set using the selected hyperparameters. Prediction performance was then reported on the held-out test set using per-output R^2.

Figure 6 demonstrates a good agreement between the predicted and calibrated data, confirming high performance. Table 2 further compares the R^2 and mean absolute percentage error (MAPE) achieved by the MLP model with those reported in a prior study [18] using six independently trained XGBoost models, demonstrating consistently higher accuracy for the MLP model – most notably for θ_p, θ_{pc}, and Λ.

Table 2. Prediction performance of the MLP model

Targets	MLP R^2	XGBoost R^2 [18]	MLP MAPE
K_e	0.9962	0.9966	2.90%
θ_p	0.9410	0.9264	5.34%
θ_{pc}	0.9666	0.9521	4.24%
M_y	0.9945	0.9929	2.38%
M_c	0.9954	0.9949	2.17%
Λ	0.9613	0.9439	5.83%

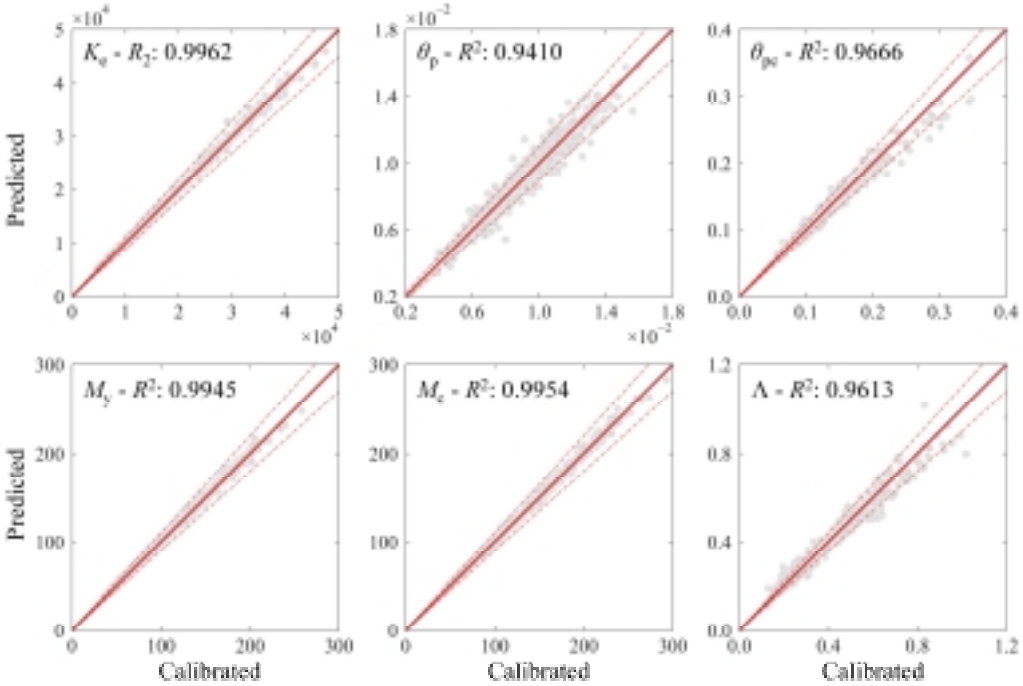


Figure 6. Comparison between the predicted and calibrated data

3.4 System-level integration

The system-level FE model was developed in OpenSees (Figure 7). The tower, blades, jacket members, and piles were represented using computationally cheaper beam elements. Following the well-established concentrated-plasticity approach, the HSTS was integrated into the system-level model using a coupled zero-length element assigned an mIMK uniaxial material whose parameters were provided by the MLP surrogate model, thereby forming an equivalent plastic-hinge representation.

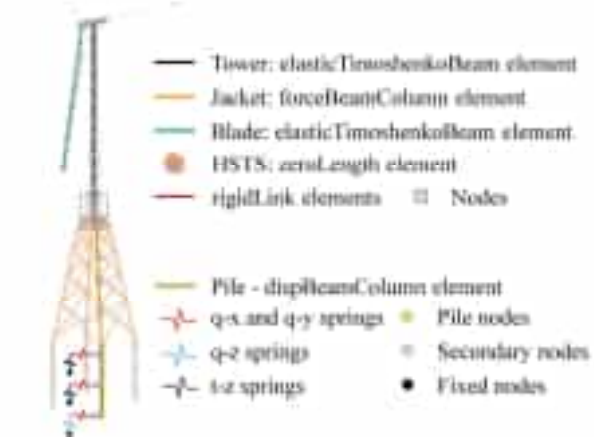


Figure 7. System-level FE model in OpenSees

4 Validation

4.1 Elastic structural dynamics

To validate the elastic structural dynamics of the developed surrogate-integrated FE model, a modal analysis was performed to obtain its periods of vibration in the fore-aft (FA) and side-side (SS) directions. Table 3 compares the periods of vibration of the FE model with the corresponding reference values for the prototype reported by von Borstel [20]. The differences are minor and the agreement is satisfactory, indicating that the proposed model captures the primary structural dynamics of the prototype with adequate fidelity.

Table 3. Periods of vibration of the surrogate-integrated FE model

Mode	Prototype (s)	FE model (s)
1st tower SS	3.49	3.52
1st tower FA	3.47	3.47
2nd tower SS	0.91	1.01
2nd tower FA	0.90	0.91

4.2 Hysteretic behaviour

The hysteretic response was finally reconstructed by feeding the predicted mIMK parameters and prescribed loading protocol into the HSTS FE model. Next, the reconstructed response was assessed against the corresponding FE-simulated response using the R^2 metric. Table 4 summarises the R^2 statistics over all 3000 samples, showing consistently accurate predictions with a mean R^2 value exceeding 0.96. In addition, approximately 3% of samples exhibit an R^2 below 0.90. Future research will prioritise the development of robust detection methods to ensure physical realism of hysteretic behaviour.

Table 4. Prediction performance of hysteretic behaviour

R^2	Train	Test	Overall
Mean	0.9649	0.9635	0.9646
Standard deviation	0.0284	0.0286	0.0284
Minimum	0.8538	0.8640	0.8539

4.3 Seismic behaviour

To validate the performance of the surrogate-integrated FE model regarding seismic assessment, a nonlinear time-history analysis (NLTHA) was conducted under the 1978 Tabas (Boshrooyeh) ground motion. Figure 8 compares tower-top displacements between the Abaqus model, comprising 19200 S4R and 1950 B31 elements, and the surrogate-integrated OpenSees model. The results demonstrate excellent agreement, with peak response discrepancies of around 8.6%.

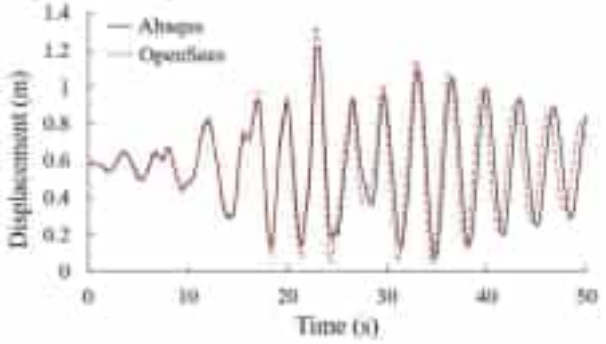


Figure 8. Seismic responses of the Abaqus model and the OpenSees model

Furthermore, to quantify efficiency gains, Table 5 summarises wall-clock times for representative analyses performed on an AMD Ryzen 9950X workstation. The surrogate model reduced hysteresis generation time from 16 minutes (Abaqus, 12-core configuration) to just 35 milliseconds. Moreover, the NLTHA was completed in 12.6 seconds via the OpenSees model, compared to 45 minutes for the Abaqus model, demonstrating its high computational efficiency for seismic assessment.

Table 5. Wall-clock times for representative analyses

Analysis	Abaqus	OpenSees
Hysteresis generation	16 min	35 ms
NLTHA	45 min	12.6 s

Overall, the proposed framework achieves a superior balance between predictive fidelity and computational efficiency. Beyond these performance gains, its modular architecture facilitates the integration of alternative ML algorithms, providing a versatile foundation for future comparative research into model performance.

5 Conclusions

This study presented a multi-fidelity surrogate-integrated modelling framework to enable efficient seismic assessment of tall energy infrastructures whose responses are governed by critical energy dissipation components.

The framework concentrates refined FE simulations on the critical component to generate a hysteresis database and trains a joint-prediction MLP surrogate to capture inter-parameter dependencies, thereby improving output consistency.

The approach was demonstrated on a representative 10-MW resilient JOWT, where an HSTS was characterised using a hysteresis database derived from 3,000 refined FE simulations and integrated into an OpenSees system-level model via a reduced-order plastic-hinge representation. The joint-prediction MLP model demonstrated consistently higher accuracy relative to independently trained XGBoost models – most notably for θ_p, θ_{pc}, and Λ. Validation results showed good agreement in modal properties and accurate hysteresis reconstruction (mean $R^2 > 0.96$), supporting the fidelity and efficiency of the proposed framework.

Acknowledgments

This work was financially supported by the National Key R&D Program of China (2025YFF0519700), the National Natural Science Foundation of China (Grant No. 52322802), the Zhejiang Provincial Natural Science Foundation (Grant No. LQN26E080033), and the Hangzhou Municipal Natural Science Foundation (Grant No. 2025SZRJJ1193).

References

[1] GWEC. Global Wind Report 2025. The Global Wind Energy Council, Lisbon, Portugal, 2025.

[2] B. Hernández, N. L. Ferber, M. Abdullah, A. Mayyas, N. Calvet, and M. Chiesa, Central receiver-based concentrated solar power plants part 1: A historical review, *Sol. Energy* 299(2025) 113693.

[3] E. I. Katsanos, S. Thöns, and C. T. Georgakis, Wind turbines and seismic hazard: a state-of-the-art review, *Wind Energy* 19(2016) 2113–2133.

[4] N. Alati, G. Failla, and F. Arena, Seismic analysis of offshore wind turbines on bottom-fixed support structures, *Phil. Trans. R. Soc. A.* 373(2015) 20140086.

[5] H. Zuo, K. Bi, and H. Hao, Using multiple tuned mass dampers to control offshore wind turbine vibrations under multiple hazards, *Eng. Struct.* 141(2017) 303–315.

[6] M. Liu, S. Li, T. Wu, Y. Li, H. Meng, and Z. Chen, Eddy-Current Tuned Mass Dampers for Mitigation of Wind-Induced Response of the Noor III Solar Tower: Design, Installation, and Validation, *J. Struct. Eng.* 147(2021) 05021009.

[7] A. Hemmati, E. Oterkus, and N. Barltrop, Fragility reduction of offshore wind turbines using tuned liquid column dampers, *Soil Dyn. Earthquake Eng.* 125(2019) 105705.

[8] Y. Li, S. Li, B. Sun, M. Liu, and Z. Chen, Effectiveness of a tiny tuned liquid damper on mitigating wind-induced responses of cylindrical solar tower based on elastic wind tunnel tests, *J. Wind Eng. Ind. Aerodyn.* 208(2021) 104455.

[9] R. Rostami and A. Tombari, A novel reduced column section approach for the seismic protection of wind turbines, *Eng. Struct.* 282(2023) 115807.

[10] K. Ke, W. Xie, X. Zhou, Y. Wang, and X. He, A novel self-centring jacket-type offshore wind turbine structure: A proof-of-concept study and seismic fragility analysis, *Thin-Walled Struct.* 205(2024) 112407.

[11] T. Wang et al., From model-driven to data-driven: A review of hysteresis modeling in structural and mechanical systems, *Mech. Syst. Signal Process.* 204(2023) 110785.

[12] S. Chen, D. Feng, and E. Taciroglu, Prior knowledge‐infused neural network for efficient performance assessment of structures through few‐shot incremental learning, *Computer aided Civil Eng.* 39(2024) 1928–1945.

[13] Y. Yan, Y. Yang, M. Bashir, C. Li, and J. Wang, Dynamic analysis of 10 MW offshore wind turbines with different support structures subjected to earthquake loadings, *Renew. Energy* 193(2022) 758–777

[14] D. Samadian, I. B. Muhit, A. Occhipinti, and N. Dawood, Surrogate models for seismic and pushover response prediction of steel special moment resisting frames, *Eng. Struct.* 314(2024) 118307.

[15] G. Georgiou and A. Elkady, ANN-Based Model for Predicting the Nonlinear Response of Flush Endplate Connections, *J. Struct. Eng.* 150(2024) 04024034.
[16] C.-L. Ning, L. Wang, and W. Du, A practical approach to predict the hysteresis loop of reinforced concrete columns failing in different modes, *Constr. Build. Mater.* 218(2019) 644–656.
[17] T. A. Horton, I. Hajirasouliha, B. Davison, and Z. Ozdemir, Accurate prediction of cyclic hysteresis behaviour of RBS connections using Deep Learning Neural Networks, *Eng. Struct.* 247(2021) 113156.
[18] W. Xie, X. Zhou, K. Ke, M. C. H. Yam, and Y. Yan, A resilient jacket-type offshore wind turbine system: Conception, surrogate-integrated modelling, and seismic performance evaluation, *Ocean Eng.* 349(2026) 124123.
[19] D. G. Lignos and H. Krawinkler, Deterioration Modeling of Steel Components in Support of Collapse Prediction of Steel Moment Frames under Earthquake Loading, *J. Struct. Eng.* 137(2011) 1291–1302.
[20] T. von Borstel, Design Report Reference Jacket (No. D4.3.1). INNWIND.EU, 2013.

Automated Scan-to-BIM Method for City-Scale Building Roof Reconstruction Using UAV Photogrammetry and Geospatial Building Footprints

Boan Tao[1], Frédéric Bosché [1] and Jiajun Li[1]

[1]School of Engineering, University of Edinburgh, Edinburgh, Scotland, UK

boan.tao@ed.ac.uk, f.bosche@ed.ac.uk, Jiajun.Li@ed.ac.uk

Abstract -

The digital transformation of existing building stock through Building Information Modelling (BIM) remains a significant goal and challenge in the Architecture, Engineering, and Construction (AEC) industry. This paper presents an automated Scan-to-BIM framework specifically designed for generating IFC-compliant roof model of buildings within a large urban environment from Unmanned Aerial Vehicle (UAV) photogrammetry data. The proposed method leverages geospatial building footprint data as semantic priors to intelligently filter the input large-scale 3D mesh and model the roofs of all individual buildings. A robust segmentation pipeline combining Taubin smoothing, region-growing algorithms, and adaptive height filtering successfully extracts planar roof surfaces. The extracted geometries are reconstructed into Industry Foundation Classes (IFC) compliant BIM objects using alpha-shape boundary regularisation, which preserves non-convex features such as conical spires and L-shaped roofs. Experimental validation is first conducted on a large building with a complex roof structure composed of 431 extracted roof patches, for which a patch-level correctness precision of 98.8% was achieved. The framework is subsequently demonstrated at the scale of a university campus to assess scalability and robustness.

Keywords -

Scan-to-BIM, Semantic Segmentation, Industry Foundation Classes (IFC), Digital Twin, Automated Roof Extraction, Geospatial Data Fusion.

1 Introduction

1.1 Background

A substantial proportion of the existing building stock lacks accurate and up-to-date digital representations. While new construction projects increasingly benefit from *born-digital* Building Information Models (BIM), many existing buildings—particularly those constructed prior to the widespread adoption of BIM—remain poorly documented in structured digital formats. This digital deficit limits the effectiveness of facility management, maintenance, retrofitting, and long-term asset monitoring.

Roofs constitute critical building elements that require regular inspection due to their continuous exposure to environmental stressors. Conventional inspection methods are labour-intensive, costly, and often pose safety risks. In recent years, Unmanned Aerial Vehicles (UAVs) equipped with high-resolution cameras have emerged as a practical alternative, enabling rapid and safe roof inspection at scale. Contemporary photogrammetry pipelines can transform UAV imagery into dense three-dimensional representations with centimetre-level geometric accuracy.

Despite these advances, a persistent *semantic gap* remains between raw geometric reconstructions—such as images, point clouds, and meshes—and structured, semantically-rich building models suitable for BIM-based workflows. Photogrammetric outputs are inherently unstructured and lack explicit semantic information, making it difficult to distinguish architectural elements such as roofs from surrounding terrain, vegetation, or urban clutter. As a result, manual interpretation and modelling remain common practice, leading to significant time and cost overheads.

Although extensive research has addressed large-scale processing and semantic labelling of aerial LiDAR and photogrammetric data, many Scan-to-BIM workflows still focus on individual processing stages, such as surface reconstruction or classification, rather than enabling end-to-end automated generation of IFC-compliant building components. In particular, approaches capable of robustly extracting roof geometry from dense photogrammetric meshes and directly translating it into interoperable BIM representations at scale remain limited. These challenges are further amplified by the size and geometric complexity of high-resolution UAV-derived datasets, underscoring the need for scalable and fully automated methods that bridge the gap between unstructured 3D data and structured digital building models. Importantly, the limitation is not the absence of individual algorithms, but the lack of an integrated and scalable framework that can bridge unstructured 3D data and structured smenatically-rich (BIM) representations in a fully automated manner.

2 Related Work

2.1 UAV Photogrammetry for Building Reconstruction

Unmanned Aerial Vehicle (UAV) technology is revolutionising building surveying and 3D reconstruction work-

flows. Besides, Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithms first enabled fully automated 3D reconstruction pipelines from UAV imagery. Recently, depth-supervised neural radiation fields have also been growing rapidly to produce navigable 3D reconstructions from images [1]. Despite these advances, the resulting reconstructions—whether expressed as dense polygonal meshes, point-based representations, or implicit radiance fields—remain largely unstructured and lack explicit semantic information, limiting their direct integration into BIM-oriented workflows.

2.2 Roof Extraction from Aerial 3D Data

When processing aerial 3D data (3D LiDAR or photogrammetrically-reconstruction point clouds and/or meshes) of built environments, region growing is a commonly adopted strategy for roof plane extraction, where planar surfaces are incrementally expanded from initial seed regions according to geometric similarity criteria. To overcome the sensitivity of traditional region-growing approaches to seed selection and boundary leakage, Li et al. [2] introduced a hierarchical roof segmentation framework for airborne LiDAR data that combines clustering-based initialisation with boundary relabelling. Their method refines roof plane boundaries through multi-level segmentation, improving robustness in separating adjacent roof facets and handling complex roof structures.

Chen et al. [3] proposed a contour extraction method for UAV-derived point clouds that combines multi-level region-growing plane segmentation and plane fusion, followed by neighbourhood geometric feature analysis on the fused planes to robustly extract both roof and façade (wall) contours, including sparse structural boundaries.

Overall, existing plane segmentation strategies primarily rely on iterative grouping mechanisms—such as region growing, RANSAC-based fitting, or clustering—to aggregate coplanar points based on local geometric similarity. While these approaches are effective for extracting dominant planar structures, their performance is often sensitive to parameter selection, point density variations, and surface completeness, which can lead to boundary leakage or over-segmentation in complex roof and façade geometries.

2.3 Geospatial Data Fusion and Building Footprint Utilisation

The integration of geospatial building footprint data with 3D point clouds provides an effective means for spatial filtering and semantic guidance in large-scale urban reconstruction tasks. While numerous studies focus on extracting building footprints from remote sensing data, many practical applications——particularly in urban environments——benefit from the availability of pre-existing footprint datasets derived from cadastral maps or open geospatial repositories. Previous studies have demonstrated that integrating aerial imagery with LiDAR-derived elevation information can significantly improve building footprint delineation in dense urban environments [4, 5].

In contrast to these extraction-oriented approaches, the present work aims to leverage readily-available accurate geospatial footprint data directly as semantic priors to process new aerial 3D data (UAV photogrammetry). By using footprints to constrain spatial queries and filter large-scale UAV-derived meshes, the proposed framework efficiently isolates each individual building and corresponding roof regions, providing a reliable foundation for subsequent segmentation and IFC-compliant BIM reconstruction. The benefits are: (1) reduced computational demand (building-by-building processing reduces computational overload and removes the need for merging processes in tiling-based processing strategies); (2) enhanced quality where roofs are not searched in areas outside building footprints, reducing the risk of false positives.

2.4 Scan-to-BIM Methodologies

Automated Scan-to-BIM pipelines have been widely investigated, particularly for converting terrestrial laser scanning or indoor point cloud data into IFC-compliant BIM models. For example, Cloud2BIM [6] introduced a fully automated pipeline for reconstructing building elements from large-scale point clouds, while BridgeTwin [7] proposed a semi-automated workflow tailored to bridge infrastructure. They are predominantly designed for ground-based acquisition scenarios and structured geometries, relying on interior layouts, predefined templates, or significant user interaction. As such, they are not directly applicable to UAV-derived photogrammetric meshes.

2.5 Research Gap and Contribution

While significant progress has been made in individual elements of the Scan-to-BIM pipeline, several critical gaps remain:

1. **Geometric Complexity**: Many existing roof modelling and Scan-to-BIM approaches operate on sparse or abstracted representations, such as planar segments derived from aerial LiDAR or simplified point clouds. As a result, these methods typically favour simplified parametric roof models (e.g., gable or hip roofs), limiting their ability to represent complex roof morphologies, including conical spires, L-shaped roofs, and non-convex boundaries. Leveraging dense UAV-derived photogrammetric meshes offers the potential to capture roof geometry at a higher level of detail, but scalable methods capable of exploiting this richness while remaining fully automated remain limited.

2. **BIM-compliant Output**: Few automated pipelines generate IFC-compliant models preserving non-convex geometries suitable for downstream digital twin applications, with most focusing on simple primitives.
3. **End-to-End Automation**: Most workflows require manual intervention for quality control, parameter tuning, or geometric editing, preventing deployment at scale across large building portfolios.

This work explores a fully automated Scan-to-BIM framework for large-scale reconstruction of building roofs from UAV photogrammetry. Its main contribution is the use of geospatial building footprints as semantic priors to guide instance-level extraction of roof structures from dense photogrammetric meshes. This footprint-guided design improves the efficiency and robustness of roof reconstruction in complex urban environments and enables the automatic generation of IFC-compliant roof models, including non-convex geometries, without manual intervention.

3 Method

The proposed framework is designed as an end-to-end automated Scan-to-BIM pipeline that processes dense UAV-derived photogrammetric meshes together with 2D geospatial building footprints as semantic priors, in order to directly generate IFC-compliant roof models without manual intervention.

By leveraging geospatial footprint data to guide building-level segmentation, the method enables efficient and scalable processing of large-scale urban scenes, bridging the gap between unstructured 3D reconstructions and structured BIM representations.

The method works in three stages:

1. Building Photogrammetric Data Segmentation,
2. Planar Primitive Segmentation, and
3. Automated IFC Reconstruction.

3.1 Building Photogrammetric Data Segmentation

The proposed framework operates on two heterogeneous but spatially consistent data sources: a dense 3D photogrammetric mesh generated from UAV imagery and a set of 2D geospatial building footprints. Both datasets are georeferenced within a common projected coordinate system, as UAV photogrammetry workflows typically employ RTK-GNSS positioning and Ground Control Points (GCPs) to ensure accurate global geolocation.

Given this shared spatial reference, no additional geometric alignment between the photogrammetric mesh and the footprint data is required. Instead, the building footprints are directly exploited as semantic priors to spatially filter the mesh, enabling the isolation of roof-level geometry corresponding to known building extents while discarding unrelated scene elements such as terrain, vegetation, and surrounding infrastructure.

The input photogrammetric mesh is denoted as

$$M = (V, F),$$

where $V = \{v_1, v_2, \ldots, v_n\}$ represents the set of mesh vertices and $F = \{t_1, t_2, \ldots, t_m\}$ denotes the set of triangular faces. Each vertex $v_i \in V$ is defined in a Cartesian coordinate system consistent with the georeferenced UAV reconstruction. Each face $t_i \in F$ is associated with a surface normal vector $\vec{n}_i$ and a local adjacency neighbourhood $\mathcal{N}(i)$ defined by shared edges.

The building footprint dataset consists of a set of polygonal geometries

$$P = \{P_b\},$$

where each polygon P_b represents the known footprint of a building of interest b and is defined in a global geodetic reference frame, reprojected as necessary to ensure coordinate consistency with the mesh.

For each building b, its corresponding footprint polygon P_b is used to segment the overall UAV-derived mesh and extract the subset of geometry associated with that building. A mesh vertex $v_i \in M$ is retained if and only if its orthogonal projection onto the horizontal plane lies within the footprint polygon P_b:

$$M_b = \left\{ v_i \in M \;\middle|\; \mathrm{Contains}\big(P_b, v_i^{xy}\big) \right\}.$$

This footprint-guided cropping step removes non-building geometry at an early stage, substantially reducing data volume and computational complexity prior to subsequent geometric segmentation. By constraining processing to semantically relevant regions, the approach improves robustness and scalability when applied to large-scale UAV photogrammetry datasets.

3.2 Planar Primitive Segmentation

For each building of interest b, individual roof panels are extracted from the cropped mesh M_b using a multi-stage segmentation pipeline designed to be robust to noise, irregular sampling, and complex roof morphologies.

3.2.1 Geometric Smoothing and Decimation

UAV photogrammetric meshes typically contain high-frequency noise that can adversely affect surface normal estimation and segmentation stability. To suppress high-frequency noise in the photogrammetric mesh, Taubin smoothing [8], a non-shrinking signal-processing–based filtering method, is applied ($\lambda = 0.5$, $\mu = -0.53$, $N = 20$).

The selected parameters were chosen to suppress high-frequency noise while maintaining sharp roof edges and planar integrity. The smoothing strength (λ) controls noise attenuation, while the negative scaling factor (μ) compensates for contraction effects introduced by the Laplacian operator. Performing $N = 20$ iterations was empirically found to provide sufficient noise reduction for stable surface normal estimation without over-smoothing architectural features critical for subsequent planar segmentation.

Subsequently, mesh simplification is applied using Quadric Error Metric (QEM) decimation to reduce geometric complexity while preserving roof morphology. Rather than enforcing a fixed target triangle count, the level of decimation is adaptively determined based on the footprint area of each building. Specifically, the target triangle budget N_b for building b is defined proportionally to its footprint area A_b as

$$N_b = \lceil \alpha \cdot A_b \rceil ,$$

where α is a global density parameter controlling the number of retained triangles per unit area. This adaptive allocation ensures that larger buildings retain proportionally higher geometric resolution than smaller ones, while maintaining consistent surface fidelity across the dataset.

QEM decimation proceeds by iteratively collapsing edges while minimising the squared geometric error between the original and simplified surfaces, thereby preserving salient geometric features such as planar roof facets and sharp ridgelines [9]. By reducing mesh density in a controlled manner, this adaptive simplification significantly improves computational efficiency in subsequent processing stages, particularly during surface normal estimation and region-growing segmentation, without compromising the geometric fidelity required for accurate roof plane extraction.

3.2.2 Region Growing Segmentation

Planar roof segments are identified using a region-growing algorithm driven by surface normal consistency. Starting from an initial seed triangle t_i, a region is iteratively expanded to adjacent triangles $\{t_j \mid j \in \mathcal{N}(i)\}$ if the angular deviation between their surface normal vectors satisfies the following planarity criterion:

$$\arccos(\vec{n}_i \cdot \vec{n}_j) < \theta_{\text{th}},$$

where $\vec{n}_i$ and $\vec{n}_j$ denote the unit normal vectors of triangles t_i and t_j, respectively. The threshold θ_{th} is set to 10°, empirically selected to balance robustness to minor surface undulations against effective separation of adjacent roof facets with distinct orientations.

This process yields a set of planar roof regions

$$\mathcal{R}_b = \{\mathcal{R}_b^1, \mathcal{R}_b^2, \ldots, \mathcal{R}_b^K\},$$

where each region $\mathcal{R}_b^k$ consists of one or more connected triangles representing an individual roof panel of building b.

3.2.3 Adaptive Semantic Filtering via Height Analysis

To discriminate true roof surfaces from vertical façades, ground artifacts and secondary structures such as balconies, an adaptive height-based semantic filtering strategy is employed. Unlike fixed global height thresholds, which are unsuitable for scenes with varying terrain elevation, this approach dynamically adapts to the vertical extent of each individual building.

For each building b, a relative height threshold is defined as a fixed proportion of its vertical span:

$$z_b^{\text{th}} = z_b^{\max} - 0.3 \left(z_b^{\max} - z_b^{\min} \right),$$

where $z_b^{\max}$ and $z_b^{\min}$ denote the maximum and minimum elevations of all mesh vertices associated with building b, respectively.

Each segmented roof region $\mathcal{R}_b^k$ whose centroid elevation $z_c < z_b^{\text{th}}$ is classified as a non-roof element and discarded. This adaptive criterion ensures that only the primary roof structure is retained, even in the presence of multi-level buildings, terraces, or uneven terrain.

3.3 Automated IFC Reconstruction

The final stage converts the validated planar roof segments into georeferenced, IFC-compliant BIM entities without manual intervention.

3.3.1 Plane Estimation via Principal Component Analysis

For each retained roof patch P_b^k, an optimal planar representation is estimated using Principal Component Analysis (PCA). The covariance matrix of the patch vertices is computed, and the eigenvector corresponding to the smallest eigenvalue is selected as the surface normal $\vec{n}_k$. The patch vertices are subsequently projected onto the corresponding local 2D plane to facilitate robust boundary extraction.

3.3.2 Boundary Regularisation Using Alpha Shapes

To reconstruct complex and non-convex roof boundaries, such as L-shaped footprints or conical turrets, boundary regularisation is performed using the Alpha Shape algorithm. The algorithm operates on the Delaunay triangulation of the projected roof points, removing triangles whose circumradius exceeds a threshold α. This filtering yields polygonal boundaries that closely follow the true roof geometry while suppressing noise-induced artifacts [10].

In this work, the parameter α is set to 2.0 m, empirically selected to balance boundary smoothness and geometric fidelity. Smaller values preserve fine architectural details at the cost of boundary fragmentation, whereas larger values lead to over-smoothing and loss of concave features. The chosen value enables robust reconstruction of irregular roof boundaries while suppressing noise-induced artifacts, ensuring suitability for IFC-compliant BIM representation.

3.3.3 IFC Entity Instantiation and Georeferencing

The regularised 2D roof boundaries are reprojected into 3D space and instantiated as `IfcRoof` entities. Geometry is represented using `IfcShellBasedSurfaceModel`, which supports complex and non-manifold roof surfaces and ensures compatibility with downstream BIM applications.

Georeferencing information is inherited directly from the UAV-derived photogrammetric mesh, which is originally reconstructed in a projected coordinate system based on onboard GNSS and flight metadata. To preserve numerical precision, model geometry is maintained in a local Cartesian coordinate system, while global positioning is encoded explicitly using the `IfcMapConversion` entity. The associated `IfcProjectedCRS` defines the transformation to the global reference system, enabling consistent spatial alignment with GIS datasets and ensuring interoperability across BIM and geospatial platforms.

4 Experimental Results

The proposed framework is evaluated using UAV-derived photogrammetric data wit <2 cm GSD acquired over the University of Edinburgh Central Campus, a dense urban environment with diverse building typologies and complex roof geometries (fig. 1). To assess both accuracy and scalability, a two-stage evaluation strategy is adopted, consisting of detailed validation on a representative individual building followed by deployment across the full campus.

For single-building evaluation, the Edinburgh Futures Institute (EFI) is selected due to its complex roof topology (fig. 2a). The corresponding photogrammetric mesh comprises 11,772,857 vertices and 23,527,453 triangular faces, providing a high-resolution geometric representation suitable for detailed geometric assessment.

Then, to evaluate the performance of our method in large scale environments, the complete campus dataset (fig. 1) is processed. This dataset contains approximately 57 million vertices and 114 million triangular faces, covering an area of 0.476 km^2. The campus-scale evaluation includes multiple building types and roof configurations, enabling a robust assessment of scalability and consistency.



Figure 1. UAV-derived photogrammetric reconstruction of the University of Edinburgh Central Campus.

Table 1. Quantitative results summary.

Metric	Value
Roof Segmentation (section 4.1)	
Extracted patches	431
Precision	98.8%
Geometric Alignment (section 4.2.1)	
RMSE to BIM surfaces	0.84 m
Mean deviation	0.68 m
Min deviation	0.002 m
Max deviation	~2.0 m
Matched area (τ = 2.0 m)	99.0%
Object Consistency(section 4.2.2)	
Matched entities	270 / 292
Matching rate	92.5%
Mean consistency	0.66
Scalability(section 4.2.3)	
Mesh size	~57M V, ~114M F
Spatial extent	0.476 km^2

4.1 Roof Segmentation Results

The proposed pipeline successfully processed the raw photogrammetric mesh and extracted individual roof surfaces in a fully automated manner. For the EFI building, a total of 431 unique roof patches were identified across the study area. Patch-level precision is computed by manually checking whether each extracted patch corresponds to a valid roof surface.

Qualitative inspection further indicates that the segmentation achieves a clean separation of complex roof structures into distinct planar elements. Non-roof components, including ground terrain, facades, balconies and surrounding vegetation, were effectively removed during the semantic filtering stage, leaving only architecturally relevant roof surfaces.

As illustrated in fig. 2b, the resulting segmentation provides a noise-free and topologically coherent representation of the complex building roof. Individual roof panels are clearly delineated, which forms a suitable basis for subsequent geometric regularisation and IFC-based BIM reconstruction.

4.2 Geometric Accuracy Evaluation

The evaluation in this study is conducted at three complementary levels. First, patch-level precision assesses whether each extracted roof patch corresponds to a valid roof surface after segmentation and filtering. Second, surface-level geometric alignment evaluates the Euclidean distance between sampled points on the generated IFC roof surfaces and the nearest reference BIM roof surfaces. Third, object-level consistency evaluates centroid-based



(a) UAV-derived photogrammetric reconstruction.



(b) Automated roof detection and segmentation results.

Figure 2. UAV-derived photogrammetric reconstruction and automated roof detection and segmentation results for the EFI building.

association between extracted roof regions and BIM roof entities.

4.2.1 Evaluation of Global Geometric Alignment

To assess the geometric fidelity of the extracted roof surfaces, the spatial alignment between the generated IFC roof geometry and the corresponding roof surfaces in the reference BIM model was evaluated using a surface-based distance analysis. Specifically, geometric deviation was measured as the Euclidean distance between sampled points on the generated IFC roof surfaces and their nearest points on the reference BIM roof surfaces.

Minor geometric discrepancies between the reference BIM and the UAV-derived photogrammetric mesh are expected, as the two datasets were generated from different data sources and represent different abstraction levels (as-designed versus as-built geometry).

Within this context, the global alignment analysis between the generated IFC roof surfaces and the reference BIM roof geometry yielded a root mean square error (RMSE) of 0.84 m and a mean error of 0.68 m, with a minimum of 0.002 m and a maximum of approximately 2.0 m. It should be noted that a systematic vertical offset was observed between GPS-derived UAV elevations and the local coordinate reference system used in the BIM model, indicating a datum inconsistency of approximately 0.7 m.

To robustly identify valid extracted roof surface geometry under these conditions, a **detection threshold** of $\tau_{\text{match}} = 2.0$ m was adopted. A sampled point x_{gen} on a generated IFC roof surface is considered successfully matched if its distance to the nearest reference BIM roof surface S_{gt} satisfies

$$d(x_{\text{gen}}, S_{\text{gt}}) \leq 2.0 \text{ m}.$$

This threshold accommodates the observed systematic vertical offset, potential horizontal geolocation errors, and local geometric deviations at roof edges, while avoiding the exclusion of valid roof surfaces. Under this criterion, approximately 99.0% of the extracted roof surface area was successfully matched to roof elements in the reference BIM model, indicating that the derived roof geometry predominantly corresponds to genuine roof structures rather than noise.

4.2.2 Evaluation of Local Surface Fidelity and Higher-Level Object Consistency

Beyond patch-level precision and surface-based deviation analysis, we further assessed object-level consistency by associating extracted roof regions with reference BIM roof entities. Because the extracted roof regions are typically more fine-grained than the BIM representation, surface-level correspondence does not yield a one-to-one mapping. We therefore perform centroid-based association: each extracted roof region is matched to the nearest BIM roof panel entity within a search radius of $r_{\text{search}} = 5.0$ m. The larger tolerance (compared to the surface-distance threshold) accounts for the observed global offset and centroid displacement caused by region fragmentation.

For each matched pair, an object-level geometric consistency score is computed from area and slope agreement:

$$S_{\text{obj}} = 0.7 \cdot S_{\text{area}} + 0.3 \cdot S_{\text{slope}},$$

where the area similarity term is defined as

$$S_{\text{area}} = 1 - \frac{|A_{\text{gen}} - A_{\text{gt}}|}{\max(A_{\text{gen}}, A_{\text{gt}})},$$

penalising over-fragmentation, and the slope similarity term is defined as

$$S_{\text{slope}} = 1 - \frac{|\theta_{\text{gen}} - \theta_{\text{gt}}|}{45^\circ},$$

Using this evaluation, 270 out of 292 extracted roof panel entities were successfully associated with a BIM roof panel element (92.5%), with an average S_{obj} of 0.66. Slope agreement was consistently high, supporting the accuracy of PCA-based plane estimation, whereas lower area agreement mainly reflects differences in modelling granularity (a BIM roof panel versus multiple extracted planar regions).

4.2.3 Scalability and Campus-Scale Deployment

To evaluate scalability, the proposed framework was applied to the complete UAV-derived photogrammetric reconstruction of the University of Edinburgh Central Campus. Geospatial building footprints covering the entire campus were used as semantic priors to guide building-level mesh filtering and automated roof extraction.

Figure 3 shows (a) the spatial distribution of building footprints and (b) the corresponding campus-scale roof extraction results. The framework successfully processed all buildings in a fully automated manner, handling diverse architectural typologies without manual intervention or parameter adjustment.

These results demonstrate that the footprint-guided processing strategy enables robust and efficient scaling from individual buildings to large urban areas, confirming the feasibility of the proposed workflow for campus- and city-scale roof modelling applications.



(a) Geospatial building footprints.



(b) Automated roof extraction results.

Figure 3. Geospatial building footprints and automated roof extraction results for the University of Edinburgh Central Campus.

To provide a concise overview of the quantitative findings reported across Sections 4.1–4.2.3, Table 1 summarises the key performance metrics of the proposed framework.

5 Discussion and Future Work

The proposed method demonstrates robust performance in automated geometric roof reconstruction. Building on this, several directions for future research can be identified to further enhance the method's performance as well as the semantic richness, generality, and applicability of the resulting BIM models.

5.1 Automated Defect Mapping via Texture Projection

The current framework produces geometrically accurate but textureless roof representations, effectively generating 'white-box' models. A natural extension is the retexturing of those dense meshes and then evaluate and record the condition of each roof panel. Using state-of-the-art computer vision methods (e.g., Mask R-CNN or YOLO-based detectors), the panels could be analysed to identify sub-components [11], individual slates [12], and their defects. As a result, defects could be spatially mapped back to their precise 3D locations and tracked over time (over multiple inspections), thereby transforming the static geometric BIM into a dynamic *digital twin* capable of supporting condition monitoring and predictive maintenance workflows.

5.2 Integration with Deep Learning–Based Segmentation

The current framework adopts a geometry-driven segmentation strategy based on footprint-guided filtering, surface-normal consistency, and adaptive height analysis. This design offers several practical advantages for large-scale Scan-to-BIM deployment. First, it avoids the need for large annotated training datasets, which remain scarce for UAV-derived roof meshes with IFC-oriented labels. Second, the method is computationally lightweight and interpretable, making it easier to deploy across new urban areas without retraining. Third, the reliance on explicit geometric constraints is well suited to planar and piecewise-planar roof typologies, which remain common in many existing building stocks.

Nevertheless, the present approach also has limitations. Because it relies on planarity assumptions and empirically selected thresholds, its performance may degrade for curved, free-form, or highly irregular roof structures, as well as in areas affected by noise, incomplete reconstruction, or weak geometric boundaries. Under these conditions, the method may produce over-segmentation of smoothly curved surfaces into multiple planar patches, incomplete recovery of roof boundaries, or simplified geometric approximations that do not fully preserve the original roof form. These limitations are inherent to the current geometry-driven design and should be interpreted within

the intended application domain of predominantly planar and piecewise-planar roof reconstruction. In such cases, deep learning-based approaches may provide complementary benefits by learning higher-level geometric and contextual features directly from data.

Future work will investigate hybrid strategies in which deep learning is used to support semantic pre-classification and boundary refinement, while the current geometry-based reconstruction and IFC generation stages are retained for interpretable and BIM-compliant model generation. In particular, multimodal learning on textured photogrammetric meshes could combine 3D geometric information with radiometric appearance cues, improving robustness for challenging roof typologies such as domes, vaults, spires, and curved shells.

6 Conclusion

This paper presented a fully automated Scan-to-BIM framework for large-scale roof modelling from UAV-based photogrammetric data. By exploiting readily available geospatial building footprints as semantic priors, the proposed approach enables efficient isolation of roof geometry from dense urban reconstructions and supports robust processing of complex roof structures.

The framework directly generates georeferenced, IFC4-compliant roof models attributed to a known set of buildings from unstructured photogrammetric meshes, preserving non-convex roof geometries through alpha-shape–based boundary regularisation and avoiding manual intervention. This establishes a practical link between reality capture outputs and interoperable BIM representations suitable for downstream digital twin and GIS applications.

Experimental results on a university campus dataset demonstrate both accuracy and scalability. For a representative complex building, 431 roof panels were correctly extracted with a precision of 98.8%, while campus-scale experiments confirmed the feasibility of applying the workflow to large urban environments.

Overall, the proposed method provides a scalable and fully automated Scan-to-BIM workflow for reconstructing planar and complex non-convex roof geometries from UAV photogrammetric meshes using geospatial building footprints as semantic priors. The contribution of this study lies not in replacing learning-based roof understanding methods, but in demonstrating that a geometry-driven and IFC-oriented pipeline can already provide robust and interpretable performance for large-scale deployment without training data or manual modelling. At the same time, the framework is not intended to address all roof typologies equally well. Curved and free-form structures remain a clear limitation and motivate future hybrid approaches that combine learned semantic understanding with the current geometric reconstruction and BIM generation stages.

Acknowledgement

This work was supported by the Centre for Net Zero High Density Buildings funded by UKRI (UKRI238), as well as a donation by Bentley Systems.

References

[1] Y. Ge, B. Guo, P. Zha, S. Jiang, Z. Jiang, and D. Li. 3d reconstruction of ancient buildings using uav images and neural radiation field with depth supervision. *Remote Sensing*, 16(3), 2024. doi:10.3390/rs16030473.

[2] L. Li, J. Yao, J. Tu, X. Liu, Y. Li, and L. Guo. Roof plane segmentation from airborne lidar data using hierarchical clustering and boundary relabeling. *Remote Sensing*, 12 (9), 2020. doi:10.3390/rs12091363.

[3] X. Chen, Q. An, B. Zhao, W. Tao, T. Lu, H. Zhang, X. Han, and E. Ozdemir. Contour extraction of uav point cloud based on neighborhood geometric features of multi-level growth plane. *Drones*, 8(6), 2024. doi:10.3390/drones8060239.

[4] S. Mola Ebrahimi, H. Arefi, and H. Rasti Veis. Aerial imagery and lidar data fusion for unambiguous extraction of adjacent level-buildings' footprints. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W4:179–183, 2017. doi:10.5194/isprs-archives-XLII-4-W4-179-2017.

[5] P. Dabove, M. Daud, and L. Olivotto. Revolutionizing urban mapping: deep learning and data fusion strategies for accurate building footprint segmentation. *Scientific Reports*, 14:13510, June 2024. doi:10.1038/s41598-024-64231-0.

[6] S. Zbirovský and V. Nežerka. Open-source automatic pipeline for efficient conversion of large-scale point clouds to ifc format. *Automation in Construction*, 177:106303, 2025. doi:10.1016/j.autcon.2025.106303.

[7] Y. Schatz and B. Domer. Semi-automated creation of ifc bridge models from point clouds for maintenance applications. *Frontiers in Built Environment*, 10:1375873, June 2024. doi:10.3389/fbuil.2024.1375873.

[8] G. Taubin. Curve and surface smoothing without shrinkage. In *Proceedings of IEEE International Conference on Computer Vision*, pages 852–857, 1995. doi:10.1109/ICCV.1995.466848.

[9] M. Garland and P.S. Heckbert. Simplifying surfaces with color and texture using quadric error metrics. In *Proceedings Visualization '98 (Cat. No.98CB36276)*, pages 263–269, 1998. doi:10.1109/VISUAL.1998.745312.

[10] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, January 1994. doi:10.1145/174462.156635.

[11] J. Li, B. Tao, F. Bosché, C. X. Lu, and L. Wilson. Extracting roof sub-components from orthophotos using deep-learning -based semantic segmentation. In *Proceedings of the 41st ISARC*, pages 675–682, June 2024. doi:10.22260/ISARC2024/0088.

[12] J. Li, F. Bosché, C. X. Lu, L. Wilson, and B. Tao. Slate detection in orthophotos of building roof panels. In *Proceedings of the 42nd ISARC*, pages 49–56, July 2025. ISBN 978-0-6458322-2-8. doi:10.22260/ISARC2025/0008.

Feasibility Assessment of a BIM-Driven Robotic Rebar Cage Assembly Framework for Ballastless Railway Track Slab Construction

Yeddula Ashritha Reddy[1,2] and Samson Mathew[1]

[1] Department of Civil Engineering, National Institute of Technology (NIT) Tiruchirappalli, Tamil Nadu, India
[2] Larsen and Toubro Constructions, India
ashritha161619@gmail.com, sams@nitt.edu

Abstract –

Ballastless track slab systems are increasingly adopted in metro and high-speed rail projects due to their superior geometric stability, reduced maintenance requirements, and long-term performance. Construction of such infrastructure is more mechanised, driven by stringent requirements on construction speed, dimensional precision, and repeatability in processes such as material handling, batching, concrete distribution, and slab handling. Despite this high level of mechanisation, rebar cage assembly for track slabs remains largely manual, time-consuming, and ergonomically demanding, and continues to represent a critical bottleneck in ballastless railway track construction. This challenge is further exacerbated by constrained possession windows and limited spatial envelopes typical of metro environments.

This paper presents a feasibility assessment of a BIM-driven process planning and robotic assembly framework for rebar cages used in ballastless railway track slab construction. The study adopts a process-level automation perspective, recognising that track slab reinforcement exhibits manufacturing-like characteristics with repetitive yet bounded geometric variability across a project. Parametric BIM models are used to extract rebar geometry and connectivity, which are translated into robot-executable assembly sequences while accounting for geometric feasibility, interim cage stability, and robotic reachability. Feasibility is evaluated using a high-fidelity digital twin implemented in CoppeliaSim, where collision-free robotic pick–move–place operations are demonstrated under representative railway track slab geometries and spatial constraints. The results show that BIM-driven process planning and simulation provide a technically viable pathway for deploying robotic rebar cage assembly systems in metro and high-speed rail construction, subject to further physical validation.

Keywords –

Construction robotics, rebar cage assembly, ballastless track slabs, BIM-to-robot integration, digital twin, automation

1 Introduction

The growing need for fast and precise construction, as well as safety and scalability, in the expansion of metro and high-speed rail systems around the globe has led to the adoption of ballastless track slab systems like plinth track systems and precast slab track systems over ballasted track systems, especially in tunnels, viaducts, and urban areas. As a result, railway track construction has achieved a relatively high level of mechanization compared to conventional building projects, with activities such as rail welding, slab placement, and material handling being extensively mechanized.

Despite this progress, rebar cage assembly for track slabs remains largely manual and continues to form a critical bottleneck in otherwise mechanized workflows. In current practice, rebar cage preparation for cast-in-situ (CIS) track slabs requires large labor crews operating over extended durations. Similar challenges persist in precast slab manufacturing yards, where upstream processes such as batching and concrete distribution are automated, and Computer Numerical Control (CNC) machines are widely used for cutting and bending of rebars, yet cage assembly and tying remain predominantly manual. Figure 1 illustrates the typical manual preparation of rebar cages on site, highlighting the labor-intensive nature of reinforcement assembly that persists despite extensive mechanization in other track slab construction activities.

Unlike conventional building slabs, ballastless railway track slabs are designed with limited flexibility in geometry and detailing, as they must satisfy strict requirements related to track alignment, stiffness continuity, and long-term structural performance. Consequently, track slab reinforcement layouts are

highly standardized within a project. Even where horizontal curves, transitions, turnouts, or station zones are present, the resulting variations are typically confined to a finite and well-defined set of slab and reinforcement configurations. This characteristic renders track slab construction closer to a manufacturing process than a one-off site activity, with repetition and uniformity dominating over ad hoc geometric variation.



Figure 1 Manual preparation of rebar cage assembly on site and prefabrication yard respectively

From an automation perspective, this bounded variability is advantageous. The repetitive nature of track slab reinforcement enables reuse of motion planning strategies and robotic behaviors across multiple slab instances, reducing the need for continuous reconfiguration. While this study evaluates a representative track slab case, the emphasis is on assessing whether a BIM-driven robotic approach is technically feasible and operationally compatible at the project scale, where multiple slab variants typical of metro construction can be accommodated through automated adaptation rather than manual reprogramming, rather than on optimizing productivity for a single instance.

At the same time, large-scale railway projects operate under strict possession windows with multiple construction fronts active simultaneously. Manual assembly of rebar cage under such conditions introduces schedule uncertainty, quality variability, and safety risks, as labor-intensive tasks are difficult to scale across parallel work zones. Recent studies report that physically demanding and repetitive construction work, such as reinforcement work, is one of the least preferred jobs in the industry, which has resulted in the use of migrant workers in large-scale construction projects. Such conditions further emphasize the need to explore the possibility of applying automation techniques in railway construction.

Recent developments in construction robotics, digital twins, and BIM-based automation offer promising pathways to address these challenges. Nevertheless, current robotic technologies are mostly restricted to single-task automation, most prominently rebar tying, or are applicable only in stationary factory environments. Fully integrated, BIM-driven cage assembly frameworks that explicitly address the repetitive yet spatially constrained nature of railway track slab construction remain underexplored.

Against this background, this paper investigates the feasibility of a BIM-driven process planning and robotic assembly framework for rebar cages used in ballastless railway track slab construction. Productivity aspects are explored only at a conceptual level and are not intended as definitive performance predictions while economic evaluation is not part of this paper. Building upon existing studies on BIM-driven robotic reinforcement assembly and feasibility-oriented sequencing algorithms, the objective of this work is to evaluate whether such a framework can be technically realized and deployed within railway-specific geometric, spatial, and operational constraints. The study employs digital process planning and high-fidelity simulation to assess reachability, sequencing feasibility, and collision-free execution, thereby providing an informed basis for future physical implementation in prefabrication yards and, potentially, on-site metro construction using alignment-guided deployment platforms.

2 Literature review

Automation in the construction sector has continued to receive considerable attention from research and industry fields over the last two decades, owing to the challenges of low productivity growth, labour shortages, safety concerns, and variability in quality related to traditional construction methods [1]; [2]. In contrast to manufacturing, the construction sector remains largely site-based, labour-intensive, and fragmented, making it difficult to scale traditional productivity enhancement methods. In this scenario, reinforcement work has been identified as a domain of particular relevance for automation, involving repetitive, physically demanding, and time-consuming activities with well-defined geometric and structural constraints [3].

Availability of labour has emerged as a key bottleneck in the delivery of infrastructure projects across the globe. Recent global studies have pointed out that the construction sector is facing one of the most serious shortages of skilled labour compared to other key sectors, due to an ageing workforce, reduced entry of youth into the skilled trades, rising labour costs, and increased regulatory and safety demands [1]. Such issues are even more acute in large metro and railway projects, which are implemented with very tight possession windows, limited site space, and intense schedule pressures. Under such conditions, automation is increasingly being considered not only as a productivity solution, but also as a response to labour dependence, safety, and delivery

risks.

Rebar cage assembly is a key bottleneck in reinforced concrete construction. Manual handling, placement, and binding of rebars are prone to serious ergonomic hazards, introduce variability in quality, and often lie on the critical path of construction activities [4]. The physical workload associated with reinforcement installation has been shown to increase with spatial complexity and congestion, further reducing productivity and increasing execution time [5]. These limitations are amplified in slab-type infrastructure elements, such as metro and high-speed railway track slabs, where dense reinforcement layouts must be produced rapidly and within strict dimensional tolerances.

Automation of upstream processes of reinforcement preparation such as cutting, bending, and batching is relatively mature. CNC-controlled machines and digitally integrated fabrication systems are widely deployed in precast factories and rebar yards, and significant gains have been realized in terms of productivity, precision, and material efficiency [6]. However, the downstream processes, particularly sorting, sequencing, and rebar cage assembly, remain predominantly manual, offsetting gains achieved during earlier stages of mechanisation.

Significant advances have been made in task-oriented robotic systems, especially in the area of rebar tying.

Vision-based robotic tying systems using RGB-D sensing and learning-based perception have demonstrated high accuracy and industrial feasibility [7] [8], while commercial systems such as TyBOT confirm productivity and safety benefits for large planar reinforcement meshes [4]. More recent research has extended robotic tying to non-planar configurations using six-degree-of-freedom pose estimation [9]. However, these systems typically assume manual placement of rebars and do not address the assembly of rebar cages as an integrated process.

Robotic assembly of rebar cages has been investigated primarily under controlled laboratory or factory conditions. Early studies demonstrated feasibility for beams and columns [3], while more recent work proposed gantry-based robotic production cells validated through simulation and experimental trials [10]. A key challenge identified across these studies is the determination of a feasible installation sequence that avoids geometric blocking and ensures interim cage stability. Digital twin–based sequencing approaches, including depth-first search heuristics, have been proposed to address this challenge [11]. However, validation has largely been limited to generic or building-scale rebar cages and stationary production environments.

BIM-enabled robotic frameworks have further advanced automation by providing machine-readable representations of reinforcement geometry and enabling digital continuity from design to robotic execution. Several studies have demonstrated BIM-driven robotic workflows validated in simulation environments such as CoppeliaSim [10], [12]. Despite these advances, most BIM–robot pipelines still require manual refinement of sequencing logic and calibration, limiting their scalability for infrastructure-scale applications.

Importantly, existing studies do not explicitly address the specific characteristics of ballastless railway track slab construction, which is governed by elongated slab geometries, repetitive yet bounded reinforcement layouts, linear deployment conditions, and strict spatial constraints imposed by railway infrastructure. Unlike conventional building structures, track slab reinforcement exhibits limited design variability within a project, resulting in manufacturing-like repetition along the alignment. This characteristic suggests that automation feasibility in railway applications is governed less by geometric complexity and more by effective process planning, sequencing logic, and integration with construction workflows.

Against this backdrop, the present study builds on established principles of BIM-driven robotics and feasibility-oriented sequencing, but repositions rebar cage assembly within the context of railway track slab construction. The study focuses on integrating BIM-based data extraction, constraint-informed installation order assignment, and high-fidelity digital twin validation to evaluate whether robotic rebar cage assembly can be practically deployed under railway-specific spatial, operational, and constructability constraints.

3 Methodology

Figure 2 illustrates the overall methodology adopted in this study, showing the end-to-end workflow from BIM-based data extraction and installation sequence assignment to robotic simulation and digital validation. The workflow begins with parametric data extraction from BIM models, where reinforcement geometry and connectivity are automatically retrieved from Bar Bending Schedule (BBS) inputs. A feasible rebar installation sequence is then assigned using constraint-based & Depth-first-search (DFS) logic informed by prior robotic assembly research. The extracted data and sequencing information are subsequently integrated into a robotic simulation environment, where inverse kinematics, collision-aware motion planning, and gantry-based robot control are implemented. The complete process is validated using a digital twin to assess reachability, collision-free execution, and operational feasibility under railway track slab constraints.

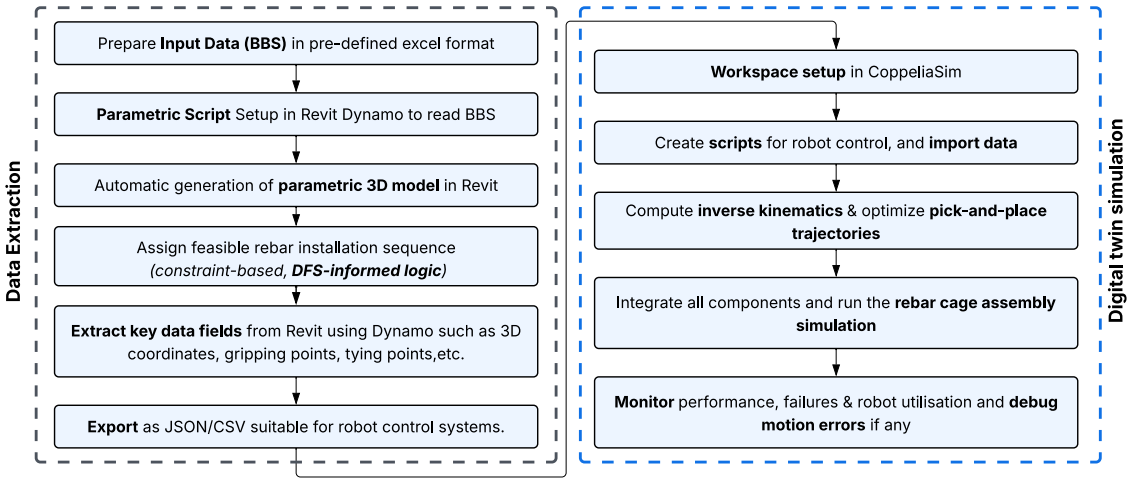


Figure 2 Methodology flow chart

4 BIM-Based Data Extraction for Railway Track Slab Reinforcement

The proposed framework relies on BIM as the authoritative source of reinforcement information for robotic process planning and simulation. A parametric rebar cage model corresponding to a ballastless railway track slab is developed as shown in figure 3 in accordance with typical BBS (figure 4) used in metro construction. The model captures the geometric definition of individual rebars, including bar diameter, length, bending shape, spatial coordinates, and layer information, as well as connectivity at intersections.

Assignment of the rebar installation order is formulated as a process planning problem rather than an optimisation task. In robotic rebar cage assembly, this sequencing problem can be represented as a constrained graph traversal problem, where individual rebars are treated as nodes and feasibility constraints such as geometric accessibility, collision avoidance, and interim structural stability-define precedence relationships. A Depth-First Search (DFS) based logic is employed to explore this solution space and generate a feasible installation sequence, consistent with established practices in robotic reinforcement assembly research. Similar DFS-based sequencing concepts have been previously reported for automated rebar cage fabrication in International Symposium on Automation and Robotics in Construction (ISARC) literature [11], where feasibility-oriented installation orders were shown to be sufficient for robotic execution.

In this study, rebars are grouped into layers based on their vertical position relative to the track slab base, with installation proceeding from lower to upper levels. Within each layer, the sequence is determined based on the distance from reference fixtures to ensure accessibility and avoid geometric blocking. The DFS algorithm incrementally builds the sequence while respecting constraints such as layer-wise progression, non-blocking placement, and support-first logic.

Compared to rule-based or layer-wise approaches, DFS offers greater flexibility in handling complex connectivity and non-uniform layouts, particularly where multiple valid sequences exist, while ensuring identification of at least one feasible, robot-executable sequence. The approach focuses on generating a consistent and repeatable sequence rather than outperforming human practices, enabling adaptability across varying slab geometries without manual reprogramming. Although DFS does not guarantee optimality in execution time or motion efficiency and may be sensitive to initial conditions, it is adopted here as the objective is feasibility validation. Given the bounded and repetitive nature of railway track slab reinforcement, identification of a single valid sequence is sufficient, making the approach computationally robust and well-suited for integration with BIM-derived data and digital twin environments.

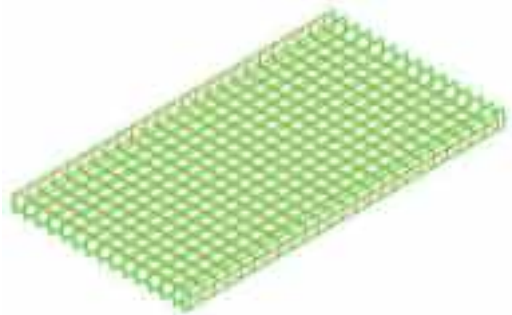


Figure 3 3D model of rebar cage

Slab Logitudinal/Transverse Reinforcement							
Description	Diameter (mm)	Spacing (mm)	a(mm)	b(mm)	c(mm)	Length (m) = a+b+c	No. of bars
Longitudinal Slab Top	12	200	150	4820	150	5.12	15
Longitudinal Slab Bottom	12	200	150	4820	150	5.12	15
Transverse Slab Top	10	200	150	2620	150	2.92	25
Transvere Slab Bottom	10	200	150	2620	150	2.92	25

Figure 4 BBS of the rebar cage

This combined BIM-based data extraction and sequencing approach ensures design-to-execution continuity while accommodating the repetitive yet constrained nature of railway track slab reinforcement. Following the assignment of a feasible installation order using DFS-based sequencing logic, the required reinforcement data are automatically extracted from the BIM environment using parametric scripting implemented through Dynamo within Autodesk Revit. This process eliminates manual interpretation of reinforcement drawings and manual data entry for robotic execution.

The extracted dataset includes unique rebar identifiers, bar diameter, assigned installation order, three-dimensional coordinates of grasp points defined at L/3 and 2L/3 locations along the rebar length for two-point gripping and the corresponding tie or placement points required for assembly so that deflection at ends of the rebar is minimum when picked up. These attributes are derived directly from the parametric BIM model, ensuring consistency between design intent and robotic process planning. The extracted information is structured and exported in a machine-readable format (JSON), enabling direct integration with robotic control and simulation environments without intermediate manual processing.

In addition to data extraction, Dynamo is used to parameterise the reinforcement model itself, allowing installation logic and attribute assignment to be embedded within the 3D model. This enables efficient handling of multiple slab variants with bounded geometric differences, such as those arising from curves, transitions, or station zones, by automating both parameter assignment and data extraction. As a result, clerical effort associated with repetitive data preparation is significantly reduced, and the framework supports scalable application across project-level variations in track slab geometry without the need for manual reprogramming of robotic tasks.

5 Simulation based evaluation for railway track slab

The proposed BIM-driven robotic rebar cage assembly framework is evaluated through simulation in CoppeliaSim using a high-fidelity digital twin configured for ballastless railway track slab construction. The digital twin models a gantry-based robotic system equipped with three six-degree-of-freedom (6-DOF) robotic manipulators mounted on a movable trolley, enabling movement in both longitudinal and lateral directions of the frame through linear actuators to ensure reachability within the assigned workspace. Of these, two robotic arms are dedicated to rebar pick–move–place operations, while the third arm is equipped with a specialised tying end-effector for executing reinforcement tying tasks. No vision sensors are required for tying as exact coordinates of tying points are being extracted from 3D model enabling accurate execution. This configuration supports coordinated operation while avoiding excessive workspace congestion.

The gantry–trolley system provides coarse positioning across the slab length and width, while the robotic arms perform localised manipulation tasks. This arrangement reflects the linear nature of railway track slab construction and enables efficient access to reinforcement locations distributed along elongated slab geometries without repositioning the entire system.

Motion planning is implemented using a hybrid strategy. First, the gantry and trolley perform coarse positioning to bring the robotic arm into the vicinity of the target rebar location in the longitudinal and/or lateral directions. Subsequently, fine positioning is achieved through inverse kinematics–based control of the 6-DOF robotic arm to execute precise grasping, placement, and tying operations. This layered approach reduces unnecessary arm motion, improves reachability, and reflects established practice in robotic reinforcement assembly research. The motion planning logic and kinematic modelling are inspired by previously reported robotic cage assembly frameworks, particularly the feasibility-oriented approaches presented by Momeni et al., while being adapted to the geometric scale and spatial constraints of railway track slabs. Conservative velocity and acceleration profiles are applied to ensure stable handling.

When importing BIM-derived reinforcement data into the simulation environment, a coordinate transformation formulation is applied to map BIM global coordinates into the local reference frame used in CoppeliaSim. This transformation accounts for differences in origin, orientation, and scale between the BIM model and the simulation environment, ensuring that rebar positions, grasp points, and placement targets extracted through scripting are accurately represented during robotic execution within the digital twin.

The simulation setup reflects typical railway site constraints, including limited lateral workspace, longitudinal deployment along the track direction, and interaction with rail-seat regions and embedded components. Spindles are planned to be temporarily placed at designated locations to fix the rail-seat base during casting and to ensure that reinforcement does not infringe upon the designed spacing. To accommodate railway-specific spatial restrictions, the complete robotic system including gantry structure, trolley, manipulators, and end-effectors is constrained within a predefined kinematic envelope representing the worst-case available construction space. This envelope is defined to prevent infringement with adjacent railway infrastructure such as overhead electrification (OHE) systems, cable trays, parapets, and other permanent or temporary civil works commonly present along metro alignments.

Collision-aware motion planning is enabled throughout the simulation to ensure safe interaction between robotic arms, partially assembled reinforcement cages, fixtures, and surrounding structural elements. The simulation experiments focus on evaluating robotic reachability, sequencing feasibility, and motion stability under conditions representative of metro construction, rather than on optimising cycle times or factory throughput. In addition to prefabrication yard scenarios, the simulation environment is intentionally constrained to reflect conditions applicable to cast-in-situ (CIS) track slab construction. By ensuring that the entire robotic system operates within the railway kinematic envelope, the framework demonstrates potential compatibility with deployment on alignment-guided platforms such as rail-mounted or road–rail vehicles (RRVs). This highlights the flexibility of the proposed approach to support both stationary prefabrication environments and mobile, track-based operations, which are essential for automation in linear railway infrastructure projects. To maintain focus on process-level feasibility and ensure comparability with existing robotic reinforcement studies, the following assumptions are adopted in the simulation:

1. The BIM-derived rebar geometry and connectivity accurately represent the as-designed reinforcement cage, and rebars are assumed to be geometrically rigid during handling and placement.
2. A feasible rebar installation sequence exists, and identification of a single valid sequence is sufficient for evaluating automation feasibility, recognising that multiple valid sequences may be possible.
3. Installation order assignments follow DFS-based, constraint-informed logic, which is not optimum but one feasible order
4. The robotic system operates in a controlled environment with known geometry; perception uncertainty, sensing noise, and real-time error correction are not considered at this stage.
5. Temporary fixtures and supports required to maintain interim cage stability are assumed to be available and correctly positioned throughout the assembly process.
6. Fixtures are provided on both directions of the table setup so that two different rebar can be placed on fixtures in their repsection directions. At any given fixture location, rebars are fed one by one to the system to eliminate the risk of pose uncertainty.

These assumptions are consistent with common practice in early-stage construction robotics research and allow the study to isolate the feasibility of BIM-driven process planning and robotic execution under railway-specific geometric and operational constraints.

When these assumptions are relaxed, several practical uncertainties arise. Rebar deflection during handling may influence placement accuracy; however, two-point gripping at one-third span locations reduces deflection, and the use of jigs helps maintain the rebars in a straight alignment during placement. Tolerance accumulation across sequential placements may affect overall cage geometry, although jigs used for spacing and cover control mitigate this effect under bounded variability. Pose uncertainty remains a limitation due to the absence of perception systems, while fixture misalignment and reinforcement congestion may further affect stability, accessibility, and placement precision. These factors indicate that incorporation of sensing, compliance control, and tolerance management strategies will be necessary for reliable real-world deployment. Key robot, gantry parameters used in the digital twin are summarised in Table 1 to ensure clarity and reproducibility.

Table 1 Robot and Gantry System Parameters Used in Simulation

Parameter	Specification	Remarks
Robotic arm-payload capacity	25 kg	Selected base on rebar + end effector weight including dynamic effect
Maximum reach of robotic arm	1.3 m	Sufficient for placement within slab region with the help of linear actuators
Gantry longitudinal travel (X-axis)	6 m	Covers slab length + margin
Gantry lateral travel (Y-axis)	3 m	Covers slab width + margin

6 Results and Discussion

Simulation results confirm that the proposed framework enables collision-free robotic rebar placement for railway track slab cages across all planned installation steps. All BIM-defined rebar target poses are reachable within the gantry workspace while maintaining compliance with railway spatial constraints. The process planning and sequencing logic successfully prevents geometric blocking and ensures interim cage stability, which is critical for elongated slab-type reinforcement assemblies typical of ballastless track systems.

Robotic motion remains stable throughout the simulation, with no kinematic singularities or unintended interference between manipulators or previously placed reinforcement. These results demonstrate that BIM-driven sequencing and digital validation are essential enablers for adapting robotics to railway reinforcement workflows, where precision, repeatability, and strict adherence to possession windows are more critical than maximising raw production speed. The framework is shown to be adaptable to both precast track slab manufacturing yards and on-site cast-in-situ (CIS) construction, subject to project-specific deployment strategies. A summary of key simulation-based performance indicators is presented in Table 2.

Table 2 Summary of Simulation-Based Performance Indicators

Parameter	Value
Number of rebars	80
Assembly completion	100 %
Average pick–move–place time (per bar)	~35 s
Total placement time	~2800 s (~47 min)
Number of ties	~750
Estimated tying time (per tie)	~ 2- 3 s (Based on literature/industrial systems)

The pick–move–place cycle times observed in the simulation are indicative and scenario-specific, as task duration is influenced by payload, travel distance, motion constraints, and placement accuracy requirements. Since the simulation assumes idealised conditions, observed cycle times are considered moderately reliable for comparative and feasibility assessment, but not as direct predictors of field performance.

In conventional practice, reinforcement cage assembly for ballastless track slabs is reported to achieve approximately 0.125 track m per man-hour, reflecting the labour-intensive and sequential nature of manual operations under constrained site conditions. Based on the simulation results for the representative slab configuration, the equivalent system-level production rate corresponds to approximately 2.5 track m per hour under idealised conditions. However, this value represents a theoretical upper bound, as the simulation does not account for real-world factors such as handling delays, coordination with other site activities, perception uncertainties, and operational interruptions. Therefore, the comparison should not be interpreted as a direct productivity gain, but rather as an indication of the potential for improved execution efficiency and reduced variability through automation. The primary contribution of this study remains the demonstration of technical feasibility, while quantitative productivity validation requires further experimental and field-based investigation.

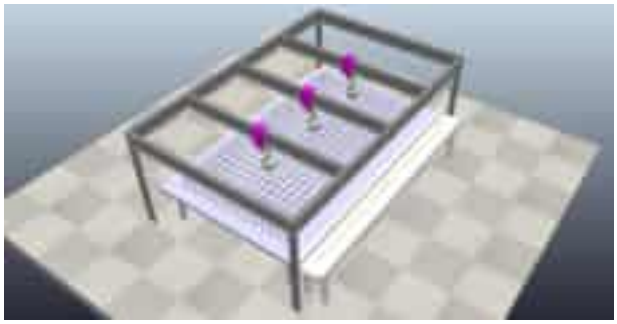


Figure 5 Simulation of rebar cage of metro track slab in CoppeliaSim

It is therefore emphasised that the objective of the simulation study is to evaluate the technical feasibility and deployability of the proposed robotic system under railway-specific spatial and operational constraints. Detailed cycle time optimisation and productivity assessment are outside the scope of this work, as reliable estimation would require physical prototyping, perception integration, and field validation. By constraining the system within the railway kinematic envelope, the framework demonstrates potential compatibility with deployment on alignment-guided platforms, such as rail-mounted or road–rail vehicles. When combined with coordinated logistics and alignment-guided mobility, the proposed approach offers a viable pathway for integrating robotic rebar cage assembly into metro construction workflows.

7 Conclusions and Future Work

This study presents a railway-specific feasibility assessment of a BIM-driven process planning and robotic assembly framework for rebar cages in ballastless track slab construction. It addresses a key gap in metro and high-speed rail projects, where rebar cage assembly remains largely manual despite mechanisation of other

activities. By adapting established BIM-driven robotic sequencing approaches to railway-specific geometric and spatial constraints, the study reframes rebar cage assembly as a process planning and digital integration problem.

Simulation using a high-fidelity digital twin confirms the feasibility of collision-free and stable robotic assembly under representative metro track slab conditions. The results demonstrate that BIM-derived data, feasibility-oriented sequencing, and hybrid motion planning can be effectively integrated to enable robotic execution without geometric blocking or kinematic conflicts. The findings highlight that feasibility in railway applications is governed primarily by sequencing logic, spatial compliance, and digital workflow continuity rather than optimisation of hardware or cycle times. The framework is applicable to linear railway infrastructure projects characterised by repetitive yet bounded reinforcement layouts and strict possession constraints. Compatibility with prefabrication yards and on-site cast-in-situ construction is demonstrated, with potential for deployment using alignment-guided platforms such as rail-mounted or road–rail vehicles and automated guided vehicles (AGVs), enabling scalable and mobile robotic operations.

Future work will focus on physical prototyping, integration of perception-based feedback for tolerance management, and validation under real or semi-controlled site conditions. Further investigation is required to evaluate coordinated operation with alignment-guided systems and to assess productivity, safety, and constructability impacts. A key limitation of this study is its reliance on simulation-based validation under controlled conditions. While the digital twin enables evaluation of reachability, sequencing feasibility, and collision-free execution, real construction environments introduce additional uncertainties, including geometric tolerances, bar deflection, fixture misalignment, and operational variability. These factors may influence positioning accuracy and system performance. Consequently, the applicability of the proposed framework to real-world deployment requires further validation through physical prototyping and field implementation.

References

[1] McKinsey & Company. Reinventing construction: A route to higher productivity. McKinsey Global Institute, 2023.

[2] T. Bock and T. Linner. Robot-Oriented Design: Design and Management Tools for the Deployment of Automation and Robotics in Construction. Cambridge University Press, Cambridge, UK, 2015.

[3] B. Dolinšek and J. Duhovnik. Robotic assembly of rebar cages for beams and columns. Automation in Construction, 8:195–207, 1998.

[4] C. Brosque and M. Fischer. Safety, quality, schedule, and cost impacts of ten construction robots. Construction Robotics, 6:163–186, 2022.

[5] F. Pfitzner, A. Braun, A. Borrmann, and F. Bosché. Spatial analysis and complexity evaluation for predicting rebar installation duration. Automation in Construction, 179:106462, 2025.

[6] G. C. Ertemli, E. G. Atalan and G. Tsakov. Robot process automation: Automation approaches for rebar manufacturing. Construction & Robotics: Research Driven Project, International M.Sc. Programme, Vol. 5, pages 7-27, 2024.

[7] J. Jin, W. Zhang, F. Li, M. Li, Y. Shi, Z. Guo, and Q. Huang. Robotic binding of rebar based on active perception and planning. Automation in Construction, 132:103939, 2021.

[8] X. Tan, L. Xiong, W. Zhang, Z. Zuo, X. He, Y. Xu, and F. Li. Rebar-tying robot based on machine vision and coverage path planning. Robotics and Autonomous Systems, 182:104826, 2024.

[9] M. Liu, J. Guo, L. Deng, S. Wang, and H. Wang. Enhanced vision-based 6-DoF pose estimation for robotic rebar tying. Automation in Construction, 171:105999, 2025.

[10] M. Momeni, J. Relefors, A. Khatry, L. Pettersson, A. V. Papadopoulos, and T. Nolte. Automated fabrication of reinforcement cages using a robotized production cell. Automation in Construction, 133:103990, 2022.

[11] M. Momeni, J. Relefors, L. Pettersson, A. V. Papadopoulos, and T. Nolte. On the bar installation order for the automated fabrication of rebar cages. In Proceedings of the 39th International Symposium on Automation and Robotics in Construction (ISARC), pages 508-511, 2022.

[12] S. Du, F. Teng, Z. Zhuang, D. Zhang, M. Li, H. Li, and Y. Weng. A BIM-enabled robot control system for automated integration between rebar reinforcement and 3D concrete printing. Virtual and Physical Prototyping, 19(1): e2332423, 2024.

[13] T. Sun, B. Han, S. Rusinkiewicz, and Y. Shao. Rebar grasp detection using a synthetic model generator and domain randomization. Automation in Construction, 176:106252, 2025.

Smartphone-Only High Accuracy Indoor Self-Localization and Navigation system for Non-GPS Environments with Patrol Free Floor Plan Based Map Creation

Bikash Lamsal[1], Masato Higo[1], Ryota Toki[2], Masato Oka[2], Bimal Kumar KC[3], Matteo Sardellitti[3] and Naofumi Matsumoto[3]

[1]Department of Creative Science and Engineering, Waseda University, Japan
[2]Kajima Technical Research Institute, Kajima Corporation, Japan
[3]NAiT Corporation, Japan
bikash@aoni.waseda.jp, higo_15@akane.waseda.jp, tokir@kajima.com, mst@kajima.com, bimal@nait-altg.com, matteo@nait-altg.com, n.matsumoto@nait-altg.com

Abstract –

Accurate indoor self-localization remains a fundamental challenge in environments where GPS signals are unavailable, such as construction sites, underground facilities, indoor spaces, and tunnels. This paper presents a smartphone-only indoor self-localization and navigation system that uses only sensors embedded in standard smartphones. The proposed system utilizes smartphone cameras and inertial sensors to achieve continuous indoor positioning without relying on external infrastructure or additional hardware. A key contribution is a patrol free floor plan-based map creation method using Map Creator, which generates navigable indoor maps directly from existing 2D floor plans without prior site surveys or environment scanning. Based on these maps, a smartphone-based indoor navigation application was developed to support real-time navigation and self-localized image acquisition with embedded position and orientation information. Experimental evaluations conducted in multiple indoor environments demonstrate an average relative localization error of 0.5–0.9%, confirming that high-accuracy indoor positioning can be achieved using only a standalone smartphone without external infrastructure. The proposed system offers a lightweight, scalable, and easily deployable solution for indoor localization and navigation in construction and facility management applications.

Keywords – Autonomous drone, Floor-plan mapping, Indoor localization, Indoor Navigation, Non-GPS, Smartphone-only

1. Introduction

Accurate indoor self-localization and navigation are essential for a wide range of construction-related activities, including inspection, progress monitoring, and safety management. While the Global Positioning System (GPS) is widely used for outdoor positioning, its performance degrades significantly in indoor and underground environments due to signal attenuation and blockage. As a result, reliable positioning in non-GPS environments such as construction sites, tunnels, and large buildings remains a critical research challenge [1,2].

To address this issue, various indoor localization technologies have been developed. These include wireless-based approaches such as Wi-Fi, Bluetooth, and ultra-wideband (UWB), as well as vision- and LiDAR-based methods such as simultaneous localization and mapping (SLAM) and feature-based localization [3–7]. In addition, visual–inertial sensor fusion using cameras and inertial measurement units (IMUs) has been increasingly adopted for mobile indoor positioning [8].

Recent studies have explored smartphone-based indoor localization as a practical and cost-effective alternative to infrastructure-dependent approaches, leveraging the widespread availability of consumer devices equipped with cameras and inertial sensors. These approaches have demonstrated sufficient accuracy for applications such as indoor navigation and spatially indexed image acquisition in non-GPS environments [9].

However, despite these advances, many existing systems still require additional preparation, including artificial marker-based initialization, deployment of wireless positioning infrastructure, or environment-dependent configuration processes such as prior feature database construction and repeated site patrol for mapping [10,11].

Such requirements limit scalability and practical

deployment, particularly in dynamic construction environments where indoor layouts frequently change and system reconfiguration is required. Therefore, achieving scalable and easily deployable indoor localization using only widely available consumer smartphones, without reliance on external infrastructure or site-specific preparation, remains an important research challenge.

To address this limitation, this paper proposes a smartphone-only indoor self-localization and navigation system that integrates floor-plan-based map creation with visual–inertial localization. The proposed approach enables navigation-ready indoor maps to be generated directly from existing 2D architectural floor plans without requiring site scanning or patrol-based mapping [12]. By constraining visual–inertial localization results using floor-plan-derived spatial information, the system improves localization accuracy while reducing setup effort and enhancing adaptability to layout changes.

Experimental evaluations conducted in representative indoor environments demonstrate that the proposed system achieves sub-1% relative localization error. These results confirm that high-accuracy indoor positioning can be achieved using standalone consumer smartphones without external infrastructure, making the proposed approach suitable for construction and facility management applications in non-GPS environments.

2. System Overview

This section presents the architecture and operational concept of the proposed smartphone-only indoor self-localization and navigation system. A standard consumer smartphone is defined as a device equipped with an RGB camera, inertial measurement unit (IMU), and onboard processor for visual–inertial computation. The proposed system performs indoor positioning and navigation in non-GPS environments using only these embedded sensors, without relying on external infrastructure or additional hardware.

To demonstrate device independence and applicability to widely available consumer hardware, experiments were conducted using both a recent smartphone model; iPhone 15, and a compact consumer-grade device: iPhone SE, confirming that the proposed approach does not depend on high-end sensing capabilities.

2.1 System Architecture and Components

The proposed system consists of three components: (1) a floor-plan-based map creation module, (2) a smartphone-based localization and navigation module, and (3) a smartphone-centered movement control interface. Their relationships are illustrated in Figure 1.

The first component, Map Creator, generates navigation-ready indoor maps directly from 2D architectural floor plans. Floor-plan-based mapping approaches have been shown to reduce deployment effort compared with scan-based or patrol-based methods, particularly in environments with evolving layouts [13,14]. In the proposed framework, Map Creator defines spatial reference parameters such as scale, orientation, and anchor points to derive walkable regions and connectivity relationships.

In practical construction environments, discrepancies often arise between design drawings and as-built layouts due to ongoing changes during construction. To address this issue, the proposed patrol-free map creation approach allows users to directly update walkable paths, connectivity relationships, and reference information on the floor plan without requiring additional site surveys or remapping. This enables efficient adaptation to layout changes while maintaining coordinate consistency.

The second component is a smartphone-based indoor localization and navigation module that performs continuous positioning using visual–inertial sensor fusion. Smartphones equipped with cameras and inertial measurement units (IMUs) have been shown to provide sufficient sensing capability for indoor motion estimation without external infrastructure [15,16]. In the proposed system, device pose is estimated by integrating visual features tracking from image sequences with inertial measurements, and the resulting trajectory is constrained using floor-plan-derived spatial topology to ensure consistency with indoor layouts.

The third component is a smartphone-centered movement control interface that converts localization and navigation output into motion commands for mobile platforms. Smartphone-centric control architectures have been shown to improve portability and system flexibility by decoupling localization from platform-specific control systems [17,18]. The proposed interface enables localization results generated on the smartphone to be reused across different navigation and autonomous movement applications without modification of the core algorithm.

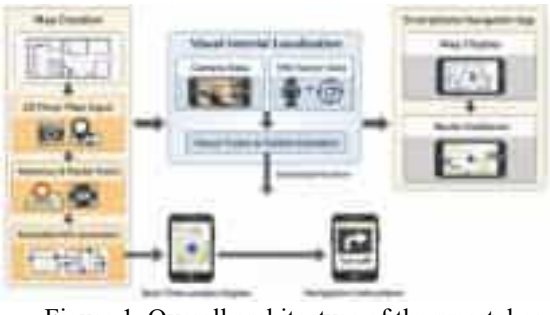


Figure 1. Overall architecture of the smartphone-only indoor localization and navigation system.

2.2 System Operational Flow

The operational flow ensures integration between

map creation, localization, and navigation. The sequence is illustrated in Figure 2.

The process begins with map creation using Map Creator, where a 2D floor plan is converted into a map by defining spatial reference information, walkable paths, and connectivity. The map is then transferred to the smartphone application.

During operation, the smartphone initializes its coordinate system and performs localization using camera and inertial data. The estimated pose is continuously updated and constrained using floor-plan-based spatial information. Based on the updated localization, navigation routes are computed and guidance is provided, as illustrated in Figure 2.

The operational flow of the proposed system is designed to ensure seamless integration between map creation, localization, and navigation execution. The overall operational sequence is illustrated in Figure 2.

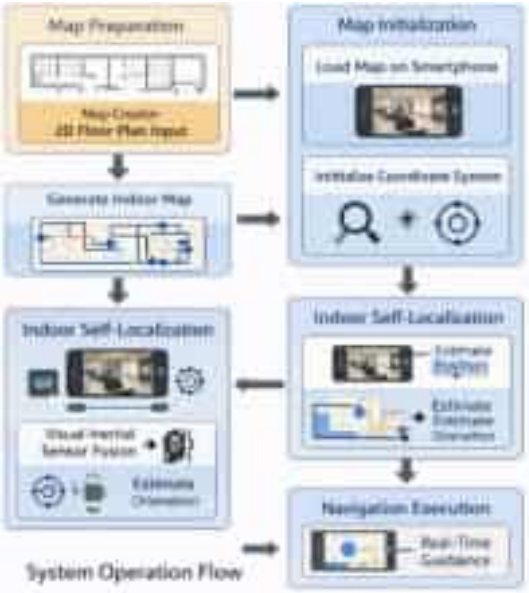


Figure 2. Operational sequence of the proposed smartphone-only indoor navigation system.

2.3 Data Flow and Information Exchange

The data flow separates static spatial information from dynamic sensor-driven data. The map creation module outputs static map data, while the smartphone modules handle real-time sensing, localization, and navigation.

The map creation module provides:

- Indoor coordinate reference information
- Walkable path networks and connectivity graphs
- Reference marker or anchor point data

These data are loaded into the smartphone application, where camera and inertial inputs are used to update localization results during operation. The interaction between static map data and dynamic localization is illustrated in Figures 1 and 2.

3. Patrol-Free Floor Plan based Map Creation System

This section presents a patrol-free indoor map creation method that generates indoor maps directly from 2D floor plans without site scanning or traversal. Spatial information is explicitly defined on architectural drawings, enabling efficient and repeatable map creation suitable for dynamic indoor environments. The workflow is shown in Figure 3.

To enable systematic indoor navigation representation, the walkable indoor environment derived from the 2D floor plan is formally modelled as a connectivity graph $G = (V, E)$, where V denotes discretized navigable nodes and E denotes feasible movement transitions between adjacent nodes. The graph is constructed from user-defined route segments on the floor plan, and connectivity is determined through reachability analysis. This graph-based representation supports route computation and serves as a constraint for localization by restricting estimated positions to valid navigable regions.

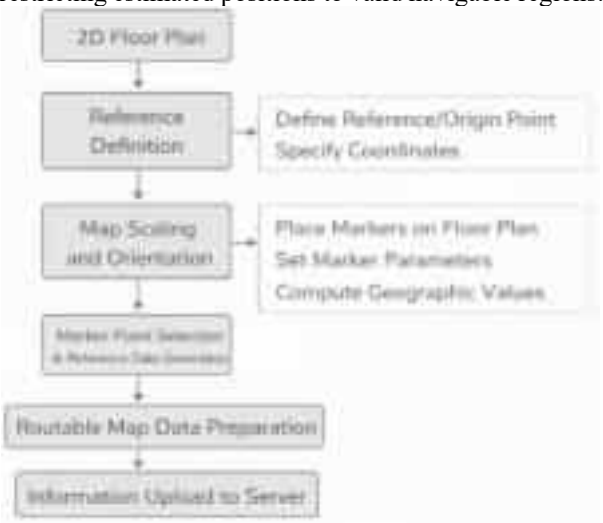


Figure 3. Patrol free indoor map creation workflow using Map Creator.

3.1 Floor-Plan-Based Spatial Definition and Coordinate Alignment



Figure 4. Map Creator screen with Floor Plan and associated geographic information.

Map Creator uses a 2D floor plan as the input for indoor map creation, where spatial relationships are derived directly from image geometry rather than sensor measurements [19].

A reference point is defined on the floor plan to establish the indoor coordinate system and enable alignment with real-world coordinates. The reference point is selected on the floor plan, and associated spatial parameters such as geographic position, orientation, and scale are defined in Figure 4.

By associating this reference point with known geographic coordinates, the indoor map is transformed into a globally consistent coordinate system without requiring on-site calibration [20].

Once the reference point is defined, all subsequent spatial elements, including marker points, walkable routes, and navigation coordinates are computed relative to this origin. This explicit reference-based approach enables repeatable and scalable map creation while maintaining consistency across devices and sessions

3.2 Marker Point Configuration and Reference Data Generation

Map Creator supports marker points that act as spatial anchors for indoor localization by linking the indoor coordinate system to real-world coordinates.

A single marker is sufficient for localization initialization; however, multiple markers can be used to improve robustness [21]. Marker configuration and reference data generation are illustrated in Figure 5. The system supports both artificial and natural visual features as markers, allowing flexible deployment without requiring dedicated marker installation [21].

Marker points are defined on the floor plan, and their spatial attributes are used to automatically compute geographic coordinates and orientation based on the reference point and map parameters. Marker data are stored for localization initialization. After initialization, continuous operation does not require persistent marker visibility, maintaining patrol-free operation.

3.3 Routable Map Generation

After defining the indoor coordinate system and marker reference information, Map Creator generates routable map data that represents valid movement paths within the indoor environment. Walkable routes are defined on the 2D floor plan as a red line segment to represent traversable paths as shown in Figure 6. These routes are connected to form a network of walkable paths, distinguishing traversable and non-traversable regions.

Once the walkable routes are defined, Map Creator applies a maze-based algorithm to analyze the connectivity of the red line network. The algorithm treats the connected red lines as traversable cells within a maze-like structure and automatically determines reachability between path segments. Through this process, valid routes, intersections, and alternative paths are identified based on the connectivity shown in Figure 6 [22].

The resulting routable map includes path nodes and connectivity relationships that are later used by the indoor localization and navigation application for route computation. This approach enables patrol-free and easily updatable route generation. When the indoor layout changes, the user can simply modify the red route lines on the 2D map, and the routable data can be regenerated without additional data collection or site traversal.

4. Smartphone-Based Indoor Localization and Navigation

This section describes the smartphone-based indoor self-localization and navigation system developed in this study. The system is designed according to a localization-first concept, in which indoor positioning is treated as the fundamental function, while navigation is implemented as an optional layer that operates on top of localization results. This design allows the system to function both as a standalone indoor positioning system and as a navigation system when route guidance is required. The flowchart illustrating the smartphone based indoor localization and navigation is shown in Figure 7.



Figure 5. Marker point configuration and automatic generation of reference coordinates in Map Creator.



Figure 6. Routable path generation in Map Creator using the 2D floor plan.

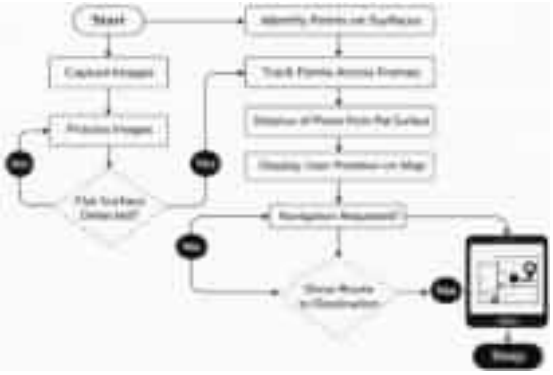


Figure 7. Flowchart illustrating the smartphone based indoor localization and navigation



Figure 8. Smartphone-based indoor self-localization results visualized on the floorplan.

4.1 Indoor Self-Localization

Indoor self-localization is formulated as a visual–inertial pose estimation problem in which the smartphone trajectory is incrementally estimated using sensor fusion. Relative motion between consecutive positions is computed by tracking sparse visual features extracted from sequential monocular RGB images, while inertial measurements from the accelerometer and gyroscope provide short-term motion propagation and orientation stabilization. Localization updates are performed at approximately 2 Hz, corresponding to the experimental image acquisition interval of 0.5 m.

At the start of operation, localization is initialized using a single reference marker whose known position establishes alignment between the smartphone coordinate frame and the floor-plan coordinate system. After initialization, continuous localization proceeds without requiring persistent marker visibility.

To suppress long-term drift, the estimated pose pt is projected onto the nearest feasible location defined by the floor-plan-derived walkable graph:

$$p_t' = \arg\min_{p \in \Omega} \|p_t - p\|$$

where Ω represents the set of valid walkable positions. This constraint ensures topological consistency between localization results and indoor layout structure while reducing accumulated positional deviation during extended indoor trajectories. The overall localization pipeline therefore consists of feature tracking, inertial motion propagation, pose fusion, coordinate initialization, and floor-plan feasibility projection.

Localization results are visualized directly on the indoor map, as shown in Figure 8, where the blue point represents the real-time estimated smartphone position and the green square box with arrow inside represents the position of the image captured in the direction of arrow.

4.2 Indoor Navigation using Smartphone

Based on the indoor localization results described in section 4.1, the proposed system performs smartphone-based indoor navigation using floor-plan-derived connectivity information. Navigation is executed directly on the smartphone, and all route planning, guidance generation, and position updates are computed using the estimated device position on the indoor map.

The navigation process follows a structured workflow, as illustrated in Figure 9. First, the user selects a target floor map and specifies a destination point on the map. After the destination is set, the system searches for reference markers and generates an optimal navigation route based on predefined walkable paths and connectivity information. Once the route is generated, indoor navigation starts, and the device position is continuously updated in real time as the user or mobile platform moves. During navigation, the system provides guidance until the destination point is reached.

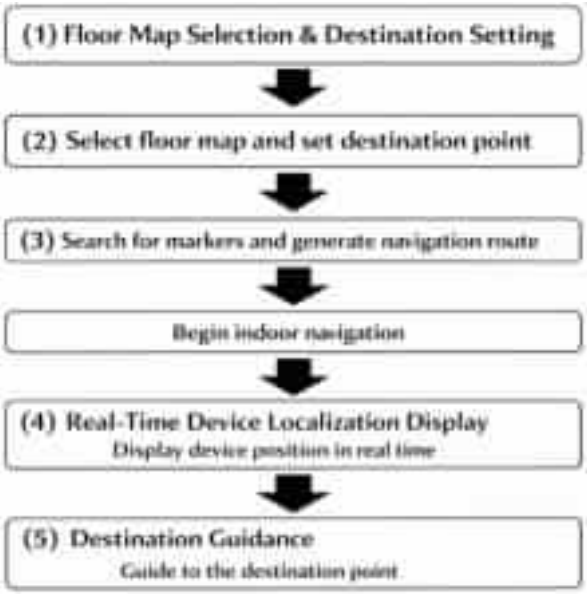


Figure 9. Workflow of smartphone-based indoor localization and navigation

The real-time navigation behaviour is visualized on the smartphone interface, as shown in Figure 10. The indoor floor plan is displayed together with navigation-related information, including the current estimated position, destination point, reference marker positions, and the planned route.

In Figure 10, the blue dot inside a blue circle represents the real-time estimated position of the smartphone, which is continuously updated based on the localization results. The destination point is shown in blue dot inside a red circle, the red dot inside a blue circle is the reference marker point. The route connecting the current position to the destination point is displayed on the map with blue line, providing intuitive visual guidance.



Figure 10. Smartphone screen showing the interface for indoor navigation

5. Experimentation and Evaluation

5.1 Experimental Setup and Localization Accuracy Results

Experimental evaluation was conducted to assess the accuracy and robustness of the proposed smartphone-only indoor self-localization and navigation system in non-GPS environments. Experiments were performed in two representative indoor environments at Waseda University. Environment A was a confined classroom space (Building 65, Room 504), while Environment B was a large open atrium area (Building 55, first floor). The experimental environments and predefined measurement paths are illustrated in Figure 11.

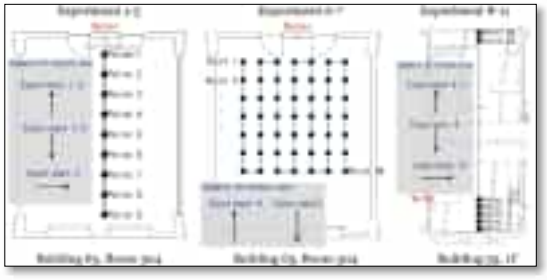


Figure 11. Experimental environments and predefined measurement paths used for accuracy evaluation

For localization initialization, a reference marker was placed at a known position in each environment. During experiments, images were captured at regular intervals of approximately 0.5 m along predefined walking paths. Most experiments were conducted using an iPhone 15, while one experiment was performed using an iPhone SE to examine the influence of device characteristics on localization performance.

A total of eleven experiments were designed to evaluate localization accuracy under different motion patterns and spatial scales. Experiments 1–5 focused on short straight-line motion with a total travel distance of 4 m. Experiments 6–7 evaluated grid-based motion following a 7 × 7 grid pattern to analyze accumulated positional drift. Experiments 8–11 involved long-range straight-line motion with a total travel distance of 32 m, representing extended indoor navigation scenarios.

Localization accuracy was evaluated using two quantitative metrics. First, relative localization error (%) was computed by comparing estimated step distance with the ground-truth step distance of 0.5 m. Second, accumulated positional drift was evaluated by measuring deviation from ideal grid reference positions in grid-based experiments. In addition, absolute positioning accuracy was quantified using the two-dimensional Euclidean error between estimated coordinates and tape-measured ground-truth positions.

The experimental results are summarized in Table 1, while detailed localization accuracy results for Environment A and Environment B are presented in Table 2 and Table 3, respectively. Figure 12 illustrates the distribution of relative localization errors across all experiments.

Table 1. Summary of Experiments for Indoor Localization

Exp. No.	Environment	Walking pattern	Travel distance	Purpose
1– 4	Bldg. 65, Room 504	Straight (forward/backward)	4 m	Short-range accuracy
5	Bldg. 65, Room 504	Straight (sideways)	4 m	Direction sensitivity
6	Bldg. 65, Room 504	7×7 grid	24 m	Drift evaluation
7	Bldg. 65, Room 504	7×7 grid	24 m	Camera orientation effect
8	Bldg. 55, Atrium	Straight (forward)	32 m	Long-range accuracy
9	Bldg. 55, Atrium	Straight (backward)	32 m	Motion robustness
10	Bldg. 55, Atrium	Straight (sideways)	32 m	Motion robustness
11	Bldg. 55, Atrium	Straight (forward)	32 m	Device dependency

Table 2: Localization Accuracy Results in Environment A (Room 504)

Experiment	Motion type	Evaluation metric	Result
1–4	Forward / backward	Total relative error	0.6–0.9 %
5	Sideways	Total relative error	1.53 %
6	Grid	Mean step-wise error	0.88 %
7	Grid	Mean step-wise error	0.87 %

Table 3: Localization Accuracy Results in Environment B (Atrium)

Exp. No.	Motion type	Device	Total relative error
8	Forward	iPhone 15	0.61 %
9	Backward	iPhone 15	0.18 %
10	Sideways	iPhone 15	0.27 %
11	Forward	iPhone SE	0.94 %

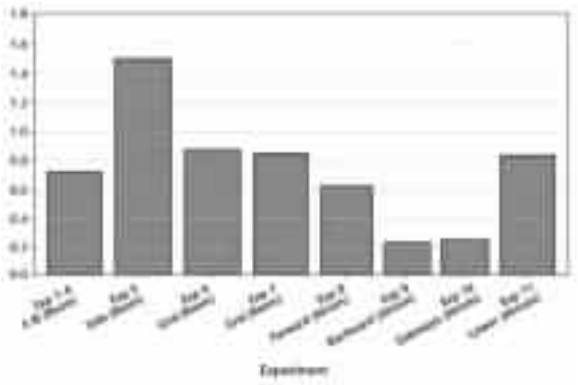


Figure12: Localization Accuracy from Experiment 1-11.

5.2 Baseline Comparison and Drift Growth Analysis

To further contextualize the localization performance observed in both experimental environments, a baseline comparison was conducted with unconstrained smartphone visual–inertial odometry (VIO). The localization accuracy results reported in Table 4 for Environment A and Table 5 for Environment B indicate that the proposed floor-plan-constrained approach maintains low relative error and controlled drift across different spatial configurations. These results are therefore compared with typical drift characteristics of infrastructure-free smartphone VIO to analyze the effectiveness of spatial constraint integration.

Because visual–inertial odometry provides relative pose estimation, accumulated positional drift generally increases with travelled distance. Engineering benchmarks suggest that consumer smartphones may exhibit approximately 0.3-0.8 m positioning error over short indoor trajectories (5-10 m), increasing to approximately 1-3 m after several tens of meters depending on motion conditions and visual feature availability.

To further validate drift growth behavior, an additional corridor experiment exceeding 50 m travel distance was conducted. Under stable motion conditions, the proposed method demonstrated approximately 0-1 m maximum drift relative to ground-truth references, whereas increased dynamic disturbances resulted in drift values approaching 3 m. The results indicate that floor-plan-derived spatial constraints significantly suppress drift accumulation during extended indoor navigation.

Table 4: Drift Growth Comparison between Baseline VIO and Proposed Method

Travel Distance	Baseline Smartphone VIO Drift (m)	Proposed Method Drift (m)
5-10 m	0.3-0.8	<0.2
20-30 m	0.8-2.0	0.2-0.4
>50 m	1-3	0-1

Table 5: Representative Indoor Localization Accuracy Comparison

Method	**Typical Absolute Horizontal Error**	**Remarks**
Smartphone GPS / generic mobile positioning	10-20 m	Indoor signal attenuation
Wi-Fi / BLE beacon systems	2-5 m	Requires infrastructure
Smartphone visual–inertial odometry	1-3 m (after ~30-50 m)	Drift accumulation
Proposed method	**~0-1 m**	Infrastructure-free

6. Discussion and Conclusion

This paper presented a smartphone-only indoor self-localization and navigation system for non-GPS environments that integrates a patrol-free, floor-plan-based map creation method with smartphone-based visual–inertial localization. Experimental results demonstrated that the proposed system consistently achieved sub-1% relative localization error across both confined and large indoor environments without relying on external positioning infrastructure, environment scanning, or specialized hardware. The results further showed that localization accuracy is influenced by motion direction and camera performance, while floor-plan-based spatial constraints effectively suppress accumulated drift during navigation.

In addition, the feasibility of extending the proposed smartphone-based localization and navigation framework to autonomous indoor mobile platforms was verified using a commercial off-the-shelf drone, demonstrating the platform-independent nature of the approach. Overall, the proposed system provides a lightweight, scalable, and easily deployable solution for

indoor localization, navigation, and spatial data acquisition, making it well suited for construction, facility management, and other indoor operational environments.

References

[1] A. Sesyuk, K. I. Sergeyev, and A. V. Shcherbakov. A survey of 3D indoor localization systems and technologies. *Sensors*, 22(23):9380, 2022.

[2] M. Zieliński, P. Węgrzyn, and K. Sławiński. An improved step detection algorithm for indoor navigation using smartphone sensors. *Sensors*, 25(20):6358, 2025.

[3] J. Dai, H. Zhang, and L. Liu. A survey of latest Wi-Fi assisted indoor positioning on smartphones. *Sensors*, 23(18):7961, 2023.

[4] J. Cho, H. Park, and S. Kim. Real-time indoor localization and safety applications using UWB in construction. *KSCE Journal of Civil Engineering*, 29(8):100164, 2025.

[5] V. H. Le, T. Nguyen, and D. Tran. A survey of visual SLAM based on RGB-D images using deep learning. *Algorithms*, 18(7):394, 2025.

[6] H. Jiang, Y. Wu, and Z. Zhang. GPS-denied LiDAR-based SLAM—A survey. *IET Cyber-Systems and Robotics*, 2025.

[7] H. Qu, Z. Li, and S. Wang. Point-line feature-based VSLAM systems: A survey. *Expert Systems with Applications*, 2025.

[8] J. Li, Y. Zhao, and H. Liu. A review of indoor localization methods leveraging smartphones. *Sensors*, 24(21):6956, 2024.

[9] B. Lamsal et al. Development of an indoor localization and positioning system in non-GPS environment using standalone smartphones and its implementation in construction site photo management application. *DFBI Conference Proceedings*, 2025.

[10] N. Kayhani, A. Teizer, and M. König. Tag-based visual–inertial localization of UAVs in GPS-denied indoor construction environments. *Automation in Construction*, 135:104110, 2022.

[11] G. C. La Delfa et al. Floor Tag: A hybrid indoor localization system based on floor-deployed markers and smartphone sensing. *Electronics*, 14(24):4819, 2025.

[12] H. Jang, J. Lee, and S. Choi. Generating spatial knowledge graphs with 2D indoor spatial data. *ISPRS International Journal of Geo-Information*, 13(2):52, 2024.

[13] W. Pan et al. Floor-plan-assisted indoor positioning using pedestrian dead reckoning. *IEEE Internet of Things Journal*, 12(4):3120–3133, 2025.

[14] A. Ayub et al. Comparative analysis of floor-plan-based indoor positioning approaches. *Scientific Reports*, 15:34514, 2025.

[15] S. Hou et al. Smartphone-based visual–inertial localization for indoor navigation applications. *Sensors*, 24(21):7123, 2024.

[16] Z. Zhou et al. Visual–inertial fusion for smartphone-based indoor positioning. *IEEE Internet of Things Journal*, 11(14):23592–23603, 2024.

[17] T. Nguyen et al. Smartphone-based control frameworks for autonomous indoor mobile systems. *Robotics and Autonomous Systems*, 178:104807, 2024.

[18] Y. Chen et al. Vision-based indoor navigation and control for autonomous mobile platforms. *Automation in Construction*, 158:105148, 2024.

[19] J. Liu et al. Interactive floor-plan-based indoor map generation without site surveying. *Automation in Construction*, 156:105084, 2024.

[20] K. Yamamoto et al. Decoupled indoor map creation and navigation frameworks. *Automation in Construction*, 155:105020, 2024.

[21] R. Martinez et al. Human-in-the-loop indoor mapping using architectural drawings. *Advanced Engineering Informatics*, 58:102112, 2024.

[22] D. Silver, S. Singh, and R. Sutton. Graph-based path extraction in grid environments. *IEEE Transactions on Systems, Man, and Cybernetics*, 53(4):2341–2354, 2023.

A Time- and Memory-Efficient Method for Axis Alignment of Very Large Point Clouds using Orthogonality in Buildings

Longyong WU*, Sou-Han CHEN, Meng SUN and Fan XUE

Department of Real Estate and Construction, The University of Hong Kong, Hong Kong (China SAR)

wulongyong@connect.hku.hk, hankchen@connect.hku.hk, sunmhku01@connect.hku.hk, xuef@hku.hk

Abstract -

In construction and operational stages of buildings and in computer vision and construction robotics, accurate alignment of LiDAR point clouds to a building's 3D coordinate axes is an essential preprocessing step. However, existing point cloud processing tools handle a point cloud in a single batch, which can lead to high memory usage or even failures for very large clouds. This study presents a chunk-based, orthogonality-enabled method for aligning very large building LiDAR point clouds to the building's coordinate in parallel computing. Our method leverages the random-access characteristic of LiDAR's LASer (*.las*) format to divide a cloud into multiple chunks, then processes each chunk, with the number of concurrently running tasks being limited. Subsequently, the global orientation is estimated by extracting the dominant directions of the building's interior and exterior orthogonal structures to rotate and align the input cloud. Experiments on a 22 GB real-world LiDAR point cloud demonstrated the effectiveness of our proposed method: It reduced peak memory consumption from exceeding 60.5 GB to 4.1 GB and achieved a 3.85-fold speedup over the sequential method. (The source code is shared at: `https://github.com/chunibyo-wly/laspy-parallel`)

Keywords -

Building and Construction; LiDAR; Point cloud processing; Orthogonal structure; Parallel computing

1 Introduction

Light Detection and Ranging (LiDAR) point clouds have become increasingly essential data sources in construction [1], surveying [2], computer vision [3], and robotics [4], thanks to LiDAR's high precision in measuring the 3D built environment. Yet, a LiDAR point cloud's local coordinate axes are usually inconsistent with the target building's or project's coordinate axes in 2D drawings, footprint on the map, or 3D building information modeling (BIM).

Axis alignment refers to aligning a point cloud to the project's local coordinate axes [5]. Axis alignment is an important step to enable many construction digitalization and automation applications, including existing condition modeling, 3D coordination, registration of 2D computer-aided design (CAD) drawings [6], scan-to-BIM [7], scan-vs-BIM for change detection [8], and construction progress monitoring [9]. In computer vision, many deep learning models for point cloud processing have been proposed and trained well on coordinate-aligned datasets [10], and some robots for building operation and maintenance also relied on axis alignment.

However, existing tools for aligning LiDAR cloud coordinates with buildings and construction often encounter challenges. Building floor areas and construction sites can span multiple thousand square meters, and modern LiDAR scanners produce 10,000–400,000 points per square meter; these factors result in LiDAR clouds that can reach up to billions of points and tens or even hundreds of gigabytes (GB). Existing 3D point cloud processing tools, such as `Open3D` [11] and `pypcd4`, always load the full cloud data into computer memory in a single batch. This one-batch processing of very large LiDAR clouds is memory-intensive and sometimes exceeds the capacity of computers, which are equipped with no more than 32 GB Random-access memory (RAM) in most cases. When memory demand exceeds system capacity, the operating system may terminate the process, causing the computation to fail and limiting the scalability of LiDAR data analysis.

To mitigate the problem of exceeding system memory capacity in axis alignment, this study presents a chunk-based, orthogonality-enabled parallel computing method. The method aims to achieve relatively efficient axis alignment for very large point clouds. The method leverages the random-access characteristic of LiDAR's LASer (*.las*) format to divide an input LiDAR point cloud into partitioned chunks, then processes them in parallel with controlled concurrency. To ensure balanced orthogonality detection, the point distances in each chunk are normalized (down-sampled) to 2 cm, and point normals are estimated from the normalized points. Finally, the method automatically estimates and optimizes the dominant orthogonal structural directions of LiDAR point clouds, in order to rotate the point cloud and align the coordinates. Overall, this parallel computing utilizing orthogonality in buildings can achieve better time and memory efficiency on daily per-

sonal computers.

The contribution of this paper is twofold:

1. A novel chunk-based, orthogonality-enabled parallel computing method for axis alignment of very large LiDAR point clouds, and

2. Preliminary evidence on the boosted time and memory efficiencies of the present method from experiments on real-world very large data.

2 Related Works

LiDAR has been widely applied in the construction industry [12, 13] for various applications, ranging from semantic enrichment of BIM models to derivative-free optimization-based (DFO) 3D reconstruction of architectural structures. For instance, Xue et al. [14] is the first to transform architectural symmetry detection from LiDAR point clouds as a nonlinear optimization problem and proposed a DFO approach to solve it. Furthermore, Xue et al. [15] extended the DFO method to register a small-scale indoor LiDAR point clouds with BIM components, enabling the reconstruction of detailed indoor scenes.

Although existing studies have demonstrated LiDAR point clouds' applications in 3D modeling and semantic enrichment, most have been limited to small-scale or exterior-level point clouds [14, 16], due to computational and memory limits. In practice, LiDAR point clouds collected from construction sites often contain millions of interior points, posing significant challenges for preprocessing using existing methods. Nevertheless, these interior point clouds are valuable for analyzing architectural symmetries and skeletons.

Orthogonality, a common characteristic of real-world architectural forms and designs, offers researchers an opportunity to understand and reconstruct urban point clouds [17]. For instance, Wu et al. [12] applied orthogonality and symmetry constraints to efficiently reconstruct 3D models from facade point clouds, and later extended this framework to support morphological analysis. This inherent architectural characteristic enables the development of chunk-based, orthogonality-enabled methods for the axis alignment of very large LiDAR point clouds.

In summary, it is of interest to design a chunk-based, orthogonality-enabled method for axis alignment of very large LiDAR point clouds, for LiDAR and construction digitalization practitioners who handle very large indoor or outdoor scenes in a time- and memory-efficient manner.

3 Our Method for Processing Very Large Point Clouds

3.1 LiDAR Scanning Principles and the LASer Format

LiDAR has been widely used in construction, robotics and mapping. It actively emits laser pulses and measures the time delay of their reflections from surfaces to estimate distances. Compared with passive sensors such as cameras, LiDAR is less affected by lighting conditions and can provide depth information even in poor lighting. After obtaining raw distance measurements, LiDAR point clouds can be combined with simultaneous localization and mapping algorithms to automatically estimate the scanner's position and orientation in unknown environments, reconstructing 3D environment with rich textures and high-precision geometric details [18]. This capability makes LiDAR a powerful tool for change detection, building modeling, and progress monitoring [13].

To efficiently manage various and massive LiDAR point clouds, The American Society for Photogrammetry and Remote Sensing defined an open format, LiDAR (with ".las" file extension), for the exchange and storage of LiDAR data. This format typically consists of four components: Public Header Block, Variable-Length Records, Point Data Records, and Extended Variable-Length Records. The header block contains the metadata indicating the byte offset to the beginning of point data and the byte size of each point, which can be used to access each point efficiently [19]. This benefits in parallel and distributed algorithms for very large point cloud processing.

3.2 Chunk-based Parallel Processing

Existing point cloud processing tools, such as `Open3D` and `pypcd4`, always load a LiDAR point cloud in a single batch. However, this approach is limited in handling very large LiDAR data, as loading the whole file in a single batch can easily cause out-of-memory failures. For instance, `Open3D`, a well-known point cloud processing library, stores point coordinates, colors, and normals using `std::vector<Eigen::Vector3d>` in memory, which is memory-intensive. Loading 789-million points into memory at once requires approximately 53 GB before any further processing. Such memory demand exceeds the capacity of computers and even some workstations. Therefore, it is essential to develop a memory-efficient method to deal with massive point clouds while maintaining time efficiency.

Leveraging the structure of the LASer format, a point cloud can be randomly accessed using the following formula:

$$\text{Offset} = \text{Offset}_{\text{header}} + i \times \text{TypeSize} \tag{1}$$

where i denotes the index of the specific point to be accessed directly, and TypeSize denotes the number of bytes used to store each point (coordinates and attributes) as defined in the LASer header. This random-access property enables data access without sequential reading, allowing the LiDAR point clouds to be divided into smaller chunks to trade off between memory efficiency and processing time.

However, processing these chunks in parallel straightforwardly introduces a new challenge: each thread requires dedicated memory allocation before execution. If all chunks are loaded and processed simultaneously, it will still raise an out-of-memory error. This is because earlier tasks may not finish in time to release their allocated memory before later tasks begin to require additional memory. To address this issue, our chunk-based parallel processing approach divides tasks into multiple batches (Fig. 1) and limits the number of concurrent processes. The next batch is launched after the slowest task in the previous batch finishes.

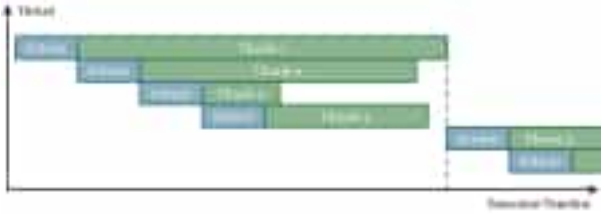


Figure 1. Illustration of chunk-based parallel processing

3.3 Axis Alignment of LiDAR point clouds

This section details the three steps of the axis alignment method that can avoid massive memory demands.

Normal estimation (Parallel computing). In the first step, we compute the normal vectors for each point. Given a point p_i, the set of neighboring points within a radius r are defined as $P = [p_1, p_2, ..., p_i, ...p_k]$. The covariance matrix of point p_i is estimated as

$$C(p_i) = \frac{1}{k} PP^T \tag{2}$$

The corresponding normal vector n_i is then obtained using Singular Value Decomposition (SVD):

$$n_i = \text{SVD}(C) \tag{3}$$

Rotation Estimation. Since the gravity direction is usually accurately provided by LiDAR scanners, the 3D rotation in axis alignment problem can be simplified to a 2D rotation estimation only around the z-axis. Specifically, by projecting each surface normal $\mathbf{n}i = (nxi, n_{yi}, n_{zi})$ onto the xy-plane, we can describe its horizontal orientation using a single rotation angle:

$$\theta_i = \arctan \frac{n_{yi}}{n_{xi}} \times \frac{180}{\pi}. \tag{4}$$

The whole angles range from $-180°$ to $180°$. We quantized the `float64` values to `uint16` integers by scaling with a factor of 100 to reduce memory usage. This is because $2^{16} = 65,536 > 360 \times 100$, all the possible angles can be represented with two-decimal precision. Therefore, three floating-point values representing normals can be compactly stored as integers without exceeding memory limits.

After chunk-based parallel processing, all rotation angles are collected in the range $[-180°, 180°)$. Since structural components such as walls, windows, and facades are often designed to fit orthogonal layouts (Fig. 3), their surface normals are distributed around 4 principal directions separated by $90°$ in the xy-plane. Leveraging this central-symmetric property, the angle distribution can be folded into the reduced range $[0°, 90°)$ to aggregate geometrically equivalent orientations, as illustrated in Fig. 2.

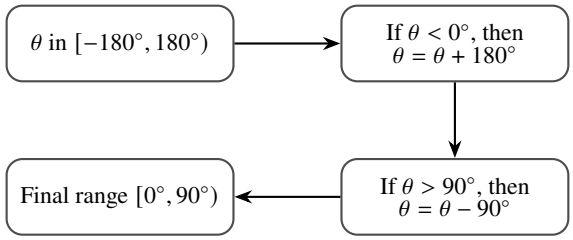


Figure 2. Range reduction from $[-180°, 180°)$ to $[0°, 90°)$

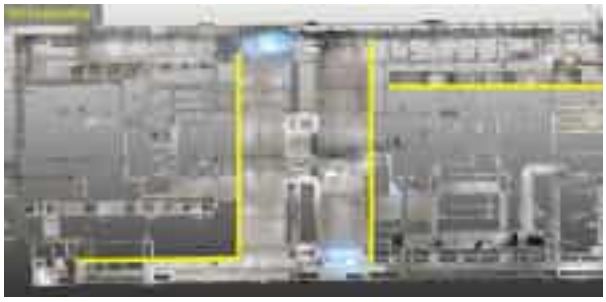


Figure 3. Illustration of the orthogonal structural components (e.g., walls) in LiDAR point clouds

Peak Direction Estimation In the final step, we extract the peak angle with the highest frequency in the range

$[0°, 90°)$. This angle denotes the dominant rotation angle between the current point cloud orientation and the global xy-axis. By rotating the point cloud around the z-axis by this estimated angle, the orthogonal building component can be parallel to the xy-axis. Note that a finer resolution in the bins of degrees can lead to more accurate peak direction estimation.

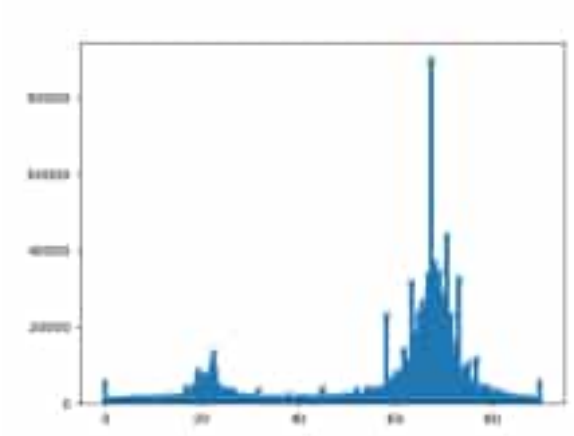


Figure 4. Degrees distribution of LiDAR point clouds normals in xy-plane

4 Experiments

4.1 Experimental Setup

To evaluate the proposed chunk-based parallel processing method for massive LiDAR point clouds, we conducted experiments on a real-world LiDAR scan dataset. The case study involves a real-world LiDAR point cloud dataset collected from the Hang Tai Road, Ma On Shan Area 86B Phase 2 construction project, a public housing development under the Hong Kong Housing Authority[1]. The scanned area covers approximately 3,392 m^2. The data were acquired using a NavVis VLX 3 system, which provides 5 mm accuracy and a capture rate of $2 \times 640,000$ points/s. The resulting LAS file is 22 GB in size and contains 789,448,875 points, with an average point density of approximately 232,738 points/m^2.

All experiments were evaluated on an HP Z2 workstation equipped with an Intel Core i7-13700K CPU, 64 GB RAM with 16 GB swap memory, running under the Windows Subsystem for Linux (WSL) environment.

Our evaluation focuses on the pipeline from reading the point cloud from disk to rotating it into alignment with the global coordinate axes. Specifically, there are three consecutive stages: (1) data loading, (2) normal estimation, and (3) axis alignment. We focus on these stages because they account for most of the memory demand throughout the point cloud preprocessing pipeline and often serve as the foundation for subsequent point cloud analyses. Therefore, evaluating these three stages can provide a comprehensive assessment of both memory efficiency and computational efficiency in very large point cloud processing.

[1]Approved Planning Brief from the Hong Kong Housing Authority

4.2 Results

We compared the proposed method with two baseline implementations, `Open3D`[2] and `pypcd4`[3], as well as a sequential chunks version of our method. As shown in Table 1, our proposed method significantly improves memory usage while maintaining computational efficiency.

Table 1. Comparison of methods for aligning the 789-million point cloud

Method	Runtime ↓	Peak RAM ↓
`Open3D` Baseline	N.A.†	>64 GB†
`pypcd4` Baseline	27.0 mins	60.5 GB
Sequential Chunks	98.6 mins	4.9 GB
Ours (this paper)	**25.6 mins**	**4.1 GB**

↓: Smaller is better; †: Failed at 23 mins.

Specifically, as illustrated in Fig. 5, `Open3D` takes around 30 minutes to load the point cloud into memory, occupying approximately 53.4 GB the system's available memory, and continues to allocate a large amount of memory, causing the process to be terminated by the system process manager. Compared with `Open3D`, which uses the `float64` format, `pypcd4` stores point coordinates in `float32` format in RAM for reducing both the memory demand (60.5 GB) and the time (27 mins) spent on memory allocation during evaluation.

The synchronization chunk processing method sequentially processes the chunk's point cloud, significantly reducing peak memory usage (Fig. 5). However, the total runtime is around 1 h 38 mins. In contrast, our proposed parallel chunk-based method achieves a trade-off between memory and time efficiency (Fig. 5). The whole process finishes within 26 minutes, which is 3.85-fold speedup compared with the synchronization chunk-based method. And the peak memory usage is also below 5 GB, considerably lower than the system's limits.

In addition to total runtime and peak RAM usage, Fig. 6 breaks the runtime into two stages, (1) data loading, (2) normal estimation and axis alignment. The results show

[2]https://github.com/isl-org/Open3D
[3]https://github.com/MapIV/pypcd4

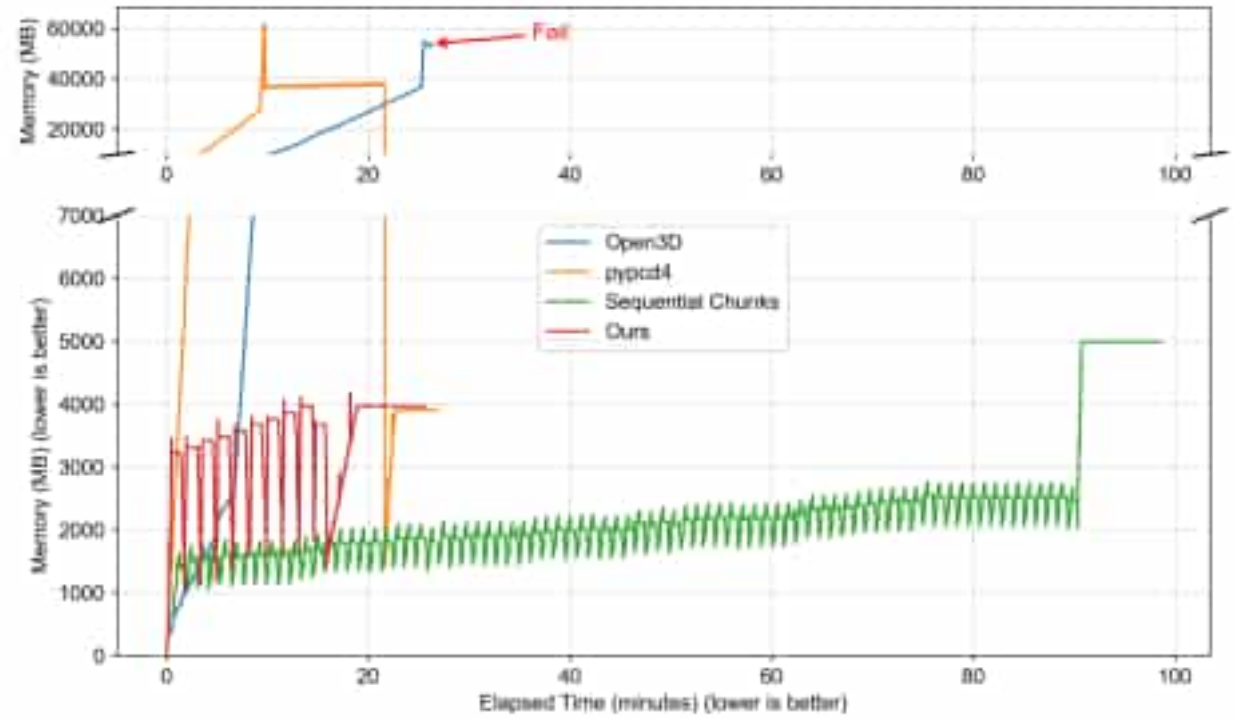


Figure 5. Memory usage comparison of `Open3D` (single batch), `pypcd4` (single batch), sequential chunk processing, and our proposed method

that most of the runtime is spent on data loading, indicating that I/O is the major bottleneck in this pipeline. Compared with the sequential chunk way, our proposed method significantly reduces the time spent on point cloud loading from disk, achieving a 4.98-fold speedup. This demonstrates that our proposed method efficiently avoids memory accumulation while maintaining time efficiency.

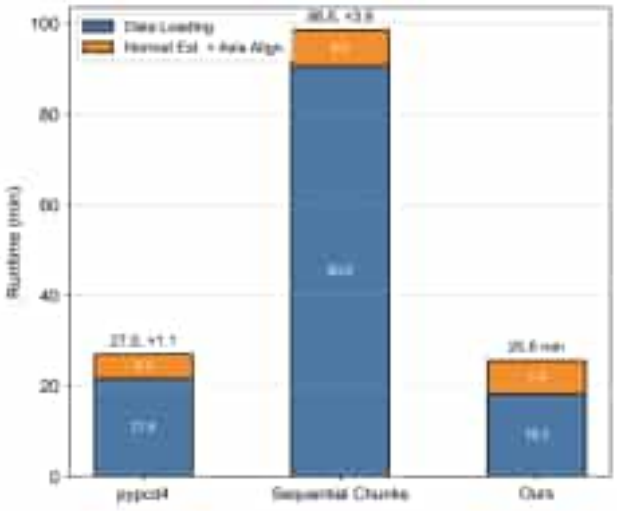


Figure 6. Runtime break down comparison of `pypcd4` (single batch), sequential chunk processing, and our proposed method

Axis Alignment Result After parallel processing, the point cloud was successfully aligned to the global 3D axis, as illustrated in Fig. 7. The red, green, and blue arrows represent the xyz-axis, respectively. The result demonstrates that our estimated dominant rotation angle enables efficient alignment of very large LiDAR point clouds to global coordinates.

Fig. 8 shows the registration result of LiDAR point clouds with the 2D drawings. As the point clouds have already been globally axis-aligned, the registration process only requires solving 2 degrees of freedom (translation along the xy-axes), significantly reducing human effort compared with full 3D registration.

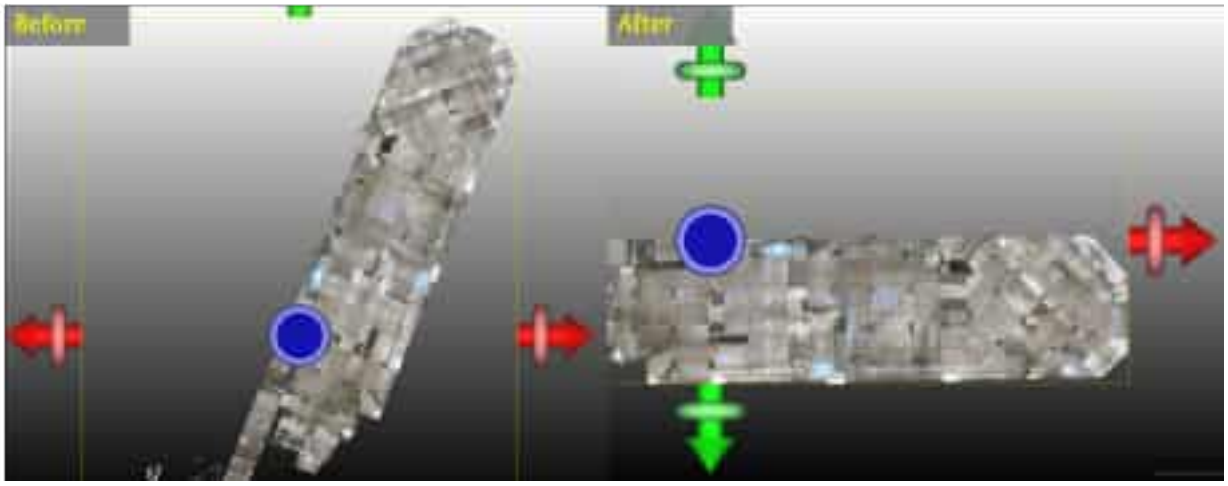


Figure 7. Axis alignment result of the LiDAR point clouds.



Figure 8. Registration result of 2D drawings with LiDAR point clouds. For better visualization, a sector along the z-axis is selected.

5 Discussion and Limitations

The experimental results demonstrated that the proposed chunk-based, orthogonality-enabled parallel computing method effectively estimates the global orientation of very large LiDAR point clouds and achieves both time and memory efficiencies. Compared with the single-batch loading approach in `pypcd4`, our method significantly reduces peak memory usage from over 60.5 GB to approximately 4.1 GB while maintaining comparable data loading time. These findings demonstrate that loading a very large point cloud into memory at once is inefficient when memory is limited. In contrast, aggregating local geometric cues from chunks without full memory access to the full point cloud can be a more efficient way to estimate global orientation and preprocess.

However, several limitations still exist. First, the proposed method currently supports only the LASer file format, although ".PLY" and ".PCD" also have random-access properties. Second, the hyperparameters, such as the chunk size and the number of concurrent processes, are set manually, leading to underutilization of RAM and CPU resources. Last but not least, the method was not tested across diverse building and construction scenes. More importantly, the assumption of dominant orthogonality may not fit for all buildings (e.g., curved facades and non-Manhattan layouts). In such cases, the degree distribution of surface normals may have multiple peaks, potentially degrading the global orientation estimation and failing the axis alignment. In addition, this study does not provide experiments on how chunk size affects runtime and memory usage, which should be further explored in future work.

6 Conclusion

This study presents a chunk-based, orthogonality-enabled method for automatically aligning the global axes of very large LiDAR point clouds. By leveraging the random-access property of the LASer file and controlling parallel concurrency, our proposed method efficiently achieves massive point cloud loading, normal estimation, and global axis alignment on very large LiDAR point clouds on computers. Experiments are conducted on a 22 GB real-world LASer file. The experimental results demonstrated that our method drastically reduces memory consumption and achieves a 3.85-fold speedup compared with the sequential chunk-based approach. The parallel computing mechanisms in this axis alignment method can also benefit other applications of LiDAR point clouds, including learning-based point cloud analysis, point cloud-to-CAD/BIM registration, change detection, and construction progress monitoring.

For future work, one can (1) extend the support for point cloud exchange formats, such as `pypcd4`'s ".PCD" and Stanford polygon ".PLY", (2) analyze the influence of chunk size and concurrency on the memory and CPU utilization, and (3) evaluate the method on a wider range of very large scanning datasets.

Acknowledgment

The data used in this study were collected and provided by Geosys Hong Kong Limited. The work described in this study was supported by the Hong Kong Research Grants Council (No. 17201325) and, in part, by the Hong Kong Innovation and Technology Commission (No. ITP/004/23LP).

References

[1] Mingyu Zhang, Yushu Yang, Shuai Han, Heng Li, Dongliang Han, Xiaotong Yang, and Nan Guo. Lidar-based framework for accurate positioning and robust tracking of multiple construction workers. *Journal of Computing in Civil Engineering*, 39(3): 04025027, 2025.

[2] Paul Sestras, Gheorghe Badea, Ana Cornelia Badea, Tudor Salagean, Sanda Roșca, Shuraik Kader, and Fabio Remondino. Land surveying with uav photogrammetry and lidar for optimal building planning. *Automation in Construction*, 173:106092, 2025.

[3] Wenyao Liu, Jinhua Chen, Zemin Lyu, Rui Feng, Tong Hu, and Lu Deng. Automatic tile position and orientation detection combining deep-learning and rule-based computer vision algorithms. *Automation in Construction*, 171:106001, 2025.

[4] Andrew Yarovoi and Yong Kwon Cho. Review of simultaneous localization and mapping (slam) for construction robotics applications. *Automation in Construction*, 162:105344, 2024.

[5] Feng Li, Wenzhong Shi, Yunlin Tu, and Hua Zhang. Automated methods for indoor point cloud preprocessing: Coordinate frame reorientation and building exterior removal. *Journal of Building Engineering*, 76:107270, 2023.

[6] Lin Bie, Shouan Pan, Siqi Li, Yining Zhao, and Yue Gao. Graphi2p: Image-to-point cloud registration with exploring pattern of correspondence via graph learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22161–22171, 2025.

[7] Ke You, Ke Chen, and Fan Xue. Automated scan-to-bim for construction digital transformation: Conceptual framework, processing methods and best-

practice guidelines. *Automation in Construction*, 182:106755, 2026.

[8] Iris De Gélis, Thomas Corpetti, and Sébastien Lefèvre. Change detection needs change information: Improving deep 3-d point cloud change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–10, 2024.

[9] Dong Liang, Longyong Wu, Meng Sun, Ruibo Hu, Lingming Kong, Yipeng Pan, and Fan Xue. Transfer learning from building information model-based synthetic data for three-dimensional module detection in point clouds of modular-integrated construction hoisting. *Engineering Applications of Artificial Intelligence*, 164:113243, 2026.

[10] Maxim Kolodiazhnyi, Anna Vorontsova, Anton Konushin, and Danila Rukhovich. Oneformer3d: One transformer for unified point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20943–20953, 2024.

[11] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018.

[12] Yijie Wu, Fan Xue, Maosu Li, and Sou-Han Chen. A novel building section skeleton for compact 3d reconstruction from point clouds: A study of high-density urban scenes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 209:85–100, 2024.

[13] Dong Liang, Sou-Han Chen, Zhe Chen, Yijie Wu, Louis YL Chu, and Fan Xue. 4d point cloud-based spatial-temporal semantic registration for monitoring mobile crane construction activities. *Automation in Construction*, 165:105576, 2024.

[14] Fan Xue, Weisheng Lu, Christopher J Webster, and Ke Chen. A derivative-free optimization-based approach for detecting architectural symmetries from 3d point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148:32–40, 2019.

[15] Fan Xue, Weisheng Lu, Ke Chen, and Anna Zetkulic. From semantic segmentation to semantic registration: derivative-free optimization–based approach for automatic generation of semantically rich as-built building information models from 3d point clouds. *Journal of Computing in Civil Engineering*, 33(4): 04019024, 2019.

[16] Yijie Wu, Fan Xue, Liangliang Nan, Longyong Wu, Jantien Stoter, and Anthony GO Yeh. Morphcut: an efficient convex decomposition method of 3d building models for urban morphological analytics. *International Journal of Geographical Information Science*, pages 1–23, 2025.

[17] Xiaolu Zeng, Yang Hu, Xiaopeng Yang, Zixiang Yin, and Shichao Zhong. A multi-source fusion system for through-wall radar compensation using lidar and slam-based 3d reconstruction. *IEEE Sensors Journal*, 2025.

[18] Chunran Zheng, Wei Xu, Zuhao Zou, Tong Hua, Chongjian Yuan, Dongjiao He, Bingyang Zhou, Zheng Liu, Jiarong Lin, Fangcheng Zhu, et al. Fast-livo2: Fast, direct lidar-inertial-visual odometry. *IEEE Transactions on Robotics*, 2024.

[19] Andre Samberg. An implementation of the asprs las standard. In *ISPRS Workshop on Laser Scanning and SilviLaser*, pages 363–372, 2007.

Energy-Intelligent Forecasting for Taiwanese Office Buildings: Developing a BIM-Derived Dataset and Integrating Advanced AI Models

Ngoc-Mai Nguyen[1], Minh-Tu Cao[2*] and Wei-Chih Wang[2]

[1]Department of Civil Engineering and Environmental Informatics, Minghsin University of Science and Technology, Taiwan
[2]Department of Civil Engineering, National Yang Ming Chiao Tung University, Taiwan

nnmai@must.edu.tw, mtcao@nycu.edu.tw, weichih@nycu.edu.tw

Abstract

Reliable and rapid energy forecasting is essential for early-stage decision-making in office building design, yet data-driven modeling in Taiwan is hindered by the lack of region-specific datasets that reflect local building typologies and climates. This study addresses this gap by introducing TOBE (Taiwanese Office Building Energy), a BIM-derived dataset developed for Taiwanese office buildings and designed to reflect regional characteristics. 1,006 simulated office-building cases were created to provide three energy targets: annual electricity (AE), energy use intensity (EUI), and lifecycle electricity (LE). To enable high-accuracy prediction without manual hyperparameter tuning, four machine learning learners (RBFNN, LSSVR, AdaBoost, and XGBoost) were hybridized with the Forensic-Based Investigation (FBI) metaheuristic optimizer and evaluated using 10-fold cross-validation. Results show that boosting-based learners consistently outperform NN- and SVM-based baselines. The best models achieved strong generalization, including AE with FBI_AdaBoost (MAPE = 3.45%; R^2 = 0.944), EUI with FBI_XGB (MAPE = 1.86%; R^2 = 0.954), and LE with FBI_AdaBoost (MAPE = 3.51%; R^2 = 0.945). The proposed BIM-to-AI pipeline offers a practical route to scalable, data-driven energy assessment tailored to Taiwanese office buildings.

Keywords –

BIM-derived dataset; Taiwanese office buildings; building energy prediction; boosting algorithm; Forensic-Based Investigation (FBI).

1 Introduction

Building energy consumption accounts for approximately 17.5% of global greenhouse gas emissions and around 30% of annual carbon emissions from the construction sector [1]. Worldwide, building energy utilization is projected to increase by 1.3% annually between 2018 and 2050, indicating a long term upward trend that challenges global decarbonization efforts [2]. In Taiwan, service and residential buildings consumed roughly 15% of the nation's total energy demand in 2020, with office buildings contributing significantly due to intensive cooling and electrical loads [2]. These figures highlight the pressing need for accurate and scalable prediction tools that can support energy efficient office building design and long-term sustainability planning in Taiwan's dense urban environments.

Traditional simulation based tools such as EnergyPlus and TRNSYS provide detailed insights into building energy performance, yet they require extensive input specifications and long computation times that limit their practicality during early design stages [3-5]. A single detailed simulation may require hours to complete, making these tools unsuitable for evaluating large numbers of design alternatives. Machine learning (ML) methods address these limitations by rapidly modeling nonlinear interactions among building features, and studies have shown that ML informed design decisions can yield potential energy savings of up to 30% [6]. However, the development of robust ML models in Taiwan remains constrained by the lack of large scale, high resolution, and region-specific datasets that accurately represent local office building characteristics.

Understanding the physical and operational characteristics of Taiwanese office buildings is essential for developing reliable datasets. Typical buildings in Taiwan feature rectangular or square floor plans, range from 9 to 26 stories, and incorporate façade systems such as glass curtain walls, aluminium panels, and brick that reflect contemporary architectural practice and green building requirements [7, 8]. Climatic variations across northern, central, and southern Taiwan further influence cooling loads, underscoring the importance of incorporating geographic diversity into dataset

construction [9]. Building Information Modeling (BIM) offers an effective solution by enabling detailed 3D representations of building geometry, materials, and operational parameters. Leveraging BIM alongside cloud-based simulation tools makes it possible to generate thousands of consistent models, ensuring comprehensive coverage of design variables relevant to energy analysis [10-12].

Recent advances in building energy prediction have shown that ensemble learning models consistently outperform single model approaches. For example, XGBoost has achieved mean absolute percentage errors as low as 1.92% for electricity consumption prediction, while Random Forest models have reached classification accuracies near 78.7% for Energy Use Intensity (EUI) analysis [13, 14]. Achieving optimal performance nevertheless depends on effective hyperparameter tuning, which often involves searching for large and complex parameter spaces. Metaheuristic optimization algorithms have been employed to address this challenge, and the Forensic Based Investigation (FBI) algorithm has demonstrated strong stability and fast convergence without requiring algorithm specific tuning parameters [15, 16]. These characteristics make FBI a promising optimization strategy for developing hybrid ML models aimed at building energy prediction.

Despite these advancements, no existing study has proposed a Taiwanese office buildings dataset to support ML based energy prediction. To address this research gap, this study introduces a comprehensive TOBE (Taiwanese Office Building Energy) dataset, generated from detailed BIM models that capture geometric, material, and environmental attributes of real-world office buildings in Taiwan. Using this dataset, hybrid machine learning models optimized with the FBI algorithm are developed to predict the three key energy indicators for Taiwanese office buildings, including annual electric (AE), EUI, and lifecycle electric (LE). The proposed framework provides a scalable and data driven approach that strengthens early-stage decision making and supports long term energy planning for Taiwan's commercial building sector, contributing a region specific and comprehensive foundation for advancing sustainable office building design.

2 Research Methodology

This research was guided by two primary objectives. The first objective was to leverage BIM techniques to create a novel and robust dataset of energy consumption patterns for typical office structures in Taiwan - the TOBE dataset. The second objective was to develop and train AI models to predict related energy consumption patterns in order to provide accurate and early-stage insights into energy consumption to minimize future building energy demand. This approach contributes to a more comprehensive and sustainable framework for construction. The entire research process is presented in **Fig. 1**.

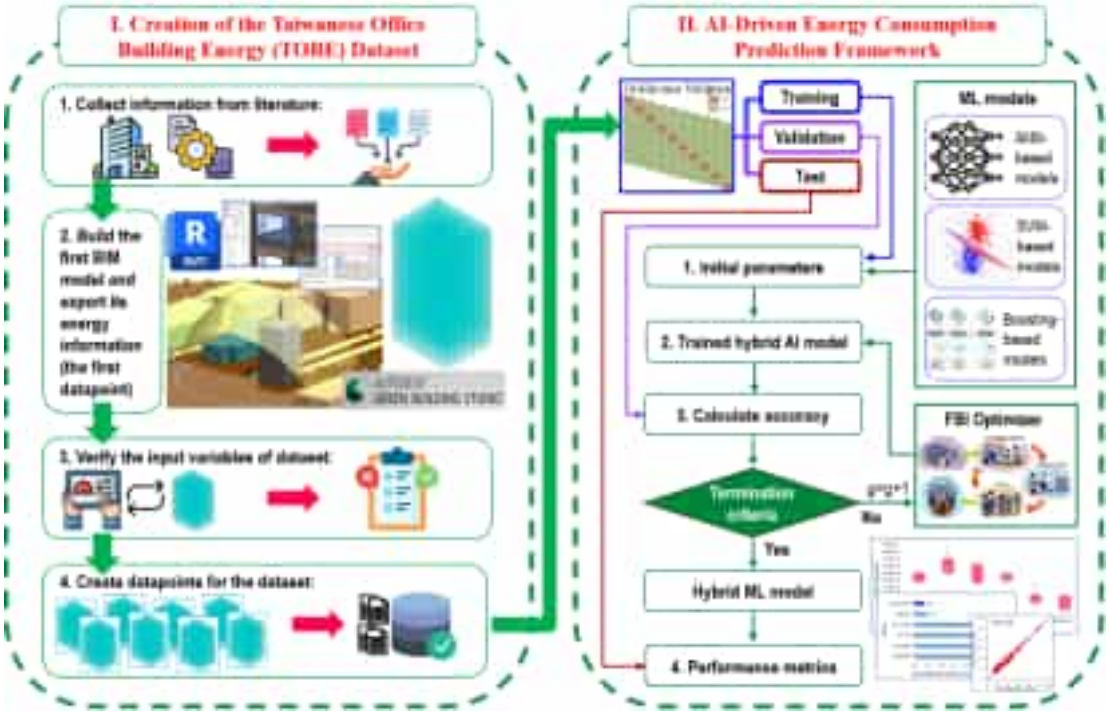


Fig. 1. Research process

2.1 TOBE Dataset Development

The TOBE dataset was established to represent the characteristics of contemporary office buildings in Taiwan systematically and to support machine learning based energy prediction. To identify the building attributes most relevant to energy consumption, a comprehensive review of previous studies and design guidelines, together with field surveys, was conducted for Taiwanese office buildings. Insights from literature, representative built examples, and field surveys highlighted consistent patterns in building geometry, façade systems, glazing configurations, and region specific climatic considerations. **Figure 2** illustrates some typical office buildings in Taiwan, and **Table 1** lists the initial set of variables derived from this investigation along with the typical ranges observed in Taiwanese office buildings. These variables capture the physical diversity required to construct realistic simulation scenarios for dataset creation.

A typical office building was modeled in Autodesk Revit using the characteristics identified earlier. All the input information required for this process was considered as input variables for the dataset. This initial model incorporated sixteen candidate variables, including: Floor area, number of stories, floor height, basement number, beam size, column size, floor thickness, stairs, roof thickness, walls, window materials, window-to-wall ratio, ceiling material, location, orientation, and HVAC system. After the BIM model was completed, the model's energy-related information was extracted using cloud-based Green Building Studio (GBS) software. Each variable was then independently modified and re-simulated to evaluate its impact on energy use, allowing non influential factors such as structural thicknesses, ceiling type, and stair configurations to be excluded.

Through this sensitivity based screening process, nine variables were identified as significant contributors to energy performance. These include floor area, floor height, number of stories, number of basements, exterior wall type, window type, window to wall ratio, geographic location, and orientation, as summarized in **Table 2**. By sampling across the ranges of these validated variables, a total of 1006 BIM models were generated and simulated to produce the final energy dataset. The value ranges of the three energy outputs derived from this dataset are also summarized in **Table 2**. Notably, these 1,006 TOBE cases were not produced through full-factorial enumeration; instead, constrained random sampling was used to achieve broad yet practical coverage of the feasible design space without generating an excessive number of combinations. As a result, the resulting TOBE dataset captures diverse combinations of geometry, envelope systems, and climatic conditions, ensuring a realistic and consistent foundation for training machine learning models aimed at early-stage energy assessment.



Chia Hsin Office, Chiayi city *Office complex building, Taipei* *ITRI office building, Hsinchu city*

Fig. 2. Some typical office buildings in Taiwan

Table 1. Key factors influencing energy consumption in Taiwan office buildings.

No.	Name	Information	Reference
1	Floor area	$480m^2 \sim 1470m^2$	[7]
2	Height of floor	3.2m ~ 4m	[7, 9]
3	Number of stories	9FL-26FL	[7]

<table>
<tr><td>4</td><td>External wall materials</td><td>Typical design of external walls in Taiwan office buildings<table><tr><th>Main material</th><th>Structure</th><th>U (W/m2 K)</th></tr><tr><td>Brick</td><td>Porcelain tiles: 10mm, Mortar: 15mm, Brick: 230mm, Mortar: 10mm</td><td>2.144</td></tr><tr><td>Aluminium plate</td><td>Aluminium plate: 6mm, Mineral wool panel: 20mm, Air layer, Calcium silicate: 25mm</td><td>1.251</td></tr><tr><td>Glass curtain</td><td>Toughened glass: 8mm, Air layer, Fiber surfboard: 4mm, Mineral wool insulation board: 32mm, Fiber surfboard: 4mm</td><td>0.981</td></tr><tr><td>Steel curtain</td><td>Enamel: 6mm, Steel plate: 3mm, Mineral wool panel: 20mm, Air layer, Fiber surfboard: 4mm, Mineral wool insulation: 32mm, Fiber surfboard: 4mm</td><td>0.711</td></tr></table></td><td>[17]</td></tr>
<tr><td>5</td><td>Window materials</td><td>Typical structure of glass windows in Taiwan office buildings<table><tr><th>Glass type</th><th>Thermal transmittance (W/m^2.K)</th><th>Shading coefficient (SC)</th><th>Visible light transmittance (VT)</th></tr><tr><td>Reflective glass: 10mm green-colored glass</td><td>5.67</td><td>0.57</td><td>0.21</td></tr><tr><td>Single layer low E glass: 6mm clear glass + G insulation paper</td><td>5.77</td><td>0.60</td><td>0.69</td></tr><tr><td>Two-layer low E glass: 6mm green glass + low E coating + 6mm clear glass</td><td>1.72</td><td>0.35</td><td>0.52</td></tr><tr><td>Three-layer low E glass: 6mm green glass + 9AS+ PET low E film + 9AS + 6mm clear glass</td><td>1.47</td><td>0.33</td><td>0.52</td></tr></table></td><td>[18]</td></tr>
<tr><td>6</td><td>WWR</td><td colspan="2">(Window-to-wall ratio) 20% ~ 76%</td><td>[7]</td></tr>
<tr><td>7</td><td>Location</td><td colspan="2">Taipei, Hsinchu, Taichung, Tainan, Kaohsiung, Taitung, Hualien</td><td>[9]</td></tr>
<tr><td>8</td><td>Orientation</td><td colspan="2">North, Northeast, East, South, Southeast, Southwest, West, Northwest</td><td></td></tr>
</table>

Table 2. Input and output variables of TOBE dataset.

No.	**Variable name**	**Value range**	**Note**
Input 1	Floor area	480 ~ 1470	Unit: m^2
Input 2	Floor height	3.2 ~ 4	Unit: m
Input 3	Number of stories	9 ~ 26	Natural number
Input 4	Basement number	0 ~ 3	Natural number
Input 5	External walls	1, 2, 3, 4	1 to 4 corresponding to brick, aluminium plate, glass curtain, and steel curtain, as presented in **Table 1**
Input 6	Windows	1, 2, 3, 4	1 to 4 corresponding to reflective glass, single-layer low E glass, double-layer low E glass, and three-layer low E glass, respectively, as presented in **Table 1**
Input 7	WWR	20% ~ 76%	If the type of external wall is a glass curtain, the WWR is equal to 0. For the others, WWR range is 20% ~ 76%
Input 8	Location	1, 2, 3, 4, 5, 6, 7	1 to 7 corresponding to Taipei, Hsinchu, Taichung, Tainan, Kaohsiung, Taitung, and Hualien, respectively
Input 9	Orientation	1, 2, 3, 4, 5, 6, 7, 8	1 to 7 corresponding to North, Northeast, East, South, Southeast, Southwest, West, and Northwest, respectively
Output 1	AE	713651 ~ 10842300	Unit: kWh
Output 2	EUI	518 ~ 1443	Unit: MJ/m²/year
Output 3	LE	21409.5 ~ 235.27	Unit: MW

2.2 AI-Based Energy Prediction

To develop an effective energy forecasting tool for Taiwanese office buildings using the TOBE dataset, four machine learning techniques were employed: Radial basis function neural network (RBFNN) representing

neural network (NN)-based models, least squares support vector regression (LSSVR) representing support vector machine (SVM)-based models, and two boosting algorithms (AdaBoost, XGBoost). These models were selected based on prior literature indicating their high performance in energy prediction tasks. Boosting methods are especially suitable here because the TOBE dataset is structured tabular data with a moderate sample size (1,006 cases), where tree-based ensembles often generalize well with minimal preprocessing and can efficiently capture nonlinear interactions among BIM-derived geometric, material, and climatic variables. Meanwhile, the NN- and SVM-based models were retained as reference learners to ensure a balanced comparison.

To improve predictive accuracy, the hyperparameters of each model were optimized using the FBI algorithm, a parameter free metaheuristic designed to iteratively refine candidate configurations through an adaptive search strategy. The FBI optimizer evaluates each candidate model using validation data and updates the hyperparameters until convergence or until a stopping criterion is met. This automated tuning process eliminates the need for manual adjustments and enhances the robustness of model calibration.

Model training followed a 10 fold cross validation procedure to minimize sampling bias, prevent overfitting and improve generalizability. Throughout the optimization and training process, model performance was assessed using standard regression metrics, including mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). These metrics provide complementary insights into prediction accuracy, relative error magnitude, and model explanatory power, ensuring a comprehensive evaluation of each algorithm.

3 Results and Discussion

3.1 Annual Electric

AE is one of the important outputs in TOBE dataset. It defined as the total annual electricity consumption of a building. Offering insights into electricity consumption over a year, this metric helps determine building operational costs and environmental impact.

Table 3. Performance of the four hybrid models.

Method		Testing data			
		RMSE	MAPE	MAE	R^2
FBI_ RBFNN	Mean	1091754	28.80%	777582	0.700
	Std.	142479	3.49%	82746	0.063
FBI_ LSSVR	Mean	1082352	22.60%	632261	0.701
	Std.	108131	3.57%	62177	0.070
FBI_XGB	Mean	489026	5.01%	145527	0.933
	Std.	172772	1.51%	43359	0.048
FBI_ AdaBoost	**Mean**	**396941**	**3.45%**	**98827**	**0.944**
	Std.	**235577**	**1.78%**	**43953**	**0.067**

Table 3 presents the performances of the four hybrid models. The results reveal that boosting algorithm-based models (FBI_XGB and FBI_AdaBoost) outperformed both NN-based and SVM-based models in all metrics. FBI_AdaBoost achieved the highest accuracy, with the lowest testing RMSE (396,941 kWh), MAPE (3.45%), and the highest R^2 (0.944). FBI_XGB followed closely, maintaining high predictive reliability and narrow error distribution. In contrast, NN-based models, particularly FBI_RBFNN, recorded the highest MAPE and lowest R^2, indicating limited suitability for this dataset.

Performance comparisons are illustrated in **Figs. 3-5**. MAPE and R^2 plots (**Figs. 3-4**) highlight the superior consistency of the BA-based models. Boxplots of MAE (**Fig. 5**) show tighter distributions and lower medians for FBI_AdaBoost and FBI_XGB, confirming their robustness and reduced estimation bias.

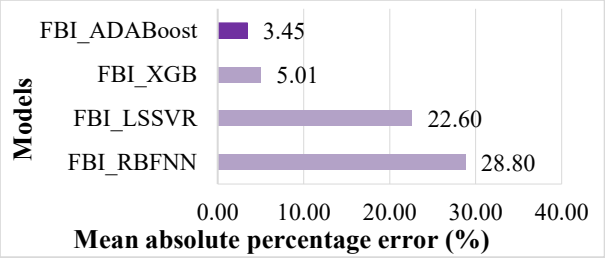


Fig. 3. Average MAPEs of hybrid models

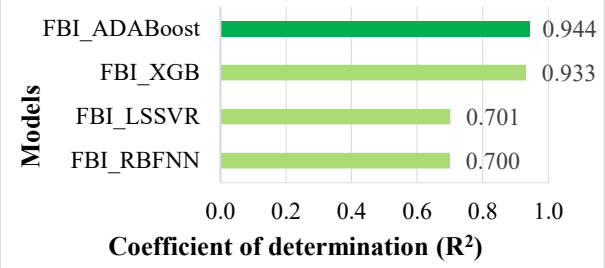


Fig. 4. Average R^2 of hybrid models

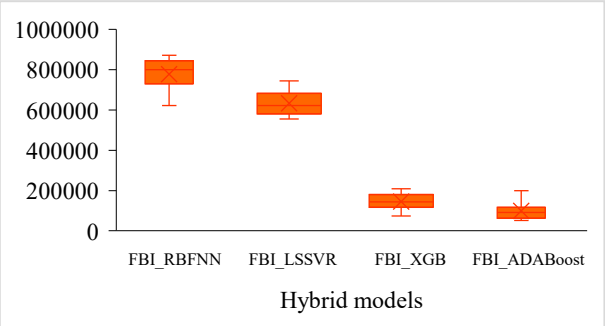


Fig. 5. MAE boxplot of hybrid models

These findings emphasize the critical role of model selection in energy prediction tasks. The BA-based models' strong generalization, high accuracy, and low variance make them well-suited for complex energy forecasting problems, particularly in data-rich environments like the TOBE dataset. Among them, FBI_AdaBoost emerges as the most effective model, offering practical utility for early-stage energy planning and decision-making in Taiwan's office building sector.

3.2 Energy Use Intensity

EUI is a straightforward but powerful measure of how efficiently a building uses energy. It is calculated by dividing the total energy consumed by a building in 1 year by its total floor area. Accurate EUI prediction models are essential for effective energy management, optimisation, and sustainability and can achieve potential energy savings of up to 30%. EUI is one of the energy performance indicators that can be extracted from GBS.

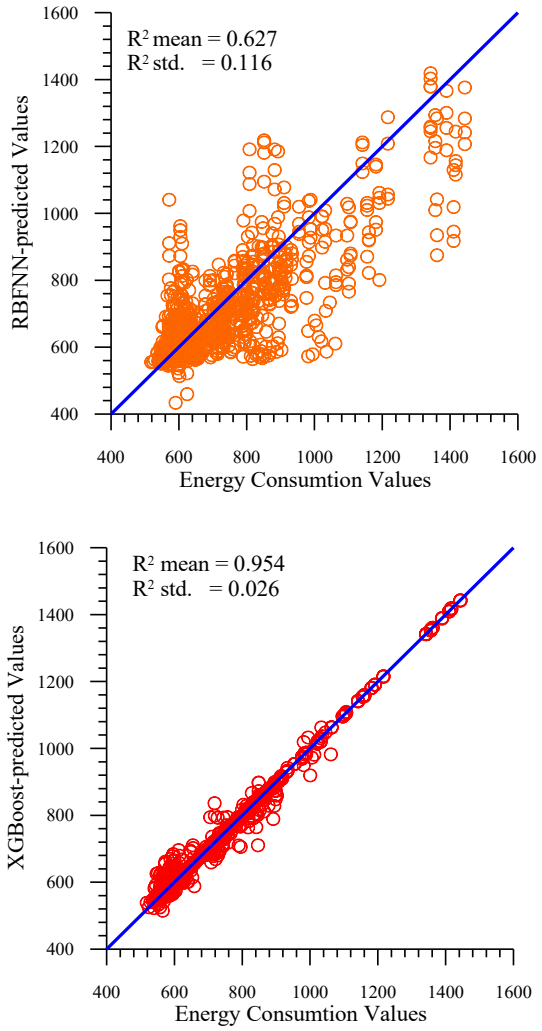


Fig. 6. Scatter plot of the EUI models

Table 4. Performance of the four hybrid models.

Method		Testing data			
		RMSE	MAPE	MAE	R^2
FBI_ RBFNN	Mean	113.22	9.66%	75.69	0.627
	Std.	14.85	1.23%	9.48	0.116
FBI_ LSSVR	Mean	92.65	8.67%	62.55	0.746
	Std.	17.14	1.14%	9.48	0.107
FBI_XGB	**Mean**	**37.98**	**1.86%**	**12.60**	**0.954**
	Std.	**11.81**	**0.50%**	**3.23**	**0.026**
FBI_ AdaBoost	Mean	46.09	2.85%	19.49	0.932
	Std.	12.91	0.79%	5.32	0.031

As seen in **Table 4**, the RBFNN and LSSVR models performed relatively poorly across all metrics. RBFNN gave the weakest results, with the highest RMSE (113.22 ± 14.85 MJ/m²/year) and MAPE (9.66% ± 1.23%) on the test set. In contrast, XGBoost achieved the best overall performance, with the lowest RMSE (37.98 ± 11.81 MJ/m²/year), lowest MAPE (1.86% ± 0.50%), and highest R2 (0.954 ± 0.026), demonstrating strong predictive accuracy and generalization. FBI_AdaBoost ranked second, showing that boosting-based methods outperformed the NN- and SVM-based models.

Figure 6 further illustrates the contrast between RBFNN, the weakest model, and XGBoost, the best-performing model, through predicted-versus-simulated energy consumption plots. RBFNN shows wide dispersion around the 45° reference line, indicating larger prediction errors and only moderate explanatory power, whereas XGBoost predictions cluster closely around the line with only minor deviations. This confirms that XGBoost provides substantially more accurate and robust energy consumption predictions than RBFNN.

3.3 Lifecycle Electric

In Autodesk GBS, the LE indicator quantifies the total electricity consumption of a building over a specified analysis period, typically aligned with its expected service life. It aggregates annual electric use from all major end uses, such as lighting, plug loads, HVAC systems, and auxiliary equipment, and projects this consumption forward using consistent operating assumptions. This indicator is often coupled with utility rate information to estimate long-term operating costs. As a result, LE provides a useful basis for evaluating the economic and environmental benefits of energy efficiency strategies across the whole building life cycle.

Based on the two preceding results, boosting-based models provide the strongest predictive performance on the TOBE dataset. Accordingly, this section employs FBI_AdaBoost and FBI_XGB to develop prediction models for lifecycle electricity in Taiwanese office buildings. **Table 5** summarizes the 10-fold cross-validation performance of both models. For visualization,

Figs. 7 and 8 compare the fold-wise test MAPE and R^2 values across the 10 folds, respectively.

Across the 10-fold cross-validation, both boosting models demonstrate strong predictive capability for LE in Taiwanese office buildings. FBI_AdaBoost achieves the best average performance, yielding lower mean errors (RMSE = 10,664; MAE = 2,891) than FBI_XGB (RMSE = 13,932; MAE = 4,146), together with a higher mean coefficient of determination (R^2 = 0.945 vs. 0.931). The low mean MAPE values (3.51% for FBI_AdaBoost and 5.02% for FBI_XGB) further indicate good relative accuracy, supporting the suitability of boosting-based learners for this dataset.

Table 5. 10-fold performances of boosting models.

FBI_XGB (testing performance)				
Fold	**RMSE (KW)**	**MAPE (%)**	**MAE (KW)**	**R2**
1	18,156.5	6.14	4,911.3	0.929
2	9,583.5	7.58	4,959.1	0.977
3	17,235.9	4.30	3,275.6	0.920
4	5,375.0	2.57	2,027.0	0.994
5	10,931.0	4.31	3,613.4	0.942
6	22,697.5	7.48	6,041.2	0.849
7	7,504.9	4.51	3,013.6	0.985
8	10,974.1	3.67	3,543.9	0.951
9	12,978.4	3.71	4,000.9	0.958
10	23,877.9	5.89	6,069.4	0.808
Mean	**13,931.5**	**5.02**	**4,145.5**	**0.931**
Std.	**6,294.8**	**1.68**	**1,322.3**	**0.060**

FBI_AdaBoost (testing performance)				
Fold	**RMSE (KW)**	**MAPE (%)**	**MAE (KW)**	**R2**
1	15,842.1	5.01	3,714.9	0.946
2	5,737.8	4.11	1,952.4	0.992
3	17,689.7	4.22	3,316.0	0.916
4	5,992.8	1.98	1,611.8	0.993
5	7,406.1	2.93	2,375.7	0.973
6	6,107.6	2.94	2,688.0	0.989
7	3,341.9	2.60	1,402.2	0.997
8	7,598.2	2.93	2,511.1	0.976
9	5,698.4	2.22	1,765.2	0.992
10	31,229.0	6.16	7,572.6	0.671
Mean	**10,664.4**	**3.51**	**2,891.0**	**0.945**
Std.	**8,590.8**	**1.33**	**1,802.1**	**0.099**

Nevertheless, robustness differs across folds. FBI_XGB exhibits more stable explanatory power (std. R^2 = 0.060) and does not show extreme deterioration, although weaker generalization is observed in a small number of splits. A brief diagnostic comparison of **Table 5** shows that Fold 10 is the most difficult test partition for both boosting models: FBI_XGB records its lowest R^2 (0.808) and highest RMSE (23,877.9) in this fold, while FBI_AdaBoost shows an even larger drop (R^2=0.671; RMSE = 31,229.0; MAE = 7,572.6). This concurrent degradation suggests that Fold 10 likely contains harder-to-predict or more extreme cases than the remaining folds. Therefore, the Fold 10 result is interpreted as evidence of a challenging partition, while the larger decline of FBI_AdaBoost indicates that it is more sensitive than FBI_XGB to such cases, despite achieving better average LE prediction performance overall.

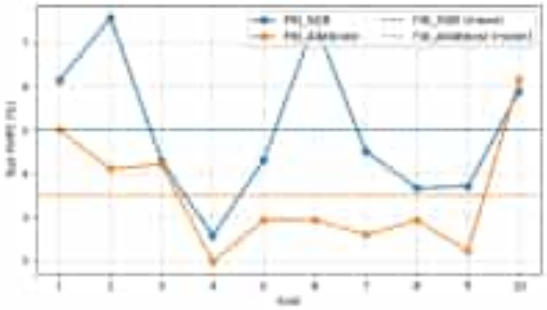


Fig. 7. MAPE comparison across 10 folds

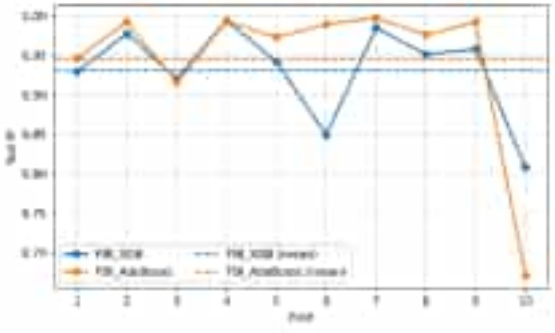


Fig. 8. R^2 comparison across 10 folds

4 Conclusions

This study aims to (i) establish a region-specific, simulation-consistent dataset for Taiwanese office buildings and (ii) develop a high-accuracy prediction framework suitable for early design without labor-intensive manual calibration. The proposed TOBE dataset was generated via a BIM-to-simulation workflow, in which representative office-building geometries were modelled in Autodesk Revit and evaluated using Green Building Studio. A sensitivity-based screening procedure reduced the input space to nine influential variables that are physically meaningful and readily extractable from standard BIM deliverables. The final dataset comprises 1,006 cases and includes three electricity-related targets: annual electricity, energy use intensity, and lifecycle electricity.

Four learning algorithms (RBFNN, LSSVR, AdaBoost, and XGBoost) were then optimized using the FBI metaheuristic and assessed through 10-fold cross-validation, with results indicating that boosting-based models are consistently the most effective for TOBE. From a feasibility and application standpoint, the proposed pipeline is designed for real-world deployment in early-stage decision-making. The nine retained inputs can be obtained directly from BIM or basic design

documentation, enabling rapid estimation of AE, EUI, and LE without repeatedly running computationally expensive simulations.

Overall, the combination of TOBE and the FBI-optimized boosting models provides a scalable foundation for data-driven electricity forecasting to support energy-efficient design and planning in Taiwanese office buildings. Because the inputs were pre-screened through BIM-based sensitivity analysis, the framework already focuses on energy-relevant variables. Nevertheless, enhancing model interpretability remains an important next step. Future research should quantify the relative influence of these variables using feature-importance techniques such as XGBoost gain and SHAP values, thereby enabling the predictive framework to provide more explicit and actionable support for design decision-making. In addition, the framework should be further validated using measured operational data, when available, to assess its real-world applicability and transferability.

Acknowledgments

The authors would like to thank the National Science and Technology Council, Taiwan for financially supporting this research under contract 114-2221-E-A49 -020 -.

References

[1] X. Wu *et al.*, "BIM-driven building greenness evaluation system: An integrated perspective drawn from model data and collective experts' judgments," *Journal of Cleaner Production,* vol. 406, p. 136883, 2023/06/20/ 2023, doi: https://doi.org/10.1016/j.jclepro.2023.136883.

[2] W.-T. Tsai and C.-H. Tsai, "Interactive analysis of green building materials promotion with relevance to energy consumption and greenhouse gas emissions from Taiwan's building sector," *Energy and Buildings,* vol. 261, p. 111959, 2022/04/15/ 2022, doi: https://doi.org/10.1016/j.enbuild.2022.111959.

[3] C. N. Antoniadis and G. Martinopoulos, "Optimization of a building integrated solar thermal system with seasonal storage using TRNSYS," *Renewable Energy,* vol. 137, pp. 56-66, 2019/07/01/ 2019, doi: https://doi.org/10.1016/j.renene.2018.03.074.

[4] F. Bu, D. Yan, G. Tan, and J. An, "A novel approach based on equivalent sky radiative temperature for quick computation of radiative cooling in building energy simulation," *Renewable Energy,* vol. 221, p. 119820, 2024, doi: https://doi.org/10.1016/j.renene.2023.119820.

[5] P. Shirazi, A. Behzadi, P. Ahmadi, M. A. Rosen, and S. Sadrizadeh, "Comparison of control strategies for efficient thermal energy storage to decarbonize residential buildings in cold climates: A focus on solar and biomass sources," *Renewable Energy,* vol. 220, p. 119681, 2024/01/01/ 2024, doi: https://doi.org/10.1016/j.renene.2023.119681.

[6] A. Colmenar-Santos, L. N. T. de Lober, D. Borge-Diez, and M. Castro-Gil, "Solutions to reduce energy consumption in the management of large buildings," *Energy and Buildings,* vol. 56, pp. 66-77, 2013, doi: https://doi.org/10.1016/j.enbuild.2012.10.004.

[7] K.-T. Huang and H.-T. Lin, "Development of Simplified Estimation Method of Chiller Energy Use for Office Buildings in Taiwan," presented at the ASHRAE Transactions, 2007.

[8] Y.-S. Tsay, C.-H. Yang, and C.-Y. Yeh, "Evaluation of Expanded Metal Mesh Applied on Building Facades with Regard to Daylight and Energy Consumption: A Case Study of an Office Building in Taiwan," *Buildings*, vol. 12, no. 8, doi: https://doi.org/10.3390/buildings12081187.

[9] N.-T. Ngo, "Early predicting cooling loads for energy-efficient design in office buildings by machine learning," *Energy and Buildings,* vol. 182, pp. 264-273, 2019/01/01/ 2019, doi: https://doi.org/10.1016/j.enbuild.2018.10.004.

[10] D. L. Luong, N.-S. Truong, N.-T. Ngo, and N.-Q. Nguyen, "Developing the hybrid BIM-BEM and jellyfish search optimization system for optimizing energy consumption and building installation costs," *Scientific Reports,* vol. 14, no. 1, p. 17186, 2024/07/26 2024, doi: https://doi.org/10.1038/s41598-024-68021-6.

[11] Y. Pan *et al.*, "Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies," *Advances in Applied Energy,* vol. 10, p. 100135, 2023/06/01/ 2023, doi: https://doi.org/10.1016/j.adapen.2023.100135.

[12] H. X. Li, Z. Ma, H. Liu, J. Wang, M. Al-Hussein, and A. Mills, "Exploring and verifying BIM-based energy simulation for building operations," *Engineering, Construction and Architectural Management,* vol. 27, no. 8, pp. 1679-1702, 2020, doi: https://doi.org/10.1108/ECAM-06-2019-0314.

[13] J. H. C. G. Tang and Z. Lu, "Use of publicly available data to create a dataset for data-driven urban commercial building energy intensity classification: Model accuracy, interpretation, and implications of an open data framework in Hong Kong," *Sustainable Cities and Society,* vol. 100, p. 105063, 2024/01/01/ 2024, doi: https://doi.org/10.1016/j.scs.2023.105063.

[14] X. Li, Z. Wang, C. Yang, and A. Bozkurt, "An advanced framework for net electricity consumption prediction: Incorporating novel machine learning models and optimization algorithms," *Energy,* vol. 296, p. 131259, 2024/06/01/ 2024, doi: https://doi.org/10.1016/j.energy.2024.131259.

[15] J.-S. Chou and N.-M. Nguyen, "FBI inspired meta-optimization," *Applied Soft Computing,* p. 106339, 2020/04/28/ 2020, doi: https://doi.org/10.1016/j.asoc.2020.106339.

[16] H. M. Moghaddam *et al.*, "Shear modulus prediction of landfill components using novel machine learners hybridized with forensic-based investigation optimization," *Construction and Building Materials,* vol. 411, p. 134443, 2024/01/12/ 2024, doi: https://doi.org/10.1016/j.conbuildmat.2023.134443.

[17] S.-J. Guo and T. Wei, "Cost-effective energy saving measures based on BIM technology: Case study at National Taiwan University," *Energy and Buildings,* vol. 127, pp. 433-441, 2016/09/01/ 2016, doi: https://doi.org/10.1016/j.enbuild.2016.06.015.

[18] C.-M. Lai and Y.-H. Wang, "Energy-Saving Potential of Building Envelope Designs in Residential Houses in Taiwan," *Energies*, vol. 4, no. 11, pp. 2061-2076doi: https://doi.org/10.3390/en4112061.

Accurate Geometric Reconstruction and High-Fidelity Photorealistic Rendering of Reflective Glass Façades using Gaussian Splatting with a Novel Reflection MLP

Zhenyu Liang[1 2], Xiao Zhang[1], Boyu Wang[3], Ang Li[1], Zhaolun Liang[1], Jeff Chak Fu Chan[1], Mingzhu Wang[4], Jack Chin Pang Cheng[1]

[1]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong S.A.R. (China)
[2]Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA
[3]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[4]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong S.A.R. (China)
zhenyu.liang@connect.ust.hk, xiao.zhang@connect.ust.hk, bw3005@nyu.edu, lionel_a.li@connect.ust.hk, zliangaq@connect.ust.hk, cfchanay@connect.ust.hk, m.wang@cityu.edu.hk, cejcheng@ust.hk.

Abstract –

Generating building digital twins from UAV imagery via 3D reconstruction is significant. However, reflective glass façades remain particularly challenging: their high-frequency, view-dependent appearance often causes photogrammetry to yield distorted geometry and corrupted textures. While Gaussian Splatting supports view-dependent rendering, its spherical harmonics color model mainly captures low-frequency components and fails to reconstruct high-frequency specular reflections. In addition, vanilla Gaussian Splatting tends to "explain" reflections by reconstructing ghost geometry behind the glass surface and producing a visually plausible view-dependent appearance, rather than recovering the true façade geometry. To address these issues, this paper proposes a novel Gaussian Splatting–based reconstruction and rendering framework for reflective glass façades, enabling both accurate geometry recovery and high-fidelity photorealistic rendering. First, UAV images are processed using photogrammetric reconstruction to obtain a coarse façade mesh with distorted geometry. Second, individual glass panels are segmented in 2D images using structural regularities and then projected onto the coarse mesh to correct geometric distortions in the façade. Third, using the corrected façade mesh as a geometric prior, a Gaussian Splatting model is trained with a rendering formulation that considers both diffuse color and specular reflection. For specular reflection, a novel reflection Multilayer Perceptron (MLP) is developed that takes the 3D surface position and the reflection direction as input to efficiently model near-field illumination with low computational and memory costs. Experiments on two buildings in Hong Kong show that our method reliably reconstructs reflective façade geometry and outperforms state-of-the-art approaches in rendering high-frequency, view-dependent appearances.

Keywords –

Gaussian Splatting; Reflective Glass Façades; 3D Reconstruction; Photorealistic Rendering; Reflection MLP; Digital Twin Modelling

1 Introduction

Using UAV-acquired imagery to reconstruct a building's 3D model can serve as a foundation for a digital twin [1-4]. Such a model can be further converted into a BIM representation [5-7], or used to inspect and annotate façade-level defects. This research direction is therefore highly valuable. However, for buildings with reflective glass façades, this pipeline faces challenges.

Optics-based passive 3D reconstruction, commonly referred to as photogrammetry, fundamentally relies on extracting and matching image keypoints across views and leveraging epipolar geometry to recover both the 3D coordinates of scene points and the camera poses [1]. A core assumption is multi-view consistency, i.e., that the same physical point exhibits similar appearance in overlapping images and can thus be reliably matched. For reflective surfaces, this assumption is violated: the observed color at the same surface point can differ significantly between adjacent viewpoints because the reflection directions change significantly. As a result, photogrammetry-based reconstructions often suffer from

severe geometric and textural distortions when applied to reflective façades.

For such view-dependent textures on reflective glass, conventional point cloud or mesh representations are inadequate, as they can only represent static textures. In recent years, learning-based 3D reconstruction and rendering methods such as Neural Radiance Fields (NeRF) [8] and Gaussian Splatting (GS) [9] have emerged. NeRF trains an implicit MLP that takes a 5D input (3D position and 2D viewing direction) and outputs color together with a density value representing geometry. In contrast, GS represents scenes as an explicit set of Gaussian geometries and renders novel views via rasterization. Compared with NeRF, GS enables real-time rendering and faster training, making it more suitable as a base representation for building digital twins.

GS typically models view-dependent color using spherical harmonics (SH). While SH can capture view dependence to some extent, it is primarily effective for low-frequency effects (i.e., appearance changes that vary smoothly with camera pose.) Specular reflections, however, are high-frequency view-dependent effect and can change dramatically with viewpoint. Consequently, SH provides an inadequate representation for reflections on glass façades.

Recent studies have attempted to enhance GS for reflection modeling. One line of work introduces a 2D environment map to represent global illumination and queries it using reflection directions to reproduce high-frequency, view-dependent appearance [10-11]. However, such queries inherently assume illumination originates from infinity. In real-world 3D environments, reflections on glass façades are induced by a near-field light field, where contributing sources lie at varying distances. Representing this with a 2D environment map leads to parallax errors. Another line of work attempts to reconstruct the 3D environmental light field [12], e.g., by fitting millions of "environment Gaussians" to approximate where in 3D space light of different colors is emitted and how it reaches and reflects off the façade. Although this explicitly approximates the reflection formation process, inferring a complete 3D environment solely from reflections is highly complex, and representing the entire surrounding scene with Gaussians leads to substantial computation and storage costs.

To address these limitations, this paper proposes a new approach: we train a reflection Multilayer Perceptron (MLP) that takes as input the 3D position on the reflective surface and the reflection direction computed from the viewing direction and the surface normal, and outputs the corresponding color. In effect, the MLP learns a surface light field, enabling efficient modeling of near-field illumination effects without reconstructing the entire environment, thereby providing an effective and computationally efficient solution.

Effective training of the reflection MLP requires accurate surface geometry and normal directions of the reflective façade. Photogrammetry often produces distorted and incomplete geometry for reflective surfaces. Notably, vanilla GS also tends to place Gaussians behind reflective surfaces to reproduce the illusion of reflection in rendering, rather than reconstructing the true reflective surface itself. Therefore, we further propose a new geometric correction method for reflective glass façades by combining (i) the distorted coarse façade mesh obtained from photogrammetry and (ii) individual 3D panel models derived from façade panel segmentation, resulting in an accurate reconstruction of the reflective glass façade surfaces.

In summary, our main contributions are as follows: (1) proposing a novel model that integrates GS with a reflection MLP to efficiently model near-field specular reflections. By decomposing diffuse color and reflection via a shading formulation, our method enables high-fidelity photorealistic rendering of reflective glass façades; (2) developing a mesh correction approach that combines photogrammetry with a façade panel segmentation model, enabling accurate reconstruction of reflective glass façade geometric surfaces; (3) validating the proposed method on two real-world buildings with reflective glass façades in Hong Kong, and the experimental results demonstrate that our method outperforms state-of-the-art methods in rendering high-frequency, view-dependent facades' appearance.

2 Preliminary of Gaussian Splatting

This paper adopts 2D Gaussian Splatting (2DGS) [13] as the basic primitive for reconstructing and rendering reflective glass façades. Unlike 3D Gaussian Splatting (3DGS), which represents scenes using volumetric Gaussian ellipsoids, 2DGS represents objects using Gaussian elliptical disks. This representation improves multi-view consistency in both surface normals and depth, facilitating accurate querying of façade surface geometry from different camera viewpoints.

Each 2D Gaussian is parameterized by a central point p_k, two principal tangential vectors t_u and t_v, two scaling vectors s_u and s_v, an opacity α, a color coefficient vector c represented using spherical harmonics. In the local tangent plane defined in world space, the 2D Gaussian can be formulated by Eq. (1) and Eq. (2). Given local coordinates $\mathrm{u} = (u, v)$ on the tangent plane, the corresponding Gaussian value is computed by Eq. (3).

$$P(u,v) = p_k + s_u t_u u + s_v t_v v = H(u,v,1,1)^{\top} \quad (1)$$

$$H = \begin{bmatrix} s_u t_u & s_v t_v & 0 & p_k \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} RS & p_k \\ 0 & 1 \end{bmatrix} \quad (2)$$

$$\mathcal{G}(\mathrm{u}) = \exp\left(-\frac{u^2 + v^2}{2}\right) \quad (3)$$

Based on these definitions, rendering in 2DGS amounts to establishing the correspondence between a camera pixel $\mathrm{x} = (x, y)$ and the local tangent-plane coordinates $\mathrm{u} = (u, v)$ of each 2D Gaussian. This can be computed via a ray–splat intersection, where $W \in \mathbb{R}^{4\times4}$ denotes the transformation matrix from world space to screen space:

$$\mathrm{x} = (xz, yz, z, 1)^{\top} = WH(u, v, 1, 1)^{\top} \quad (4)$$

After obtaining the mapping from pixel $(x, y)$to the local coordinates (u, v), the Gaussian contribution to that pixel is evaluated using Eq. (3), and the intersection depth z is obtained from Eq. (4). The final pixel color is then computed by rasterization using volumetric alpha blending:

$$\mathrm{c(x)} = \sum_{i=1} c_i \alpha_i \hat{\mathcal{G}}_i(\mathrm{u(x)}) \prod_{j=1}^{i-1} (1 - \alpha_j \hat{\mathcal{G}}_j(\mathrm{u(x)})) \quad (5)$$

where $\hat{\mathcal{G}}$ denotes a low-pass filtered version of the Gaussian kernel, introduced to mitigate unstable optimization caused by strongly slanted viewpoints.

3 Methodology

The proposed Gaussian Splatting–based reconstruction and rendering framework is illustrated in Fig. 1. It consists of two main stages: (1) Geometric Reconstruction and Correction, and (2) Gaussian Splatting Reconstruction and Rendering.

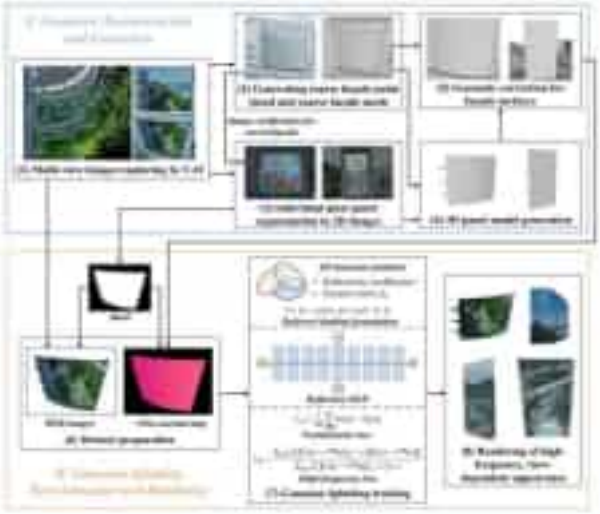


Fig. 1 Overall framework for the proposed method

3.1 Geometric Reconstruction and Correction for Reflective Glass Façade

The UAV-captured images were first used to generate a geometrically distorted coarse façade mesh and the corresponding coarse façade point cloud via photogrammetric 3D reconstruction. Notably, the coarse façade mesh exhibits uneven and broken glass panel surfaces, while the coarse façade point cloud contains missing points on the glass panels, as illustrated in Fig. 2.

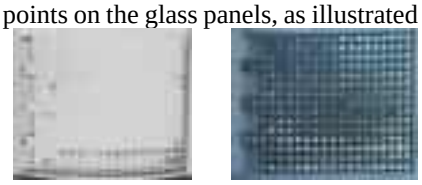


Fig. 2 Coarse façade mesh (left) and coarse façade point cloud (right) from photogrammetric 3D reconstruction

3.1.1 Image Rectification for Curved Façades

In the 2D images, the glass-panel segmentation method proposed in the authors' prior work is adopted. Leveraging the structural regularities of glass panels, specifically linear features and repetitive patterns, this method formulates a maximum a posteriori (MAP) estimation problem to extract panel corner points for the segmentation of the individual glass panels (see [14] for details). This method is effective in mitigating the adverse impact of reflections and background clutter on panel segmentation, and it performs well on frontal-view façade images. However, because it relies on line-based features, it is directly applicable to planar (flat) façade surfaces. For curved façades, an image rectification step is required to warp the curved surface into a planar surface in the image before segmentation.

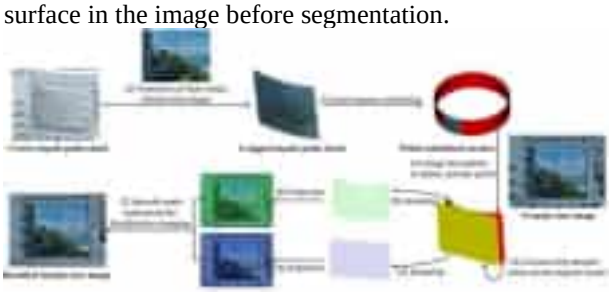


Fig. 3 Image rectification for curved façade surfaces

Fig. 3 illustrates the image rectification process for curved façades. First, a close-range frontal-view image is captured near the façade surface. This image contains only the façade, without any background clutter. Using the 3D-to-2D projection relationship between the coarse façade point cloud (3D points (X, Y, Z)) and the image pixels (u, v), as given in Eq. (6), the subset of 3D points that project into the image is extracted to form a cropped point cloud. A cylindrical surface is then fitted to the cropped point cloud via least-squares estimation, yielding the cylinder equation that represents the curved façade geometry.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6)$$

where K is the camera intrinsic matrix; R and T are extrinsic rotation matrix and translation vector, respectively.

For image that requires panel segmentation, the image boundaries are used to determine the corresponding extent of the fitted cylindrical surface, thereby defining a patch on the cylinder (highlighted in red in the figure). This cylindrical patch is first sampled to obtain a set of surface points (green points). A tangent plane is then constructed at the angular center of the patch (yellow plane), and the sampled surface points are orthogonally projected onto this plane to form a planar point set (blue points).

Both the original cylindrical surface points (green) and their planar counterparts (blue) are projected onto the target image. In this projection, the green points indicate the pixel locations corresponding to the true 3D geometry of the curved façade, whereas the blue points indicate the pixel locations after unfolding the curved façade into a planar façade. Consequently, rectifying a curved-façade image into a planar-façade image amounts to estimating a mapping from the green-point locations to the blue-point locations in the image. A seventh-order polynomial is used to fit this mapping, enabling effective image rectification for curved façades.

3.1.2 Glass Panel Segmentation and Projection for 3D Panel Model Generation

For frontal-view image of planar façade and rectified frontal-view image of curved façade, glass-panel corner points are extracted using structural regularities-based MAP estimation. By pairing the four corner points belonging to the same panel, each glass panel can be segmented. For curved façades, the corner points extracted in the rectified image are further mapped back to the unrectified image using the fitted seventh-order polynomial. Fig. 4 shows the glass-panel segmentation results for both planar and curved façades.

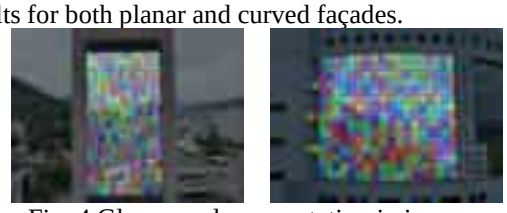


Fig. 4 Glass panel segmentation in images

The extracted glass-panel corner points are further projected onto the corresponding coarse façade mesh via ray casting, producing the 3D locations of the panel corners. Because corner extraction in the image may incur subtle pixel-level errors, which can be amplified after ray casting over tens of meters, the resulting 3D panel corners are regularized: for each row of corners, the z-coordinates are averaged, and for each column, the x- and y -coordinates are averaged, thereby reducing segmentation-induced deviations. After pairing the regularized 3D corner points, the reconstructed 3D panel models are obtained, as shown in Fig. 5.

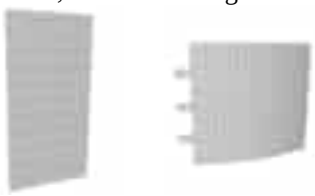
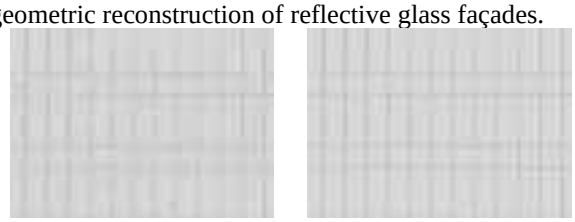


Fig. 5 3D panel models for planar façade (left) and curved façade (right)

3.1.3 Geometric Correction on Coarse Façade Mesh

Based on the obtained 3D panel models, geometric correction can be applied to the coarse façade mesh. Notably, simply overlapping the 3D panel models with the coarse mesh still results in an uneven façade surface (Fig. 6(a)), because the coarse mesh often contains local protrusions and irregularities on the panel regions. Conversely, directly replacing the corresponding mesh regions with the full 3D panel models would remove the surrounding frames.

To address this, each 3D panel is slightly shrunk in-plane about its center, and a small cuboidal region is defined along the panel normal. The portion of the coarse façade mesh within this cuboid is then replaced with the shrunk 3D panels. This procedure produces the corrected façade surface shown in Fig. 6(b), enabling accurate geometric reconstruction of reflective glass façades.



(a) (b)

Fig. 6 Geometric correction by (a) simply overlapping and (b) cuboidal region replacement

3.2 Gaussian Splatting Reconstruction and Rendering for Reflective Glass Façade

As the mesh models are unable to represent the high-frequency, view-dependent appearance of reflective glass façades, the final digital representation is constructed using the 3D model reconstructed by Gaussian Splatting. The accurate façade geometry obtained in Section 3.1 is incorporated as a geometry prior in the training of the Gaussian Splatting model.

3.2.1 Gaussian Augmentation and Deferred Shading Function

For each Gaussian, in addition to the parameters introduced in Section 2, two attributes are further included: a reflectivity coefficient r and a 3-dimensional

Fresnel vector f_0 for calculating the Fresnel term. The Fresnel term modulates specular reflectance as a function of the viewing angle, capturing the characteristic behaviour of glass surfaces, i.e., reflectance increases toward grazing angles and decreases near normal incidence. During rendering, different appearance components are first rendered into 2D images independently and then composited, following a deferred shading strategy.

For an image I, the deferred shading function is defined in Eq. (7). The diffuse color D is computed using Eq. (5) and represented with spherical harmonics. The specular reflection R is estimated by the reflection MLP described in Section 3.2.2. The reflectivity $refl$ is computed using the same volumetric alpha blending formulation as in Eq. (5). Since the surface roughness of reflective glass façades is negligible and can be approximated as ideal specular reflection, the Fresnel term F can be calculated using Schlick's approximation in Eq. (8), parameterized by F_0obtained via volumetric alpha blending of the per-Gaussian $f_0$values.

$$I = (1 - refl) \cdot D + refl \cdot F \cdot R \quad (7)$$

$$F = F_0 + (1 - F_0) \cdot (1 - \cos\theta)^5 \quad (8)$$

where θ denotes the angle between the surface normal n and the outgoing radiance direction ω_o (i.e., the direction opposite to the viewing direction).

3.2.2 Reflection MLP

To model near-field specular reflections, we design a reflection MLP that predicts the specular reflected color R from the surface position X_r and the reflection direction ω_r at that position (computed using Eq. (9)).

$$\omega_r = \omega_o - 2\omega_o^{\top} n \cdot n \quad (9)$$

The architecture of the reflection MLP is illustrated in Fig. 7. It comprises eight fully connected layers with ReLU activations and 512 hidden units per layer, followed by a final sigmoid-activated output layer. Since standard MLPs exhibit a spectral bias toward low-frequency signals whereas specular reflections contain high-frequency details, the inputs X_r and ω_r are first mapped to high-dimensional features using Fourier feature encoding, as defined in Eq. (10). Before encoding, X_r is normalized to the range $[-1,1]$, and ω_r is enforced to be a unit direction vector. We set $L = 12$ for both $\gamma(X_r)$ and $\gamma(\omega_r)$ to ensure sufficient capacity for modeling high-frequency content.

$$\gamma(p) = (\sin(2^0\pi p), \cos(2^0\pi p), \cdots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p)) \quad (10)$$

The encoded position $\gamma(X_r)$ is first processed by the first four layers alone, rather than being concatenated with the direction input from the beginning, so that the network can learn position-dependent material characteristics and extract relevant features. This staged fusion helps disentangle spatially varying properties from direction-dependent reflection. At the fifth layer, the learned position features are concatenated with the encoded reflection direction $\gamma(\omega_r)$, and $\gamma(X_r)$ is re-injected via a skip connection. The skip connection facilitates gradient flow and preserves positional information, mitigating over-smoothing in deep networks. Finally, a sigmoid activation is applied to produce the specular reflected RGB color R.

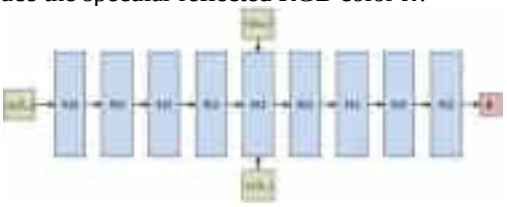


Fig. 7 Architecture of reflection MLP

3.2.3 Regularization

To effectively leverage the accurate façade geometry obtained in Section 3.1 and to ensure that high-frequency appearance details are properly learned, we introduce two additional regularization terms: a normal-prior loss and a high-frequency loss.

Normal-prior loss. As shown in Eq. (11), we compute an $\ell_1$loss between the rendered surface normals n and the reference normals $\hat{n}$ derived from the accurate façade geometry under the same viewpoint. This term enforces geometric fidelity of the reconstructed façade surface.

$$\mathcal{L}_{np} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \|n(p) - \hat{n}(p)\|_1 \quad (11)$$

where Ω denotes the set of valid pixels, and p indexes image pixels.

High-frequency loss. As defined in Eq. (12), we focus the supervision on pixels whose color error exceeds a threshold τ. Specifically, we select the subset of pixels with large $\ell_1$residuals between the rendered color C and the ground truth C^{gt}, and average the residuals over this subset. Empirically, we observe that using only the standard RGB reconstruction loss $\mathcal{L}_c$ in Gaussian Splatting tends to distribute errors uniformly, yielding overly smooth reconstructions that fail to recover sharp, high-frequency reflections. The proposed high-frequency loss explicitly corrects these high-error regions.

$$\mathcal{L}_{hf} = \frac{\sum_{p \in \Omega} \mathbb{I}(\|C(p) - C^{\mathrm{gt}}(p)\|_1 > \tau) \, \|C(p) - C^{\mathrm{gt}}(p)\|_1}{\sum_{p \in \Omega} \mathbb{I}(\|C(p) - C^{\mathrm{gt}}(p)\|_1 > \tau) + \varepsilon} \quad (12)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, ε is a small constant to avoid division by zero, and τ is initialized to 0.2 and decreased in a stepwise manner over training iterations to a final value of 0.09.

Finally, the overall objective is given in Eq. (13).

Here, $\mathcal{L}_c$ combines an ℓ_1 term with the D-SSIM term, and $\mathcal{L}_n$ denotes the normal-consistency loss; both follow the original 2DGS definitions.

$$\mathcal{L} = \mathcal{L}_c + \lambda_n \mathcal{L}_n + \lambda_{np} \mathcal{L}_{np} + \lambda_{hf} \mathcal{L}_{hf} \quad (13)$$

where $\lambda_n = 0.05$, $\lambda_{\text{np}} = 1$, and λ_{hf} is initialized to 1 and increased in a stepwise manner over training iterations, reaching a final value of 10.

4 Experiment

The proposed Gaussian Splatting–based reconstruction and rendering framework for reflective glass façades is validated on two buildings in Hong Kong: the HKUST Library and the Sha Tin Office Building.

4.1 Experimental Setup

For the HKUST Library, 600 multi-view images were collected, while 560 multi-view images were captured for the Sha Tin Office Building. Since this study focuses on the reconstruction and rendering of reflective glass façades, the 3D panel models of both buildings were first projected onto the images to generate façade masks, which remove background regions outside the façades. The prior normal maps used for training were likewise restricted to the masked regions.

All images were captured using a DJI M3E UAV equipped with an RTK module, providing centimeter-level GPS positioning. Due to interference from view-dependent appearance, camera poses estimated by COLMAP can exhibit substantial errors. Therefore, the captured images were first processed in DJI Terra to perform photogrammetry by incorporating GPS positions and camera orientations. This procedure supports the geometric reconstruction and correction in Section 3.1 while providing more accurate camera poses. The resulting poses were converted to the COLMAP format for Gaussian Splatting training using 100,000 epochs.

The proposed method is compared with vanilla 3DGS [9], 2DGS [13], and state-of-the-art reflection-aware Gaussian models, including GaussianShader [10] and Ref-Gaussian [11]. The proposed model is trained for a total of 2×10^5 iterations, while all baseline methods are run with their default hyperparameter settings.

In addition, different designs of the reflection MLP are explored and compared to demonstrate the rationality of our proposed architecture. Performance is evaluated using PSNR, SSIM, and LPIPS. All experiments are conducted on an NVIDIA A40 GPU.

4.2 Baseline Comparison

Figs. 8 and Table 1 illustrate the baseline comparison results on the HKUST Library dataset, while Figs. 9 and Table 2 show the results on the Sha Tin Office Building dataset. Quantitatively, the proposed method achieves the best performance across all metrics, and improves PSNR by more than 5 dB over the SOTAs on both datasets.

As illustrated by the renderings in Figs. 8 and 9, the proposed method faithfully reconstructs the view-dependent textures of reflective glass façades while preserving high-fidelity high-frequency details, such as the circular fountain in the center of the top row in Fig. 8, the reflected building in the lower-right corner of the bottom row in Fig. 8, the blue sports court in the lower-left region of the top row in Fig. 9, and the fine details of the bridge and vehicles in the bottom row of Fig. 9.

For Ref-Gaussian, it can recover view-dependent appearance to some extent; however, it represents global illumination using a 2D environment map, which causes parallax errors for real-world 3D near-field reflections, leading to noticeably blurred high-frequency details. GaussianShader also relies on a 2D environment map, but uses 3D Gaussians as primitives and defines normals using the minor-axis direction of each 3D Gaussian. This representation results in inaccurate façade-surface geometry, which in turn yields incorrect reflection estimation. Consequently, Fig. 8 shows locally chaotic appearance in the middle of the façade, and Fig. 9 exhibits unstable appearance across large façade regions, both due to poor geometric reconstruction. In addition, vanilla 3DGS and 2DGS employ spherical harmonics to model color, which captures only low-frequency view-dependent variations and fails to reproduce high-frequency specular reflections.

Overall, these comparisons demonstrate that our method reconstructs façade surfaces with 2D Gaussian primitives, estimates near-field specular reflections with a reflection MLP, and incorporates a corrected façade mesh as a geometric prior, together providing an effective solution for high-quality reconstruction and rendering of reflective glass façades.

Table 1 Quantitative comparison of HKUST Library

Model	PSNR↑	SSIM↑	LPIPS↓
Ours	**31.66**	**0.9237**	**0.0695**
Ref-Gaussian	26.65	0.9057	0.0930
GaussianShader	22.62	0.8402	0.1599
3DGS	22.70	0.8521	0.1278
2DGS	22.40	0.8511	0.1346

Table 2 Quantitative comparison of Sha Tin Office

Model	PSNR↑	SSIM↑	LPIPS↓
Ours	**30.17**	**0.9339**	**0.0594**
Ref-Gaussian	25.02	0.9139	0.0797
GaussianShader	21.40	0.8350	0.1638
3DGS	21.98	0.8566	0.1135
2DGS	21.10	0.8474	0.1250

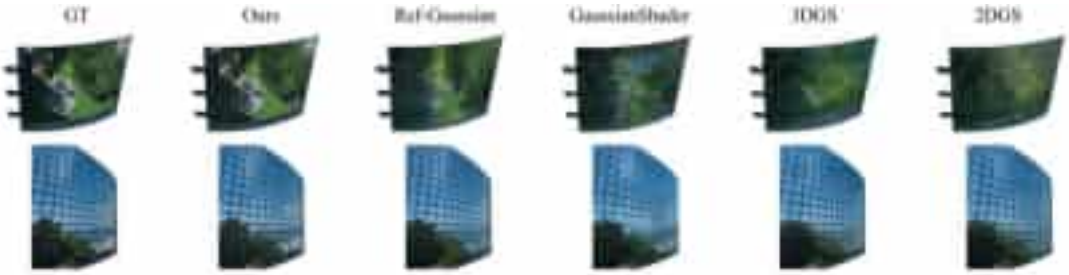


Fig. 8 Qualitative comparison for HKUST Library

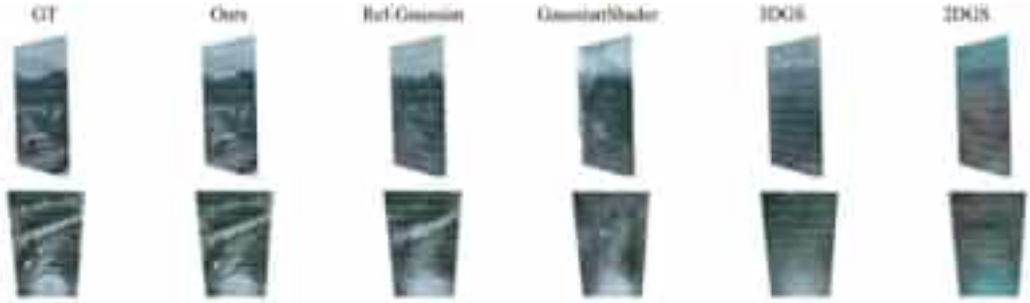


Fig. 9 Qualitative comparison for Sha Tin Office Building

4.3 Reflection MLP Architecture Evaluation

One of the core contributions of this work is the proposed reflection MLP, whose architectural design has a substantial impact on performance. The architecture is primarily governed by four factors: (1) the Fourier feature encoding order L, (2) the network depth (number of layers), (3) the hidden-layer width (hidden dimension), and (4) whether to incorporate a skip connection that re-injects the encoded position features.

In the Sha Tin Office Building, different reflection MLP architectures are evaluated, as summarized in Table 3. Comparing Cases 1, 2, and 4 indicates that increasing the hidden-layer width and network depth benefits the reconstruction of high-frequency, view-dependent textures. This is because the reflection MLP effectively serves as a high-capacity function approximator that maps position and reflected direction to reflected color; wider layers improve feature expressiveness, while deeper networks enhance the modeling of nonlinear interactions between spatial location and direction.

Comparing Cases 2 and 3 shows that introducing a skip connection that re-injects the encoded position features improves near-field reflection estimation. Additionally, Cases 3, 5, 6, and 7 suggest that moderately increasing the Fourier feature encoding order helps the MLP represent high-frequency components, whereas excessively large encoding orders degrade performance, because high-order encodings introduce very high-frequency basis functions that amplify sensitivity to small pose/calibration errors and pixel-level noise.

5 Conclusion

This paper proposes a novel Gaussian Splatting–based framework for reconstructing and rendering reflective glass façades, enabling accurate geometric reconstruction and high-fidelity photorealistic rendering. The main contributions are twofold: (1) a geometry reconstruction and correction strategy that integrates segmented 3D glass panel models with photogrammetric coarse façade mesh to refine the façade surface geometry; and (2) an enhanced 2D Gaussian representation equipped with a reflection MLP, which supports

Table 3 Comparison of reflection MLP architecture

Case	Architecture	PSNR↑	SSIM↑	LPIPS↓
1	$L = 10$, 8 layers, 256 hidden dim, w/o skip connection	29.08	0.9267	0.0636
2	$L = 10$, 8 layers, 512 hidden dim, w/o skip connection	29.68	0.9310	0.0608
3	$L = 10$, 8 layers, 512 hidden dim, w/ skip connection	29.95	0.9323	0.0603
4	$L = 10$, 6 layers, 512 hidden dim, w/ skip connection	29.77	0.9308	0.0618
5	$L = 12$, 8 layers, 512 hidden dim, w/ skip connection	30.17	0.9339	0.0594
6	$L = 14$, 8 layers, 512 hidden dim, w/ skip connection	29.84	0.9308	0.0615
7	$L = 16$, 8 layers, 512 hidden dim, w/ skip connection	29.02	0.9235	0.0672

reconstruction and rendering of high-frequency, view-dependent façade appearance.

Future work will focus on leveraging multimodal large language models (MLLMs) to further optimize the overall pipeline and to enable more natural interaction with the generated façade digital models [15–17], as well as extending Gaussian-based models to downstream simulation and analysis tasks [18–20].

References

[1] Z. Liang, L. Yang, Z. Liang, J. C. F. Chan, Z. Zhang, M. Wang, and J. C. Cheng. Optimized UAV view planning for high-quality 3D reconstruction of buildings using a modified sparrow search algorithm. *Adv. Eng. Inform.*, *65*, 103344, 2025.

[2] Z. Liang, J. C. F. Chan, J. Zhang, Z. Liang, B. Wang, M. Wang, and J. C. Cheng. Multimodal LLM-driven language-embedded 3D gaussian splatting for semantic and realistic digitization of historical buildings. *Automat. Constr.*, *181*, 106628, 2026.

[3] Z. Zhao, L. Ju, L. Zhang, H. Zou, J. He, Y. Lv, T. Xiao and Cai, G. Three-dimensional geological modeling with multisource data fusion. *J. Geotech. Geoenviron.*, *152*(1), 04025164, 2026.

[4] L. Y. Ju, T. Xiao, J. He, W. F. Xu, S. H. Xiao, and L. M. Zhang. A simulation-enabled slope digital twin for real-time assessment of rain-induced landslides. *Eng. Geol.*, 108116, 2025.

[5] B. Wang, F. Lin, M. Li, Z. Liang, Z. Chen, M. Wang, and J. C. Cheng. Informative As-Built Modeling as a Foundation for Digital Twins Based on Fine-Grained Object Recognition and Object-Aware Scan-vs-BIM for MEP Scenes. *Adv. Eng. Inform.*, *65*, 103382, 2025.

[6] B. Wang, F. Lin, M. Li, Z. Liang, H. Yue, Q. Wang, and J. C. Cheng. Aligning as-built and as-designed: Local point cloud to BIM registration via hybrid visibility map encoding for construction digital twins. *Automat. Constr.*, *180*, 106551, 2025.

[7] Y. Lv, Y. Pang, L. Ju, W. Xu, Z. Liang, J. He, and L. Zhang. Automating digitalization of engineered slopes to support interactive digital twin visualization. *Automat. Constr.*, *181*, 106564, 2026.

[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision* (pp. 405-421), Glasgow, UK, 2020.

[9] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, *42*(4), 139-1, 2023.

[10] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5322-5332), Seattle, WA, USA, 2024.

[11] Y. Yao, Z. Zeng, C. Gu, X. Zhu, and L. Zhang. Reflective Gaussian Splatting. In *The Thirteenth International Conference on Learning Representations*. Singapore, 2025

[12] T. Xie, X. Chen, Z. Xu, Y. Xie, Y. Jin, Y. Shen, S. Peng, H. Bao, and X. Zhou. Envgs: Modeling view-dependent appearance with environment gaussian. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 5742-5751). Nashville, TN, USA, 2025.

[13] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers* (pp. 1-11). Denver, CO, USA, 2024.

[14] Z. Liang, X. Zhang, Z. Liang, B. Wang, M. Wang, and J. C. Cheng. A Novel Framework for Generating Realistic and Semantic Digital Model of Reflective Glass Façades Using RefGlass-3DGS. In *Computing in Civil Engineering 2025: Computational and Intelligent Technologies* (pp. 914-923), New Orleans, LA, USA, 2025.

[15] A. Li, P. K. Y. Wong, X. Tao, J. Ma, and J. C. Cheng. An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent. *Adv. Eng. Inform.*, *66*, 103375, 2025.

[16] C. F. Chan, P. K. Y. Wong, X. Guo, J. C. Cheng, J. P. C. Chan, P. H. Leung, X. Tao. Context-aware vision-language model agent enriched with domain-specific ontology for construction site safety monitoring. *Automat. Constr.*, *177*, 106305, 2025.

[17] K. X. Guo, P. K. Y. Wong, J. C. Cheng, C. F. Chan, P. H. Leung, and X. Tao. Enhancing visual-LLM for construction site safety compliance via prompt engineering and Bi-stage retrieval-augmented generation. *Automat. Constr.*, *179*, 106490, 2025.

[18] W. Liu, X. Tao, C. Mao, and W. He. Scheduling optimization for production of prefabricated components with parallel work of serial machines. *Automat. Constr.*, *148*, 104770, 2023.

[19] W. Liu, Z. Wu, X. Tao, H. H. Kwok, and J. C. Cheng. Automated simulation modeling and sustainable performance evaluation for flat layout design. *Automat. Constr.*, *180*, 106556, 2025.

[20] Yin, C., Yue, H., Han, Q., Hu, D., Liang, Z., Lin, F., ... & Cheng, J. C. (2026). Industrial3D: A Terrestrial LiDAR Point Cloud Dataset and CrossParadigm Benchmark for Industrial Infrastructure. *arXiv preprint arXiv:2603.28660*.

An Automated Framework for Railway Sleeper Detection and Spacing Measurement using UAV Imagery

Bhanu Pratap Singh[1] and Sahil Garg[2]

[1]Department Of Civil and Environmental Engineering, Indian Institute of Technology Delhi, New Delhi, India
[2]Department Of Civil and Environmental Engineering, Indian Institute of Technology Delhi, New Delhi, India
cec242814@iitd.ac.in, gargsahil@iitd.ac.in

Abstract –

Railway sleepers form the structural interface between the rails and the ballast, governing track geometry stability through load transfer and positional restraint, making them a critical track component. Regular monitoring of sleeper condition and spacing contributes to maintaining track condition and supports safe railway operation for goods and passenger traffic. In current practice, sleeper spacing inspection on ballasted track relies on manual tape measurements and visual judgement, which are time consuming, labour-intensive, and difficult to apply consistently over long construction sections or during short night possessions.

This paper proposes an automated approach for sleeper spacing measurement using unmanned aerial vehicle imagery and a hybrid computer vision-based processing framework. The method applies a YOLOv11m object detection model to localise individual sleepers in near-nadir UAV images, followed by mask refinement using the Segment Anything Model to delineate sleeper regions. Sleeper centres and rail edges are extracted from the resulting masks, and pixel distances are converted to physical spacing through calibration based on the known effective track gauge.

The framework was evaluated on a dataset of straight ballasted track sections acquired under routine field conditions. Quantitative validation demonstrated a mean absolute error of 6.56 mm and a root mean square error of 7.67 mm, with 93.75% of predicted spacing values falling within ±10 mm of reference measurements. These results indicate that the proposed workflow enables reliable and physically interpretable sleeper spacing measurements from UAV imagery.

Keywords –
Ballasted Track, UAV-Based Inspection, YOLO-Based Object Detection, Image Segmentation, Sleeper Spacing.

1 Introduction

Railway networks rely on consistent track geometry to maintain safe operations, control dynamic wheel-rail interactions, and limit long-term degradation of sleepers, ballast, and fastening systems. Within this broader geometry envelope, the longitudinal spacing of sleepers plays a very practical role in how loads are distributed, how ballast consolidates, and how maintenance limits are defined in track standards and inspection manuals [1]. In day-to-day practice, however, sleeper spacing for ballasted track is still often checked through manual gauge and visual inspections or only at discrete locations using track recording cars, which can miss localised anomalies, construction non-conformities, or gradual shifts that emerge between scheduled runs [2]. For railways that are expanding or upgrading lines with limited track access, this combination of manual point-based measurements and visual checks is increasingly difficult to reconcile with the demand for more frequent, fine-grained quality control.

Image-based inspection offers a potential route toward denser geometric assessment. An ideal scenario would use nadir images of the track acquired from a UAV or similar platform under controlled conditions, with known camera parameters and stable flight geometry. In such a scenario, each sleeper could be detected, cleanly segmented from ballast and rails, and used together with calibrated reference dimensions to compute spacing in physical units in a transparent and interpretable manner. In practice, the conditions are far less controlled. UAV imagery is affected by variable altitude, camera tilt, motion blur, and changing lighting; tracks themselves may present occlusions due to vegetation, ballast pockets, or construction materials; and, in many projects, full camera calibration or ground control is either unavailable or only loosely specified. These factors make it difficult to move from visually plausible sleeper detection to reliable metric spacing measurements that can be trusted for acceptance and compliance checks [3], [4].

Recent vision-based railway work has advanced along several strands that do not always converge. Object

detection models, particularly YOLO variants, identify sleepers, fasteners, and surface defects at high frame rates in inspection vehicles or UAVs, prioritising throughput, and automation rather than geometric boundary precision [5], [6]. UAV imagery has supported track localisation, ballast assessment, and overhead line inspection, showing aerial data can integrate into workflows when operational constraints are managed. Visual metrology systems for gauge and other parameters achieve sub-millimetre to millimetre-scale engineering accuracy under careful calibration, but typically with fixed cameras, dedicated targets, and constrained environments rather than mobile UAV platforms.

Sleeper-level geometry highlights the trade-off between flexibility and rigour. UAV-based approaches using template matching or simple image processing report centimetre level spacing estimates, but over limited, visually clean segments and under stable scale/pose assumptions. Higher accuracy gauge or profile systems rely on controlled illumination, rigid setups, or track-mounted artefacts that are difficult to reproduce over long aerial surveys. This leaves a gap between fast, detection-centric pipelines treating sleepers as coarse pixel objects and metrology-centric systems handling geometry carefully but in constrained configurations. Despite these developments, a clear understanding of how detection-driven segmentation workflows perform as measurement-oriented tools under practical UAV inspection conditions remains limited.

The present study addresses this gap by focusing explicitly on the geometric reliability of sleeper spacing measurement from UAV imagery using a hybrid detection and segmentation framework. Rather than proposing a new detection architecture, the contribution lies in evaluating whether segmentation-refined detections combined with explicit geometric calibration can produce spacing outputs that are geometrically consistent and practically interpretable under typical UAV inspection conditions. The framework employs a YOLO-based model to localise sleepers in near-nadir UAV images, followed by the Segment Anything Model to refine detections into detailed masks. This refinement improves boundary definition and supports reliable centre estimation for spacing measurement. In addition, the workflow integrates an explicit calibration strategy using track gauge as a geometric reference, enabling conversion from pixel distances to physical units without requiring external ground control. The study focuses on defined straight track sections to establish a controlled baseline and assess whether such an integrated approach can support spacing measurement for construction verification and routine maintenance inspection. This study evaluates the reliability of a hybrid detection and segmentation workflow for sleeper spacing measurement.

2 Related Work

Vision-based inspection has become an important approach in railway infrastructure monitoring, enabling large networks to be inspected with reduced reliance on manual surveys and specialised geometry vehicles. Camera-based systems mounted on inspection vehicles, wayside installations, and mobile platforms are widely used to detect rails, sleepers, fasteners, and ballast conditions. Many of these systems rely on deep learning-based object detection models, particularly YOLO variants, which prioritise rapid localisation and automation rather than precise geometric boundary representation [5], [6].

UAV imagery has further expanded inspection coverage and deployment flexibility, particularly in construction zones and remote locations. Prior studies demonstrate the use of UAV data for track localisation, ballast assessment, and geometric parameter estimation such as track alignment or gauge using photogrammetric or model-based approaches [7]. While these methods confirm the feasibility of metric measurement from aerial imagery, they often rely on dense ground control, careful mission planning, or computationally intensive three-dimensional reconstruction, which limits their practicality for routine sleeper spacing measurement over short track sections.

A parallel body of work focuses on image-based rail geometry and metrology. Visual measurement systems for track gauge and wheel rail position show that millimetre level accuracy is achievable when camera calibration, reference geometry, and error modelling are treated explicitly [8]. However, such systems typically operate in controlled environments with fixed or stereo camera setups and do not readily transfer to UAV imagery acquired under variable pose and lighting conditions. Deep learning models such as YOLO and Mask R-CNN-based segmentation has been applied to estimate inter-sleeper distances from image data, demonstrating feasibility under controlled conditions [9].

Taken together, existing studies either prioritise detection performance without explicit treatment of geometric interpretation or achieve precise measurements under controlled configurations that are difficult to reproduce in routine UAV operations. As a result, the geometric reliability of sleeper spacing measurement derived from overhead imagery, particularly when calibration resources are limited and track gauge is the primary physical reference, remains insufficiently understood. This limitation motivates the present study, which evaluates whether segmentation-refined detection combined with practical geometric calibration can support reliable sleeper spacing measurement under realistic field conditions.

3 Data Acquisition and Study Area

UAV imagery used in this study was collected along straight ballasted railway sections in the Mau-Aunrihar corridor of the Varanasi Division, India, using a DJI Mavic 3 unmanned aerial vehicle equipped with an onboard RGB camera. Image acquisition was carried out in a near-nadir configuration to minimise perspective distortion and maintain a consistent geometric view of sleeper placement along the track. Straight sections were selected intentionally to isolate geometric reliability before extension to curved alignment.

All images were captured from a near-nadir viewpoint under natural daylight conditions without artificial markers, calibration targets, or ground control points. Natural variations in ballast distribution, sleeper surface condition, and surrounding track environment were retained to reflect routine field inspection conditions. Images were stored at full resolution and processed without geometric rectification beyond standard onboard stabilisation. Track gauge, clearly visible in the imagery, was used as the sole physical reference for pixel-to-millimetre scale conversion.

Table 1. Dataset Characteristics

Parameter	Description
UAV Platform	DJI Mavic 3
Imaging Orientation	Near-Nadir (Top-Down)
Flight Altitude	12-15 m
Image Resolution	4000 × 2250 pixels
Track Geometry	Straight, Ballasted Track
Track Coverage	Two parallel tracks
Sleeper Type	Concrete Sleepers
Scale Reference	Track Gauge

For this study, two hundred images were selected for experimental evaluation. The imagery was collected along an approximately 4 km operational railway stretch including both open track and yard regions. These environments introduced variation in ballast condition, sleeper surface wear, shadow intensity, and minor vegetation presence. Two different concrete sleeper types were present within the surveyed section, introducing additional geometric variability.

While the dataset was obtained from a single operational corridor, the extended spatial coverage, environmental diversity, and presence of multiple sleeper types provide a representative baseline for evaluating geometric measurement reliability under realistic field variability. Further validation across multiple corridors and track geometries will be required to establish full operational generalisation.

3.1 Data Pre-Processing and Annotation

Annotation was performed manually using rectangular bounding boxes for two object classes, sleeper and rail. Bounding box annotation was selected to ensure reliable localisation of sleeper regions rather than precise boundary delineation. Partially obscured sleepers affected by ballast, shadows, or surface wear were annotated using visible portions. Rail annotations were included only to support geometric reference and track gauge estimation in later processing stages.

The annotated dataset consisted of 200 images divided into training, validation, and test subsets using a 70 percent, 20 percent, and 10 percent split, respectively. This partitioning enabled model training, parameter tuning, and independent performance evaluation. Figure 1 presents representative annotated UAV images.



Figure 1. Sample of Annotated UAV Images.

To improve robustness under operational variability, the training data were augmented using controlled geometric, photometric, and noise-based transformations, as summarised in Table 2. After augmentation, the effective size of the training dataset [12] increased to 496 images.

Table 2. Augmentation applied to the training dataset.

Parameter	Range	Purpose
Rotation	−15° to +15°	To handle slight changes in camera orientation
Hue	−10° to +10°	To account for colour changes caused by lighting
Brightness	−15% to +15%	To capture differences in illumination
Gaussian blur	up to 0.5 px	To represent slight blur in some images
Random noise	≤ 0.1% pixels	To reflect minor noise during image capture

4 Methodology

This section describes the methodology for automated railway sleeper detection and spacing measurement using UAV imagery. The proposed hybrid

framework integrates detection, segmentation, and geometric processing to derive sleeper spacing from single-view nadir images using track gauge as the physical reference. The workflow is designed to ensure reliable localisation under variable visual conditions and consistent geometric handling prior to spacing estimation.

The framework is designed with two central considerations. First, reliable sleeper localisation must be achieved under variable visual conditions, including ballast overlap, shadows, surface wear, and changes in illumination. Second, spacing estimation requires explicit geometric handling of sleeper positions rather than direct use of object detection outputs. To address these requirements, detection and segmentation are integrated with a structured geometric processing pipeline that supports consistent centre estimation and longitudinal alignment prior to spacing measurement.

4.1 Overview

The proposed framework consists of four important sequential stages: sleeper detection, segmentation and mask refinement, sleeper centre estimation and alignment, and sleeper spacing measurement. The UAV image serves as the input, and spacing values are obtained independently for each track present in the scene. Track gauge is used as the sole physical reference for image (pixel)-to-ground scale conversion.

Deep learning models such as YOLO and the Segment Anything Model (SAM) are widely used for object detection and instance segmentation in complex visual environments. These models learn hierarchical [13] feature representations directly from image data, enabling robust localisation and boundary extraction under the variability present in UAV-based railway imagery.

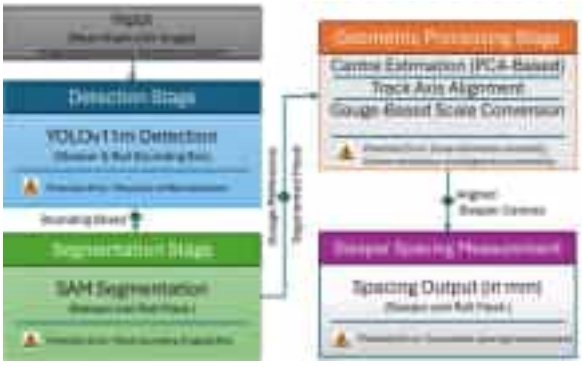


Figure 2. Proposed workflow illustrating detection, segmentation, geometric processing, and spacing measurement stages, with representative sources of uncertainty contributing to cumulative measurement error.

The framework adopts a hybrid deep learning approach that combines a single stage object detector with an instance segmentation model. YOLO operates on a convolutional neural network based single stage detection architecture to provide efficient coarse localisation of railway sleepers. SAM operates on a transformer-based segmentation architecture and uses detector outputs as prompts to generate refined instance level masks. An overview of the complete processing workflow is illustrated in Figure 2.

Each stage of the workflow introduces positional uncertainties that propagate into the final spacing measurement. Detection errors may slightly shift bounding box placement, while segmentation inaccuracies caused by ballast overlap or shadows influence boundary definition. These effects impact centre estimation and alignment, and any error in rail localisation affects the pixel-to-metric scaling process. The influence of these uncertainties is reflected in the spacing error statistics reported in Section 5.3.

4.2 Detection Model Training

Nvidia Tesla T4 GPUs with 16 GB of memory were utilised for model training within the Kaggle notebook environment. The YOLOv11m architecture was employed, initialised with pretrained weights from the COCO dataset. Using the annotated dataset described in Section 3.1, training was conducted on high resolution UAV imagery originally captured at 4000 × 2250 pixels. During training, images were resized to 1024 × 1024 pixels for computational consistency. Bounding box annotations were provided for two object classes: sleeper and rail.

The model was trained for 100 epochs with a batch size of 4 using the AdamW optimiser and an initial learning rate of 1×10^{-4}. Early stopping was enabled based on validation loss stabilisation to prevent overfitting. Data augmentation included controlled geometric and photometric transformations. The confidence threshold during inference was set to 0.25, and non-maximum suppression was applied with an intersection over union threshold of 0.50. Model performance was monitored using validation loss and mean average precision during training to ensure stable convergence.

Within the overall processing pipeline, the role of the detection model was intentionally limited. Detection outputs were not used directly for sleeper spacing measurement or geometric analysis. Instead, sleeper bounding boxes were used to initialise subsequent segmentation, while rail detections provided geometric references for track gauge estimation and track axis definition. All geometric accuracy related to sleeper centre estimation, alignment, and spacing measurement was handled in later processing stages.

4.2.1 SAM Vit-H Configuration

The Segment Anything Model (SAM Vit-H) was used in inference mode with pretrained weights and without additional fine-tuning. Bounding box detections generated by the YOLOv11m model were used as prompt inputs to initialise segmentation for each detected sleeper instance. Standard inference configuration was adopted to ensure reproducibility, with automatic mask generation applied to each prompted region. Predicted masks were filtered by retaining the largest connected component to remove small-isolated regions and noise. This post-processing step ensured consistent boundary representation for subsequent centre estimation and spacing measurement.

4.3 Sleeper Detection

In the first stage, individual sleepers are localised using a YOLOv11m-based object detection model trained on domain specific UAV imagery of railway tracks. Bounding box annotations for sleepers were prepared to support detector preparation. The detector processes the complete image and outputs bounding boxes corresponding to sleeper instances across both tracks. No assumptions are made regarding uniform spacing, sleeper orientation, or symmetry between tracks. The bounding boxes generated at this stage provide coarse spatial localisation, and the identified image regions are then analysed in the next stage. They are not used directly for geometric measurement; instead, segmentation is performed using the regions identified in the detection stage.

4.3.1 Rail Localisation for Geometric Reference

Rail instances were detected to establish track gauge and define the longitudinal track axis required for centre alignment and spacing computation. Rail outputs were used only for geometric reference.

4.4 Segmentation and Mask Generation

Segmentation was applied to refine the initial sleeper detections using the Segment Anything Model (SAM Vit-H). Detector outputs were used as prompts to generate binary masks representing sleeper regions. Segmentation improves boundary definition compared to bounding boxes and provides geometrically consistent regions for centre estimation. The resulting masks were used directly in subsequent geometric processing.

4.5 Sleeper centre estimation and alignment

For each segmented sleeper, a centre point is derived from the spatial distribution of its mask pixels, after which principal component analysis (PCA) is used to estimate the sleeper's longitudinal axis. The sleeper centre is defined as the midpoint along this axis.

To ensure consistent longitudinal measurement, the centres of individual sleepers are projected onto an estimated track axis derived from their overall arrangement along each rail. This alignment step reduces lateral variation caused by perspective effects, minor detection inaccuracies, and sleeper skew, and simplifies the spacing analysis by representing sleeper positions along a single one-dimensional track direction.

Following alignment, each sleeper centre is represented by a single scalar coordinate s_i corresponding to its projected position along the estimated track axis. Sleepers are then ordered sequentially along this axis, with indexing performed separately for the left and right tracks.

4.6 Sleeper spacing measurement

Following centre estimation and alignment, sleeper spacing is computed independently for each track using the one-dimensional representation established in Section 4.5. After alignment, each sleeper centre is characterised by a scalar coordinate s_i, representing its projected position along the estimated track axis in pixel units. Sleepers are ordered sequentially along this axis for the left and right tracks separately.

The sleeper spacing is defined as the longitudinal separation between the aligned centres of two consecutive sleepers. Accordingly, the pixel domain spacing between sleeper i and sleeper i+1 is computed as

$$d_i^{px} = s_{i+1} - s_i \quad (1)$$

To express the spacing in physical units, a pixel-to-metric scale factor is derived using the known nominal track gauge visible in the image. Let the physical track gauge be denoted by G_{mm} (in millimetres) and let G_{px} denote the corresponding gauge measured in pixel units from the detected rail positions. The pixel-to-metric conversion factor is then defined as

$$\alpha = \frac{G_{mm}}{G_{px}} \quad (2)$$

The sleeper spacing in real-world units is obtained by applying this scale factor to the pixel-domain spacing as

$$d_i^{mm} = \alpha d_i^{px} \quad (3)$$

Track gauge is used as the physical reference for scale conversion, enabling spacing values to be expressed in real world units without reliance on external ground control points, calibration targets, or multi-view reconstruction.

The complete sequence of processing stages, from (a) sleeper detection, (b) segmentation, (c) centre alignment, and (d) spacing measurement, is shown in Figure 3.

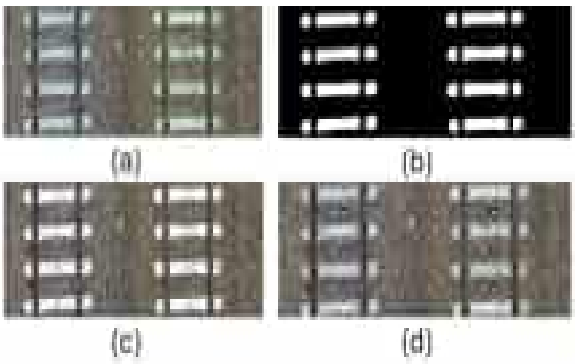


Figure 3. Stages of the proposed framework

5 Result and Discussion

5.1 Detection Performance Evaluation

The detection component of the framework was evaluated to verify that sleepers and rails could be identified reliably for use in later segmentation and spacing analysis. Performance was assessed on the test dataset using precision, recall, and mean average precision at an intersection over union threshold of 0.50.

Intersection over union is used to quantify the spatial agreement between a predicted bounding box B_p and its corresponding ground truth annotation B_g by comparing the area of overlap to the total area covered by both regions. It is defined as

$$IoU = \frac{area\ B_p \cap B_g}{area\ B_p\ \cup B_g} \tag{4}$$

Based on a predefined IoU threshold, detections are categorised as correct or incorrect, forming the basis for computing precision and recall following standard object detection evaluation practice [10]. Precision reflects the proportion of predicted objects that correspond to valid detections, while recall represents the proportion of ground truth objects that are successfully identified by the model [13] and are defined as

$$p = \frac{TP}{TP + FP} \tag{5}$$

$$r = \frac{TP}{TP + FN} \tag{6}$$

where TP, FP and FN denote true positive, false positive, and false negative detections, respectively. Mean average precision provides an overall measure of detection reliability across confidence thresholds and object classes [10], [11].

Performance on the test dataset shows a precision of 93.4% and a recall of 90% for the two object classes. The corresponding mean average precision at an intersection over union threshold of 0.50 is 95% [14]. With a recall of 0.90, the model captured nearly all sleeper and rail instances visible in the test imagery. These results confirm that the detection stage provides consistent and repeatable object localisation under near-nadir UAV imaging conditions.

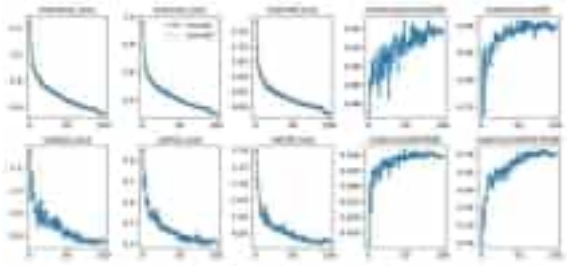


Figure 4. Training performance of the YOLOv11m detection model showing loss convergence and evolution of mean average precision across training epochs.

Figure 4 illustrates the training behaviour of the detection model, showing convergence of training and validation losses together with the progression of mean average precision across epochs.

Sleeper detection outputs are shown in Figure 3(a), where bounding boxes are overlaid on sleepers across both tracks in a near-nadir UAV image. The detections demonstrate consistent localisation across the image extent and provide reliable initialisation for subsequent segmentation and geometric processing stages.

5.2 Comparative Evaluation of Segmentation Models

To provide comparative context for segmentation performance, the proposed YOLOv11m-SAMvit-H workflow was evaluated alongside a Mask R-CNN instance segmentation model trained using the same dataset partitions. The Mask R-CNN model was trained for 50 epochs using the same GPU hardware and notebook environment as the YOLO-SAM implementation to ensure consistent training conditions.

The YOLO-SAM workflow achieved a mean average precision of 70% across IoU thresholds from 0.50 to 0.95, with a mean average precision of 92.78% at IoU = 0.50. The corresponding precision and recall values were 91% and 86%, respectively.

The Mask R-CNN model achieved a mean average precision of 64.2% across IoU thresholds from 0.50 to 0.95, with a mean average precision of 95% at IoU = 0.50 and a recall value of 70.9%. A summary of the segmentation performance comparison between the evaluated models is presented in Table 3.

Table 3. Segmentation performance comparison between the proposed YOLO-SAM workflow and Mask R-CNN evaluated on the same dataset partitions.

Model	Precision (%)	Recall (%)	mAP@0.50 (%)	mAP@ [0.50:0.95] (%)
YOLO + SAM	91.0	86.0	92.78	70.0
Mask R-CNN	88.4	70.9	95.0	64.2

The results indicate that the proposed YOLO-SAM workflow provides segmentation performance comparable to Mask R-CNN while supporting modular integration within the proposed measurement pipeline. Since sleeper spacing measurement in the proposed workflow depends directly on segmentation-derived centre locations, comparative segmentation performance provides an appropriate basis for evaluating the relative suitability of alternative models within the measurement pipeline.

5.3 Quantitative Measurement Validation Using Ground Truth Sleeper Spacing

Quantitative validation of the proposed measurement framework was performed using manually collected ground truth sleeper spacing data. A total of forty-eight consecutive sleeper spacings were measured manually under limited track access constraints to provide reference values for accuracy assessment. Ground measurements were obtained using a calibrated measuring tape with millimetre level resolution. Each value represented the centre-to-centre distance between adjacent sleepers along the track direction. Measurements were collected across both tracks to ensure representative coverage of the inspected section.

Corresponding spacing values were independently generated using the proposed UAV based pipeline. Predicted spacing values were derived from aligned sleeper centre coordinates and converted into millimetre units using the scale factor obtained from track gauge-based calibration. Predicted and reference spacing values were paired for direct comparison.

The comparison demonstrates consistent agreement between predicted and reference spacing values. Most predicted values remained within the allowable engineering tolerance range of ±10 mm, indicating reliable performance under practical inspection conditions. The observed spacing error distribution reflects the influence of uncertainties introduced across sequential processing stages illustrated in Figure 2. In particular, small positional deviations during detection influence segmentation boundary definition, which subsequently affects centre alignment accuracy, while uncertainty in rail localisation directly impacts the pixel-to-metric scaling factor. The cumulative effect of these sequential uncertainties is reflected in the final spacing error metrics summarised in Table 4.

Table 4. Summary of Quantitative Spacing Measurement Accuracy

Metric	Value
Number of samples	48
Mean Ground Truth Spacing (mm)	599.5
Mean Predicted Spacing (mm)	596.4
Mean Bias Error (mm)	−3.14
Mean Absolute Error (mm)	6.56
Root Mean Square Error (mm)	7.67
Standard Deviation of Error (mm)	6.99
Percentage Within ±10 mm (%)	93.75

5.4 Discussion

The results demonstrate that the proposed geometry aware pipeline enables reliable sleeper localisation and spacing measurement from UAV imagery. Consistent detection of sleepers and rails under realistic field conditions provided stable inputs for segmentation and geometric processing, supporting reliable centre estimation across consecutive sleepers.

A key contribution of the framework lies in treating detection outputs as intermediate geometric representations rather than final inspection results. Bounding box detections were refined into segmentation-based masks and aligned centre points, enabling spacing to be computed from structured positional data. Comparative segmentation results support the use of the YOLO-SAM workflow, which achieved comparable accuracy to Mask R-CNN while maintaining higher recall. This improved continuity across consecutive sleepers, which is necessary for reliable spacing computation.

Quantitative validation showed strong agreement between predicted and reference spacing values under near nadir imaging conditions. The mean absolute error of 6.56 mm and root mean square error of 7.67 mm indicate millimetre scale engineering accuracy suitable for spacing verification within practical tolerance limits. The high percentage of measurements within ±10 mm further confirms that cumulative uncertainties remained controlled across the processing stages.

From an operational perspective, the framework demonstrates potential to support preliminary inspection and construction verification tasks over short track sections. By reducing reliance on manual tape

measurements and enabling spacing evaluation from UAV imagery, the proposed workflow provides a scalable basis for extending automated geometric inspection to more complex track environments.

5.5 Future Work

Future work will focus on extending the dataset to multiple railway corridors with varying sleeper types, ballast conditions, and track geometries, including curved track sections. Additional validation will be carried out using independently measured field data collected under different environmental conditions to evaluate the robustness of the proposed framework. Further development will investigate integration of the workflow into routine inspection processes through automated data handling and reporting tools to support practical railway maintenance applications.

6 Conclusion

This paper presented a hybrid computer vision framework for railway sleeper detection and spacing measurement from UAV imagery, integrating a YOLOv11m-based object detector with the Segment Anything Model (SAM Vit-H) for segmentation driven geometric refinement. The study evaluated whether segmentation-refined localisation and geometric alignment could support reliable sleeper spacing measurement using track gauge as the sole physical reference.

Quantitative validation demonstrated consistent agreement between predicted and reference spacing values, with millimetre-scale engineering accuracy achieved across consecutive sleeper intervals. These results confirm that segmentation-refined localisation combined with explicit geometric alignment enables reliable spacing measurement under near-nadir UAV conditions.

The present study establishes a controlled baseline for spacing measurement using straight track sections and near-nadir UAV imagery. Extension to curved track geometries and more variable imaging conditions remains an important direction for future investigation. Overall, the findings demonstrate the feasibility of structured geometric measurement from aerial imagery for scalable railway inspection applications.

7 References

[1] Indian Railways Institute of Civil Engineering (IRICEN). Role of Sleeper Spacing in Maintenance of Points. Ministry of Railways, Government of India. On-line: https://iricen.gov.in, Accessed: 10/01/2026.

[2] Indian Railways Permanent Way Manual (IRPWM). Indian Railways Permanent Way Manual Chapter II: Maintenance of Permanent Way. Ministry of Railways, Government of India. On-line: https://indianrailways.gov.in, Accessed: 10/01/2026.

[3] Z. Peng, H. Wang, Y. Jiang, and Y. Wang. Railway infrastructure inspection using unmanned aerial vehicles: A review. Automation in Construction, 125:103599, 2021.

[4] G. Morgenthal, N. Hallermann, T. P. Kersten, and H.-G. Maas. Surveying and monitoring of railway infrastructure using unmanned aerial vehicles. ISPRS Journal of Photogrammetry and Remote Sensing, 126:1–15, 2017.

[5] Q. Li, H. Zhang, Y. Ding, Y. Sun, and Y. Wang. Vision-based inspection of railway infrastructure: A survey. IEEE Transactions on Intelligent Transportation Systems, 22(10):6477–6497, 2021.

[6] T. Ye, H. Liu, Y. Wang, and Z. Wang. Railway object detection and fault diagnosis based on deep learning: A review. Sensors, 20(3):755, 2020.

[7] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter. Deep convolutional neural networks for detection of rail surface defects. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2016.

[8] C. Alippi, E. Casagrande, F. Scotti, and F. Piuri. Composite real-time image processing for rail inspection. IEEE Transactions on Instrumentation and Measurement, 49(3):559–564, 2000.

[9] A. Kumar and R. Singh. Measurement of railway sleeper spacing using Mask R-CNN. In Proceedings of the International Conference on Decision Aid Sciences and Applications (DASA), 2022.

[10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge," International Journal of Computer Vision, vol. 88, no. 2, pp. 303–338, 2010.

[11] T. Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common objects in context," in Proceedings of the European Conference on Computer Vision, 2014, pp. 740–755.

[12] J. S. Choi and J. K. Lee, "Effects of data augmentation on the Nine-Axis IMU-Based orientation estimation accuracy of a recurrent neural network," Sensors, vol. 23, no. 17, p. 7458, Aug. 2023, doi: 10.3390/s23177458.

[13] ReportBench: Evaluating deep research agents via academic survey tasks. (n.d.). https://arxiv.org/html/2508.15804v1

[14] Ultralytics. (2025, December 12). Performance Metrics Deep Dive. Ultralytics YOLO Docs.

Reflective Twins for Deep Excavation using Physics-Informed Machine Learning

Nandeesh Babanagar[1] and Brian Sheil[1]

[1]Department of Engineering, University of Cambridge, United Kingdom
nb710@cam.ac.uk, bbs24@cam.ac.uk

Abstract

Basement construction projects often experience low productivity, schedule delays, cost escalation, safety risks, and high embodied carbon. These challenges largely arise from uncertainty in ground conditions and construction-stage soil-structure interaction, which cannot be fully resolved during design. The observational method offers a structured means of managing this uncertainty by updating design assumptions using monitored performance during construction. However, its adoption remains limited due to the lack of digital tools that robustly integrate monitoring data with predictive and reflective models within construction timescales.

This paper proposes a digital twin framework for deep excavations that embeds the observational method within a layered architecture to enable stage-wise model updating and adaptive decision-making. This structure supports progressive automation from descriptive to prescriptive operation while clarifying interactions between monitoring, data handling, modelling, and decision support. The central contribution of the paper is the development of a reflective twin enabling back-analysis engine, based on physics-informed neural networks. The engine enables continuous updating of the virtual model using field observations, allowing physically consistent inference of loads and deformation profiles from sparse and noisy monitoring data. Validation using canonical cantilever retaining wall problems demonstrates robust reconciliation between observed and predicted behaviour. This capability enables rapid model updating and adaptive construction control, thereby operationalising the observational method within a digital twin environment.

Keywords –
digital twins; observational method; Physics-informed machine learning;

1 Introduction

Urban densification is driving a rapid increase in deep basement and underground construction, including station boxes, high-rise foundations, and transport or energy infrastructure. These projects are typically delivered within highly constrained urban environments and close to existing assets, where construction-stage performance control is critical. A persistent challenge in such works is uncertainty arising from limited site investigation data, complex soil behaviour, and modelling idealisations. To manage this uncertainty, designs are often conservative, leading to inefficient construction sequences, increased material use, and elevated embodied carbon.

The observational method in geotechnical engineering provides a structured framework for managing uncertainty during construction through a learn-as-you-go process [1]. It enables continuous refinement of design assumptions using monitored performance, supported by back-analysis and predefined contingency measures. In deep excavation projects, the observational method is used to control construction-stage risk, optimise temporary works and excavation sequencing, validate soil-structure interaction models, and adjust support systems based on observed behaviour. Its application has demonstrated tangible benefits in construction control, safety, and efficiency, in several complex underground works and high-profile infrastructure projects [2], [3].

Despite its strengths, the practical implementation of observational method is constrained by analytical and computational limitations. The method relies on repeated back-analysis of non-linear ground-structure interaction, often using sparse and indirect measurements such as wall deflections. This process is typically manual and time-intensive, limiting its ability to operate within construction timescales and preventing monitoring data from being fully leveraged in predictive models.

Digital twins offer a pathway to overcome these limitations by integrating monitoring data with

computational models to create continuously evolving representations of underground systems. While sensing, data platforms, and visualisation tools have advanced significantly, a robust back-analysis engine capable of translating sparse field observations into an updated, physically consistent state remains a key bottleneck. The current study addresses this gap by developing a physics-informed back-analysis framework that forms the reflective core of a digital twin for deep excavation. Beyond a conceptual link between digital twins and the observational method, it provides a practical workflow for converting sparse monitoring data into an updated, mechanically consistent system state for stage-wise decision-making.

The main contributions of this work are:

- A digital twin framework tailored to observational method in deep excavations.
- A parsimonious inverse physics-informed neural networks (PINNs) for physically consistent back-analysis from sparse and noisy data.
- Validation on canonical examples, including a retaining-wall idealisation, to demonstrate feasibility in a reduced-order setting.

2 Related work

2.1 Digital twins and observational method

Digital twins are virtual representations of physical systems that are continuously updated using field data to reflect current behaviour and support prediction and decision-making [4]. For underground construction, Babanagar et al. (2025) [5] proposed a layered digital twin architecture (Figure 1) with increasing levels of intelligence, ranging from reflective twins that update models in real time to prescriptive twins that actively support decision-making. In this framework, the physical system, including ground, structures, and sensors, feeds into a data layer that manages monitoring and simulation outputs. These data are linked to modelling and visualisation tools such as BIM and numerical models, while an application layer delivers functions including model updating, stability assessment, risk evaluation, and decision support. While this architecture establishes a robust conceptual structure, it also highlights the opportunity to further develop practical guidance for digital twin frameworks tailored to specific requirements.

Recent developments in the observational method research have explored advanced sensing technologies and data-driven techniques to improve data processing and quantify uncertainty in monitoring data. However, in practice, the observational method still relies heavily on manual interpretation and expert judgement for model updating and construction-stage decisions, which limits responsiveness and consistency across projects.

Digital twins provide a natural platform to address these limitations by enabling automated data integration, real-time model updating, and continuous assessment of construction performance. Integrated analytics and visualisation allow alternative construction scenarios to be evaluated efficiently, supporting proactive risk management and cost control. The principles of observational method align closely with these capabilities. It requires continuous monitoring, comparison of observed and predicted behaviour, and timely adjustment of construction strategies when deviations arise. Digital twins directly support these steps through automated data pipelines, rapid model updating, and efficient evaluation of alternative responses [6]. Despite this strong conceptual alignment, the integration of observational method and digital twins have largely been unexplored in research [5]. Observational method studies typically treat monitoring, interpretation, and decision-making as loosely coupled processes, while digital twins research in underground construction has focused mainly on data integration, visualisation, and asset management, with limited attention to structured observational method workflows.

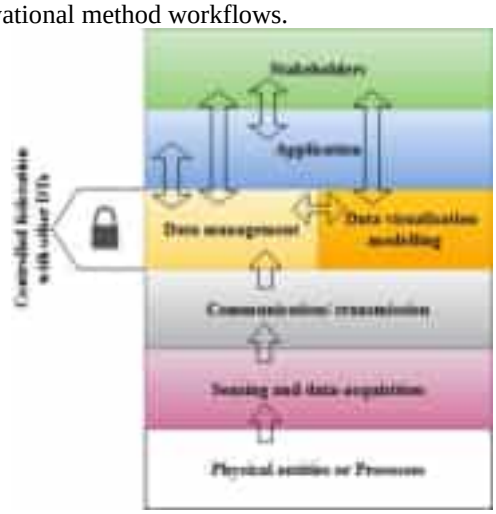


Figure 1: Layered architecture of digital twins for urban underground space.

A key barrier to effective integration is back-analysis, or system identification. Updating model parameters and load states from monitoring data is often iterative and computationally intensive, and typically requires specialist input, which limits its suitability for near real-time use. Without fast and reliable back-analysis, digital twins remain largely reflective and cannot fully support the adaptive decision-making that is central to the observational method. At the same time, the requirements of the observational method vary across underground construction types, including tunnels, deep basements, shafts, and caverns, due to differences in governing mechanisms, monitoring strategies, data characteristics, and decision timelines. This creates a

need for a tailored digital twin framework for the observational method that defines appropriate information flows, modelling and back-analysis techniques, and decision processes for each construction type, and that explicitly map these requirements across digital twin maturity levels, from reflective to predictive and prescriptive operation.

2.2 Back-analysis

Back-analysis of deep excavations has evolved from heuristic trial-and-error calibration and empirical case histories, through deterministic optimisation, towards probabilistic and hybrid data-driven approaches. Early practice relied on engineer judgement and empirical correlations between wall deflections and excavation geometry or ground conditions [2]. Formal optimisation-based back-analysis then coupled gradient-based or evolutionary algorithms with finite element models to minimise misfit between measured and simulated responses [7], improving objectivity but remaining computationally demanding and sensitive to local minima. Bayesian frameworks introduced probabilistic updating of soil parameters using staged monitoring data or related schemes, yet their cost can be prohibitive when combined with advanced constitutive models [8] . More recently, surrogate-assisted and deep learning-based methods have sought to reduce computational effort and exploit richer datasets, at the expense of reduced physical transparency [9]. In all these approaches, inferred parameters remain strongly dependent on the chosen constitutive soil model, so model-form error can be unwittingly hidden by best-fit parameters and propagated in future predictions. The evolution of back-analysis in the literature highlights a shift towards data-driven machine learning methods, while underlining that most existing techniques are either too costly for frequent updating or rely on black-box models that are difficult to trust in safety-critical decisions [10]. To address these issues, there is a need to adopt physics consistent methods such as physics-enhanced machine learning, which integrates domain knowledge and physical laws into data-driven models for greater transparency and reliability.

3 Digital twins for the observational method in deep excavations

Retaining walls supporting deep excavations are subjected to complex, stage-dependent earth pressures that evolve with excavation progression, support installation, and groundwater fluctuations. Temporary props or struts are designed conservatively based on stage-wise numerical models. These models often predict higher wall deflections and lateral loads than observed in the field, creating opportunities to optimise support systems, reduce prop requirements, and improve construction efficiency. Conversely, unforeseen variations in soil behaviour, groundwater conditions, excavation sequences, or environmental conditions can result in larger-than-anticipated wall movements, posing risks to wall stability and adjacent assets. Among these environmental factors, temperature fluctuations can have a significant influence on prop forces, as thermal expansion or contraction alters strut loads and can indirectly affect wall deflections. Accurate estimation of these effects is therefore essential to support adaptive construction decisions. Managing both opportunities and risks requires a system capable of continuously estimating the evolving structural state and forecasting future behaviour, which can be achieved by embedding the observational method within a digital twin framework tailored for deep excavation projects.

Practical application of the observational method in deep excavations is constrained by several factors. Monitoring networks are typically sparse and indirect, often comprising a few inclinometers, load cells, and piezometers, which cannot fully resolve the wall-soil-support system. Back-analysis is inherently ill-posed, as multiple combinations of lateral loads and support reactions may produce similar wall deflection patterns. Conventional approaches rely on iterative finite element calibration or empirical adjustments, which are often time-consuming, sensitive to model assumptions, and require dense instrumentation. In contrast, a physics-informed learning approach can directly incorporate governing mechanical laws, providing a data-efficient and physically consistent solution to the inverse problem. Moreover, repeatedly updating non-linear numerical models is computationally intensive, making stage-wise decision-making within construction timescales difficult. These limitations define the following specific requirements for a digital twin framework:

(i) data efficiency: ability to reconstruct structural states from sparse and noisy measurements.

(ii) computational speed: stage-wise updates and forecasts within construction timescales.

(iii) physical consistency: mechanically realistic and interpretable state estimates.

(iv) decision alignment: outputs directly mapped to pre-defined thresholds and construction controls.

3.1 Integrated layered architecture and maturity levels

The proposed digital twin framework for deep excavations is structured around four interdependent layers (Figure 2), enabling progressive levels of automation from descriptive to prescriptive operation. The physical layer captures the retaining wall, excavated soil, temporary supports, groundwater, and

instrumentation network, providing real-time measurements of structural response and environmental variables, including temperature, at each excavation stage. Measurements are acquired alongside construction metadata, such as excavation stage identifiers, prop configurations, and support layouts.

The data management layer consolidates measurements and metadata from multiple sources, performing data quality control, stage identification, time alignment, and contextual tagging. It acts as a central database where all project data are stored and enables federation of data with all stakeholders. The modelling and visualisation layer hosts numerical and surrogate models of the wall-soil-support system and maintains outputs from finite-element software (e.g., PLAXIS) and BIM-based structural representations. It integrates data from instrumentation dashboards, monitoring systems, and environmental sensors, ensuring that all observations, including temperature and water table variations are coherently visualised with the excavation sequence. Updated geometries, back-analysed lateral pressure distributions, and deflection profiles, are visualised through dashboards, providing interpretable insights for engineers. These outputs support assessment of wall performance against design criteria, trigger thresholds, and planned excavation stages. While these components are well established in conventional basement construction, they primarily provide descriptive or visual capabilities and do not inherently perform stage-wise back-analysis or forecasting.

The application layer forms the operational core and addresses the critical bottleneck of real-time system identification. It incorporates the inverse modelling or system identification engine, which performs reflective back-analysis to reconstruct a low-dimensional structural state vector describing wall deflections, lateral pressures, prop reactions, and environmental influences such as temperature-induced load variations. Forward simulation modules then use this state to predict system responses under planned excavation sequences, support adjustments, and environmental variations. Forecast envelopes, stage-wise load estimates, and trigger indicators guide adaptive construction decisions specific to deep excavation, including:

- Selective adjustment or removal of props when measured deflections and temperature-influenced loads remain within safe limits, optimising material use and schedule.
- Modification of excavation sequences, including stage height, duration, or order, based on current wall-soil system behaviour and anticipated temperature effects.
- Reconfiguration of temporary supports, such as strut repositioning or force redistribution by adjusting the hydraulic jacks in the props, to maintain stability under evolving loads and thermal variations.
- Groundwater control interventions, including controlling dewatering rate in response to changing hydrogeological conditions.
- Trigger-based contingency actions, such as temporary support installation, monitoring intensification, or staged excavation pauses, when forecasted deflections or temperature-influenced prop forces approach thresholds.

This layered organisation is aligned with digital twin maturity levels. At the descriptive level, monitoring data, dashboards, and PLAXIS/BIM visualisations allow engineers to track system behaviour and environmental variations. At the reflective level, the inverse engine reconciles observed behaviour with the virtual model to produce a physically consistent structural state. The predictive level propagates this state to forecast responses for subsequent excavation stages, supporting risk-aware planning under evolving conditions. At the prescriptive level, outputs guide adaptive decisions on excavation sequencing, prop management, support modifications, and groundwater interventions, enabling safe, efficient, and optimised construction.

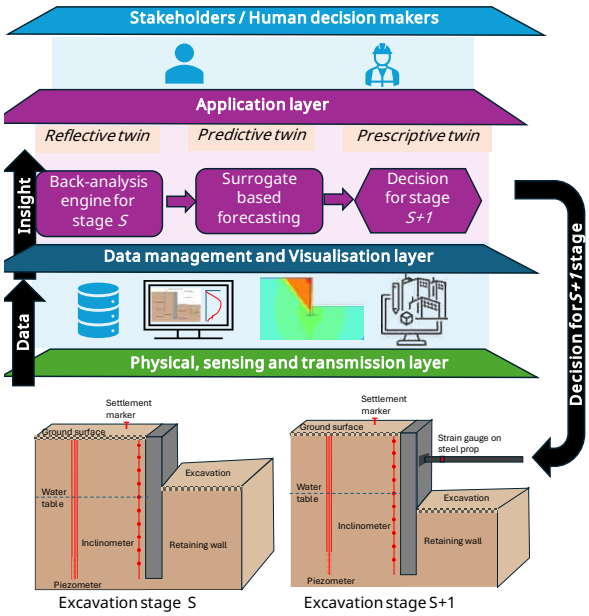


Figure 2: Layered architecture for digital twins of deep excavation to enable observational method.

As a first step in operationalising the framework, the back analysis or the inverse modelling engine serves as the critical component of the reflective twin, translating sparse monitoring data into a physically consistent representation of the wall-soil-support system. This structural state forms the foundation for both predictive forecasting and prescriptive decision-making within the

digital twin. To realise this capability efficiently and at scale, the framework employs PINNs, which provide a computationally effective and physically constrained approach for back-analysis in retaining-wall systems.

4 PINNs for back-analysis

The aim of the physics-informed neural networks is to perform inverse analysis i.e., to infer the lateral loads acting on retaining walls from sparse measurements of wall deflections. This provides the critical information required for reflective updating of the digital twin and supports stage-wise forecasting and adaptive decision-making in deep excavation projects.

4.1 Canonical cantilever beam

To develop and validate the core methodology, the PINNs-based inverse analysis is first applied to a canonical cantilever beam problem. This simplified model serves as a representative analogue of a retaining wall, capturing the essential mechanics of wall bending under lateral earth pressures. In this abstraction, the wall's flexural stiffness governs the deflection profile, allowing back-analysis to focus on inferring a small set of physically interpretable load parameters from sparse monitoring data. To simulate propped excavation conditions, temporary supports are modelled as discrete point loads applied at prop levels, representing measured prop reactions and further constraining the inverse problem. This approach maintains physical interpretability, computational efficiency, and practicality for stage-wise model updating in deep excavation projects.

In the canonical beam problem, the beam geometry, flexural stiffness EI, and boundary conditions were assumed to be known, while an unknown load distribution $q(x)$, is to be back-analysed from the sparse measured deflections, where x denoted the longitudinal coordinate along the beam axis. The transverse deflection $w(x)$ is governed by the Euler-Bernoulli beam equation which directly links the load distribution to the deflection profile and provides the core physical constraint in the PINN:

$$\mathrm{EI}\,.\,\frac{d^4 w(x)}{dx^4} - q(x) = 0 \qquad (1)$$

A synthetic benchmark was constructed by solving the beam analytically under a combination of a uniform load q and a mid-span point load P, and then sampling sparse deflection values corrupted with Gaussian noise to emulate inclinometer data (Figure 3). The inverse task is to recover both the continuous deflection field $w(x)$ and the underlying load distribution $q(x)$ from these noisy observations. Direct recovery of $q(x)$ by differentiating sparse, noisy deflections is ill-posed due to numerical instability of high-order derivatives, motivating a joint learning formulation where deflection and load are inferred simultaneously within a shared physics-informed network.

To regularise the problem and embed prior engineering knowledge, the distributed load is parameterised as a superposition of a uniform component and a localised (concentrated) component, representing lateral earth pressure and surcharge effects. In case of retaining walls, classical earth pressure formulations such as limit equilibrium method can be adopted as prior engineer knowledge.

The deflection *w(x)* is represented by a feedforward neural network with two hidden layers of 30 neurons and Swish activation. Automatic differentiation is employed to compute the high-order derivatives required by Equation (1), enabling stable enforcement of the governing equation during training.

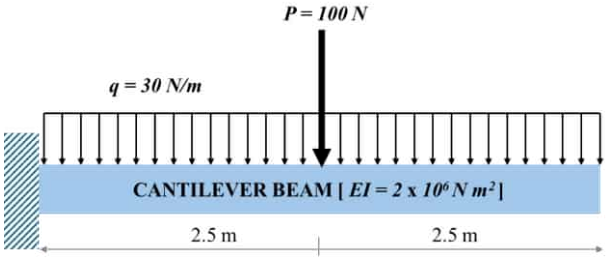


Figure 3: Canonical cantilever beam setup used for demonstrating PINNs framework.

The total training loss ($\mathcal{L}_{\text{total}}$) is defined as

$$\mathcal{L}_{total} = \lambda_d \cdot \mathcal{L}_d + \lambda_{phy} \cdot \mathcal{L}_{phy} + \lambda_{bc} \cdot (\mathcal{L}_{bc}) \qquad (2)$$

$$\mathcal{L}_{bc} = w(0)^2 + w'(0)^2 + w''(L)^2 + w'''(L)^2 \qquad (3)$$

Where $\mathcal{L}_d$ is the data loss enforcing agreement between predicted and observed deflections, $\mathcal{L}_{phy}$ is the physics loss penalising residuals of from Equation (1) at collocation points within the domain, $\mathcal{L}_{bc}$ is the boundary condition loss enforcing Dirichlet (deflection and slope at fixed end) and Neumann boundary conditions (Bending moment and Shear force at free end) of the beam (Equation (3). The weights λ_d, λ_{phy}, and λ_{bc} control the relative influence of data fidelity, physics consistency, and boundary enforcement. Two Adam optimisers are used: one updates the network weights, and the other updates the trainable load parameters (including the point load), constrained to be non-negative via a soft plus transformation.

The magnitude of the weights assigned to the data, physics, and boundary condition loss terms directly influences the accuracy of the inverse PINN solution. To identify the most suitable weighting scheme, a sensitivity analysis was performed in which 30 different weight

combinations were tested, and their results were compared. Solution accuracy was evaluated using two indicators: (i) the coefficient of determination R^2 (deflection) between the predicted deflection profile and the analytical reference, and (ii) the accuracy of the load distribution predictions. To enable balanced comparison, both deflection metric (R^2) and the load-accuracy metric were min-max normalised across all runs to place them on a common 0-1 scale. A single performance score was then defined as the arithmetic mean of the normalised values. Using only 16 sparse deflection measurements with 10% Gaussian noise, each model was trained for 8000 epochs. Figure 4 presents the sensitivity analysis results, from which the optimal weight combination ($\lambda_d, \lambda_{phy}, \lambda_{bc}$)=(0.4,0.4,0.2) were selected based on the highest overall score, indicating a balanced trade-off between data fit, physics enforcement, and boundary satisfaction. All computations were performed on a laptop with an Intel Core Ultra 7 165H CPU and 32 GB RAM; for the canonical inverse-analysis setup trained for 8000 epochs, the runtime was approximately 63 s per case

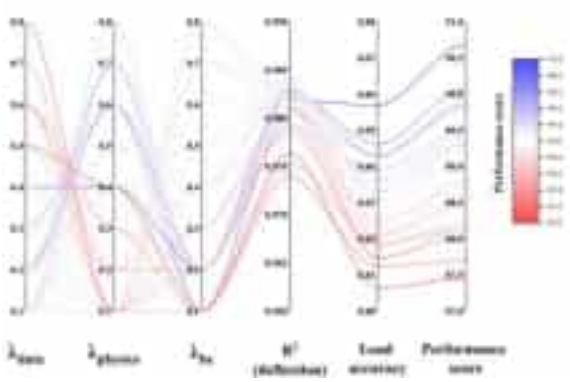
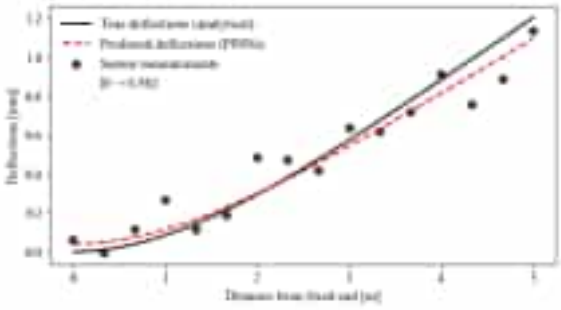


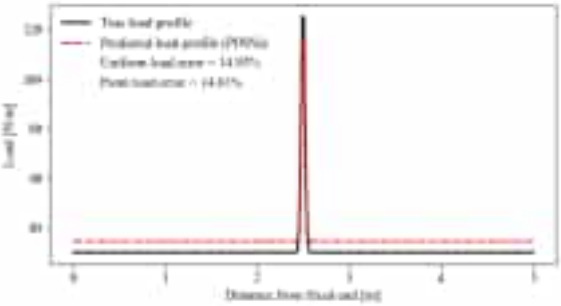
Figure 4: Sensitivity of PINN loss weights on predictive performance, evaluated by R2 for deflections, load prediction error, and overall performance score.

Figures 5a and 5b present the predicted deflection profile and the inferred load distribution obtained using the best-performing weight combination identified in the sensitivity study. The deflection predicted by the PINN closely matches the analytical solution, achieving a coefficient of determination of $R^2 = 0.982$. Despite the sparse and noisy measurement inputs, the incorporation of governing beam mechanics through the physics loss constrains the solution space, preventing overfitting and ensuring mechanically consistent behaviour along the beam. The recovered load distribution $q(x)$captures both the uniform and localised components of the applied loading with good agreement. The estimated uniform load exhibits an error of approximately 14.9%, while the recovered point load shows an error of 14.6%.

For comparison, Figure 6 shows the results obtained using a purely data-driven neural network of identical architecture and trained for the same number of epochs, using deflection data only. In this case, the predicted deflections exhibit a lower agreement with the analytical solution, with $R^2 = 0.913$. This comparison highlights the benefit of embedding physical constraints within the learning process, particularly when measurements are sparse and noisy, as is typical in field monitoring of retaining structures.



(a)



(b)

Figure 5: a) (a) Deflection and (b) Load profile - recovered from PINN framework compared with true values for cantilever beam canonical problem.

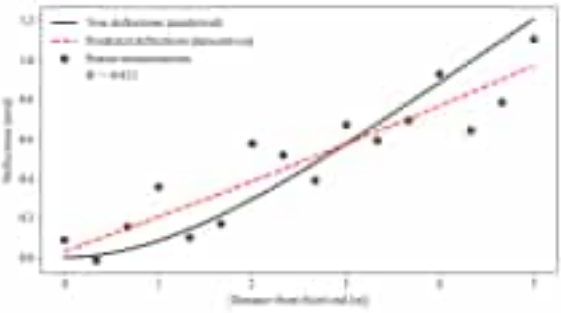
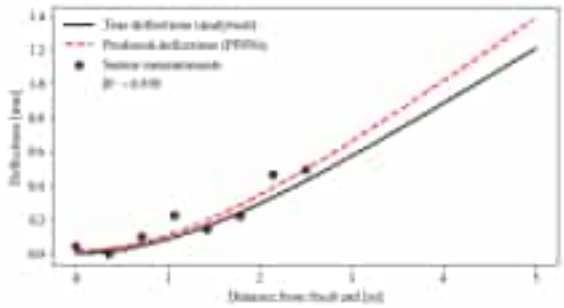
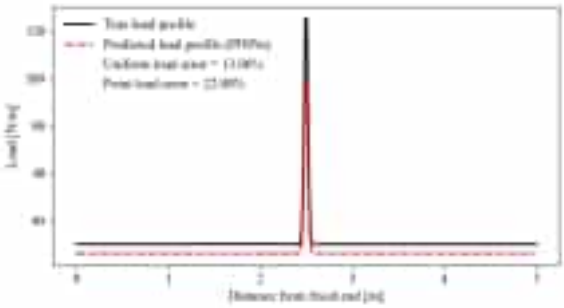


Figure 6: Deflections recovered from purely data-driven method compared with true deflections for cantilever beam canonical problem.

The extrapolation capability of the PINN was further assessed in an experiment where only eight sensors, each with 10% Gaussian noise, were assumed to be placed on the beam up to its mid-span. As shown in Figure 7a, the PINN still predicts the deflection profile with good accuracy ($R^2 = 0.939$), while Figure 7b shows that both the uniform load (13.1% error) and point load (22.8% error) are reasonably recovered. In contrast, the purely data-driven network (Figure 8) performs significantly worse, with a reduced accuracy ($R^2 = 0.850$) and clear signs of overfitting. This demonstrates that the PINN maintains stable and physically consistent inference even with sparse and spatially limited measurement.



(a)



(b)

Figure 7: a) Deflection and (b) Load profile - recovered from PINN framework for extrapolation task on cantilever beam canonical problem.

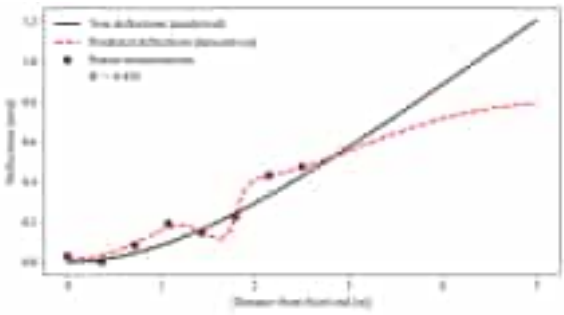


Figure 8: Deflections recovered from purely data-driven method for extrapolation task on cantilever beam canonical problem.

4.2 Back-analysis on a retaining wall

Building on the cantilever beam example, the inverse PINN framework is extended to a canonical propped retaining wall representative of deep-excavation behaviour. To keep the inverse problem tractable while preserving physical interpretability, the lateral earth pressures are represented by classical linear depth-dependent active and passive pressure distributions scaled by mobilisation coefficients m_a and m_p. These coefficients, bounded as $m_a, \mathrm{m_p} \in [0,1]$, act as lumped parameters capturing partial mobilisation and provide an inductive bias that stabilises inversion from sparse measurements.

Figure 9 illustrates the wall configuration, with total height H, retained height h, and a prop force P applied below ground level; wall deflections are observed at eight sensor locations along the depth. The inverse task is to recover m_a and m_p from these sparse deflection measurements, optionally incorporating measured prop force where available.

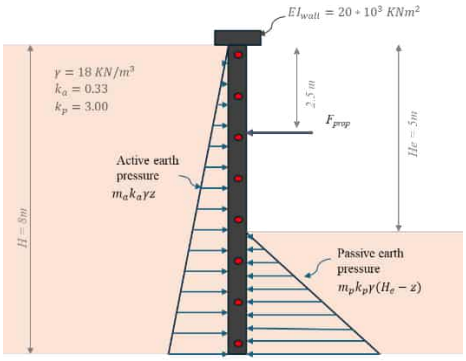


Figure 9: Illustration of the propped retaining wall canonical example.

Synthetic training data are generated by prescribing m_a, m_p, and P, solving the corresponding boundary-value problem, and perturbing the sampled deflections with 10% Gaussian noise. The network architecture and training strategy follow the cantilever-beam case, with depth as input, deflection as output, and automatic differentiation used to enforce the Euler-Bernoulli residual. Using the best-performing loss weights identified previously, the results in Figure 10 show that the PINN can recover excavation-relevant reduced-order state variables from sparse and noisy data. This retaining-wall example therefore provides a more application-relevant intermediate step between the beam analogue and future field deployment.

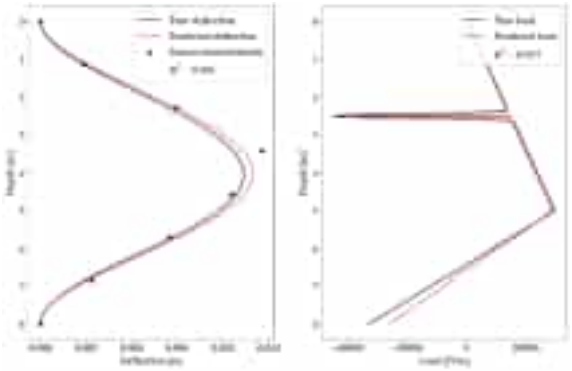


Figure 10: (a) Deflection and (b) load profiles recovered by PINN inverse analysis for the retaining wall.

5 Conclusions

This paper has presented a digital twin framework that embeds the observational method within a layered architecture for deep excavations, clarifying the roles of monitoring, data management, modelling, and application layers, and identifying key requirements including data efficiency, computational speed, physical consistency, scalability, and alignment with trigger-based decision-making.

Within this framework, a parsimonious inverse PINN was developed as the reflective engine of the twin. Validation on a canonical cantilever-beam analogue of a propped retaining wall showed that the method can recover deflection profiles and mobilised load distributions from sparse and noisy measurements, while outperforming a purely data-driven neural network and retaining mechanical interpretability and numerical stability.

Together, the framework and inverse PINN provide a basis for reflective digital twins that support stage-wise model updating and integration with construction monitoring and decision workflows. By enabling physically consistent interpretation of sparse monitoring data, the approach offers a pathway toward predictive and prescriptive digital-twin capabilities for deep excavation management.

Several important areas for future work include, validation using real excavation monitoring data, broader benchmarking against optimisation-based and probabilistic back-analysis approaches, and extension to more realistic retaining-wall idealisations and higher-fidelity 2D/3D soil-structure interaction models. Further work is also needed to assess robustness and computational performance for routine project deployment

6 References

[1] S. Hardy, D. Nicholson, P. Ingram, A. Gaba, Y. Chen, and G. Biscontin, "New observational method framework for embedded walls," *Geotechnical Research*, vol. 5, no. 3, pp. 122–129, Sep. 2018, doi: 10.1680/jgere.18.00013.

[2] H.-C. Yeow, D. Nicholson, C.-L. Man, A. Ringer, P. Glass, and M. Black, "Application of observational method at Crossrail Tottenham Court Road station, UK," *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, vol. 167, no. 2, pp. 182–193, Apr. 2014, doi: 10.1680/geng.13.00035.

[3] A. J. Powderham and A. O'Brien, *The observational method in civil engineering: maximising value, minimising risk*, First edition. Boca Raton: CRC Press, 2021.

[4] AIAA, "Digital twin: Definition & Value," Aerospace Industries Association, Dec. 2020.

[5] N. Babanagar, B. Sheil, J. Ninić, Q. Zhang, and S. Hardy, "Digital twins for urban underground space," *Tunnelling and Underground Space Technology*, vol. 155, p. 106140, Jan. 2025, doi: 10.1016/j.tust.2024.106140.

[6] H. Amthiou, M. Arioua, and T. Benbarrad, "Digital Twins in Industry 4.0: A Literature Review," *ITM Web Conf.*, vol. 52, p. 01002, 2023, doi: 10.1051/itmconf/20235201002.

[7] C. Ou and Y. Tang, "Soil parameter determination for deep excavation analysis by optimization," *Journal of the Chinese Institute of Engineers*, vol. 17, no. 5, pp. 671–688, Jul. 1994, doi: 10.1080/02533839.1994.9677634.

[8] HseinC. Juang, Z. Luo, S. Atamturktur, and H. Huang, "Bayesian Updating of Soil Parameters for Braced Excavations Using Field Observations," *J. Geotech. Geoenviron. Eng.*, vol. 139, no. 3, pp. 395–406, Mar. 2013, doi: 10.1061/(asce)gt.1943-5606.0000782.

[9] Y. M. A. Hashash, C. Marulanda, J. Ghaboussi, and S. Jung, "Novel Approach to Integration of Numerical Modeling and Field Observations for Deep Excavations," *J. Geotech. Geoenviron. Eng.*, vol. 132, no. 8, pp. 1019–1031, Aug. 2006, doi: 10.1061/(asce)1090-0241(2006)132:8(1019).

[10] C. G. Chua and A. T. C. Goh, "Estimating wall deflections in deep excavations using Bayesian neural networks," *Tunnelling and Underground Space Technology*, vol. 20, no. 4, pp. 400–409, Jul. 2005, doi: 10.1016/j.tust.2005.02.001.

Inferring Spatial Interface Dependencies from Co-change Records in Civil Infrastructure Design Models

Ryoko Arashida[1] and Masahide Horita[2]

[1] Institute of Engineering Innovation, School of Engineering, The University of Tokyo, Tokyo, Japan
[2] Department of Civil Engineering, School of Engineering, The University of Tokyo, Tokyo, Japan
arashida@i-con.t.u-tokyo.ac.jp, horita@g.ecc.u-tokyo.ac.jp

Abstract

This paper presents a data-driven framework for inferring spatial interface dependencies in large-scale civil infrastructure design using empirically observed co-change as a proxy for coordination-related interactions. Using tender and as-built 3D models from an expressway project, object-level changes are identified and distance-based inter-category features are constructed to predict change occurrence. Among the evaluated classifiers, LightGBM achieved the best performance (average G-mean = 0.903). The results show that co-change is better explained by neighborhood-level spatial interactions than by single nearest-neighbor relationships. Distance cutoff analysis indicates that predictive performance saturates within limited spatial ranges, typically within about 30 m in this dataset. Furthermore, Individual Conditional Expectation (ICE) analysis reveals five distinct distance-dependent co-change patterns, while also identifying cases where distance reflects broader spatial heterogeneity rather than direct interface relationships. These findings suggest that spatial distance captures a substantial portion of co-change behavior and can support prioritization of interface checks and selection of search radii in infrastructure design coordination.

Keywords –

Co-change; Hard/soft clash; Building Information Modeling; Machine learning; Design change; Change Propagation

1 Introduction and Related Works

In large-scale Architecture, Engineering, and Construction (AEC) projects, resolving all inconsistencies during design is unrealistic because design and construction are separated, site conditions are uncertain, and production is largely project-specific [1, 2]. Consequently, design changes are unavoidable and are major drivers of cost overruns and schedule delays [3, 4]. The key challenge is therefore not to eliminate changes, but to clarify when they propagate across elements and to limit cascading coordination efforts.

Change propagation is governed by interface constraints between elements, particularly geometric relationships such as connectivity, proximity, and clearance [5]. Under these constraints, local modifications often require adjustments to related elements, which are observed as coordinated updates among multiple components. In this study, such coordinated updates are referred to as co-change.

In software engineering, several studies have analyzed version histories to identify software artifacts that tend to be modified together, often described as logical coupling, evolutionary coupling, or change patterns [6–8]. However, these studies primarily focus on temporal co-change relationships in software repositories and do not explicitly consider spatial relationships between physical components. In contrast, spatial dependencies are central in civil infrastructure systems and are explicitly addressed in this study through geometric distance-based analysis.

With the widespread adoption of Building Information Modeling (BIM), automated detection of interface-related inconsistencies has become widely used. Model-checking research has expanded from hard/soft clash detection to workspace conflicts in construction planning [9–11], and compliance checking against regulations and design rules [12–14]. However, these approaches depend on explicitly specified constraints and dependency rules, making large-scale deployment costly to develop and maintain. Even soft clash detection often requires clearance rules for each object pair, which is burdensome in practice [15].

In parallel, computer graphics and geometric learning have developed methods to extract implicit geometric relationships from large shape collections by learning regularities in component arrangements and encoding them as reusable knowledge or constraints [16–20]. However, these approaches typically assume stable component decompositions and limited interaction ranges, limiting their applicability to the AEC domain. While ML-based generative design has been actively

studied in AEC using drawings and photographs, existing work largely focuses on early-stage, single-building problems and does not adequately capture relationships among heterogeneous elements, due to the knowledge- and reasoning-intensive nature of AEC design [21].

These limitations motivate the use of co-change as an empirical signal for inferring interface dependencies. By treating change records as observable outcomes of coordination processes, spatial relationships between elements can be analyzed without explicitly defining all interface rules.

In this study, object-level changes **derived** from differences between tender and as-built models are used to analyze co-change among structural categories and its dependence on object-to-object distance. **Focusing** on geometric interface **conditions, distance**–co-change relationships are identified and organized into interpretable patterns using Individual Conditional Expectation (ICE) analysis.

2 Methodology

2.1 Workflow Overview

Figure 1 illustrates the workflow of this study. The inputs are the tender and as-built models of an expressway infrastructure project. As a preprocessing step, both models are segmented into category-specific analysis objects using predefined rules, and each object is assigned a binary change label indicating whether it has changed between the two models.

Based on these labeled objects, spatial interfaces between categories are represented using distance-based features. Specifically, multiple inter-category distance representations are constructed to characterize the spatial relationship between a target object and potentially related objects in other categories. These distance-based representations serve as explanatory variables for predicting object-level change occurrence.

The analysis proceeds in four stages. First, a suitable classification model is selected through comparative evaluation. Second, alternative distance-feature representations are assessed to examine how spatial information should be encoded. Third, effective distance cutoffs are determined to identify the spatial ranges within which distance information contributes meaningfully to change prediction. Fourth, important distance features are identified through feature selection to obtain a compact and informative representation.

For the selected distance features, Individual Conditional Expectation (ICE) is employed to examine how changes in distance are associated with changes in predicted change probability. By jointly interpreting the shapes of the mean ICE curves and the corresponding distance distributions, co-change behavior between categories is organized into interpretable interface patterns.

The outputs of the framework are (1) effective distance ranges for predicting change occurrence in each target category and (2) a typology of distance-dependent, interface-based co-change patterns.

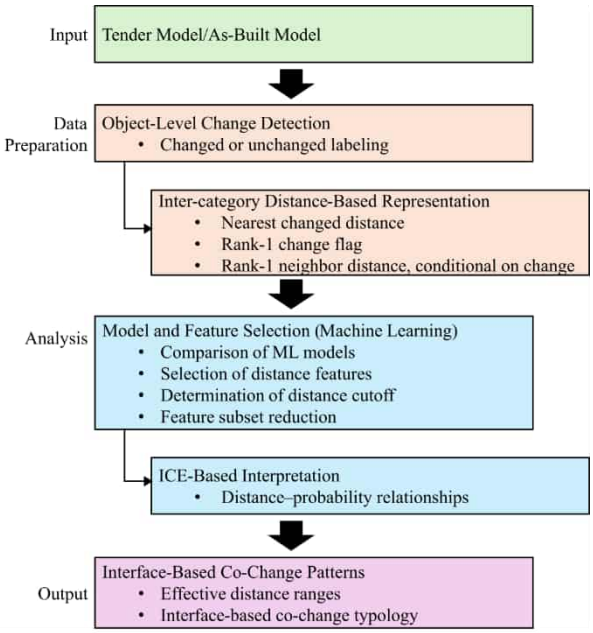


Figure 1. Overview of the proposed analysis framework.

2.2 Data and Preprocessing

2.2.1 Target Models, Categories, and Object Segmentation

The dataset comprises the tender and as-built models of an expressway infrastructure project. The study area covers approximately 350 m × 2100 m and includes bridge and embankment sections, excluding cut sections and tunnels. A total of 25 design categories are analyzed.

Because civil infrastructure models often include large continuous elements whose native BIM representations do not correspond to analysis units, category-specific segmentation rules were defined to obtain consistent object-level units reflecting structural boundaries and management units. Continuous elements were segmented based on structural or alignment-based rules. For bridge-related categories, segmentation followed structural boundaries, while for road and earthwork categories, objects were segmented along the road centerline at 20 m intervals, corresponding to the cross-section spacing in the tender drawings. In contrast, discrete components were treated as single elements without further subdivision. Table 1 summarizes the category definitions and sample sizes, and Figure 2

illustrates the bridge segmentation scheme.

Table 1. Summary of categories, sample sizes, and change rates.

Category	Sample Size	Change Rate
Expressway Pavement	**317**	**21%**
Service Road Pavement	93	95%
Earthwork Crest	38	26%
Earthwork Toe	41	49%
Deck Slab	**1,343**	**35%**
Main Girder	1,356	96%
Cross Girder	**1,332**	**30%**
Pier Shaft	107	3%
Pier Cap	66	5%
Foundation	70	4%
Bearing	199	2%
Bearing Pedestal	199	6%
Fall Protection	20	70%
Barrier Wall	**2,874**	**28%**
Bridge Inspection Walkway	93	99%
Bridge Drainage	100	89%
Detention Pond	11	100%
Drainage	**848**	**40%**
Temporary Support	207	99%
Construction Yard	75	28%
Girder Storage	401	100%
Crane Position	170	100%
Access Road	65	9%
River Diversion	7	100%

Note: Categories shown in **bold** indicate the target categories selected for analysis (Section 2.3).

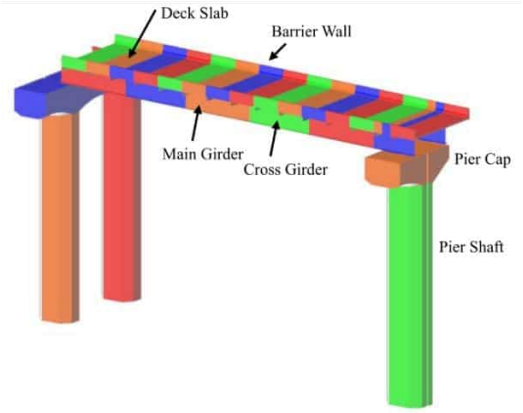


Figure 2. Example of object segmentation for bridge components.

2.2.2 Definition of Change Occurrence

Objects from the tender and as-built models are matched at the object level after projecting their coordinates onto a common 2D plane. An object is labeled as changed if (1) it appears in only one model, or (2) if present in both models, its centroid displacement is at least 10 cm. All remaining objects are labeled as unchanged.

The 10 cm threshold is used as a practical tolerance to mitigate minor positional noise introduced during model creation and integration while preserving substantively meaningful changes. Category-wise change rates are reported in Table 1.

2.3 Target Category Selection

Among the 25 categories, only categories that meet the following criteria are analyzed to ensure stable model training and interpretable results: at least 240 objects in the tender model and a change rate between 20% and 80%. For each target category, distance features are constructed using the other 24 categories as predictor sources. The minimum sample size of 240 was set as ten times the number of explanatory categories, and the change-rate range was used to avoid severe class imbalance that can destabilize the classifier and obscure distance–change relationships. Consequently, although only a subset of categories is used as prediction targets, all categories are retained as sources of explanatory distance features, meaning that their spatial influence on the analyzed targets remains represented in the model.

Based on Table 1, five categories satisfy these criteria: expressway pavement, deck slab, cross girder, barrier wall, and drainage.

2.4 Distance-Feature Construction

For each object in each of the five target categories, neighboring objects are extracted from all other categories. All objects are projected onto a common 2D plane prior to distance computation. In civil infrastructure models, most design categories occupy relatively consistent vertical bands (e.g., road-surface elements versus ground-level components). As a result, situations in which objects appear close in plan view but are vertically separated tend to occur systematically between specific category pairs rather than randomly at the object level. Because the model uses category-based distance features, such systematic vertical offsets are reflected consistently in the learned relationships rather than introducing random noise. To partially account for vertical structure, Earthwork and Pier categories were subdivided into two components each, representing distinct interaction zones even when distances are computed on a 2D projection.

For each target object and each non-target category, 50 candidate neighbors are first identified based on centroid distances. Exact boundary-based distances are then computed on the 2D plane, and the 20 closest neighbors are retained together with their distances and

change labels.

For each target category, four datasets are constructed using distance-based features derived from other categories. If no changed object is found within the 20 nearest neighbors, a sentinel value (maximum observed distance + 10 m) is assigned. This value is intentionally placed beyond the range of typical neighbor distances so that it functions as an indicator of "no nearby change" rather than a meaningful spatial distance, and therefore does not affect the learned splits of the classifier. The datasets differ in how distance information is represented:

- Dataset 1: the minimum distance to the nearest changed object in each category.
- Dataset 2: a binary indicator of whether the nearest neighbor (rank 1) is changed.
- Dataset 3: distance to the nearest neighbor, conditional on that neighbor being changed.
- Dataset 4: all features from Datasets 1–3 combined.

To assess whether nearest-neighbor relationships alone can capture design coordination effects, the rank-1 datasets (Datasets 2 and 3) are compared with Dataset 1, which can capture changed neighbors beyond the nearest object.

2.5 Machine Learning (ML) Setup for Change Prediction

For each target category, the prediction models use the change label of the target object as the response variable, while the distance-based features derived from the remaining categories (Datasets 1–4) are used as explanatory variables. Change prediction is formulated as a binary classification task, in which each object is labeled as either changed or unchanged. The following algorithms are compared:

- Logistic regression (L1/L2/elastic net)
- Random Forest
- Extra Trees
- HistGradientBoosting
- LightGBM

All models are implemented in Python using scikit-learn and LightGBM with default hyperparameters. For each target category, the data are split into 90% for training and 10% for testing. This procedure is repeated for 10 different random seeds, and evaluation metrics are averaged across seeds to improve robustness and reproducibility. To account for class imbalance, performance is evaluated primarily using the G-mean, defined as the geometric mean of sensitivity and specificity.

2.6 Distance Cutoff, Feature Selection, and ICE-Based Patterning

A distance cutoff sensitivity analysis is performed to identify the spatial range over which distance features contribute to explaining co-change. An upper distance limit is imposed on all features and progressively reduced, while classification performance is evaluated. Distances exceeding the limit are replaced with a large value, and the effective cutoff is defined as the smallest limit that retains at least 95% of the baseline performance.

Feature selection is then conducted using the selected ML model. Features are ranked by LightGBM gain and added sequentially while monitoring performance. The final feature set is defined as the smallest set that retains at least 95% of the baseline performance.

ICE is then computed for each target category and each retained distance feature using models retrained on the final feature set. For each feature, object-level ICE curves and their mean are derived to characterize the distance–change probability relationship. ICE is evaluated from 0 m to the effective distance cutoff or the maximum observed distance, whichever is smaller.

To stabilize estimation, distances are discretized into 30 equal-width bins. Bins with fewer than five samples are excluded, and features with fewer than three valid bins are removed as unreliable. In addition, ICE curves with less than 1% probability variation are excluded, as they indicate negligible sensitivity to distance. Such low-variation curves correspond to nearly flat relationships and primarily introduce noise rather than additional pattern types.

Co-change interface patterns are identified from the mean ICE curves and their associated distance distributions. Classification considers both the shape of the ICE curves (e.g., monotonic, plateau, or peak behavior) and the distribution of observed distances, and category pairs are grouped into patterns exhibiting similar distance-dependent behavior.

3 Results

3.1 Comparison of Classification Algorithms

Across all target categories and dataset settings, LightGBM achieved the highest mean G-mean (0.903) and showed stable performance across random seeds. Although performance varied by category, LightGBM generally maintained G-mean ≥ 0.80 and was selected for subsequent analyses.

3.2 Comparison of Feature Representations and Dataset Selection

Using LightGBM, Dataset 4 yielded the best overall performance, while Dataset 1 achieved a comparable G-

mean with substantially fewer features. As shown in Figure 3, Dataset 1 provided a stable and interpretable baseline across target categories, whereas Datasets 2 and 3 exhibited greater variability. Therefore, Dataset 1 was adopted for subsequent analyses. The baseline G-mean values with Dataset 1 were 0.817 (expressway pavement), 0.938 (deck slab), 0.942 (cross girder), 0.978 (barrier wall), and 0.889 (drainage).

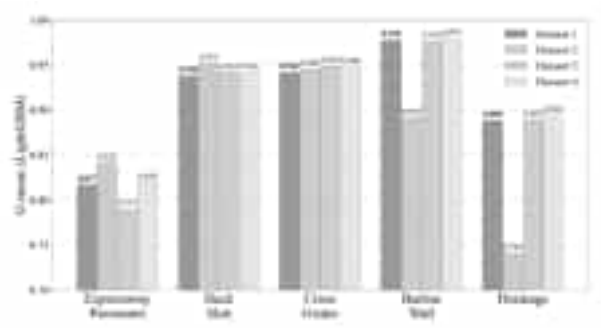


Figure 3. Comparison of distance-based feature representations using G-mean (LightGBM).

3.3 Effective Distance Ranges

Distance cutoff sensitivity analysis indicated that performance improvements saturated within a limited spatial range for most target categories. The resulting effective distance cutoffs (defined in Section 2.6) differed by category and are summarized in Figure 4: 15.5 m (expressway pavement), 0.5 m (deck slab), 0.5 m (cross girder), 2.0 m (barrier wall), and 28.5 m (drainage).

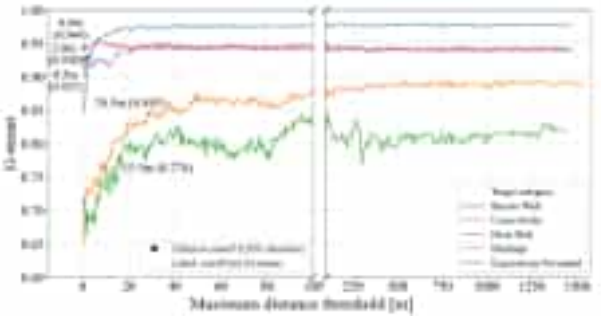


Figure 4. Distance Cutoff Sensitivity of G-mean Performance.

3.4 Important Features

Gain-based feature selection produced compact feature sets while retaining at least 95% of the baseline performance. The retained distance features are:

- Expressway Pavement (8): Barrier Wall, Girder Storage, Deck Slab, Crane Position, Earthwork Toe, Service Road Pavement, Fall Protection, Drainage
- Deck Slab (1): Cross Girder
- Cross Girder (1): Deck Slab
- Barrier Wall (4): Cross Girder, Deck Slab, Main Girder, Service Road Pavement
- Drainage (6): Detention Pond, Expressway Pavement, Service Road Pavement, Earthwork Toe, Earthwork Crest, Temporary Support

3.5 ICE Patterns of Distance-Dependent Co-Change Behavior

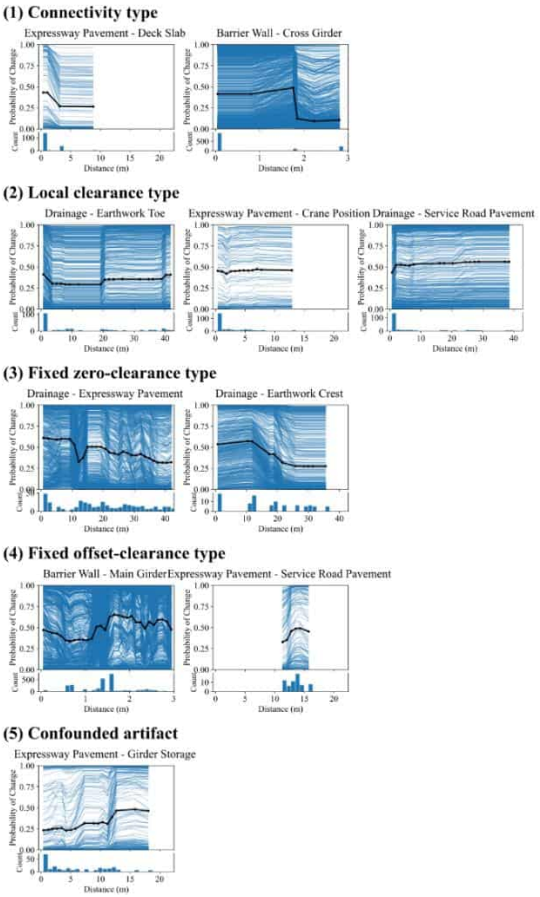


Figure 5. Classification of distance-based co-change interface patterns based on ICE analysis. *Note:* Blue lines show individual ICE curves, the black line indicates their mean ICE curve, and the histogram shows the distance distribution.

ICE analysis was applied to the features remaining after feature selection and reliability filtering, leaving only 10 category-pair distance features overall. Figure 5 presents the corresponding mean ICE curves and distance histograms. Following the criteria in Section 2.6, these profiles were grouped into five co-change interface patterns, reflecting differences in the mean ICE shapes

and associated distance distributions.

4 Discussion

4.1 Implications Beyond the Nearest Neighbor

Across the evaluated representations, Dataset 1, which uses the minimum distance to the nearest changed object in each category, provided higher performance and greater stability than Datasets 2 and 3, which rely on the change state of the single nearest neighbor (Figure 3). This indicates that co-change is not explained solely by the nearest object, but is more strongly associated with the presence of changed objects within a local neighborhood. These results suggest that spatial distance captures a substantial portion of the observed co-change behavior, highlighting spatial proximity as a key predictive signal for interface dependencies.

4.2 Implications of Effective Distance Upper Bounds

Distance cutoff sensitivity analysis showed that predictive performance was largely preserved as information from more distant objects was excluded, with performance generally saturating within approximately 30 m (Figure 4). This supports the view that coordination is primarily governed by local spatial interactions. However, the effective distance ranges varied across categories, indicating that a uniform threshold may be suboptimal and that category-specific search radii could provide a more rational basis for practical checks. In particular, categories with higher predictive accuracy were dominated by short-distance information, whereas lower-accuracy categories exhibited broader effective ranges, suggesting a greater influence of longer-range spatial constraints.

4.3 Differences in the Number of Important Features and Dependency Complexity

As shown in Section 3.4, some target categories preserved predictive performance using only a small set of distance features, suggesting that their change occurrence is strongly driven by a limited number of distinct interface relationships. In contrast, categories with lower predictive accuracy, particularly expressway pavement and drainage, required a larger set of important features, indicating more multifaceted dependencies on surrounding categories and/or the influence of factors not fully captured by distance features alone.

4.4 Interpretation of ICE-Based Patterns and Mechanism Hypotheses

As illustrated in Figure 5, ICE profiles of the important distance pairs can be grouped into five types, reflecting distinct ways in which distance is associated with co-change. Here, ICE refers to the mean ICE profile, and DD denotes the corresponding distance distribution.

1. Connectivity type (ICE: sharp drop after peak; DD: concentrated near zero)

High co-change probability at near-zero distances followed by a rapid decline suggests direct or near-contact interfaces, where coordination is driven by maintaining compatibility at connection points.

2. Local clearance type (ICE: plateau; DD: zero-peaked but widespread)

A flat ICE profile despite broad distance distributions indicates that effective clearance varies across objects, implying that uniform category-level rules may be insufficient.

3. Fixed zero-clearance type (ICE: gradual decay after peak; DD: zero-peaked and widespread)

A gradual decrease in co-change probability with distance suggests proximity-driven interactions without strict clearance requirements, primarily reflecting avoidance of hard clashes.

4. Fixed offset-clearance type (ICE: increase before peak, gradual decay after peak; DD: non-zero peak and widespread)

A peak at non-zero distances indicates characteristic clearance bands, consistent with standardized layouts or design rules.

5. Confounded artifact (non-mechanistic; reported for completeness)

Anomalous patterns driven by project-specific spatial heterogeneity rather than direct interface mechanisms are excluded from the primary typology.

4.5 Transferability and Prospective Use

The present results are derived from a single expressway project and should not be interpreted as universally transferable numerical rules. In particular, absolute distance cutoffs and detailed category-pair behaviors may remain partly project-dependent, reflecting the geometry, segmentation scheme, and revision history of the analyzed case.

However, the observed co-change relationships are unlikely to be entirely project-specific. Many spatial interfaces in civil infrastructure are governed by stable engineering conditions, such as structural connectivity, standard cross-sections, and clearance requirements, which are typically defined by design codes and common practices. As a result, while numerical thresholds may vary, the identified interface-pattern types and their

spatial tendencies are expected to generalize across comparable projects.

The framework can therefore be used to extract reusable coordination knowledge from completed projects. Patterns identified from historical revisions can inform subsequent projects by suggesting category pairs requiring closer coordination and by providing initial search ranges for spatial checks.

Transferability is expected to be highest among projects sharing similar asset types and regulatory environments, such as highway projects designed under comparable design codes and maintenance requirements. By contrast, application to substantially different infrastructure types (e.g., tunnels or coastal structures) or to projects developed under different national or organizational design regimes would require additional empirical data and revalidation.

5 Conclusion

This study proposed a data-driven framework to infer interface dependencies in a large-scale civil infrastructure project using empirically observed co-change events. Object-level changes between tender and as-built models were modeled using distance-based inter-category features, and distance–change relationships were interpreted using ICE.

Using the minimum distance to the nearest changed object within each explanatory category yielded higher accuracy than using only the change state of the single nearest neighbor, indicating that co-change is not adequately explained by one-to-one proximity, and is better characterized at the local neighborhood level. These results also indicate that spatial distance alone captures a substantial portion of the observed co-change behavior. Distance cutoff analysis further showed that most predictive power was retained within category-specific spatial ranges, generally saturating within about 30 m, supporting the use of category-dependent search radii rather than a uniform threshold. In addition, gain-based selection reduced the feature set while retaining at least 95% of baseline performance, and ICE analysis of the retained features revealed five distance-dependent co-change interface patterns, with confounded cases reported separately.

Taken together, these results suggest that the workflow can support the inference of plausible interface-related coordination mechanisms by identifying influential category pairs, the spatial ranges over which distance is informative, and characteristic distance–probability profiles. These outputs provide actionable guidance for prioritizing inter-category checks and designing distance-based search and coordination rules.

It is important to note that the numerical cutoffs and category-pair relationships identified in this study should be interpreted as case-grounded observations rather than universally applicable interface rules. Because the analysis is based on a single infrastructure project, the exact numerical values may vary depending on project geometry, design standards, and coordination practices. Nevertheless, many spatial interface conditions in civil infrastructure are shaped by recurring engineering requirements, such as structural connectivity, clearance provisions, and operation-and-maintenance needs defined in design standards. As a result, while the precise thresholds may differ across projects, the broader pattern classes and relative spatial tendencies identified in this study are expected to remain informative for similar infrastructure projects governed by comparable design regimes. In this sense, the framework provides a practical workflow through which reusable coordination knowledge can be extracted from completed projects and applied to guide interface analysis in future projects of the same type.

This study has several limitations. First, distances were computed based on 2D projections of object geometries, without considering vertical relationships. Although most infrastructure categories occupy relatively consistent vertical bands and vertically distributed elements were subdivided into separate categories, this projection may still simplify certain three-dimensional relationships. Second, object- or project-specific confounding factors, such as construction sections or work packages, were not incorporated. Third, the available dataset was limited in size, and consequently the typology of distance-based interface relationships remains incomplete. In particular, categories with extremely low or extremely high change rates could not be analyzed as prediction targets in the present dataset, although their interface behaviors may warrant investigation in larger datasets. Fourth, the analysis relied solely on spatial distance information derived from geometric changes in the models. Changes without spatial modifications (e.g., material substitutions or schedule adjustments) are therefore outside the scope of the present analysis and may not be represented in the inferred relationships. Addressing such non-geometric change dependencies would require additional data and modeling approaches beyond the scope of the current framework. Finally, the engineering interpretations assigned to the identified pattern types (e.g., connectivity or clearance relationships) should be regarded as plausible explanations of the observed statistical behaviors rather than formal validation against documented design coordination rules. Further validation through expert assessment and comparison with established design practices would help strengthen the practical interpretation of these patterns.

Future work will address these limitations by

expanding the dataset to further develop the typology and by incorporating additional parameters. In particular, beyond spatial distance, semantic information embedded in BIM will be integrated to extract interface relationships that cannot be explained by proximity alone.

Acknowledgements

This work was supported by the i-Construction System Studies Endowed Chair, Faculty of Engineering, The University of Tokyo.

References

[1] J. Freire and L. F. Alarcón. Achieving Lean Design Process: Improvement Methodology. *Journal of Construction Engineering and Management*, 128(3):248–256, 2002.

[2] L. Koskela. *Application of the new production philosophy to construction*, volume 72. Stanford university Stanford, 1992.

[3] C. W. Ibbs. Quantitative Impacts of Project Change: Size Issues. *Journal of Construction Engineering and Management*, 123(3):308–311, 1997.

[4] Y. A. Olawale and M. Sun. Cost and time control of construction projects: inhibiting factors and mitigating measures in practice. *Construction Management and Economics*, 28(5):509–526, 2010.

[5] P. J. Clarkson, C. Simons, and C. Eckert. Predicting Change Propagation in Complex Design. *Journal of Mechanical Design*, 126(5):788–797, 2004.

[6] T. Zimmermann, A. Zeller, P. Weissgerber, and S. Diehl. Mining version histories to guide software changes. *IEEE Transactions on Software Engineering*, 31(6):429–445, 2005.

[7] A. T. T. Ying, G. C. Murphy, R. Ng, and M. C. Chu-Carroll. Predicting source code changes by mining change history. *IEEE Transactions on Software Engineering*, 30(9):574–586, 2004.

[8] H. Gall, K. Hajek, and M. Jazayeri. Detection of logical coupling based on product release history. In *Proceedings. International Conference on Software Maintenance*, pages 190–198, 1998.

[9] A. K. Agrawal, Y. Zou, L. Chen, M. Abdelmegid, V. A. González, and H. Jin. Leveraging linked data for space constraints checking of mobile cranes in modular construction assembly lookahead planning. *Advanced Engineering Informatics*, 68:103778, 2025.

[10] X. Su and H. Cai. Life Cycle Approach to Construction Workspace Modeling and Planning. *Journal of Construction Engineering and Management*, 140(7):04014019, 2014.

[11] M. S. Dashti, M. RezaZadeh, M. Khanzadi, and H. Taghaddos. Integrated BIM-based simulation for automated time-space conflict management in construction projects. *Automation in Construction*, 132:103957, 2021.

[12] P. Pauwels, E. van den Bersselaar, and L. Verhelst. Validation of technical requirements for a BIM model using semantic web technologies. *Advanced Engineering Informatics*, 60:102426, 2024.

[13] R. Wang, D. A. S. Samarasinghe, L. Skelton, and J. O. B. Rotimi. A Study of Design Change Management for Infrastructure Development Projects in New Zealand. *Buildings*, 12(9):1486, 2022.

[14] L. Shi and D. Roman. From Standards and Regulations to Executable Rules: A Case Study in the Building Accessibility Domain. In *CEUR Workshop Proceedings*, 2017.

[15] S. Mehrbod, S. Staub-French, N. Mahyar, and M. Tory. Beyond the clash: investigating BIM-based building design coordination issue representation and resolution. *Journal of Information Technology in Construction*, 24, 2019.

[16] N. Fish, M. Averkiou, O. van Kaick, O. Sorkine-Hornung, D. Cohen-Or, and N. J. Mitra. Meta-representation of shape families. *ACM Trans. Graph.*, 33(4):34:1–11, 2014.

[17] E. Kalogerakis, S. Chaudhuri, D. Koller, and V. Koltun. A probabilistic model for component-based shape synthesis. *ACM Trans. Graph.*, 31(4):55:1–11, 2012.

[18] K. Genova, F. Cole, D. Vlasic, A. Sarna, W. T. Freeman, and T. Funkhouser. Learning Shape Templates With Structured Implicit Functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7154–7164, 2019.

[19] Z. Chen and H. Zhang. Learning Implicit Fields for Generative Shape Modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019.

[20] M. Kodnongbua, B. Jones, M. B. S. Ahmad, V. Kim, and A. Schulz. ReparamCAD: Zero-shot CAD Re-Parameterization for Interactive Manipulation. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–12, 2023.

[21] X. Zhuang, P. Zhu, A. Yang, and L. Caldas. Machine learning for generative architectural design: Advancements, opportunities, and challenges. *Automation in Construction*, 174:106129, 2025.

Optimizing High-Mix Low-Volume Prefabrication Scheduling: A Hybrid Simulation-Based Optimization Framework for Dual-Resource Constrained Systems

Nima Moghimi[1*], Sahar H. Shamaee[2], Qipei Mei[3], Vicente A. Gonzalez[4], Farook Hamzeh[5]

[1*]Corresponding Author, Ph.D. Student, Department of Civil and Environmental Engineering, University of Alberta
[2]Graduate Research Assistant, Department of Civil and Environmental Engineering, University of Alberta, Canada
[3]Assistant Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
[4] Hal Kvisle Professor and Tier 1 Canada Research Chair in Digital Lean Construction, Infrastructure Human Tech (IHT) Lab, Strategic Projects Insight Centre in Engineering (SPICE), Department of Civil and Environmental Engineering, University of Alberta, Canada
[5]Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
nmoghimi@ualberta.ca, hamedsha@ualberta.ca, qipei@ualberta.ca, vagonzal@ualberta.ca, hamzeh@ualberta.ca

Abstract

The shift toward Off-Site Construction (OSC) is constrained by the complexity of High-Mix, Low-Volume (HMLV) production, where geometric variability and reliance on manual labour create highly stochastic processing times. In these Dual-Resource Constrained (DRC) systems, workers act as flexible assets who move across stations to relieve bottlenecks—behaviour that static scheduling methods cannot capture. To address this, this study presents a Hybrid Simulation-Based Optimization (HSBO) framework that combines a Genetic Algorithm (GA) with a Hybrid Discrete Event/Agent-Based Simulation (DES/ABS) model. The simulation captures workforce-driven production dynamics, including cross-station assistance and bottleneck relief, enabling schedule evaluation under stochastic operating conditions. The framework was applied to a stair prefabrication line in Edmonton, Canada, to optimize the balance between Just-In-Time (JIT) delivery and internal flow efficiency. Results show that the optimized schedule reduced the external finished-goods buffer by 66% (from 3 days to 1 day), while increasing average internal lead time by approximately 20% (from 17 to 20.5 hours). These findings highlight the trade-off between delivery precision and internal flow efficiency in storage-constrained prefabrication systems and demonstrate the value of HSBO as a decision-support approach for dynamic scheduling in DRC environments.

Keywords – Off-Site Construction (OSC); Simulation-Based Optimization; Dual-Resource Constrained Scheduling; Hybrid Agent-Based Simulation; Just-In-Time (JIT) Production

1 Introduction

The construction industry is increasingly shifting toward off-site construction (OSC) and prefabrication to mitigate traditional inefficiencies [1]. Yet, unlike the high-volume, low-mix environments of automotive manufacturing [2], construction prefabrication operates under High-Mix, Low-Volume (HMLV) conditions where customized components—such as precast stairs or wall panels—produce substantial variability in processing times and resource allocation [3]. This variability intensifies logistical challenges, particularly on urban sites with limited storage, making precise Just-In-Time (JIT) delivery essential to avoid disrupting the critical path or causing site congestion [4,5]. Compounding this complexity is the Dual-Resource Constrained (DRC) nature of prefabrication, where productivity depends not only on machinery but also on a flexible, multi-skilled workforce capable of moving between stations to relieve bottlenecks [6–8]. While this flexibility can enhance efficiency, it also renders static scheduling tools (e.g., CPM, Gantt charts) ineffective, as they assume fixed resource allocations and cannot capture stochastic human performance or dynamic behaviours such as multi-station collaboration [9]. Consequently, static schedules quickly diverge from real shop-floor conditions, leading to "shop floor nervousness," widening plan–reality gaps, and reinforcing inefficiencies that deterministic models cannot resolve [6].

The optimization of production scheduling has been widely explored in manufacturing, with foundational work centered on the Job Shop Scheduling Problem (JSSP) and its Flexible variant (FJSP) [10]. Approaches in this domain range from exact methods such as Mixed-

Integer Linear Programming (MILP), which guarantee optimality but struggle to scale, to metaheuristics like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO), which efficiently navigate the combinatorial complexity of large production systems [11]. Although these algorithms excel at improving machine uptime and minimizing makespan in standardized, deterministic environments [12], they typically assume resources—both machines and workers—have fixed availability and constant performance [13,14]. Such static assumptions are incompatible with Dual-Resource Constrained (DRC) settings like construction prefabrication, where output depends equally on equipment and a flexible, multi-skilled workforce capable of shifting stations, assisting bottlenecks, and experiencing fatigue [15]. These dynamic behaviours—central in High-Mix, Low-Volume (HMLV) production—are largely absent from traditional JSSP/FJSP models [11,13], leaving a notable research gap in methods that merge the prescriptive strength of metaheuristic optimization with the behavioural realism needed to capture human-driven crew interactions in complex OSC environments.

To address this gap, following a Design-Science Research approach, this study introduces a Hybrid Simulation-Based Optimization (HSBO) scheduling artifact that couples an evolutionary search algorithm with a simulation-based behavioural model of workforce-driven production dynamics. At its core is a Hybrid Discrete Event/Agent-Based Simulation (DES/ABS), in which workers are represented as autonomous agents capable of collaboration and fatigue-driven performance changes [16,17]. This hybrid simulation serves as the fitness evaluator for a Genetic Algorithm (GA), enabling candidate schedules to be assessed based on emergent crew behaviours rather than nominal processing times. By iteratively exploring how alternative timetables influence system performance, the framework leverages worker flexibility to identify lookahead schedules that improve JIT adherence while controlling internal lead-time growth. The approach is validated using a stair prefabrication line at a modular construction facility in Edmonton, Canada, where a labour-intensive, HMLV process highlights the operational value of flexible crew coordination under real-world variability. Finally, the paper outlines a path toward integrating Reinforcement Learning to learn adaptive, policy-based scheduling—effectively a "digital scheduler" capable of generating plans across varying lookahead scenarios [18,19].

2 Methodology

This study adopts a Design-Science Research (DSR) methodology [20], which focuses on designing and rigorously evaluating prescriptive artifacts that address real operational problems. In this paper, the artifact is the proposed Hybrid Simulation-Based Optimization (HSBO) scheduling framework for Dual-Resource Constrained (DRC) prefabrication. This simulation-based optimization paradigm—where candidate decisions are iteratively evaluated through stochastic simulation and refined toward improved performance—has also been adopted in other uncertainty-driven engineering decision contexts [21,22]. Following the DSR cycle, the study proceeds through: (i) problem identification, formalizing the planning–execution gap in HMLV DRC environments; (ii) artifact design and development, coupling a Genetic Algorithm with a hybrid DES/ABS simulation evaluator to test candidate schedules under stochastic operating conditions; and (iii) demonstration and evaluation, assessing the artifact through the industrial case study and performance results reported in the subsequent section(s).

2.1 Artifact Design: HSBO Framework Overview

This research proposes an HSBO framework (Figure 1) designed to solve the dynamic scheduling problem in Dual-Resource Constrained (DRC) prefabrication. The framework utilizes a nested optimization architecture, where a Genetic Algorithm (GA) serves as the global optimizer for determining production timetables, and a Hybrid Discrete Event/Agent-Based Simulation (DES/ABS) serves as the fitness evaluation engine to test these timetables under stochastic operating conditions [23]. The simulation is implemented as a digital-twin-inspired virtual representation (digital shadow) of the physical production system that enables validation of schedules before execution, accounting for real-world variability and emergent workforce dynamics [24,25]. The artifact is demonstrated and evaluated in the subsequent case-study section by comparing its performance against baseline scheduling approaches under identical production scenarios. The artifact outputs a planned start-time timetable for projects in the lookahead window, which is then stress-tested in simulation before deployment.

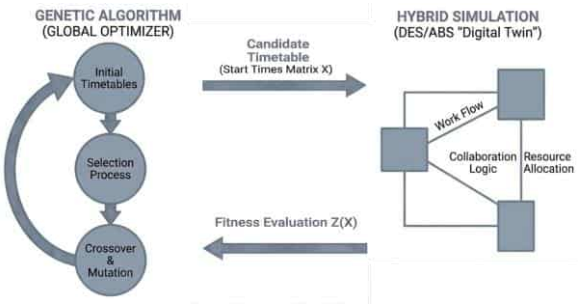


Figure 1: Schematic of the proposed HSBO framework, coupling a Genetic Algorithm with a

Hybrid DES/ABS model to generate and evaluate robust production timetables.

2.2 Problem Definition: Lookahead Scheduling

The scheduling problem is defined over a finite Lookahead Window—a bounded planning horizon representing the immediate, actionable production schedule for the factory. Specifically, let W denote the lookahead window duration (e.g., $W = 2$ weeks), and let P_{window} be the subset of projects (e.g., modular components, prefabricated stairs) released for production within this horizon. The production environment consists of a set of workstations K, where processing is constrained by both machinery availability and the flexible allocation of a multi-skilled workforce.

The objective is to assign a specific Planned Start Date ($ST_{p,k}$) for every project $p \in P_{window}$ at the initial workstation of its required assembly sequence. The scheduling problem is subject to three critical constraints. (1) Technological Precedence ensures that a downstream operation cannot commence until its upstream predecessor is complete, maintaining the assembly logic of complex prefabricated components. (2) Resource Availability dictates that a task cannot start if the required workstation or crew agents are occupied by other work. (3) Shift Constraints require that all processing must occur within valid factory shift hours and operational calendars.

Crucially, while the optimization algorithm generates Planned Start Times (ST_p), the Actual Start Times (S'_p) are stochastic, dependent on the emergent behavior and autonomous collaboration decisions of the workforce. The simulation bridges this planning-reality gap by predicting actual completion times under dynamic conditions.

2.3 Hybrid DES/ABS Simulation Model

To capture the gap between static planning and operational reality, the framework employs a Hybrid Discrete Event/Agent-Based Simulation approach. The simulation serves as a digital-twin-inspired virtual testbed, executing candidate schedules under realistic conditions to evaluate feasibility and performance.

2.3.1 Agent-Based Collaboration Logic

Unlike traditional scheduling models that assume fixed, immutable resource capacities, this framework models workers as autonomous agents capable of cross-training and dynamic collaboration [26]. Agents continuously monitor system state (e.g., queue lengths, throughput rates, bottleneck intensity) and autonomously traverse between compatible workstations to alleviate production constraints.

The simulation implements two primary behavioral rules governing crew agent decision-making:

Downstream Pull: An agent at a downstream station j may autonomously move upstream to station i to provide assistance if the feed rate drops below a critical threshold, preventing starvation (idle time) at downstream workstations.

$$\text{Count}_{\text{completed}}(j) > \text{Count}_{\text{completed}}(i) + \delta_{\text{pull}} \quad (1)$$

Upstream Push: An agent at an upstream station i may autonomously move downstream to station j to provide assistance if the intermediate buffer exceeds capacity, preventing congestion and excessive Work-In-Process (WIP) accumulation.

$$\text{QueueLength}(j) > \delta_{\text{push}} \quad (2)$$

These rules are continuously evaluated at each discrete event during simulation, enabling emergent, dynamic crew reallocation without explicit pre-planning.

2.3.2 Dynamic Processing Rates Under Collaboration

The simulation explicitly models the non-linear productivity impact of worker collaboration. When a workstation k operates with a primary crew and N_{helpers} assisting agents, the effective processing speed $v_{\text{eff}}(t)$ is formulated as:

$$v_{\text{eff}}(t) = v_{\text{base}} \times \left(1 + \eta \cdot N_{\text{helpers}}\right) \quad (3)$$

where v_{base} is the baseline processing rate of the primary crew working alone, η represents the Collaboration Efficiency Factor (e.g., $\eta = 0.5$), accounting for coordination overhead and diminishing returns, N_{helpers} is the number of agents simultaneously assisting at the station.

The parameter $\eta < 1$ reflects realistic crew dynamics: adding additional workers accelerates processing but with sublinear productivity gains due to coordination constraints and task indivisibility [27].

2.4 Optimization Model Formulation

2.4.1 Decision Variables

The decision variable X is a continuous vector of time values representing the Planned Start Time for each project in the current lookahead window:

$$\text{X} = \{ ST_p \mid p \in P_{window} \} \quad (4)$$

It is to be noted that the simulation logic determines the Actual Start Time (S'_p) as:

$$S'_p = \max\left(ST_p, \text{ResourceAvailability}_p\right) \quad (5)$$

The schedule is penalized if the planned time ST_p is infeasible (i.e., if $S'_p > ST_p$), reflecting poor schedule quality.

2.4.2 Objective Function: Multi-Criteria JIT-Aware Fitness

The fitness of a candidate schedule is evaluated by executing a complete simulation run, which returns the actual Completion Time and Start Time for each project under dynamic workforce conditions. The objective function minimizes a weighted sum of normalized Earliness, Lateness, and Maximum Lead Time, ensuring that objectives with different magnitudes do not dominate the optimization process:

$$Minimize Z(X) = \frac{\alpha}{T_d}\sum_{i=1}^{n} E_i + \frac{\beta}{T_d}\sum_{i=1}^{n} L_i + \gamma \frac{C_{max}}{N_{max}} \quad 6$$

where Earliness (E_i) penalizes early completion of project i to minimize site storage costs and outdoor exposure risks. Lateness (L_i) penalizes delays beyond the target due date to protect the critical path and maintain schedule adherence. Maximum Lead Time (C_{max}) minimizes the longest project cycle time, preventing excessive Work-In-Process (WIP) accumulation within the factory.

Normalization Factors scale raw time values into dimensionless, comparable ranges: T_d represents the maximum allowable JIT deviation (e.g., the lookahead window duration), while N_{max} represents the maximum theoretical project cycle time. This ensures that JIT and flow efficiency terms are commensurable and neither dominates due to scale differences.

Weighting Coefficients (α, β, γ) are user-defined values representing managerial priority. Typical values may include equal weight to earliness and lateness, with adjustments based on whether WIP or JIT precision is prioritized.

This multi-criteria formulation enables the GA to discover schedules that are both temporally precise (meeting JIT targets) and operationally efficient (minimizing internal WIP). The objective structure is flexible: additional metrics such as makespan, resource utilization, or crew overtime can be incorporated or substituted based on facility-specific constraints and priorities.

3 Framework Evaluation: A Case Study

3.1 Case Study Overview

This section demonstrates the framework's implementation in a prefabricated timber factory in Edmonton, Alberta. The Stair Assembly Station is used as the testbed, representing a High-Mix, Low-Volume (HMLV) system with high customization, variable processing times, and labor-intensive work Figure 2. The station operates as a Dual-Resource Constrained (DRC) flexible flow shop with four sequential phases: Stringer Construction (S), Landing Fabrication (L), Frame Building (F), and Final Cage Assembly (C). The workforce is organized into two semi-coupled teams—Upstream (S–L) and Downstream (F–C)—with the upstream team typically working one day ahead to sustain flow. When one team becomes ahead or idle, crews may move to assist the other ("partitioned flexibility"), requiring agent-based modeling to capture feasibility and throughput beyond rigid flow-shop assumptions.

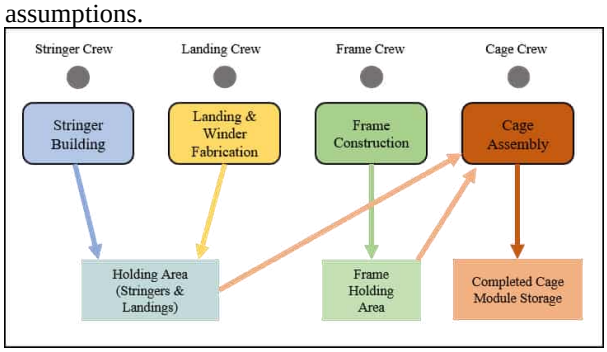


Figure 2: Schematic layout of the stair station.

3.2 Simulation Model Development

To capture the stochastic dynamics of the stair line, a Hybrid Discrete Event/Agent-Based Simulation (DES/ABS) model was developed in Python using the SimPy library. This hybrid approach enables the coupling of material flow logic (DES) with autonomous crew decision-making (ABS).

3.2.1 Data Collection and Parameterization

Historical production data were collected to characterize the variability of manual execution times. Probability distributions were fitted to the duration data for all four tasks (S, L, F, C). To ensure modeling accuracy and avoid double-counting of collaboration effects, only tasks completed independently by single crews were used for fitting; collaboration effects (which reduce duration) were excluded from the baseline distributions and instead generated dynamically by the simulation's agent logic.

3.2.2 System Constraints and Operational Logic

The simulation enforces system constraints observed during field studies. First, Resource Exclusivity ensures that each crew can process only one project at any given time. Second, Technological Precedence and Partial Dependencies are modeled to capture critical operational nuances where downstream phases may overlap with upstream work based on specific progress thresholds. Specifically, the Cage Assembly phase is permitted to commence once the upstream Framing phase reaches approximately 50% completion, and the installation of stringers and landings typically begins once the Cage assembly reaches approximately 70% completion. This partial overlap mechanism reflects the flexibility and adaptive nature of the modular construction workflow,

where work does not strictly follow finish-to-start dependencies but rather progresses dynamically based on material readiness and crew availability.

These partial-dependency rules were explicitly incorporated into the model to capture the realistic overlap between production phases. By allowing downstream tasks to begin before upstream completion, the model prevents the introduction of artificial blocking effects that would otherwise skew the simulation results and underestimate true system capacity.

3.2.3 Agent-Based Behavioural Logic

The simulation implements autonomous "collaboration" rules governing crew decisions. When the Downstream Team is idle or ahead of schedule, agents move upstream to assist the Upstream Team, and vice-versa. The model logic was validated through walkthroughs with production crews to ensure face validity, particularly regarding the fidelity of these autonomous collaboration behaviors.

3.3 Optimization Specification

The optimization problem focuses on generating an optimal production timetable for a representative Lookahead Window of two weeks. A dataset of 25 unique stair projects, sourced from the plant's operational backlog, served as the experimental sample.

3.3.1 Problem Formulation

The objective is to determine the optimal Planned Start Times (X) for both teams. The multi-objective function balances three conflicting goals:

- Lateness ($\alpha = 2$): Minimizing delays beyond target due dates to protect the critical path and maintain synchronization with site schedules.
- Earliness ($\beta = 1$): Minimizing early completions to prevent outdoor storage and weather-induced quality degradation.
- Maximum Lead Time ($\gamma = 0.5$): Minimizing internal production cycle time to enhance factory throughput efficiency.

The prioritization of JIT adherence (α, β) over lead time (γ) reflects the specific constraints of the Edmonton facility, as detailed below.

3.3.2 Managerial Rationale for Weight Assignment

This prioritization reflects specific facility constraints that distinguish this operation from standard manufacturing. The plant lacks indoor storage for finished goods; projects completed early must be stored outdoors, exposed to harsh Western Canadian weather conditions (snow, freeze-thaw cycles, temperature extremes, moisture). Such exposure poses significant quality risks to high-value prefabricated stairs, including material degradation, warping, and corrosion.

Conversely, Internal WIP (partially finished stairs in progress) remains sheltered indoors between workstations, where environmental conditions are controlled. Therefore, prioritizing a schedule that minimizes external exposure (balancing Earliness and Lateness) even at the cost of slightly longer internal lead times is logical. This reflects a realistic operational trade-off: in constrained facilities with outdoor storage vulnerabilities, *delivery precision* often outweighs *internal throughput* as the primary scheduling objective.

3.3.3 Algorithm Configuration and Search Enhancements

The Genetic Algorithm (GA) was configured with a population size of 50 individuals, an evolution horizon of 25 generations, and tournament selection with a tournament size of 3. To navigate the partitioned solution space effectively and accelerate convergence toward high-quality schedules, three domain-specific search enhancements were implemented.

Smart Initialization (Heuristic Seeding) seeded the initial population with heuristic schedules constructed using domain knowledge, where Buffer represents a conservative estimate of required lead time based on historical data. This approach accelerates convergence by providing high-quality starting points rather than relying on random initialization, allowing the optimizer to begin from a region of promising solutions.

Coupled Mutation (Topology-Aware Move) was implemented as a mutation operator that shifts linked Team 1 and Team 2 task start times simultaneously. This topology-aware move allows the optimizer to "slide" entire projects along the timeline while preserving the internal synchronization between the Stringer/Landing phase and the Framing/Cage phase. Formally, if a project's Team 1 start time is adjusted by Δ, the Team 2 start time is automatically adjusted by the same amount, maintaining the phase coupling relationships.

Deadline Snapping (Goal-Oriented Refinement) implemented a local search operator to force project start times toward their zero-penalty Just-In-Time dates. This operator refines temporal precision in the latter stages of evolution by exploiting the known good region of the solution space, thereby improving final schedule quality without excessive computational cost.

4 Results and Discussion

To evaluate the efficacy of the proposed HSBO framework, a comparative analysis was conducted

between the facility's baseline production schedule and the optimized timetable generated by the Genetic Algorithm. To ensure a rigorous comparison, the same set of stochastically sampled processing times across all four assembly stages was used for both the baseline simulation and the optimization routines. This ensures that any observed performance differences are attributable solely to scheduling decisions rather than random variation.

4.1 Baseline Schedule Performance

The baseline performance was established by simulating the 25-project sample schedule using the original planned start days and project sequence extracted from the factory's operational records. Figure 3 presents the resulting execution timeline for this baseline scenario, including the simulated project execution windows and their completion positions relative to the assigned crane dates.

Figure 3a illustrates the total project duration relative to the assigned "Crane Dates" (delivery deadlines). As shown, there is a consistent, significant buffer—averaging 2 to 3 days—between project completion and the deadline. While this buffer is strategically intended to absorb production variability, in the context of the Edmonton facility, it represents finished goods sitting outdoors in harsh weather for extended periods, incurring high quality risks.

Figure 3b provides a granular view of team-level execution. Visually, the gap between Team 1 (Stringer/Landing) and Team 2 (Framing/Cage) is minimal, indicating a tightly coupled flow with low Internal WIP. Quantitatively, the average project lead time in this baseline state is approximately 17 hours. However, comparing Figure 3a and 3b reveals a systemic imbalance: the facility maintains very low internal buffers (protected indoor WIP) but incurs massive external buffers (risky outdoor WIP).

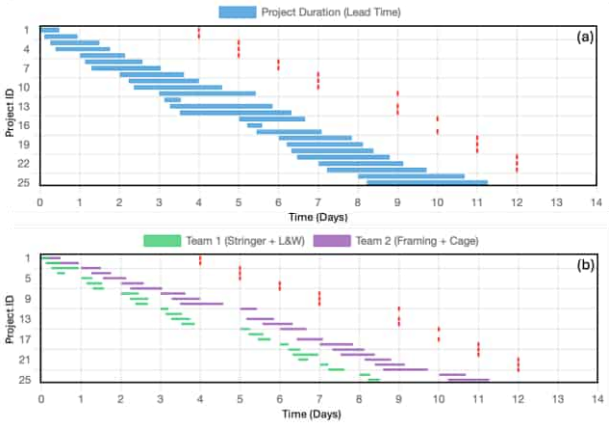


Figure 3: Baseline simulation results: (a) project execution windows relative to crane dates (red dashed lines); (b) Team 1 and Team 2 workload distribution.

4.2 Optimization Results

Following the baseline analysis, GA optimization was executed with objective weights prioritizing JIT adherence ($\alpha = 2, \beta = 1$) over internal lead-time minimization ($\gamma = 0.5$). The results are presented in Figure 4.

Compared with the baseline (Figure 3a), the optimized timetable (Figure 4a) shifts planned start times so completions align more closely with crane dates, reducing average outdoor finished-goods storage from 3 days to 1 day (66%). This does not eliminate the buffer entirely, but it substantially reduces exposure of finished stairs to harsh weather and related damage risk.

Figure 4b shows a clear structural change at the team level: increased spacing between Team 1 completion and Team 2 activity, indicating greater internal WIP (semi-finished components held indoors). Consistent with this, average project lead time increases by approximately 20% compared to the baseline (from ~17 hours to ~20.5 hours).

Overall, the optimizer trades a modest increase in safe indoor flow time for a large reduction in risky outdoor storage, with the Maximum Lead Time term ($\gamma = 0.5$) acting as a guardrail that penalizes excessive decoupling while still meeting delivery targets.

A side effect of JIT prioritization is schedule extension: the baseline executes from Day 0–11, whereas the optimized plan shifts to Day 1–13 and introduces a non-working gap at Day 12, increasing operational makespan. This occurs because delaying early projects reduces storage penalties, but the objective lacks an explicit makespan-compression term to discourage such gaps.

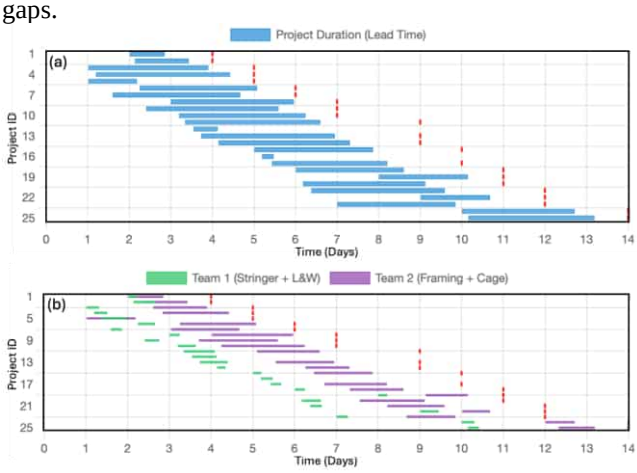


Figure 4: Optimized simulation results: (a) project execution windows relative to crane dates (red dashed lines); (b) Team 1 and Team 2 workload distribution.

From a managerial perspective, this trade-off should be interpreted as a deliberate buffer relocation rather than a simple loss in efficiency. In the case facility, reducing outdoor finished-goods storage from approximately 3 days to 1 day lowers the exposure of completed stairs to harsh weather, handling damage, and site coordination uncertainty, all of which carry quality, rework, and logistical risks. The corresponding increase in internal lead time indicates that more buffer is being held indoors as semi-finished Work-In-Process, where it is better protected and more controllable. For decision-makers, this means the proposed framework can support context-specific prioritization: in storage-constrained or weather-sensitive projects, accepting a moderate increase in internal lead time may be preferable if it substantially reduces external risk exposure. Conversely, in facilities where indoor space is more constrained or rapid flow is the dominant priority, the weighting structure of the objective function can be adjusted to place greater emphasis on lead-time compression or makespan reduction. Thus, the framework is not intended to prescribe a single best schedule universally, but to provide a decision-support tool that helps managers explicitly balance delivery precision, labor flow, and risk according to operational priorities.

4.3 Limitations and Future Directions

The results show the trade-off between delivery alignment and internal flow efficiency in HMLV systems. The 20% lead time increase indicates that reducing External WIP and improving Just-In-Time delivery required accepting some loss in factory flow performance. Two limitations remain. First, the objective function focuses on global maximum lead time, allowing local inefficiencies such as longer and more variable Team 2 flow times. Future models could include team-level penalties, though this would increase complexity. Second, the current Genetic Algorithm solves each schedule from scratch through repeated simulation, making it too slow for real-time use. Future work will explore DRL to learn reusable scheduling policies that can generate faster decisions while balancing delivery timing, lead time, and resource utilization.

5 Conclusion

This research addressed dynamic scheduling in High-Mix, Low-Volume (HMLV) off-site construction, where static planning tools struggle to represent the stochastic behavior of flexible, multi-skilled workforces. Following a Design-Science Research approach, the study developed and evaluated a prescriptive HSBO scheduling artifact that couples a Genetic Algorithm with a hybrid DES/ABS simulation model. The simulation served as a digital-twin-inspired evaluation testbed, enabling timetable optimization while capturing emergent collaboration behaviors in Dual-Resource Constrained (DRC) environments. Applied to a stair prefabrication line in Edmonton, Alberta, the optimized schedule reduced the average external finished-goods buffer from 3 days to 1 day (66%), lowering exposure risk, while increasing internal project lead times by about 20% (17 to 20.5 hours). The results also indicate that start-time decisions influence workforce dynamics by changing opportunities for upstream crews to assist downstream bottlenecks. Finally, the study highlights limitations of static evolutionary optimization: emphasizing JIT goals can increase schedule fragmentation and requires costly re-optimization as backlogs change. Future work will transition toward Deep Reinforcement Learning (DRL) to learn adaptive scheduling policies that provide real-time decision support under evolving system states.

6 Acknowledgement

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance grant ALLRP 581074–22, Discovery Grant RGPIN-2020-04420, and NSERC Alliance—Alberta Innovates Advance Program 232403286. All findings and conclusions expressed in this paper are those of the authors and do not reflect those of the contributors.

7 References

[1] Thibault D, Osman T, Brown A, Efficiency NE. Benefits and Opportunities of Off-Site Construction: Analysis of Indiana and Pennsylvania 2024.

[2] Barnes O. A Study of Lean Production Techniques in High-Volume Manufacturing Systems. American Journal Of Industrial And Production Engineering 2024;5:6–10.

[3] Amaresh A, Oooramvely K. Standardization in a High Mix Low Volume Company: A Case Study 2020.

[4] Hussein M, Zayed T. Critical factors for successful implementation of just-in-time concept in modular integrated construction: A systematic review and meta-analysis. J Clean Prod 2021;284:124716.

[5] Zhang C, Jiang J, Xia C, Fu Y, Liu J, Duan P. Dual-objective optimization of prefabricated component logistics based on JIT strategy. Sci Rep 2024;14:31267.

[6] Magalhães R, Santos F, Vieira S, Sousa JMC. Dual Resource Flexible Job Shop Scheduling Problems: The xBTF Algorithm. International Conference on Information Processing and

Management of Uncertainty in Knowledge-Based Systems, Springer; 2024, p. 115–25.
[7] Dhiflaoui M, Nouri HE, Driss OB. Dual-resource constraints in classical and flexible job shop problems: a state-of-the-art review. Procedia Comput Sci 2018;126:1507–15.
[8] Daniels RL, Mazzola JB, Shi D. Flow shop scheduling with partial resource flexibility. Manage Sci 2004;50:658–69.
[9] Krishnan M, Chinnusamy TR, Karthikeyan T. Performance study of flexible manufacturing system scheduling using dispatching rules in dynamic environment. Procedia Eng 2012;38:2793–8.
[10] Xiong H, Shi S, Ren D, Hu J. A survey of job shop scheduling problem: The types and models. Comput Oper Res 2022;142:105731.
[11] Zhang W, Bao X, Hao X, Gen M. Metaheuristics for multi-objective scheduling problems in industry 4.0 and 5.0: a state-of-the-arts survey. Frontiers in Industrial Engineering 2025;3:1540022.
[12] Shan D, Wang Y. A review of the application of genetic algorithms for solving flexible job shop scheduling problems. Proceedings of the 2nd International Conference on Information Technologies and Electrical Engineering, 2019, p. 1–5.
[13] Delgoshaei A, Ariffin MKAM, Maleki S, Leman Z. Review evolution of dual-resource-constrained scheduling problems in manufacturing systems: modeling and scheduling methods' trends. Soft Computing-A Fusion of Foundations, Methodologies & Applications 2023;27.
[14] Rahman M, Han SH. Performance Evaluation of Genetic Algorithm and Particle Swarm Optimization in Off-Site Construction Scheduling. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 41, IAARC Publications; 2024, p. 275–82.
[15] Pourrahimian E, Salhab D, Fayek AR, Hamzeh F. Fuzzy-Chaos Framework for Analyzing and Quantifying Uncertainty in Labour Productivity in Complex Construction Environments. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 42, IAARC Publications; 2025, p. 571–9.
[16] Sorbi T, Getuli V, Capone P, Rahimian FP. AGENT-BASED SIMULATION FRAMEWORK FOR ENHANCED CONSTRUCTION SITE RISK ESTIMATION AND SAFETY MANAGEMENT. Journal of Information Technology in Construction 2024;29.
[17] Eltahan A, Lee G, Hamzeh F. Exploring Cognitive Load and Task Complexity in Dynamic Tracking Tasks: Insights for Construction Workflows. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 42, IAARC Publications; 2025, p. 532–9.
[18] ElMenshawy M, Wu L, Gue B, AbouRizk S. Exploring the potential of reinforcement learning in pipe spool scheduling in industrial projects. Proc., 41st Int. Symp. on Automation and Robotics in Construction, 2024, p. 965–71.
[19] Moghimi N, Arshang O, Hamzeh F. Digital twins in offsite construction: Current implementations, challenges, and future pathways. Autom Constr 2026;183:106820.
[20] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Quarterly 2004:75–105.
[21] Moghimi N, Kashani H. A probabilistic framework to design the financing scheme for resilience enhancement of infrastructure systems. Reliab Eng Syst Saf 2025;262:111179. https://doi.org/10.1016/j.ress.2025.111179.
[22] Moghimi N, Kashani H. Probabilistic Modeling to Develop the Optimal Seismic Resilience Enhancement Strategy for Industrial Facilities. Journal of Infrastructure Systems 2026;32:04025041. https://doi.org/10.1061/JITSE4.ISENG-2754.
[23] Fu MC. Optimization for simulation: Theory vs. practice. INFORMS J Comput 2002;14:192–215.
[24] Kritzinger W, Karner M, Traar G, Henjes J, Sihn W. Digital Twin in manufacturing: A categorical literature review and classification. Ifac-PapersOnline 2018;51:1016–22.
[25] Alsakka F, Yu H, El-Chami I, Hamzeh F, Al-Hussein M. Digital twin for production estimation, scheduling and real-time monitoring in offsite construction. Comput Ind Eng 2024;191:110173. https://doi.org/10.1016/j.cie.2024.110173.
[26] Barkokebas B, Al-Hussein M, Hamzeh F. Assessment of digital twins to reassign multiskilled workers in offsite construction based on lean thinking. J Constr Eng Manag 2023;149:04022143.
[27] Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 2002;99:7280–7.

Scene Coordinate Regression–Based Visual Localization for Advanced Multi-Crane Safety Monitoring in Construction Sites

Tran Dang Khoa Vo[1], Tan-Dat Pham[1], Sujin Jin[1], Pa Pa Win Aung [1], Solmoi Park[1] and Seunghee Park[1,2]

[1] Department of Global Smart City, Sungkyunkwan University, South Korea
[2]School of Civil, Architectural Engineering and Landscape Architecture, Sungkyunkwan University, South Korea

19980407@skku.edu, phamtandat@skku.edu, kane0113@skku.edu,
ppwinaung06@skku.edu, solmoipark@skku.edu, shparkpc@skku.edu

Abstract -

Crane operations pose significant safety risks on construction sites, particularly when multiple mobile cranes operate concurrently in confined and dynamic environments. This paper presents a vision-based framework for real-time multi-crane safety monitoring using Scene Coordinate Regression (SCR). By directly regressing 3D scene coordinates from monocular images, the proposed approach enables accurate six-degree-of-freedom crane tip localization without requiring repeated camera calibration during deployment. Visual data acquired from crane-mounted cameras and aerial drone surveys are jointly leveraged to construct scene-specific spatial representations for SCR model training. The estimated crane-end poses are integrated into a BIM-aligned digital twin to support real-time visualization, operational envelope analysis, and automated near-collision detection. Experimental results demonstrate consistent centimeter-level localization accuracy, with median errors of approximately 1–1.3 cm, confirming the suitability of the proposed framework for interaction-aware multi-crane safety monitoring in construction environments.

Keywords -

Visual Localization; Crane Safety Monitoring; Computer Vision; Digital Twins

1 Introduction

Construction remains one of the most hazardous industries worldwide. In 2023, the construction sector recorded 1,075 fatal occupational injuries, the highest among all industries, according to the U.S. Bureau of Labor Statistics (BLS) [1]. Construction-related accidents also impose significant economic burdens, with workplace injuries costing nearly $1 billion per week as estimated by the Occupational Safety and Health Administration (OSHA) [2]. These statistics highlight the urgent need for proactive and technology-driven safety monitoring on construction sites.

On modern construction sites, multiple mobile cranes often operate simultaneously in confined and dynamic environments, where collision risks arise from interactions among cranes, surrounding structures, and workers. Traditional safety approaches, such as manual supervision and rule-based controls, are increasingly insufficient under these conditions, motivating the development of real-time monitoring and proactive collision avoidance systems [3].

Sensor-based approaches have been widely adopted to enhance crane operation management by integrating planning data, operational information, and monitoring technologies [4, 5]. These include laser scanning for visibility analysis [6] and multi-sensor systems for localization and collision prevention [7], often extended through IoT frameworks for coordinated safety management [8]. However, such approaches rely on dedicated hardware, calibration, and sensor fusion, resulting in high deployment complexity and cost.

In parallel, vision-based approaches have emerged as a lightweight alternative. Crane-mounted cameras enable continuous visual perception for tasks such as worker detection and hazard awareness [9], while image-based methods have been extended to 3D spatial understanding using multi-view reconstruction or depth estimation [10]. Hybrid systems combining cameras with additional sensors have also been proposed for crane localization and workspace analysis [11, 12]. Nevertheless, several limitations remain. First, many vision-based localization methods rely on explicit 2D–3D feature matching, making them sensitive to occlusions, illumination changes, and visual noise. Second, both vision- and sensor-based systems often require explicit geometric modeling and repeated calibration, limiting adaptability to dynamic construction environments. Finally, existing approaches provide limited support for unified spatial representation and interaction-aware safety reasoning in multi-crane scenarios.[12].

To address these challenges, this study proposes a vision-based crane localization framework based on Scene Coordinate Regression (SCR). By directly regressing 3D scene coordinates from monocular images, the proposed approach avoids explicit feature matching and enables efficient 6-DoF camera pose estimation from a single image.

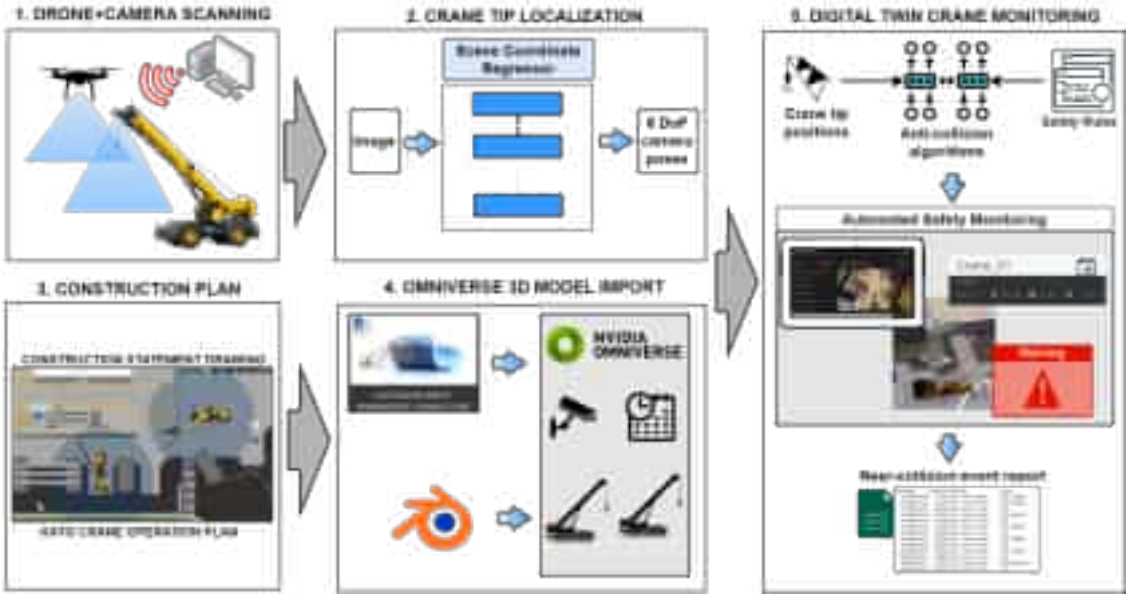


Figure 1. Overall framework of digital-twin crane collision monitoring.

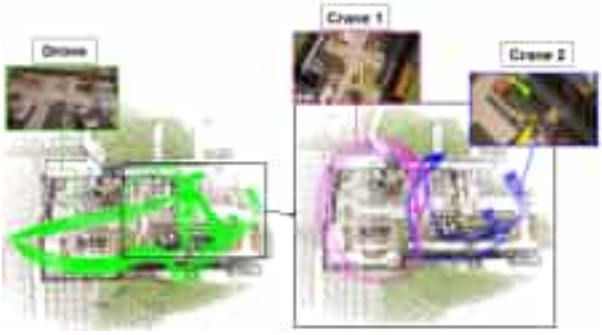


Figure 2. Drone and mobile crane operation path.

Visual data from crane-mounted cameras and drone surveys are jointly leveraged to construct scene-specific spatial representations without requiring additional sensing hardware. The estimated crane-end poses are integrated into a BIM-aligned digital twin to support real-time visualization and interaction-aware safety monitoring in multi-crane environments.

The main contributions of this study are summarized as follows:

- A real-time multi-crane safety monitoring framework that integrates vision-based localization with a BIM-aligned digital twin, enabling interaction-aware analysis of crane operations in shared workspaces.
- A unified spatial representation for multi-crane environments, where crane tip poses estimated from monocular images are directly mapped into a common BIM coordinate system for consistent safety reasoning.
- A lightweight deployment strategy using crane-mounted cameras and drone imagery, eliminating the need for additional sensing infrastructure while supporting continuous monitoring in dynamic construction sites.
- A digital twin–driven safety monitoring pipeline that enables real-time visualization, event logging, and near-collision detection for multiple cranes operating simultaneously.

2 Background

2.1 Advanced Crane Monitoring

The increasing complexity of construction sites has driven extensive research on advanced crane monitoring systems to improve operational safety and situational awareness [4, 5]. Early approaches relied on geometric and rule-based models to define crane working envelopes and collision constraints [3, 6], but their dependence on accurate modeling and frequent manual updates limits adaptability to dynamic site conditions. To improve localization accuracy, sensor-based systems incorporating GNSS, IMUs, UWB, and laser scanners have been widely adopted [7, 11], with multi-sensor fusion and IoT-based frameworks enabling real-time crane monitoring and fleet-level safety management [8]. However, these approaches rely on dedicated sensing infrastructure and precise calibration, resulting in high deployment cost and limited scalability. As a result, their applicability to rapidly changing construction environments and hetero-

geneous crane fleets remains limited. Vision-based crane monitoring has therefore emerged as a lightweight alternative, leveraging camera data and deep learning to support crane operation awareness [9]. While recent studies have extended vision-based methods toward three-dimensional spatial reasoning through image-based reconstruction or depth-assisted perception [10], many existing approaches remain constrained by explicit geometric assumptions or hybrid sensor dependencies [12, 11]. Moreover, most prior work focuses on single-crane scenarios, with limited consideration of interaction-induced safety risks in multi-crane environments [4].

2.2 Scene Coordinate Regression

Scene Coordinate Regression (SCR) is a deep-learning-based camera localization paradigm that directly regresses 3D scene coordinates from image pixels, enabling camera pose estimation without explicit 2D-3D feature matching [13]. SCR methods learn a mapping from image appearance to scene-specific spatial representations, from which six-degree-of-freedom (6-DOF) camera poses can be efficiently inferred [14]. By implicitly learning scene geometry, SCR demonstrates improved robustness to texture ambiguity, repetitive structures, and illumination variations compared with feature-based localization pipelines [15]. In addition, SCR-based localization can be operated using a single RGB image, enabling efficient fast inference for time-critical applications [16]. Although these characteristics make SCR well suited to dynamic and visually challenging environments such as construction sites, the application of SCR to construction equipment localization remains largely unexplored. In crane operation monitoring, SCR offers clear advantages. Crane-mounted and aerial cameras naturally observe consistent site geometry from diverse viewpoints, providing suitable visual input for learning scene coordinate representations. Unlike conventional vision-based crane localization methods, SCR does not require explicit camera-to-crane calibration or handcrafted geometric constraints, enabling flexible deployment across heterogeneous cranes and evolving site layouts. Furthermore, SCR-based localization can provide a unified spatial reference for interaction-aware safety analysis in multi-crane environments.

3 Proposed Approach

This section presents an end-to-end SCR-Digital Twins framework for real-time multi-crane safety monitoring, covering visual data acquisition, scene coordinate learning, crane-end localization, and digital twin–based collision analysis. Starting with multi-camera scanning, data from drone and crane-mounted cameras are then used to train the model for fully scene understanding. The trained model, which is adopted from [17] then is able to output the camera positions, which are also crane-end positions. Meanwhile, the 3D BIM model is built in Omniverse and the crane operation areas are defined based on construction plan. Finally, the anti-collision algorithms are used to analyze mobile crane-end positions with safety rules, then perform automated anti-collision monitoring and critical hazard event reports. The whole process is illustrated as in Figure. 1.

3.1 Data Processing

In order enable robust multi-crane monitoring with minimal site instrumentation, we construct a scene-specific sparse 3D representation by jointly leveraging crane-end mounted cameras that provide continuous close-range observations of the working area and a drone survey that ensures global coverage and mitigates blind spots as shown in Figure. 2. The drone flight can be conducted following site visit schedule to capture wide-baseline images of the entire operation zone and update on-site changes, while the crane-end cameras continuously record image streams during lifting operations. In the next step, a Structure-from-Motion (SfM) pipeline is then executed to recover a sparse point cloud and camera poses. Specifically, keypoints and descriptors are extracted for each image, feature correspondences are established across views, and a bundle-adjusted reconstruction yields a set of 3D points with associated multi-view feature tracks. This SfM model provides reliable 2D-3D associations that are later used as supervision for scene coordinate learning.

3.2 SCR-based Visual Localization

This subsection describes the D2S-based scene coordinate regression architecture adopted for real-time crane-end visual localization. The pipeline is divided into 2 parts: (1) Offline training where Scene Coordinate Regression model is trained with input images and their corresponding point cloud and (2) Online inference where the live-streaming data is feed directly into trained model to predict camera positions, as shown in Figure. 3.

3.2.1 Offline Training

Given the SfM reconstruction aligned with the BIM coordinate frame, each training image $\mathbf{p}_i \in \mathbf{P}$ provides a set of verified descriptor–scene coordinate pairs $\mathbf{V} = \{(d_i, \mathbf{X}_i)\}$ obtained from multi-view feature tracks. Following the D2S formulation [17], a scene coordinate regressor is trained to learn a direct mapping from local image descriptors to 3D scene coordinates. The architecture consists of a scene-agnostic local feature extractor and a scene-specific regression module equipped with self-attention to capture global descriptor context and suppress

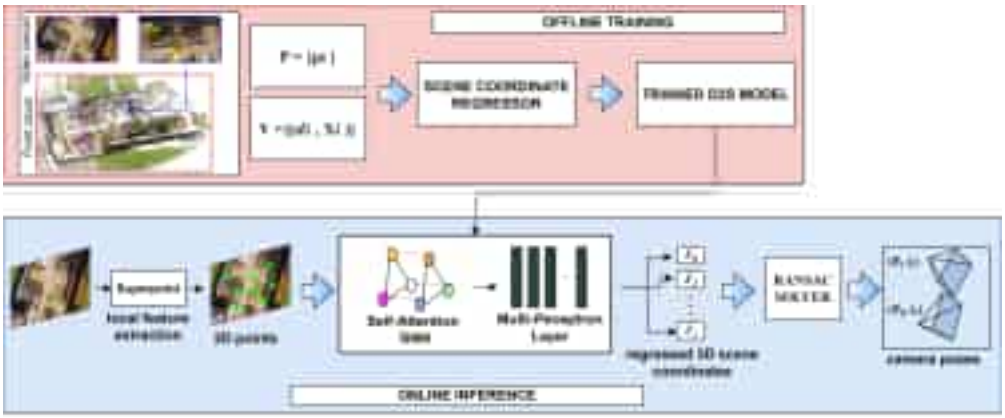


Figure 3. Overall SCR-based visual localization pipeline.

unstable regions such as moving objects or textureless areas. A shared MLP head outputs the regressed scene coordinates $\hat{\mathbf{X}}_i$ together with an optional reliability score $\hat{p}_i$. Once trained, the model is able to predict the new 3d feature coordinates from a single image without any SfM-guided descriptors.

3.2.2 Online Inference

During operation, each incoming crane-end camera frame is processed independently in a purely feed-forward manner. Sparse local features are first extracted from the image and passed through the trained D2S model, which directly regresses a set of 3D scene coordinates $\{\hat{\mathbf{X}}_i\}$ from the corresponding descriptors without explicit feature matching against a global 3D map. This regression-based formulation avoids costly nearest-neighbor search and scale-dependent matching, making the inference computationally efficient.

To ensure robustness, only the most reliable correspondences are retained based on the predicted confidence scores $\hat{p}_i$ (or by selecting the top-N correspondences). These sparse 2D–3D pairs are then fed into a RANSAC-based PnP solver to estimate the 6-DoF camera pose $(\mathbf{R}, \mathbf{t})$, followed by a lightweight nonlinear refinement. Since the camera is rigidly mounted at the crane tip, the estimated camera pose is directly transformed into the crane-end pose in the BIM coordinate frame using a pre-calibrated extrinsic transformation. This transformation is defined once and remains fixed during operation, eliminating the need for repeated calibration procedures in dynamic construction environments.

By avoiding dense reconstruction and global map matching, the proposed inference pipeline maintains low computational overhead and stable runtime behavior suitable for real-time deployment. The use of sparse descriptors, lightweight regression networks, and a single PnP optimization per frame enables real-time performance, allowing continuous crane-end pose tracking without reliance on dense reconstruction or iterative map matching. This real-time localization capability forms the basis for online multi-crane anti-collision monitoring and hazard analysis within the digital twin environment.

3.3 Digital Twin Visualization and Real-Time Monitoring

Based on the BIM-aligned scene coordinate system, a USD-based digital twin is constructed in NVIDIA Omniverse [18] to enable real-time visualization and monitoring of multi-crane operations. Each mobile crane is represented as a parametric articulated model with a fixed base and a dynamically updated arm. At each time step t, the estimated crane-end position $\mathbf{p}_i^t$ from the visual localization module is integrated by updating the corresponding USD transforms, rendering the crane arm as a line segment connecting the base $\mathbf{b}_i$ and the tip $\mathbf{p}_i^t$.

Following Algorithm 1, frame-wise crane operational metrics—including height, horizontal working radius, and azimuth angle—are computed directly in the BIM coordinate frame and compared against historical maxima to provide intuitive feedback on operating envelopes. All updates are performed in a frame-synchronous manner to ensure temporal consistency between geometry updates, metric computation, and visualization.

To assess collision risk between multiple cranes, the system evaluates the horizontal separation distance between crane tips at each frame,

$$d_t = \left\| (\mathbf{p}_1^t - \mathbf{p}_2^t)_{xz} \right\|,$$

where only the ground-plane components are considered to reflect practical collision risk during lifting operations. When the distance falls below a predefined safety threshold D_{th}, the system transitions from a *safe* to a *near-collision* state. These state transitions are logged with timestamps and frame indices, enabling post-event analysis and safety

auditing. In this study, the safety threshold D_{th} is set to 10 m as a conservative safety margin to account for crane motion uncertainties, load swing, and localization errors in dynamic construction environments, following general safety principles outlined in ISO 12480-1 [19].



Figure 4. NVIDIA Omniverse window for DT-based crane safety monitoring

In addition to geometric visualization, synchronized live camera streams from crane-end cameras are displayed alongside the digital twin, providing operators with complementary visual context. All visualization elements—including crane geometry, distance indicators, operational statistics, and safety status—are updated in real time within a unified dashboard interface, as illustrated in Figure. 4. The use of sparse crane-end pose updates, lightweight geometric primitives, and event-driven state evaluation allows the system to operate at interactive frame rates without requiring dense reconstruction or physics-based simulation. This tightly coupled integration of real-time visual localization, digital twin visualization, and rule-based safety assessment enables continuous multi-crane monitoring and forms the foundation for automated near-collision detection in complex construction environments.

Algorithm 1 Real-Time Multi-Crane Digital Twin Monitoring and Near-Collision Detection

Require: Frame-wise crane tip positions $\mathbf{p}_i^t$, crane base transforms $\mathbf{T}_i$, synchronized camera images, safety threshold D_{th}
Ensure: Real-time crane visualization, operational metrics, near-collision events
Initialize USD-based digital twin environment and crane geometries
Initialize event logger and playback parameters
for each frame t **do**
  for each crane i **do**
    Retrieve crane base position $\mathbf{b}_i$
    Update crane arm geometry from $\mathbf{b}_i$ to $\mathbf{p}_i^t$
    Compute height, horizontal range, and azimuth angle
    Update historical maximum height and range
  end for
  Compute horizontal distance

$$d_t = \|(\mathbf{p}_1^t - \mathbf{p}_2^t)_{xz}\|$$

  if $d_t \le D_{\text{th}}$ **then**
    Set system state to *near-collision*
  else
    Set system state to *safe*
  end if
  Log state transition events with timestamp if detected
  Update dashboard indicators and synchronized camera views
end for



Figure 5. Plan view of experimental construction site

4 Implementation and Results

This section illustrates the experimental setup to evaluate model performance and results of proposed method. First, a simulated construction site is built by Twin-motion platform replicating the real construction site at Central Florida Boulevard, USA, as shown in Figure. 5. In the next step, drone and mobile crane working activities are simulated to capture data for model training with image resolution of 1080x1920. The dataset consists of approximately 2000 images collected from drone and crane-mounted cameras. The data are split into training and testing sets at the sequence level to avoid frame overlap. Meanwhile, after real-time video stream is fed into train D2S model to predict mobile crane tips, the output are inputs in NVIDIA Omniverse to perform digital-twin visualization. The whole pipeline is implemented on a computer with NVIDIA RTX 4090 GPU, Intel i9-139000KF Chipset and 64GB RAM.

4.1 Visual localization results

In order to evaluate the accuracy of SCR-based D2S model in estimating spatial information of crane tips, we we establish a testbed from videos recording two KATO cranes operating and the camera poses extracted from COLMAP serving as ground-truth value to estimate the errors of model predictions. The SfM-derived camera poses are used only for evaluation and are not directly used to supervise camera pose learning.

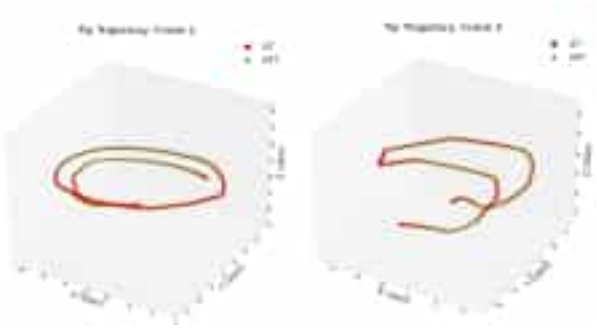


Figure 6. Qualitative results of SCR-based crane tip localization of Crane 1 (Left) and Crane 2 (Right). Axes are shown in units of 10 m for visualization clarity.

Table 1. Quantitative results of SCR-based crane tip localization.

Error threshold	Crane 1	Crane 2
$< 1\,cm$	48.89%	48.80%
$< 2\,cm$	81.89%	82.10%
$< 3\,cm$	93.11%	92.69%
$< 5\,cm$	96.89%	97.60%
Median error (cm)	**1.30**	**1.01**

The Figure. 6 compares the estimated crane tip trajectories with the reference trajectories for both cranes. The results show that the estimated positions closely follow the ground truth over the entire operation sequence, maintaining stable alignment during slewing motions, height changes, and direction transitions. In addition, the quantitative results are summarized in Table.1. For Crane 1, 81.89% of the estimated crane tip positions exhibit errors below 2 cm, increasing to 96.89% within a 5 cm threshold. Crane 2 achieves comparable performance, with 82.10% of estimates below 2 cm and 97.60% within 5 cm. The median localization errors are 1.30 cm and 1.01 cm for Crane 1 and Crane 2, respectively. These results demonstrate that the proposed SCR-based approach can provide reliable centimeter-level crane tip localization, which is sufficient to support downstream safety monitoring and near-collision detection tasks.

4.2 Digital Twin Crane Safety Monitoring

The Figure. 7 presents the digital twin–based real-time multi-crane safety monitoring results. By continuously integrating the estimated crane-end poses into a BIM-aligned digital twin, the system enables synchronized visualization of crane geometry, operational metrics, and live top-views for multiple cranes operating concurrently. During normal operation, a safe state is maintained when the inter-crane distance exceeds the predefined safety threshold, while real-time feedback on crane height, working radius, and azimuth angle is provided. After that, when the horizontal separation between crane tips falls below the safety threshold, the system automatically transitions to a near-collision state, which is immediately reflected in the digital twin visualization and monitoring dashboard. All state transitions are logged in a structured CSV format with timestamps and inter-crane distances, enabling post-operation analysis and safety auditing. The system achieves approximately 21.58 FPS on an RTX 4090 GPU, demonstrating near real-time performance for multi-crane monitoring .These results demonstrate that the proposed framework can effectively integrate vision-based localization with digital twin–based safety monitoring for real-time multi-crane operations.

Table 2. Runtime analysis of the proposed framework on i9-13900KF, RTX4090

Module	Time (ms)
Feature extraction + SCR inference	28.92
PnP + refinement	15.28
Digital Twin update	2.14
Total per frame	**46.34 ms**
Average FPS	**21.58 FPS**

5 Discussion

The proposed framework assumes a fixed rigid transformation between the camera and crane tip, as well as prior alignment between the SfM reconstruction and the BIM coordinate system. These assumptions may introduce deployment constraints in real construction sites. The results demonstrate that the SCR-based framework can support near real-time multi-crane safety monitoring using monocular visual inputs, achieving centimeter-level localization accuracy sufficient for inter-crane distance evaluation and safety threshold enforcement. Compared with feature-based localization methods, the proposed approach avoids explicit feature matching and shows improved robustness to illumination changes and occlusions. In contrast to sensor-based systems, it reduces hardware dependency and



Figure 7. Digital twin–based real-time visualization of multi-crane safety monitoring.

calibration complexity. From a methodological perspective, scene coordinate regression enables a unified scene-centric representation without explicit geometric modeling or repeated calibration, making it suitable for dynamic construction environments. The integration with a BIM-aligned digital twin further supports consistent spatial reasoning, real-time visualization, and rule-based safety assessment for interaction-aware monitoring. Nevertheless, the current implementation focuses on crane-tip separation and does not explicitly model dynamic effects such as load swing or interactions with workers. In addition, the evaluation is conducted in a simulated environment, and real-world conditions such as illumination variation, occlusions, and weather may affect performance. These limitations will be addressed in future work.

6 Conclusion

This paper presented a vision-based framework for real-time multi-crane safety monitoring using Scene Coordinate Regression (SCR) integrated with a digital twin. By directly regressing 3D scene coordinates from monocular images, the proposed approach enables accurate crane tip localization without relying on explicit feature matching or additional sensing hardware, thereby reducing system complexity, calibration requirements, and deployment overhead. Experimental results demonstrate consistent centimeter-level localization accuracy across different crane motions and configurations, indicating robust performance under dynamic operational conditions and varying viewpoints. When integrated into a BIM-aligned digital twin, the estimated crane-end poses support real-time visualization, operational envelope analysis, and automated near-collision detection among multiple cranes operating within shared workspaces, enabling interaction-aware safety assessment in complex site layouts. Furthermore, the unified spatial representation provided by the digital twin facilitates consistent safety rule enforcement and traceable event logging without manual reconfiguration. Overall, the proposed framework offers a lightweight and scalable alternative to conventional sensor-heavy crane monitoring systems and provides a practical foundation for proactive, data-driven, and interaction-aware multi-crane safety management in complex and evolving construction environments.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used GPT-5.2 in order to improve the readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) [RS-2024-00336270] and [RS-2025-02223612].

References

[1] U.S. Bureau of Labor Statistics. National census of fatal occupational injuries in 2022. BLS Census of Fatal Occupational Injuries, 2022. Accessed Feb. 19, 2025.

[2] U.S. Occupational Safety and Health Administration (OSHA). Business Case for Safety and Health. https://www.osha.gov/businesscase/costs, 2024. (accessed Jan. 2, 2024).

[3] Ziqing Yang, Jian Yang, and Enliu Yuan. Safety monitoring of construction equipment based on multi-sensor technology. In *ISARC*, pages 677–684, October 2020. doi:10.22260/ISARC2020/0095.

[4] Ali Khodabandelu and JeeWoong Park. *Integrating BIM and ABS for Multi-Crane Operation Planning through Enabling Safe Concurrent Operations*, pages 1128–1135. ASCE, 2021. doi:10.1061/9780784483893.138.

[5] Ala Nekouvaght Tak, Hosein Taghaddos, Ali Mousaei, Anahita Bolourani, and Ulrich Hermann. Bim-based 4d mobile crane simulation and onsite operation management. *Automation in Construction*, 2021. doi:https://doi.org/10.1016/j.autcon.2021.103766.

[6] Tao Cheng and Jochen Teizer. Crane operator visibility of ground operations. In *ISARC*, pages 699–705, June 2011. doi:10.22260/ISARC2011/0131.

[7] Jean-Pierre Sleiman, Emile Zankoul, Hiam Khoury, and Farook Hamzeh. *Sensor-Based Planning Tool for Tower Crane Anti-Collision Monitoring on Construction Sites*, pages 2624–2632. ASCE, 2016. doi:10.1061/9780784479827.261.

[8] Dexing Zhong, Hongqiang Lv, Jiuqiang Han, and Quanrui Wei. A practical application combining wireless sensor networks and internet of things: Safety management system for tower crane groups. *Sensors*, 2014. doi:10.3390/s140813794.

[9] Tanittha Sutjaritvorakul, Axel Vierling, and Karsten Berns. Data-driven worker detection from load-view crane camera. In *ISARC*, pages 864–871, October 2020. doi:10.22260/ISARC2020/0119.

[10] Yuntae Jeon, Dai Quoc Tran, Khoa Tran Dang Vo, Jaehyun Jeon, Minsoo Park, and Seunghee Park. Neural radiance fields for construction site scene representation and progress evaluation with bim. *Automation in Construction*, 2025. doi:https://doi.org/10.1016/j.autcon.2025.106013.

[11] Jingdao Chen, Yihai Fang, and Yong K. Cho. Real-time 3d crane workspace update using a hybrid visualization approach. *Journal of Computing in Civil Engineering*, 2017. doi:10.1061/(ASCE)CP.1943-5487.0000698.

[12] Yu Wang, Hiromasa Suzuki, and Yutaka Ohtake. Spatial maps with working area limit line from images of crane's top-view camera. *Automation in Construction*, page 104475, 2022. doi:https://doi.org/10.1016/j.autcon.2022.104475.

[13] Xiaotian Li, Jakob Verbeek, and Juho Kannala. Hierarchical joint scene coordinate classification and regression for visual localization. *CoRR*, abs/1909.06216, 2019. URL http://arxiv.org/abs/1909.06216.

[14] Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. DSAC - differentiable RANSAC for camera localization. *CoRR*, abs/1611.05705, 2016. URL http://arxiv.org/abs/1611.05705.

[15] Eric Brachmann and Carsten Rother. Learning less is more - 6d camera localization via 3d surface regression. *CoRR*, abs/1711.10228, 2017. URL http://arxiv.org/abs/1711.10228.

[16] Thuan B Bui, Dinh-Tuan Tran, and Joo-Ho Lee. Fast and lightweight scene regressor for camera relocalization. *arXiv preprint arXiv:2212.01830*, 2022.

[17] Bach-Thuan Bui, Huy-Hoang Bui, Dinh-Tuan Tran, and Joo-Ho Lee. D2s: Representing sparse descriptors and 3d coordinates for camera relocalization. *IEEE Robotics and Automation Letters*, 2024.

[18] *USD Composer*. NVIDIA Corporation, Santa Clara, CA, U.S., 2024. URL https://docs.omniverse.nvidia.com/composer/latest/index.html.

[19] Iso-12480-1:2024, Sep 2023. URL https://www.iso.org/standard/57300.html.

MiCRoadDiff: AI-driven generative design of on-site haul road for modular constructions using fine-tuned diffusion models

Fangzheng Li[1], Rongyan Li[2], and Hung-Lin Chi[3]

[1]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong
[2]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong
[3]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong
fangzheng.li@connect.polyu.hk, rongyan.li@connect.polyu.hk, hung-lin.chi@polyu.edu.hk

Abstract

Modular construction projects rely heavily on large trucks to transport prefabricated modules, making the design of on-site haul roads critical for ensuring safe and efficient logistics operations. However, the manual design process suffers from inefficiencies. To address this issue, this study proposes MiCRoadDiff, a diffusion model-based intelligent approach for modular construction on-site haul road layout generation. MiCRoadDiff, adopts a pre-trained Stable Diffusion model as an intelligent design backbone. LoRA is employed to achieve domain adaptation by injecting construction-specific knowledge, design experience, and implicit road design rules into the generative model. In addition, multiple ControlNet modules are introduced to enforce geometric and road layout constraints. A comprehensive evaluation framework that incorporates image fidelity and layout validity is proposed. Experimental results demonstrate that MiCRoadDiff exhibits strong domain adaptation capability and can rapidly generate diverse, high-quality haul road layout alternatives that satisfy practical engineering requirements. By integrating multi-source conditional information, it can robustly adapt to varying site conditions, generating coherent and feasible haul road layouts. The proposed framework significantly improves the efficiency and intelligence of construction site layout design workflows.

Keywords –

Diffusion models; Generative design; On-site haul road design; Finetuning; Modular constructions.

1 Introduction

Modular Integrated Construction (MiC) [1] is an innovative construction approach where modules are prefabricated in factories and then assembled on-site. In recent years, the Government of the Hong Kong Special Administrative Region (HKSAR) has actively promoted MiC [2] to tackle key challenges in the local construction industry, including labor shortages, site limitations, and the growing demand for environmental sustainability. The modules are transported to the construction site by trucks, entering through the main gate, being placed at designated unloading points, and then exiting the site. On-site haul roads play a critical role in ensuring the safety and efficiency of truck transportation [3]. Proper design of haul roads can significantly enhance logistic efficiency, safety, and ease of operation for truck drivers, thereby minimizing potential damage to the modules. However, traditional manual design processes face several limitations, such as low efficiency and the necessity for repetitive adjustments, which hinder overall productivity and effectiveness.

Researchers have explored classical optimization methods, including greedy algorithms [4], genetic algorithms [5], and Monte Carlo techniques [6], in various aspects of road planning. While genetic algorithms demonstrated potential in exploring design spaces, they were limited to producing near-optimal solutions and fell short of capturing the complex nature of the layout design process. Mixed-integer linear programming [3] was also employed to achieve exact solutions, but it suffered from prolonged computational times.

In recent years, advances in deep learning have significantly transformed automated engineering and construction design, enabling more efficient and

intelligent design workflows. Among these approaches, Convolutional Neural Networks (CNNs) have played a foundational role in processing spatial information and extracting meaningful features from visual data [7,8]. Their strong capability in analyzing complex spatial patterns has facilitated the automation of design tasks that traditionally relied on extensive manual effort. Subsequently, deep generative models based on convolutional architectures, particularly Generative Adversarial Networks (GANs) [9], have attracted considerable attention. Owing to their adversarial training mechanism, GANs are capable of producing designs that not only exhibit high visual realism but also satisfy certain engineering constraints. In the construction domain, GAN-based methods have been applied to layout generation [10], and defect detection [11]. Nevertheless, GAN models often suffer from training instability and limited output diversity [12], which restrict their scalability and robustness in practical engineering applications. Moreover, the rationality of a single generated design cannot be guaranteed, making model optimization and deployment challenging. In addition, existing studies on road design frequently depend on expert-based qualitative evaluation, which lacks standardized and objective metrics. Recent progress in state-of-the-art image generation techniques, particularly diffusion models [13,14] and their associated control mechanisms [15], which provide improved training stability, enhanced diversity of generated results, offers a promising pathway to overcome these limitations.

To address the aforementioned limitations, this paper presents MiCRoadDiff, a diffusion model-based intelligent approach for modular construction on-site haul road layout generation conditioned on construction plans and textual design conditions. Building upon the exceptional image synthesis capabilities of Stable Diffusion, this study aims to learn from designers' expertise to rapidly generate initial design proposals, serving as a valuable tool for assisting and streamlining the design process. This approach significantly contributes to the automation and efficiency of construction design workflows. This study integrates both textual and image design conditions to enable automated design. By utilizing the ControlNet architecture, the generated designs demonstrate strong alignment with the input conditions, facilitating the creation of customized and precise layout solutions. Domain adaptation technology, LoRA, is incorporated to improve model performance under limited data conditions. This effectively addresses the challenges of data scarcity frequently encountered in construction engineering applications. A comprehensive evaluation strategy is developed to assess the generated results, including the image quality aspect (FID, SSIM, and PSNR) and haul road layout aspect (transportation cost, accessibility, and number of corners). This approach aids in model optimization while ensuring both visual realism and compliance with engineering requirements.

2 Related Works

2.1 Deep Learning Methods for Site Layout Generation

Currently, there is limited research on using deep learning methods for site layout generation. Li et al. [10] utilized GAN-based methods to design tower crane locations for high-rise buildings, which overcomes complexity limitations and facilitates layout planning decision-making process. GAN models were adopted mostly for building layouts and floor layouts design [16]. However, GAN-based approaches encounter challenges such as training instability and limited output diversity. Additionally, existing research has yet to effectively address the critical issue of limited datasets, a challenge in the construction domain.

2.2 Diffusion Models in Image Generation

Diffusion models [13,14] are a powerful class of generative models that have recently gained significant attention for their ability to produce high-quality and diverse outputs. By employing an iterative denoising process, these models excel at capturing and modeling complex data distributions, making them particularly effective for tasks such as image synthesis and video generation.

To improve controllability and flexibility compared to GAN-based models, a multi-conditional model for residential floor plan design was developed using a diffusion model [17]. This approach demonstrated high-quality, diverse, and controllable results through experimental validation. An image-prompt diffusion model showed exceptional feature extraction capabilities and design effectiveness in shear wall design [18].

3 Methodology

3.1 Problem Formulation

MiCRoadDiff aims to automatically generate feasible and efficient haul road layouts for modular construction sites under complex spatial and operational constraints. Given a fixed construction site layout, and design conditions, the objective is to generate a temporary haul road configuration that ensures connectivity, safety, and constructability while minimizing conflicts with non-movable elements. Formally, the task is defined as a conditional image generation problem. Let A denote the site layout constraints (e.g., site boundary and building

footprints), *F* represent fixed facilities (e.g., tower crane bases and site gates), and *T* indicate textual design conditions (e.g., road width and minimum turning radius). The goal is to learn a generative model *G* that produces a haul road layout *R*: $R = G(A, F, T)$, where *R* satisfies geometric constraints imposed by *A* and *F*, and design logic encoded in *T*.

3.2 Overall Framework of MiCRoadDiff

The overall architecture of MiCRoadDiff is illustrated in Figure 1. The framework consists of three main components: (1) multi-layer conditional encoding, (2) conditional diffusion-based generation with ControlNet and LoRA, and (3) post-generation evaluation. First, for a construction site layout to be designed, multi-conditional maps (including boundary map and fixed facility map) are constructed, along with textual design requirements. These inputs are encoded and fed into a fine-tuned Stable Diffusion model. Then, the model leverages LoRA to perform experience-based haul road design, while ControlNet ensures compliance with the specified rigid constraints. Finally, the ideal generated layouts are evaluated using computer vision-related metrics and road design-specific models.

The core idea of MiCRoadDiff is to achieve flexible layout designs while satisfying rigid design constraints. Multiple condition maps are enforced through different ControlNet modules, while generalizable haul road design principles are learned via a lightweight LoRA adaptation of a pre-trained Stable Diffusion model.



Figure 1. The framework of the research

3.3 Stable Diffusion

Stable Diffusion is a latent diffusion-based generative model that synthesizes images by iteratively denoising a latent representation guided by conditional information. Instead of performing diffusion directly in pixel space, Stable Diffusion operates in a compressed latent space obtained via a variational autoencoder (VAE), which significantly reduces computational cost while preserving high-level semantic structure. During training, Gaussian noise is progressively added to latent variables, and a denoising U-Net is trained to predict and remove this noise step by step. At inference time, the model starts from random noise and gradually reconstructs an image that aligns with the provided conditions.

In the context of MiCRoadDiff, Stable Diffusion serves as the core generative engine responsible for synthesizing plausible haul road layout designs. Rather than explicitly encoding construction rules or geometric algorithms, the diffusion backbone provides strong priors for producing continuous, coherent, and visually structured layouts.

Among available Stable Diffusion variants [19], MiCRoadDiff adopts Stable Diffusion XL (SDXL) [20] as the backbone model. SDXL introduces a larger U-Net architecture, a higher-capacity text encoder, and a dual-stage conditioning mechanism compared to earlier versions such as SD v1.5. These improvements enable SDXL to better capture complex spatial relationships and long-range dependencies within an image, which are critical for construction site layout generation where global topology and local geometry must be jointly satisfied.

3.4 LoRA and ControlNet

Low-Rank Adaptation (LoRA) [21] is a parameter-efficient fine-tuning technique designed to adapt large pre-trained models to downstream tasks without updating the full set of model parameters. Instead of directly modifying the original weight matrices in the neural network, LoRA introduces a pair of low-rank matrices that approximate the weight updates. During training, only these low-rank matrices are optimized, while the original backbone parameters remain frozen. This significantly reduces memory consumption and training cost, while preserving the generalization capability of the pre-trained model. In MiCRoadDiff, LoRA plays a critical role in learning domain-specific haul road design logic that is not explicitly encoded in the condition informations. LoRA can capture higher-level design principles, including road continuity, transportation efficiency, and typical geometric patterns of temporary haul roads in modular construction sites. By integrating LoRA with the SDXL backbone, MiCRoadDiff leverages the strong generative prior of a large diffusion model while maintaining efficient and controllable domain adaptation.

ControlNet [15] is a conditional control architecture designed to extend large diffusion models with fine-grained and spatially aligned guidance, without disrupting their original generative capabilities. It introduces an auxiliary neural network that mirrors the structure of the diffusion model' s denoising U-Net and injects condition-specific features into the generation process through zero-initialized convolution layers. This design ensures that, at the beginning of training or inference, the ControlNet branch has no influence on the

output, thereby preserving the behavior of the pre-trained backbone, while gradually enabling strong conditional control as learning progresses.

In MiCRoadDiff, ControlNet is responsible for enforcing rigid spatial constraints that must be strictly preserved in construction site road layout generation. These constraints include site boundaries, building footprints, and fixed entities such as tower crane bases, ideal truck parking areas, and site entrances. Such elements are non-negotiable in practice and cannot be reliably handled by text prompts or learned implicitly through data-driven adaptation alone. MiCRoadDiff employs multiple ControlNet modules, each associated with a specific type of condition map.

3.5 Multi-conditional Image Generation

MiCRoadDiff performs haul road layout synthesis through a multi-conditional image generation process that jointly integrates multiple spatial condition maps and structured textual input within a unified diffusion framework, as shown in Figure 2. This design enables the model to handle heterogeneous constraints commonly encountered in construction site planning, where geometric feasibility, operational requirements, and design intent must be satisfied simultaneously.

Each condition map encodes a single type of constraint. Map-A encodes rigid and non-movable geometric boundaries, including site boundaries and building boundaries. Map-B encodes fixed entities including tower crane bases, ideal truck parking areas, and site entrances. This multi-conditional formulation allows MiCRoadDiff to explore diverse feasible solutions within a well-defined constraint space.

In addition to image-based conditions, MiCRoadDiff incorporates structured textual input to specify general design requirements that are difficult to encode spatially. The textual prompt follows a consistent and interpretable structure, following the format: "construction on-site haul road layout, modular construction site, top view site plan, road width approximately [Value] meters, minimum turning radius around [Value] meters, tower crane base location, truck parking area, site gate(s) location, road connects [Facilities], avoid [Obstacle]" This format ensures the generated images remain contextually and environmentally accurate by integrating both spatial and climatic factors. For example: *"construction on-site haul road layout, modular construction site, top view site plan, road width approximately 9 meters, minimum turning radius around 15 meters, tower crane base location, truck parking area, site gate(s) location, road connects site gate(s) and truck parking area, avoid buildings and crane base."* This structured prompting framework enables effective conditioning across haul road design intent. Through the joint use of multiple condition maps and structured textual input, MiCRoadDiff achieves controllable, robust, and flexible multi-conditional image generation tailored to the complex requirements of haul road layout design for MiC.



Figure 2. Multi-condition encoding for image generation

3.6 Post-generation Evaluation

To evaluate the quality and perceptual fidelity of the generated road layout images, three established metrics FID, SSIM, and PSNR are used. FID assesses realism and perceptual similarity, while SSIM and PSNR measure structural consistency and reconstruction quality. Besides, domain-specific metrics including transportation cost and accessibility, are used to evaluate layout performance in terms of efficiency and safety. Transportation cost is calculated as the product of the road centerline length and the unit construction cost. Accessibility evaluates how the designed road connects site components including gates and truck parking locations.

4 Experiment and Results

4.1 Experimental details

To train and evaluate LoRA, a dedicated dataset was constructed to reflect realistic modular construction site conditions. Ten real-world projects were sourced from renowned contractors in Hong Kong, resulting in the creation of a total of 10 data. Each data consists of image and corresponding textual description, enabling joint image–text conditioning during training. Textual descriptions are manually curated and aligned with each image sample. The constructed dataset highlights diversity in road types (one-way or two-way), road width (6m-10m) and turning radius (12m-18m), gate configurations (single or multiple), modular building shapes, and boundary shapes. This diversity allows the model to learn transferable haul road design logic rather than memorizing specific layouts. During training, the SDXL backbone parameters were kept frozen, and LoRA

modules were inserted into the attention layers of the U-Net architecture. Only the low-rank adaptation parameters were updated. No augmentation techniques were used during training. LoRA training and inference were performed on a workstation equipped with 13th Gen Intel(R) Core (TM) i7-13700KF@3.40 GHz CPU and NVIDIA GeForce RTX 4090 GPU. AdamW8bit [22] was used to optimize the low-rank decomposition of a subset of weights to minimize this loss. The batch size was set to 1. The learning rate for U-Net was set to $1e^{-4}$, while the learning rate for the text encoder network was $1e^{-5}$.

4.2 Training results of LoRA

The LoRA adaptation was trained for a total of 20 epochs, each consisting of 50 training steps. The evolution of the training loss over the entire training process is illustrated in Figure 3. The loss value shows a gradual decreasing trend with increasing training iterations, indicating stable convergence of the LoRA training process. The results generated by the SDXL model under the same text prompt under different training epochs were compared, as shown in Figure 4.

From the 5th to the 10th epoch, the generated results exhibit a clear shift toward a simplified semantic style suitable for site layout representation. By the 15th epoch, most critical layout components (haul road, tower crane base, truck parking area, site gate, and building boundaries) are correctly generated except site boundaries. While the 20-epoch model produces complete and well-organized layouts with road networks that satisfy engineering design requirements.

Overall, the LoRA-enhanced model shows a clear improvement in semantic alignment with modular construction on-site haul road planning intent. It also shows the effectiveness of the LoRA-based adaptation in capturing domain-specific haul road design logic for modular construction sites rather than memorizing specific site configurations.

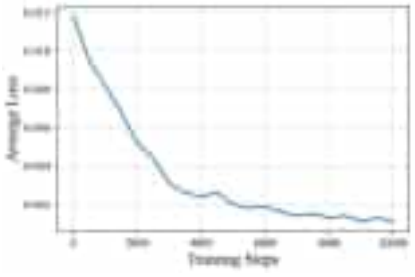


Figure 3. Training loss curves of LoRA

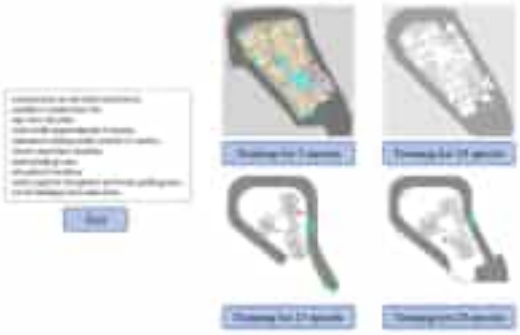


Figure 4. Comparison of model performance under different training epochs

4.3 Design results of MiCRoadDiff

The MiCRoadDiff model was built upon SDXL as the base diffusion model and incorporated a LoRA module trained for 20 epochs. In addition, two ControlNet modules were employed for image-based constraint control, with their parameters carefully tuned through multiple iterations to achieve optimal performance.

The MiCRoadDiff model was evaluated using new project cases that were not included in the training dataset. The generated haul road layouts were assessed through both qualitative and quantitative evaluations to comprehensively examine the model's generalization capability and design performance.

The project used in this study is a modular residential project located in Hong Kong. After simplification and semantic processing, the resulting design representations include condition image, design text, and corresponding haul road layouts, as shown in Figure 5. Black lines represent building boundaries and site boundaries, red squares denote tower crane locations, blue rectangles indicate truck parking areas with dimensions corresponding to real truck sizes, green symbols represent site gates, and grey regions denote haul roads, whose widths and turning radius reflect actual design dimensions. MiCRoadDiff was applied to design the haul road layout for this project, and the generated results are shown in Figure 6.

In this study, input construction drawings were pre-processed to standardize spatial resolution, enabling precise conversion of textual constraints into pixel units. Dimensional constraints are embedded into structured prompts, guiding the model toward geometrically plausible outputs during diffusion. Additionally, ControlNet ensures road layouts align with site geometry, preventing violations of spatial constraints. This ensures that the textual dimension can be accurately translated into spatial pixels.

From the design results, it can be observed that the model is capable of producing diverse haul road layout alternatives for designers' selection. The generated

layouts form relatively complete site plans with high image quality and clarity. Moreover, the generated haul roads satisfy the prescribed design conditions and are feasible from an engineering perspective.

Layout (a) was selected as the final design solution for quantitative analysis, with the results shown in Figure 7. The generated layout achieved favorable FID, SSIM, and PSNR scores, indicating high image clarity and strong structural similarity to the ground truth. In addition, the total road length of the designed haul road is 128 m, with an estimated construction cost of 385,101 HKD. The layout exhibits good connectivity by linking two site gates and effectively serving the designated truck parking areas. However, the presence of three turning corners may slightly increase driving difficulty in practice.



Figure 5. Design conditions and real on-site haul road layout of the case project

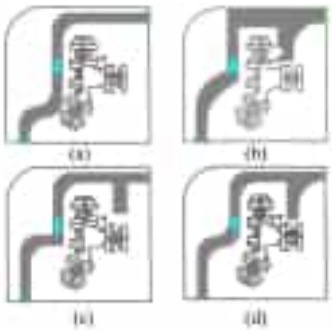


Figure 6. Design results of MiCRoadDiff

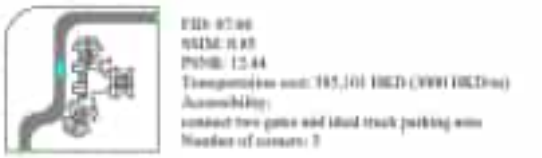


Figure 7. Evaluations on layout design result

5 Discussion

5.1 Ablation test of MiCRoadDiff

To investigate the contribution of different modules to the generation performance of MiCRoadDiff, an ablation study was conducted. The experimental results are presented in Figure 8 and Table 1.

As shown in Figure 8, when the ControlNet modules are removed, the model no longer produces engineering drawings with fixed spatial constraints. Instead, it generates layouts with varying building and site shapes, randomly arranged tower crane locations, truck parking areas, and site gates. Although these results exhibit a certain degree of creativity, the model still maintains a consistent visual style and the design experience learned during training. When the LoRA module is removed, the model shifts toward a more photorealistic visual style. While the overall building and site boundaries are still preserved, the model fails to correctly interpret the design conditions specified in the text prompts, such as road width, turning radius, and connectivity requirements. When only the SDXL baseline model is used for text-to-image generation, the generated results appear visually realistic but lack a consistent design style and fail to satisfy engineering design constraints.

Reducing each module leads to variations in site configurations or drawings that cannot be evaluated using road-related metrics, as shown in Figure 8. Consequently, only image similarity metrics were employed for quantitative comparison. The results of the compared evaluation metrics are listed in Table 1. The symbol ↓ indicates that lower values are better while ↑ the opposite is. The results show that the MiCRoadDiff model achieves the greatest performance in all three metrics. As shown in Figure 9, all three metrics exhibit clear improving trends as the model evolves from the SDXL baseline to MiCRoadDiff. Compared with SDXL, MiCRoadDiff achieves improvements of 83%, 214%, and 214% in FID, SSIM, and PSNR, respectively. The introduction of the ControlNet and LoRA modules effectively enhances overall model performance. Removing the LoRA module results in a significant decline in performance, with FID, SSIM, and PSNR dropping by 272%, 50%, and 18%, respectively. Moreover, MiCRoadDiff significantly outperforms models with a single enhancement strategy, thereby verifying the effectiveness of the proposed framework for modular construction haul road design.



Figure 8. Comparison with baseline models

Table 1 Test results of ablation experiment

Model	FID↓	SSIM↑	PSNR↑
SDXL	530.41	0.27	3.95
SDXL+CN	326.85	0.42	10.19
SDXL+LoRA	243.40	0.83	11.51
MiCRoadDiff	87.66	0.85	12.44

(SDXL+LoRA+CN)



Figure 9. Comparison between different models regarding FID, SSIM, and PSNR

5.2 Layout design with different conditions

To investigate the adaptability of MiCRoadDiff under different design conditions, a series of experiments were conducted by modifying key site constraints, including truck parking area locations, site gate positions, and conditional maps. The qualitative results are illustrated in Figure 10.

When the truck parking area is relocated, MiCRoadDiff successfully generates distinct haul road layouts that build connectivity between the site gates and the designated parking area.

When the site gate position is changed, the generated haul roads could provide feasible access to the truck parking area, while respecting site boundaries and site gates. These results indicate that MiCRoadDiff is capable of handling changes in global access conditions.

When only Map-A (Building and site boundaries) is provided, the generated haul road generally follows the site boundary but lacks precise alignment with functional facilities, resulting in ambiguous or suboptimal connectivity. Conversely, when only Map-B (Site gates, tower crane, and truck) is used, the model correctly recognizes facility locations but may produce road geometries that are less compatible with the overall site shape. These observations highlight that a single conditional map is insufficient to fully constrain the haul road design task. The combination of multiple condition maps is necessary to simultaneously enforce geometric boundaries and functional requirements.

Overall, the experimental results demonstrate that MiCRoadDiff can robustly adapt to varying site conditions. The model effectively integrates multi-source conditional information to generate coherent and feasible haul road layouts.

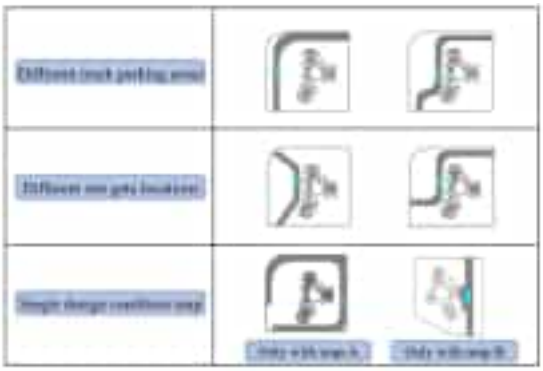


Figure 10. Comparison with different design conditions

6 Conclusion and Future Work

This paper proposed MiCRoadDiff, a novel generative design framework for automated on-site haul road layout generation in modular construction. Besides, the evaluation strategy takes into account both visual and layout-based metrics, ensuring a reliable assessment of design quality. By integrating a pre-trained Stable Diffusion backbone with LoRA-based domain adaptation and ControlNet-based geometric constraint control, the proposed method enables the efficient generation of feasible and high-quality haul road layouts under complex site conditions. The generated layouts of MiCRoadDiff form relatively complete site plans with high image quality and clarity, satisfying the prescribed design conditions and are engineering feasible. Experiments also highlight the significant contribution of LoRA to generation quality. Removing the LoRA module results in a significant decline in performance, with FID, SSIM, and PSNR dropping by 272%, 50%, and 18%, respectively. Meanwhile, ControlNet ensures the controllability of the design process, maintaining its reliability and precision. The experimental results demonstrate that MiCRoadDiff can robustly adapt to varying site conditions such as different site gate locations and truck parking areas. The proposed framework significantly reduces the manual effort required for haul road design while maintaining strong engineering feasibility. Future work will focus on extending the proposed framework to more complex construction scenarios, including larger-scale sites and irregular geometries, as well as incorporating road structural properties such as pavement strength classes. In addition, the influence of parametric variations in textual design conditions on generation outcomes will be systematically investigated. Integrating the generated haul road layouts with Building Information Modeling (BIM) platforms or professional traffic and road simulation tools provides a promising direction, which could further support engineers in preliminary design

evaluation and feasibility analysis.

References

[1] W. Pan, C.K. Hon, Briefing: Modular integrated construction for high-rise buildings, Proceedings of the Institution of Civil Engineers - Municipal Engineer 173 (2020) 64–68. https://doi.org/10.1680/jmuen.18.00028.

[2] CIC, Why MiC, (2020). https://mic.cic.hk/en/WhyMiC (accessed January 26, 2026).

[3] C. Yi, M. Lu, A mixed-integer linear programming approach for temporary haul road design in rough-grading projects, Automation in Construction 71 (2016) 314–324. https://doi.org/10.1016/j.autcon.2016.08.022.

[4] D. Xiang, H. Lin, J. Ouyang, D. Huang, Combined improved A* and greedy algorithm for path planning of multi-objective mobile robot, Scientific Reports 12 (2022) 13273.

[5] B. Huang, L. Yao, K. Raguraman, Bi-level GA and GIS for Multi-objective TSP Route Planning, Transportation Planning and Technology 29 (2006) 105–124. https://doi.org/10.1080/03081060600753404.

[6] D. Lenz, T. Kessler, A. Knoll, Tactical cooperative planning for autonomous highway driving using Monte-Carlo Tree Search, in: 2016 IEEE Intelligent Vehicles Symposium (IV), IEEE, Gotenburg, Sweden, 2016: pp. 447–453. https://doi.org/10.1109/IVS.2016.7535424.

[7] W. Fang, L. Ding, B. Zhong, P.E. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Advanced Engineering Informatics 37 (2018) 139–149.

[8] H. Luo, C. Xiong, W. Fang, P.E. Love, B. Zhang, X. Ouyang, Convolutional neural networks: Computer vision-based workforce activity assessment in construction, Automation in Construction 94 (2018) 282–289.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144. https://doi.org/10.1145/3422622.

[10] R. Li, H.-L. Chi, Z. Peng, X. Li, A.P.C. Chan, Automatic tower crane layout planning system for high-rise building construction using generative adversarial network, Advanced Engineering Informatics 58 (2023) 102202. https://doi.org/10.1016/j.aei.2023.102202.

[11] G. Zhang, Y. Pan, L. Zhang, Semi-supervised learning with GAN for automatic defect detection from images, Automation in Construction 128 (2021) 103764.

[12] D. Saxena, J. Cao, Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions, ACM Comput. Surv. 54 (2022) 1–42. https://doi.org/10.1145/3446374.

[13] J. Ho, A. Jain, P. Abbeel, Denoising Diffusion Probabilistic Models, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020: pp. 6840–6851.

[14] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, 2022: pp. 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042.

[15] L. Zhang, A. Rao, M. Agrawala, Adding Conditional Control to Text-to-Image Diffusion Models, in: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Paris, France, 2023: pp. 3813–3824. https://doi.org/10.1109/ICCV51070.2023.00355.

[16] N. Nauata, K.-H. Chang, C.-Y. Cheng, G. Mori, Y. Furukawa, House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation, (2020). https://doi.org/10.48550/arXiv.2003.06988.

[17] P. Zeng, W. Gao, J. Yin, P. Xu, S. Lu, Residential floor plans: Multi-conditional automatic generation using diffusion models, Automation in Construction 162 (2024) 105374. https://doi.org/10.1016/j.autcon.2024.105374.

[18] Y. Gu, Y. Huang, W. Liao, X. Lu, Intelligent design of shear wall layout based on diffusion models, Computer-Aided Civil and Infrastructure Engineering (2024) mice.13236. https://doi.org/10.1111/mice.13236.

[19] Stability AI, generative-models, GitHub Repository (2023). https://github.com/Stability-AI/generative-models.

[20] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, R. Rombach, SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis, (2023). https://doi.org/10.48550/arXiv.2307.01952.

[21] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-Rank Adaptation of Large Language Models, (2021). https://doi.org/10.48550/arXiv.2106.09685.

[22] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, (2017). https://doi.org/10.48550/arXiv.1412.6980.

A BIM-based Ontology Method for Knowledge Representation in Light Gauge Steel Panel Assembly

Mariana Marines Alvarado[1], Arash Hosseini Gourabpasi[2], Farzad Jalaei[2], Rafiq Ahmad[1]

[1]Smart and Sustainable Manufacturing Systems Laboratory (SMART Lab), Department of Mechanical Engineering, University of Alberta, 9211 116 Street, NW, Edmonton, AB T6G1H9, Canada
[2]National Research Council Canada, 1200 Montreal Road, Building M-58, Ottawa, Ontario K1A 0R6, Canada
marinesa@ualberta.ca, arash.hosseinigourabpasi@nrc-cnrc.gc.ca, farzad.jalaei@nrc-cnrc.gc.ca, rafiq.ahmad@ualberta.ca

Abstract –

Off-site construction offers significant potential to improve productivity and quality in the Architecture, Engineering, and Construction (AEC) industry; however, the practical adoption of Design for Manufacturing and Assembly (DfMA) remains limited due to the lack of explicit integration between product-focused Building Information Modeling (BIM) data and construction process knowledge. While BIM provides detailed geometric representations, assembly logic and process-related information are not explicitly expressed.

This paper proposes a BIM-based ontology framework for semantic knowledge representation in Light Gauge Steel (LGS) panel assembly. A BIM model is developed and exported to Industry Foundation Classes (IFC) format. Relevant component and geometric attributes are selectively extracted using IfcOpenShell and mapped to ontology individuals implemented in a semantic web-based ontology editor that supports formal knowledge modeling and reasoning. Product components and their connection relationships are explicitly modeled, and a connection-based reasoning approach is adopted to infer assembly operations, enabling a clearer linkage between BIM-derived product information and DfMA-oriented process knowledge.

Keywords –

BIM; Ontology; DfMA; Off-Site Construction; Light Gauge Steel

1 Introduction

1.1 Off-Site Construction

Off-Site Construction (OSC) refers to the manufacturing prefabricated building components in a factory or controlled facility, where elements are produced off-site under controlled conditions and then transported to the construction site for final assembly [1]. OSC has gained notable attention for its ability to address several limitations of traditional on-site construction. Major advantages include alleviating skilled labor shortages, shortening project timelines and reducing cost fluctuations, decreasing construction waste, employing environmentally friendly methods, and enhancing quality through controlled factory environments [2]. Although OSC offers several benefits, its adoption presents several challenges. Wuni and Shen [3] identified key obstacles in implementing OSC through literature reviews and categorized them into eight groups, including technical barriers related to design information management and constructability.

1.2 Role of Building Information Modeling (BIM)

Building Information Modeling (BIM) is defined in ISO 19650-1:2018 as the use of a shared digital representation of a built asset to facilitate design, construction and operation processes to form a reliable basis for decisions [4]. In practice, BIM acts as a digital model that supports planning, design, construction, and operation across the entire project lifecycle [5]. Consequently, BIM has proven to enhance decision-making in the architecture, engineering, and construction (AEC) industry by improving work efficiency, reducing modeling rework, and minimizing the repeated collection of building data for different analytical uses.

However, implementing BIM in construction practice still presents several limitations. In particular, extracting construction quantities and procedures from BIM design models requires specialized technical expertise and dedicated tools. This limitation arises because product and process information are typically structured using different schemas and classification systems: BIM product information is commonly represented using the Industry Foundation Classes (IFC) schema, whereas construction process information is usually organized according to material- and method-based classification systems employed in construction planning and management. BIM models represented using the IFC

schema describe buildings as assemblies of parametric objects with associated properties, while construction process models are structured as collections of activities organized by these classification systems [6]. This misalignment hampers the direct use of BIM data represented in IFC format for construction planning and execution, emphasizing the need to improve interoperability and integration between product and process information.

1.3 Ontologies for Knowledge Representation

Ontology is a formal approach to knowledge modeling that explicitly defines domain concepts and their relationships. An ontological model consists of classes, properties, and relationships [6][7]. In an ontology, classes represent domain concepts, properties describe their features, and relationships indicate connections among different concepts, enabling a clear and intuitive representation of complex information.

Ontologies enable custom, user-defined relationships among classes. They also allow for the explicit definition of semantic features such as symmetry, transitivity, and inversion of relationships. Based on formal logic, ontologies support automated reasoning and inference, which helps derive new knowledge and identify inconsistencies without requiring all information to be explicitly modeled [7]. Although BIM data represented in IFC can include custom property sets and attributes to describe information structures, the IFC schema itself does not provide ontology-level semantics, such as axioms and logical relationships that enable reasoning. Consequently, IFC can describe data and its structure but cannot infer new knowledge from existing relationships, whereas ontology-based representations can. This capability reduces the amount of data that must be extracted from BIM, as additional information can be inferred from existing data, and it enhances semantic interoperability by providing a shared conceptual framework through which BIM and construction-related information can be consistently integrated.

Ontology-based approaches have been successfully applied in construction to support construction-oriented quantity take-off [5], IFC-based cost modeling [8], automated verification of quality control specifications for off-site manufactured components [9], and formal representation of space and task-related constraints to enhance lean construction planning and execution control in steel fabrication projects [10]. These applications demonstrate the capacity of ontologies to structure domain knowledge and improve interoperability in BIM-driven workflows.

Beyond the construction domain, ontologies have been widely adopted to model complex cyber-physical and intelligent systems, including smart homes, smart farms, smart factories, and other Industry 4.0 environments. In these contexts, ontologies are often combined with machine learning and deep learning techniques to address interoperability challenges, support system monitoring and control, enable digital twin representations, and optimize design and operational parameters for improved performance and reduce cost [11]. This capability is particularly relevant for off-site construction, where design information, manufacturing constraints, and assembly knowledge must be systematically linked to support informed decision-making.

1.4 Design for Manufacturing and Assembly (DfMA)

Design for Manufacturing and Assembly (DfMA) is a structured design approach that incorporates manufacturing and assembly considerations early in the design process to enhance constructability, productivity, and overall project performance. In off-site construction, DfMA plays a vital role by facilitating the design of prefabricated components optimized for factory production and efficient on-site assembly [12]. Past research shows that DfMA helps reduce construction time, cut material waste, improve quality control, and boost coordination between design and construction teams [7][12]. These advantages are especially important for prefabricated systems like Light Gauge Steel (LGS) panel assemblies, where manufacturing efficiency and assembly sequencing directly influence project results.

However, the practical application of DfMA is limited by the absence of a structured and explicit way to represent manufacturability and constructability knowledge during the design process. DfMA-focused decision-making needs detailed information about component geometry, connection types, assembly logic, and construction procedures. Although BIM provides a detailed, product-centered view of building components, it does not explicitly include construction processes or assembly information needed to support DfMA-based analysis fully [6]. Several researchers have highlighted the importance of supplementing BIM with knowledge-based approaches to facilitate the integration of product and process information for prefabricated construction [6][7]. Ontology-based modeling has been recognized as an effective way to formally represent construction knowledge, establish relationships between components and processes, and enable semantic reasoning to support advanced analysis and automation [7].

In this context, this paper proposes a BIM-based ontology framework for the semantic representation of product and process knowledge in Light Gauge Steel (LGS) panel assembly. The proposed framework acts as a crucial step toward enabling DfMA-oriented analysis by connecting BIM-derived product information with construction process knowledge.

2 Related Work

Ontology-based approaches are increasingly used in the AEC industry to improve the semantic representation and use of BIM data for construction-related applications. Current studies show that ontologies can support decision-making, automate validation tasks, and enhance information sharing across different stages of the construction process, as well as exploring the use of ontologies to support Design for Manufacturing and Assembly (DfMA) in prefabricated construction. Qi and Costin introduced a BIM and ontology-based DfMA framework to help verify manufacturability and constructability requirements for prefabricated components [7]. Their work focused on representing stakeholders' needs and constraints with an ontology to enable automated checking and feedback during the design phase. This highlights the role of ontology in formalizing DfMA knowledge and improving design validation for prefabrication. Ontology-based methods have also been applied to construction-focused quantity take-off (QTO). Liu et al. created a semantic approach that combines BIM and ontology to support QTO in the light-frame building industry [5]. Their work shows how ontology can connect BIM product models with construction process information, allowing for accurate quantity extraction through semantic queries. Additionally, ontologies have been used to extend IFC-based representations for cost-related applications. Isatto proposed an IFC ontology to support process-based cost modeling, emphasizing the semantic representation of construction processes and their costs [8]. This work demonstrates the potential of IFC-based ontologies to improve interoperability and enable analysis linking product, process, and cost information.

More recently, ontology-based decision support systems have been developed to assess manufacturability in prefabricated framing systems. An et al. [13] presented a BIM-based decision support system for automated manufacturability checking of wood frame assemblies. Their system integrates BIM data input, geometric feature detection, ontology formulation, and machine eligibility determination to evaluate whether a given frame assembly can be manufactured using specific machinery. This study represents an important step toward linking BIM-derived product information with manufacturing constraints and operational feasibility.

Despite these advances, existing ontology-based approaches in the AEC industry generally focus on specific applications or isolated aspects of the construction process. While Qi and Costin [7] emphasize DfMA requirement validation, Liu et al. [5] concentrate on construction QTO, Isatto [8] addresses cost modeling, and An et al. [13] focus on manufacturability assessment for wood frame systems, these studies tend to treat product, process, and operational knowledge separately. Furthermore, limited attention has been given to explicitly linking BIM-based product geometry, component-level connection types, and the corresponding fabrication or assembly operations required for prefabricated systems. There remains a need for an integrated ontology framework that connects BIM-based product information with DfMA-oriented process knowledge, particularly for Light Gauge Steel (LGS) panel assemblies. This paper addresses this gap by proposing a BIM-based ontology framework that represents LGS components and their connections and associates them with fabrication and assembly operations, establishing a semantic foundation for knowledge representation and reasoning in prefabricated LGS construction.

3 Methodology

3.1 Overall Approach

The proposed framework creates a structured process for integrating BIM with ontology-based knowledge representation to support the semantic modeling of LGS panel assembly. It is an essential step toward DfMA-oriented analysis by systematically linking product information from BIM models with process-related construction knowledge.

The workflow starts with developing a BIM model of an LGS wall panel using Autodesk Revit. This BIM model serves as the main source of digital product information, including structural components like studs, tracks, and fasteners, along with their parametric and geometric properties. Simultaneously, a domain-specific ontology is defined using Protégé to represent relevant product and process knowledge, such as component classifications, geometric attributes, connection types, and assembly operations.

Next, the BIM model is exported from Revit to the IFC format to enable open and standardized data exchange. The IFC file is validated with BIMVision to check schema compliance and ensure accurate representation of building elements. IFC acts as an intermediary schema to facilitate interoperability and automated access to BIM data across various software platforms [2].

Relevant BIM attributes are then extracted from the IFC file using a Python-based workflow with IfcOpenShell. This tool allows computational access to IFC entities and their properties, enabling the extraction of attributes necessary for manufacturability and constructability, such as component identifiers, classifications, and geometric data [13]. The extracted information is organized to support subsequent semantic mapping and reasoning.

After data extraction, the BIM information is mapped

to ontology individuals and properties. Each LGS element becomes an individual in the ontology within Protégé, linked to relevant classes and data properties, resulting in an ontology-enhanced BIM that explicitly captures product semantics.

Connections between components are modeled explicitly in the ontology and used as a basis for linking product elements with construction processes [13]. In this approach, connections rather than component types are used to associate assembly operations. By formally representing these connection relationships within the ontology, the proposed framework builds upon the geometric and relational information available in the IFC schema and extends it with semantic constructs that enable process inference. While IFC can represent component geometry and basic relationships, it does not natively support the logical rules required to derive construction operations from connection types. The ontology, therefore, introduces a semantic layer that enhances the expressive capability of IFC to support the DfMA-oriented assembly use case presented in this study.

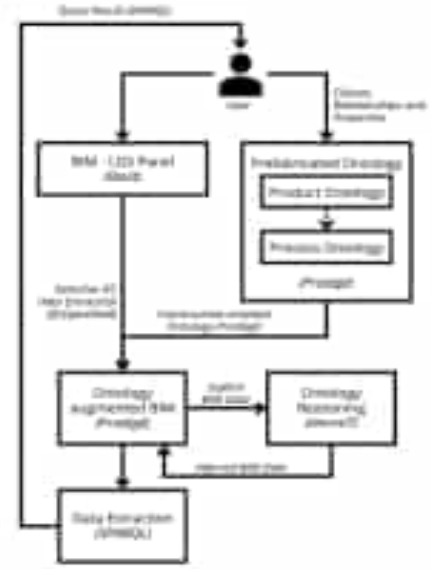


Figure 1. Overview of proposed methodology

Ontology-based reasoning is performed using the HermiT reasoner to ensure logical consistency and to infer implicit knowledge from formally defined class restrictions and object property relationships. Assembly operations are inferred through predefined axioms that associate specific connection types with corresponding construction activities; when the required semantic conditions are satisfied, individuals are automatically classified into the appropriate operation classes. This reasoning process enriches the ontology by generating product–process relationships that are not explicitly represented in the original BIM model. HermiT was selected due to its reliable support for OWL-based classification and instance reasoning, as well as its stable integration with the ontology development environment used in this study.

Finally, the populated and reasoned ontology facilitates semantic querying using the SPARQL Protocol and RDF Query Language to retrieve both explicit and inferred information related to LGS panel assembly. The query results support validation and analysis of product-process relationships and provide a structured semantic representation that can enable future developments in automated assembly planning and BIM-ontology interoperability.

3.2 Ontology Development

The ontology development process adopted in this paper follows established methodologies reported in the literature [12]. The major steps are summarized as follows:

1. Determine the purpose and scope of the ontology.

The purpose of the proposed ontology is to support BIM-based information modeling for DfMA-oriented fabrication. The scope of the ontology includes all information related to light-gauge steel panel components and their associated fabrication and assembly processes.

2. Review existing ontology.

The ontology for manufacturability and constructability of prefabricated components proposed by Qi et al. [12] is used as a reference to extract and structure product-related information from the BIM model. The shared ontology is adapted and simplified to focus on the relevant concepts needed for light gauge steel panel components. Additionally, the ontology-based knowledge modeling approach for frame assemblies proposed by An et al. [13] is reviewed to help guide the representation of relationships between product and process knowledge.

3. Enumerate main classes and build the corresponding class hierarchy.

Product-related knowledge is broken down from the product level to geometry and connection types, while process-related knowledge is shown through operation classes.

4. Construct a machine-readable ontology.

The ontology is implemented using the Protégé platform, where classes, properties, and relationships are formally defined.



Figure 2. Core ontology classes

3.2.1 Product Ontology

The product ontology describes the physical components and attributes of an LGS panel assembly. Structural elements such as studs, tracks, joists, noggings, and fasteners are modeled as subclasses of a general Structural_element class. Each component is represented as an ontology individual derived from BIM data and linked to relevant geometric and identification attributes. Geometric details, including length, height, width, thickness, and cross-sectional area, are represented through data properties and populated with values extracted from the BIM model.

In addition, component relationships are explicitly modeled through connection types (e.g., perpendicular, lateral, angled), allowing the representation of assembly logic between elements. This structured representation enables product information to be interpreted beyond purely geometric descriptions and establishes a semantic foundation for linking product and process knowledge.

It is important to note that the product ontology presented in this study represents the first iteration of the proposed framework. Further refinement and extension of the ontology structure are planned through expert feedback, validation exercises, and additional case studies to improve coverage, consistency, and applicability to broader prefabricated construction scenarios.



Figure 3. Product class hierarchy.

3.2.2 Process Ontology

The process ontology describes construction and assembly operations related to LGS panel fabrication and installation. Assembly procedures are modeled as subclasses of a general Operation class, representing actions such as positioning, aligning, and screwing.

The proposed framework adopts a connection-based modeling strategy in which assembly operations are inferred from the type of connection established between components, rather than being directly assigned to component categories. Connection types are explicitly represented in the ontology and semantically linked to corresponding fabrication or assembly operations through predefined relationships and logical restrictions.

By structuring process assignment around connection semantics, the framework enables a more precise representation of product–process relationships derived from BIM data. This approach supports DfMA-oriented reasoning by translating geometric and relational information into construction-relevant operations through formal semantic rules.



Figure 4. Process class hierarchy.



Figure 5. Object Property requiresOperation.

3.3 BIM to Ontology Mapping

The BIM-to-ontology mapping establishes a semantic link between the BIM design model and the ontology-based knowledge representation, following the ontology-driven mapping approach proposed by An et al. [13]. The LGS panel developed in Autodesk Revit is exported to Industry Foundation Classes (IFC) format as an intermediate schema for interoperable exchange. Rather than mapping the full IFC schema, only BIM attributes relevant to prefabrication and assembly are selectively extracted, consistent with the strategy in [13].

IfcOpenShell is employed to parse the IFC file and retrieve component-level data, including identifiers and geometric properties of LGS elements. The product data extraction process is based on IFC entities and property sets that represent structural components and their geometric information. Structural members are primarily retrieved from entities such as *IfcMember*, while geometric representations are accessed through shape information associated with the IFC product representation and processed using ifcopenshell.geom. Element placement and orientation are derived from local

placement information, and element attributes are accessed through property sets using IfcOpenShell utilities such as ifcopenshell.util.element.get_psets, including BaseQuantities and Identity Data, which provide dimensional and identification information. The overall implementation is carried out through the following steps:

1. Product data extraction:

A Python-based script automatically extracts element identifiers (GlobalId), semantic roles, and dimensional attributes from IFC entities and property sets using IfcOpenShell.

2. Connection detection and classification:

Geometric information derived from the IFC model is used to automatically detect potential connections between structural members using a tolerance-based proximity (AABB) method. Each detected interaction is classified into a predefined connection type (e.g., Perpendicular, Lateral, Angled) based on geometric overlap and angular relationships.

3. Ontology instantiation:

The extracted elements and classified connections are automatically transformed into RDF triples. Each component is instantiated as an ontology individual and linked to predefined classes and data properties, while connection instances are created to explicitly relate connected elements and assign their corresponding connection types. Connections between components are explicitly represented in the ontology and used to link product elements with construction operations defined in the process ontology. This connection-based mapping separates process definitions from specific component types, allowing flexible reasoning over assembly operations. Through this mapping process, BIM data are converted into structured semantic knowledge, forming the basis for ontology-based reasoning and querying in subsequent stages.

4 Validation and Results: A Case Study

A Light Gauge Steel (LGS) wall panel is selected as the case study to demonstrate the applicability of the proposed BIM-based ontology framework. The panel represents a typical prefabricated wall assembly composed of studs, tracks, and fasteners commonly used in off-site construction.

The LGS panel is modeled in Autodesk Revit using parametric definitions for its structural components. The BIM model captures geometric dimensions, spatial configuration, and connection arrangements relevant to panel fabrication and assembly. The model is exported to IFC format and processed through the proposed BIM-to-ontology mapping pipeline. Figure 6 presents the three-dimensional representation of the LGS panel used in this study.

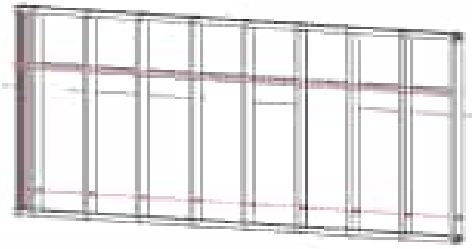


Figure 6. LGS panel case study

Through the mapping process, BIM-derived product information is transformed into a semantically structured ontology populated with both product and process knowledge. A total of 144 structural elements were extracted from the IFC model and mapped directly into ontology individuals. Each element corresponds to one ontology individual and is enriched with geometric and identification attributes derived from IFC property sets. Among these, a subset of primary structural members was considered for connection detection, including 10 studs, 2 tracks, and 2 straps. For the purpose of connection detection and process inference, only load-bearing members (i.e., studs and tracks) were used, while secondary elements such as straps were excluded from the connection detection stage.

By formalizing these relationships, the framework establishes an explicit and machine-interpretable link between product geometry and construction processes. This connection-based semantic modeling enables BIM-derived product information to be systematically translated into DfMA-oriented assembly knowledge, thereby supporting more integrated and construction-aware decision-making.

The ontology is evaluated using four validation metrics adapted from prior ontology-based studies [14]: competency, consistency, completeness, and clarity. The adopted evaluation structure is summarized in Table 1.

Table 1. Validation metrics and approaches implemented

Metrics	Approach	Tool
Competency	Competency questions	Case study + SPARQL queries
Consistency	Automated consistency checking	HermiT and Pellet reasoners
Completeness	SPARQL query	SPARQL query engine
Clarity	Structural modeling review	OOPS! (Ontology Pitfall Scanner)

4.1 Competency

Three competency questions were defined based on the purpose and scope of supporting BIM-based modeling for DfMA-oriented LGS fabrication:

1. Can the ontology retrieve geometric and identification attributes of LGS components extracted from BIM?
2. Can the ontology explicitly represent and classify connection relationships between structural elements?
3. Can the ontology infer appropriate assembly operations based on identified connection types?

The proposed ontology successfully answered all three questions. The developed ontology includes 30 classes, 6 object properties, and 22 data properties. Through the BIM-to-ontology mapping process, 144 structural component instances were automatically generated from the IFC model. Geometric attributes (e.g., length, thickness, positional data) were successfully assigned to each individual. The AABB-based geometric detection identified 20 connection instances, which were classified into predefined connection types (e.g., Crossing, Perpendicular, Lateral, Angled, Double_Angled).

To validate inference capability (Competency Question 3), the SPARQL query in Figure 7 was executed.



Figure 7. SPARQL query for validation

The query successfully retrieved connection instances, associated elements, classified connection types, and inferred assembly operations. A total of 20 operation instances were inferred automatically through ontology reasoning without manually assigning operations to specific component categories. These results confirm that assembly operations are derived semantically from connection relationships rather than explicitly encoded within the BIM model.

4.2 Consistency

Logical consistency and classification were verified using both the HermiT and Pellet reasoners. Automated reasoning confirmed the validity of class hierarchies, object property restrictions, and inference rules linking connection types to operations. The ontology was processed in 0.388 seconds using HermiT (v1.4.3.456) and 0.411 seconds using Pellet, demonstrating computational feasibility for the case study scale. No logical inconsistencies were detected.

4.3 Completeness

Completeness was evaluated through SPARQL-based query testing against the defined competency questions. Queries successfully retrieved:

- All structural component individuals and their geometric attributes.
- All detected connection instances and their classified types.
- All inferred assembly operations associated with each connection type.

The successful retrieval of expected results demonstrates that the ontology sufficiently covers the defined product and process scope.

4.4 Clarity

Clarity was evaluated through structural inspection of class hierarchies, object property definitions, and naming conventions. The ontology maintains a clear separation between product, connection, and process concepts while preserving semantic coherence. Additionally, the ontology was analyzed using the OOPS! (Ontology Pitfall Scanner) tool. The evaluation reported no critical errors and only minor warnings, indicating acceptable modeling quality for this first implementation.

4.5 Interpretation

The case study confirms that the proposed framework establishes an explicit, machine-interpretable link between product geometry and construction processes. By associating assembly operations with connection types rather than component categories, the framework addresses the research problem of integrating product-oriented BIM data with process-oriented DfMA knowledge. Although this implementation represents the first iteration of the framework, the validation results demonstrate its feasibility, logical soundness, and reasoning capability. Future work will focus on incorporating expert feedback, expanding case studies, and enhancing automation of connection recognition to further strengthen the ontology model.

5 Conclusion and Future Work

This paper introduced a BIM-based ontology framework for semantic knowledge representation in LGS panel assembly. The proposed approach combines

BIM-derived product data with ontology-based process knowledge to support DfMA-oriented analysis in off-site construction. Using IFC-based data extraction, ontology modeling, and reasoning, the framework allows for structured representation and inference of assembly-related information. The case study demonstrated the feasibility of mapping BIM elements to ontology individuals, enriching them with geometric attributes, and associating construction operations through a connection-based reasoning approach. The results confirm that the proposed framework can bridge the gap between product-centric BIM models and process-oriented construction knowledge.

Future work will focus on extending the proposed ontology framework to incorporate formal constraint modeling beyond product–process relationships. Ontology-based approaches have demonstrated strong potential for representing quality control requirements, manufacturing constraints, regulatory specifications, and task-level planning rules within BIM-driven workflows. Building on this direction, future developments of the framework will investigate the integration of design standards, regulatory requirements, and task-level constraints into the ontology, enabling automated compliance checking and enhanced decision support during LGS panel design and assembly planning. This extension will further strengthen the ontology as a knowledge base for DfMA-oriented optimization and constraint-aware design in prefabricated construction.

Acknowledgments

This study has been supported by the Construction Sector Digitalization and Productivity (CSDP) program under the project "NRCC CSTIP CSDP-018-1 – BIM-Integrated Multi-Robot Platform for Customized and Sustainable Offsite Construction Manufacturing". The authors would like to extend their sincere gratitude to all the stakeholders who participated in this research and contributed to the findings presented in this paper.

References

[1] S. Abrishami and R. Martín-Durán. BIM and DfMA: A paradigm of new opportunities. *Sustainability*, 13:9591, 2021.

[2] H. Hyun, H.-G. Kim, and J.-S. Kim. Integrated off-site construction design process including DfMA considerations. *Sustainability*, 14:4084, 2022.

[3] Wuni, I. Y., and Shen, G. Q. Barriers to the adoption of modular integrated construction: Systematic review and meta-analysis, integrated conceptual framework, and strategies. *Journal of Cleaner Production*, 249:119347, 2020.

[4] ISO. ISO 19650-1:2018 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) -Information management using building information modelling - Part 1: Concepts and principles. *International Organization for Standardization*, 2018.

[5] H. Liu, M. Lu, and M. Al-Hussein. Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry. *Advanced Engineering Informatics*, 30:190–207, 2016.

[6] B. Qi, A. Costin, and M. Razkenari. An ontology for manufacturability and constructability of prefabricated component. In *Proceedings of the International Conference on Computing in Civil Engineering 2021*, pages 745–752, Orlando, FL, USA, 2021.

[7] B. Qi and A. Costin. BIM and ontology-based DfMA framework for prefabricated component. *Buildings*, 13:394, 2023.

[8] E. Isatto. An IFC representation for process-based cost modeling. In *Proceedings of the International Conference on Computing in Civil Engineering*, pages 519–528, 2021.

[9] J. Lee, Y. Kim, and S. Lee. Automatic selection tool of quality control specifications for off-site construction manufacturing products: A BIM-based ontology model approach. *Automation in Construction*, vol. 120, 2020.

[10] M. Hosseini, R. Ahmad, and M. Al-Hussein. Development of a space constraint management ontology for lean construction: Enhancing constraint management in steel fabrication projects. *Automation in Construction*, vol. 132, 2021.

[11] Abbasi, R., Martinez, P., and Ahmad, R. An ontology model to represent aquaponics 4.0 system's knowledge. *Information Processing in Agriculture*, 9(4): 514–532, 2022.

[12] B. Qi, A. Costin, and M. Razkenari. An ontology for manufacturability and constructability of prefabricated component. In *Proceedings of the International Conference on Computing in Civil Engineering*, pages 745–752, 2021.

[13] S. An, P. Martinez, R. Ahmad, and M. Al-Hussein. Ontology-based knowledge modeling for frame assemblies manufacturing. In *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, pages 709–715, 2019

[14] A. H. Gourabpasi and M. Nik-Bakht. An ontology for automated fault detection & diagnostics of HVAC using BIM and machine learning concepts. *Science and Technology for the Built Environment*, 30(8):972–988, 2024.

Less is More: Reinforcement Learning-powered Training Data Quality Evaluation and Sampling for Enhancing Computer Vision in Construction

Ziqing Wang[1] and Jinwoo Kim[2]

[1]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
[2]Department of Civil and Environmental Engineering, Hanyang University, Republic of Korea
ziqing003@e.ntu.edu.sg, jinwookim@hanyang.ac.kr

Abstract

Deep learning-based computer vision in construction has often assumed that more training data leads to better performance. This quantity-driven approach overlooks the detrimental effects of redundant, noisy, or mislabeled samples that can degrade model accuracy while inflating computational costs. To overcome this limitation, we propose and test an alternative hypothesis: selectively removing low-quality training samples can outperform indiscriminate use of all data. Specifically, we present a reinforcement learning (RL)-powered data quality evaluation framework that scores individual training images and prioritizes the most beneficial samples for model learning. Within a unified environment, the RL agent (referred to as the RL-evaluator) interacts with an object detector by probabilistically selecting subsets of high-quality samples for model training and using validation performance as a reward signal to optimize its selection policy. Experiments on a construction benchmark dataset of varying sizes show that removing low-quality samples identified by the RL-evaluator consistently improves detection performance compared to training on the full dataset, while also reducing computational overhead. These findings demonstrate the proposed method as a scalable and robust way for enhancing computer vision in construction.

Keywords –

Construction; Computer Vision; Reinforcement Learning; Data Quantity; Data Quality

1 Introduction

Deep learning-based computer vision has emerged as a key driver in the construction industry's shift toward digital project delivery and robotic automation. Instead of relying on manual site monitoring, which is often inconsistent and labor-intensive, automated computer vision systems can systematically interpret large volumes of visual data with greater speed and reliability. Such capabilities make it possible to detect and track resources [1], analyze ongoing activities [2], and identify operational irregularities [3], thereby supporting practical needs such as safety management [4], progress monitoring [5], and efficiency optimization [6]. Computer vision also provides the perceptual input required for robotics, equipping autonomous machines with spatial awareness and adaptability in unstructured environments. This connection between visual perception and autonomous operation further contributes to the development of digital twins, which integrate physical processes with virtual models. Taken collectively, these advances signify not only incremental improvements in construction monitoring and control but also a redefinition of construction practice toward intelligent, automated, and fully digitalized production systems.

Although computer vision in construction has advanced rapidly, much of the progress is based on the widely accepted hypothesis that larger datasets lead to better model performance [7]. This reliance on sheer data quantity has led to massive image collections covering a wide range of objects, activities, and site conditions, thereby enabling state-of-the-art benchmarks. However, model performance depends not only on dataset size but also on the quality of individual training samples. Redundant, noisy, and mislabeled data can introduce bias, slow or even prevent convergence, and even reduce accuracy. Meanwhile, data redundancy imposes heavy computational and storage demands. We therefore advance an alternative hypothesis: excluding some low-quality or harmful data may lead to superior results compared with using the full dataset. Despite its intuitive appeal, this hypothesis has rarely been systematically tested for computer vision in construction. Existing methods—such as leave-one-out and Shapley-based evaluation—are computationally prohibitive for high-

dimensional visual tasks, while active learning techniques typically rely on uncertainty measures, which do not directly reflect how each sample influences the learning dynamics or final model accuracy. As a result, the lack of scalable and reliable data quality evaluation strategies continues to constrain further improvements in the field.

To investigate our hypothesis, we present an RL-powered framework that evaluates the quality of individual training images and prioritizes those most valuable for model learning, while discarding samples deemed harmful or redundant. In this setup, the RL agent—referred to as the RL-evaluator and a computer vision model are embedded in an RL training environment. The RL-evaluator assigns quality scores (potential contributions for shaping model performance) to individual images and stochastically selects subsets of high-quality samples, which are then used to update the computer vision model. The model performance on a validation dataset is then used to provide the reward signal for refining the evaluator's policy. Through these repeated interactions, the RL-evaluator progressively learns a policy that guides sample selection toward improved model accuracy and data efficiency. We validate this framework on the Moving Objects in Construction Sites (MOCS) [7] dataset of different scales. We design and compare model performance using the following data curation strategies: (i) full data, (ii) low-quality removal, (iii) high-quality removal, and (iv) random removal. As an initial attempt, our experiments focus on worker detection, a fundamental and well-represented task in construction, allowing detailed investigation of the proposed framework.

This work contributes to advancing computer vision in construction in three aspects. First, we present an RL-powered framework for object detection tasks, where the training of a detection model is jointly integrated with the RL evaluator to enable fine-grained assessment of individual training samples. Second, we challenge the prevailing belief that larger datasets yield better model performance and instead demonstrate that removing low-quality samples identified by the RL evaluator can improve detection accuracy while reducing computational cost. Third, we show that the framework achieves consistent and reliable data quality evaluation under different construction benchmark dataset sizes, highlighting its robustness and adaptability. Collectively, these contributions lay the groundwork for more data-efficient and effective computer vision model deployment in dynamic and complex construction environments.

2 Data Quality Evaluation in the AI and Construction Domains

While extensive efforts have focused on enlarging datasets to enhance computer vision model performance, such quantity-oriented strategies often overlook the quality of individual samples, leading to redundant or noisy data that undermines model efficacy and efficiency. Consequently, there is a pressing need to shift focus from mere volume to data quality. However, unlike evaluating an entire dataset—which is straightforward via aggregate metrics—assessing the actual contribution of a single training sample remains a complex challenge.

One foundational strategy for isolating the specific contribution of a training sample is the Leave-One-Out (LOO) method. Conceptually, LOO operates on a perturbation principle: it quantifies the value of a datum by excluding it from the training set, retraining the model from scratch, and measuring the resultant shift in test performance. A notable degradation in accuracy signals that the omitted sample contained unique and valuable information. However, since the computational cost of literal retraining for every sample is prohibitive, Koh and Liang [8] addressed this bottleneck by proposing the use of influence functions. By differentiating through the model's training parameters, they derived a closed-form approximation that estimates how upweighting or perturbing a specific point would affect model parameters and loss, all without the need for actual retraining. This technique has proven instrumental for diagnostic applications, such as debugging domain mismatches, explaining model behavior, and detecting mislabeled examples. Yet, despite its rigorous theoretical basis, the computational complexity of calculating Hessian-vector products remains a barrier for large-scale datasets and deep architectures [9]. Furthermore, LOO suffers from a structural deficiency known as the "masking effect": in the presence of redundant or correlated data, it tends to assign near-zero value to individual samples because removing a single instance rarely impacts global performance, thus failing to recognize their collective importance [10].

To capture these complex interactions between data points, a parallel line of research leverages the Shapley value, a solution concept originally derived from cooperative game theory [11]. While originally utilized for interpreting supervised models and performing feature attribution, this framework has been systematically extended to the realm of data valuation. Spearheading this adaptation, Ghorbani and Zou [12] modeled the training process as a cooperative game where data points act as players working together to maximize model performance. To overcome the combinatorial explosion inherent in exact Shapley value calculations, they introduced a truncated Monte Carlo sampling strategy (TMC-Shapley). This stochastic

approximation method renders the evaluation computationally feasible by estimating marginal contributions over a subset of permutations rather than the full space. However, even with such algorithmic optimizations, the computational burden remains substantial, particularly when scaling to massive computer vision datasets [13]. Moreover, a critical limitation of both Shapley-based and influence function approaches is their static nature; they typically evaluate sample importance as an offline process completely decoupled from the active training loop. This separation prevents the joint optimization of data valuation and model learning, which may lead to value estimates that are suboptimal for maximizing downstream task performance [9].

Beyond traditional greedy search methods, recent research has explored more adaptive strategies, most notably Active Learning. This paradigm prioritizes selecting the most informative unlabeled samples for annotation, typically guided by uncertainty or entropy metrics. In the construction domain, such techniques have proven effective for minimizing labeling effort; for instance, Kim et al. [14] developed a deep active learning framework that significantly reduced the data requirements for object detection by automatically querying meaningful images. However, active learning relies heavily on heuristics that often lack robustness in complex visual conditions [15]. More critically, its primary objective is to acquire labels for newly collected data rather than to retrospectively identify and remove harmful samples from existing datasets. Consequently, it falls short in unlocking the full potential of models trained on noisy or redundant data [16].

To bridge these gaps, Reinforcement Learning (RL) has emerged as a promising avenue, offering the ability to model fine-grained data valuation through dynamic trial-and-error interactions. Yoon et al. [17] pioneered this direction with Data Valuation using Reinforcement Learning (DVRL), a framework that employs the REINFORCE algorithm to learn probabilistic weights for training samples. Unlike LOO and Shapley value, DVRL's computational complexity does not scale directly with dataset size, making it highly scalable for large-scale applications. Building on this, Batanero et al. [9] introduced RLBoost, a trajectory-free approach leveraging Proximal Policy Optimization (PPO). By integrating the valuation process directly into the supervised training loop, RLBoost achieves superior evaluation performance compared to both traditional methods and DVRL, demonstrating the potential of RL for joint data-model optimization.

Despite these advancements, a critical methodological gap remains: existing RL-based data quality evaluation methods [9,17] are predominantly limited to simpler tasks such as classification or regression. Extending these frameworks to object detection is not a trivial adaptation; unlike prior methods that rely on simple classification accuracy, it requires redefining the RL reward formulation and state representations to handle significantly higher-dimensional outputs, intricate architectures, and heightened sensitivity to visual noise. This challenge is particularly acute in the construction domain, where data is inherently diverse and prone to redundancy. Moreover, the robustness of these methods under varying dataset scales and diverse visual contexts has not been rigorously examined, leaving their scalability and generalization capabilities unverified. To bridge these gaps, we propose an RL-powered data quality evaluation and sampling framework explicitly designed for object detection within construction environments. We systematically investigate its effectiveness on a representative benchmark dataset of differing sizes and characteristics. Uniquely, our approach involves the joint optimization of an RL-evaluator and a visual detector within a unified RL environment. We then conduct controlled data curation experiments to benchmark the detection performance of models trained on these selectively curated data subsets.

3 RL-powered Data Quality Evaluation Framework

3.1 Problem Definition

The overall workflow of the proposed framework is illustrated in Figure 1. It integrates two learnable components: (i) an object detector $f(;,\omega)$ and (ii) the RL-evaluator $h(;,\theta)$, trained within a unified environment. The RL-evaluator interacts with the detector by a closed feedback loop, generating quality scores for training samples, selecting subsets for model updates, and refining its policy based on the reward signal derived from the model's validation performance. Through this iterative feedback, the RL-evaluator gradually refines its scoring policy to emphasize informative samples and down-weight noisy or harmful ones, thereby enhancing the detector's performance.

Formally, we denote the training dataset as $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i is an input image and y_i is its corresponding annotation set. Each annotation consists of n_i object instances, i.e., $y_i = \{b_i^{(1)}, b_i^{(2)}, \dots, b_i^{(n_i)}\}$. Each object instance $b_i^{(j)}$ is represented as a 5-dimensional vector (c, c_x, c_y, w, h), where $c \in \{1, \dots, C\}$ is the class ID and $(c_x, c_y, w, h) \in [0,1]^4$ are the normalized bounding-box coordinates. Model performance is finally evaluated on a test dataset $D^t = \{(x_i^t, y_i^t)\}_{i=1}^{M}$, while a validation dataset $D^v = \{(x_i^v, y_i^v)\}_{i=1}^{L}$, assumed to share the same distribution as the test set, is used both to monitor performance during training and to provide

feedback to the RL-evaluator.

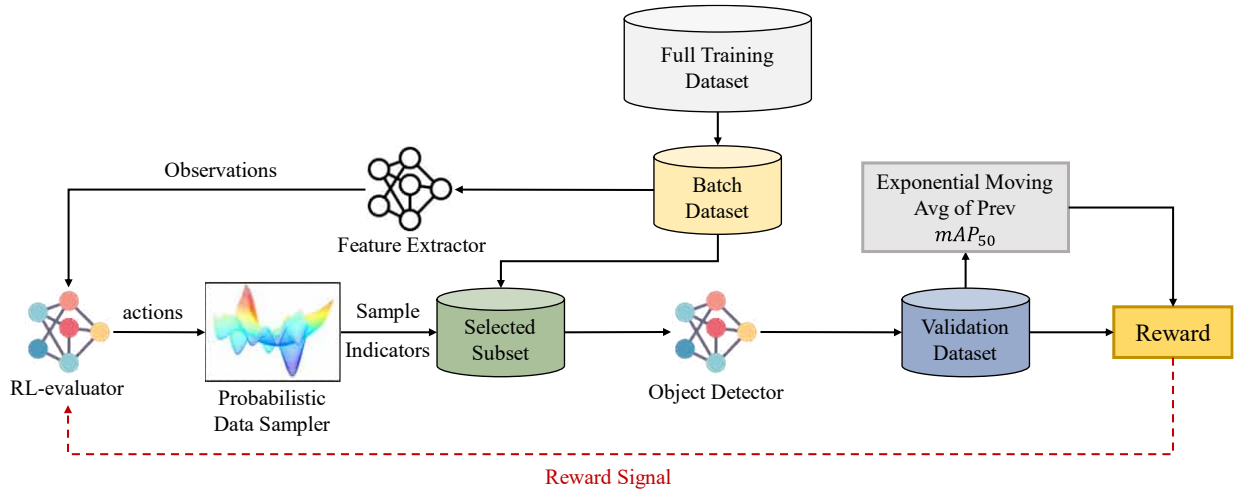


Figure 1. The overall workflow of the RL-powered data quality evaluation framework

In each training epoch, a mini batch $D^b = \{(x_i^b, y_i^b)\}_{i=1}^{B_s} \subset D$ is randomly sampled and passed to the RL-evaluator. The RL-evaluator generates quality score (i.e., selection probability) for each image, estimating its contribution to improving detection performance. A subset $S = \{(x_i^s, y_i^s)\}_{i=1}^{n} \subset D^b$ is then drawn according to these probabilities and used to update the object detector $f(;,\omega)$. The updated model is evaluated on the validation dataset D^v, and the performance is converted into a reward signal that guides the update of the evaluator's policy $h(;,\theta)$.

3.2 RL Formulation

The data quality evaluation can be formulated as Markov Decision Process (MDP), where the RL-evaluator interacts with an environment comprising the detector $f(;,\omega)$, the training data, and the reward derivation process. A standard MDP $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, T)$ involves observing a state $s \in \mathcal{S}$, taking an action $a \in \mathcal{A}$ according to the policy $\pi_\theta(a|s)$, and receiving a reward signal $\mathcal{R}$. In this study, we adopt a contextual bandit setting [18], where the evaluator assigns selection probabilities to samples within each mini-batch and receives immediate rewards from validation performance. As a result, the state transition $\mathcal{P}$, time horizon T, and discount factor γ are not relevant and are excluded from the formulation. In the following, we detail the definitions of state, action, and reward under this bandit MDP setting.

3.2.1 State

Each sample state s is its image-level feature embedding extracted via a ResNet50 [19] pretrained on ImageNet. Using features instead of raw pixels reduces input dimensionality, improves efficiency, and provides semantically rich, task-agnostic representations that enhance generalizability.

3.2.2 Action

For a sample (x_i, y_i) with state s_i, the evaluator outputs a probability $p_i = h_\theta(s_i) \in [0,1]$, which is then passed to a probabilistic data sampler characterized by a Bernoulli distribution to decide selection z_i : $z_i \sim Bernoulli(p_i)$, where $z_i \in \{0,1\}$. Here, $z_i = 1$ means the sample is chosen for training the detector, while $z_i = 0$ excludes it. Since this action directly determines the sampling behavior, the probability p_i serves as a proxy for data quality.

3.2.3 Reward

The reward signal measures the benefit of selected samples by evaluating detection accuracy on the validation set. We use mAP_{50} as the performance metric, and define the reward at iteration t as the difference between current mAP_{50} and the exponential moving average (EMA) of past values:

$$\mathcal{R}_t = mAP_{50_t} - EMA_{t-1} \quad (1)$$

$$EMA_t = \alpha \cdot mAP_{50_t} + (1-\alpha) \cdot EMA_{t-1} \quad (2)$$

where $\alpha \in (0,1)$ is the smoothing parameter for EMA.

3.3 Network Structures: Object Detector and RL-evaluator

The detector $f(;,\omega)$ can be any learnable object detection architecture. Specifically, $f(;,\omega): x_i^s \rightarrow y_i^s$ is

trained by minimizing the detection loss $\mathcal{L}_{det}$ on the selected subset S:

$$\omega \leftarrow \underset{\omega}{argmin} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_{det}(f(x_i^s; \omega), y_i^s) \quad (3)$$

The RL-evaluator $h(;,\theta)$ is optimized using the PPO algorithm [20], which generally outperforms standard Policy Gradient methods such as REINFORCE. PPO offers two key advantages: (i) a clipping mechanism to prevent excessively large policy updates and (ii) advantage estimation to evaluate the relative benefit of actions. The evaluator follows an actor–critic paradigm: the actor $\pi_\theta(a|s)$ models the sampling policy, while the critic $V_\emptyset(s)$ estimates the expected return, yielding the advantage $A = \mathcal{R} - V_\emptyset(s)$. The actor is trained using the clipped surrogate loss:

$$\mathcal{L}^{clip}(\theta) = \mathbb{E}_{(s,a)}[min(r_\theta(a|s)A, clip(r_\theta(a|s), 1 - \epsilon, 1 + \epsilon)A)] \quad (4)$$

where $r_\theta = \frac{\pi_\theta(a|s)}{\pi_{\theta old}(a|s)}$ is the probability ratio between new and old policies, and ϵ is a clipping hyperparameter that bounds the update within $[1-\epsilon, 1+\epsilon]$.

The critic network is optimized by minimizing the squared value loss:

$$\mathcal{L}^v(\emptyset) = \|\mathcal{R} - V_\emptyset(s)\|^2 \quad (5)$$

Additionally, an entropy regularization term is introduced to encourage exploration:

$$S(\theta) = -\pi_\theta(a|s) log \pi_\theta(a|s) \quad (6)$$

The final PPO objective integrates all three components into a joint loss:

$$\mathcal{L}_{PPO}(\emptyset, \theta) = \mathcal{L}^{clip}(\theta) - c_1 \mathcal{L}^v(\emptyset) + c_2 S(\theta) \quad (7)$$

where c_1 and c_2 are hyperparameters that balance the critic and entropy terms.

This formulation enables joint optimization of the detector and RL-evaluator within the proposed framework, ensuring tight coupling between data selection and task performance.

4 Experiment, Results and Discussion

We conducted an experiment to address three knowledge gaps: (i) whether removing low-quality images can improve model performance compared to using the full dataset; (ii) whether RL-powered evaluation can extend to complex tasks like object detection in construction; and (iii) how robust such methods are under varying dataset sizes.

We focused on detecting human workers, given their prevalence and importance in construction safety and productivity monitoring. Experiments were performed on the MOCS dataset [7], a publicly available benchmark which contains worker annotations collected from diverse construction sites and is well suited for evaluating robustness under varying conditions. We first filtered images containing worker instances and constructed three subsets of different sizes (500, 5,000, and 10,000 instances) to represent small, medium, and large dataset sizes. Each subset was split into training and validation sets at a ratio of 8:2. In addition, an independent test set of 5,000 worker instances was built for consistent evaluation.

YOLO11 [21] was used as the detector, chosen for its balance of accuracy and efficiency. Batch sizes were scaled with dataset size—4, 8, and 16 images for the small (500 instances), medium (5,000 instances), and large (10,000 instances) sizes. Each epoch involved 30 inner updates to provide sufficient learning while ensuring timely feedback to the RL-evaluator. Default hyperparameters were used with a learning rate of 0.01.

The RL-evaluator was trained with PPO using a two-layer actor–critic network (64 units, ReLU activations). The actor's output was initialized at mean 0.5 with standard deviation $log(0.2)$. PPO used Adam with a learning rate of 1×10^{-4}, clipping ratio $\epsilon = 0.1$, and an EMA smoothing factor $\alpha = 0.05$. Regarding hyperparameter sensitivity, empirical tests indicated that a tight clipping ratio and a low EMA factor were crucial for maintaining stable convergence. Mini batch sizes B_s were 100, 200, and 400 for the small, medium, and large datasets, respectively. PPO training was run for 1,000 epochs, while all other hyperparameters followed standard defaults. The joint training of the RL-evaluator required approximately 48 hours on a single NVIDIA RTX 4090 GPU. To account for the stochastic nature of RL training, experiments were conducted across three independent runs for each dataset size.

The RL-evaluator's output probabilities $p_i = h_\theta(s_i) \in [0,1]$ were used as image quality scores. After sorting image qualities based on these scores, four data utilization strategies were compared:

1. Full data: training on the entire dataset as a baseline.
2. Low-quality removal: progressively removing the lowest-scoring 5%, 10%, … of samples.
3. High-quality removal: removing the top-scoring samples at the same rates.
4. Random removal: removing the same proportions randomly.

Model performance was measured by average precision (AP) at IoU 0.5 (AP_{50}) and IoU 0.5–0.95 (AP_{50-95}), reflecting both detection accuracy and localization quality.

We first report the validation results on the MOCS dataset to verify whether the RL-evaluator can reliably distinguish data quality during training. As shown in

Figure 2, removing a small proportion of the lowest-quality samples consistently improved performance compared to training with the full dataset, whereas removing random samples led to gradual performance degradation. Moreover, progressively removing high-quality samples resulted in the greatest performance reduction. These trends confirm that the RL-evaluator learns an effective policy for differentiating useful from harmful samples.

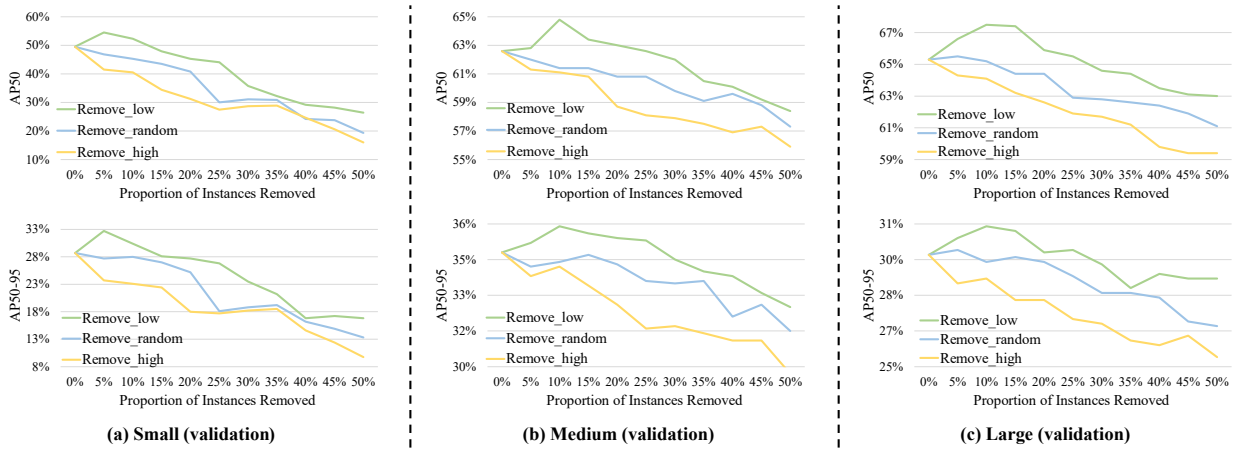


Figure 2. Validation results on MOCS dataset across different dataset sizes

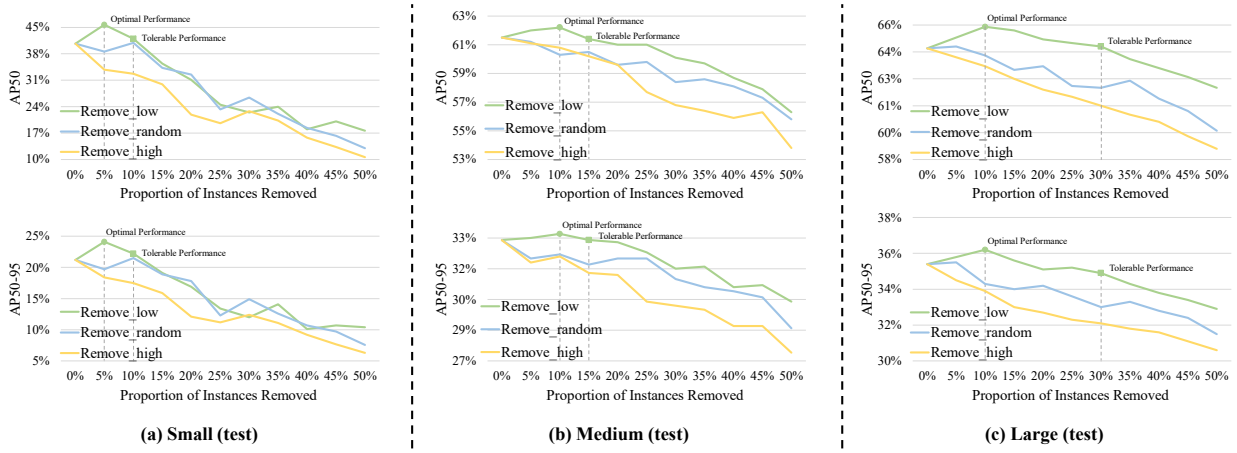


Figure 3. Test results on MOCS dataset across different dataset sizes

We then evaluate our framework on the test dataset. As summarized in Figure 3, the test outcomes are consistent with the validation results. For example, in the small dataset setting, training on the full dataset achieved an AP_{50} of 40.7%. After removing 5% of the lowest-quality samples identified by the RL-evaluator, detector performance improved to 45.7%. In contrast, random removal of the same proportion reduced accuracy to 38.6%, while discarding the top 5% of high-quality samples led to the most severe drop (–6.9%). Similar patterns were also evident in the medium and large dataset settings. Moreover, the benefits were not limited to AP_{50} along: consistent gains were also observed in AP_{50-95}, indicating that the learned data quality scores capture both overall accuracy and localization precision, showing the framework's reliability even with a coarse reward metric. These findings first confirm the conventional belief that "more data leads to better performance," as randomly discarding samples consistently degraded model accuracy. However, this principle does not hold universally. The results validate our alternative hypothesis: selectively removing low-quality samples can in fact yield higher performance than using the full dataset. This counterintuitive outcome underscores the heterogeneous nature of real-world training data—where mislabeled, redundant, or noisy samples can obscure gradient signals and impede learning. By removing such harmful data, the training process becomes clearer and more focused, enabling the detection model to leverage high-quality data more

effectively. Conversely, discarding high-quality samples deprives the model of essential visual and contextual information, resulting in sharp performance degradation.

We further observe that the proportion of removable low-quality samples varies with dataset size. In general, both the optimal removal rate—at which performance peaks—and the tolerable removal rate—at which performance remains comparable to the full dataset—increase as dataset scale grows. For example, on the large MOCS dataset, training with the full data yielded 64.2% AP_{50} , while discarding 10% of the lowest-quality samples achieved optimal performance to 65.4%. Moreover, performance remained close to full data baseline even after removing up to 30% of such samples. By contrast, the optimal removal rates for the small and medium sets were 5% and 10%, with tolerable rates of 10% and 15%, respectively. These results highlight a consistent trend: larger datasets allow more aggressive pruning without degrading performance, demonstrating that our framework improves data utilization efficiency by enabling superior accuracy with fewer, higher-quality samples.

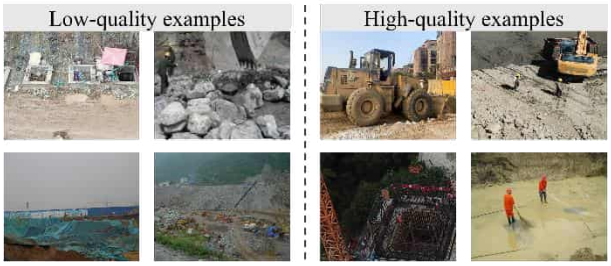


Figure 4. Visual examples of low- and high-quality images in MOCS

Figure 4 shows examples of low- and high-quality images identified by the RL-evaluator on MOCS. It can be observed that low-quality samples often suffered from blur, low resolution, or small and occluded worker instances that provided limited learning signals. In contrast, high-quality samples generally contained clearly visible workers, meaningful context, and diverse construction scenes representative of real conditions. Nevertheless, some exceptions were also observed—certain visually acceptable images contributed little to model performance due to high information redundancy, while some visually challenging ones (e.g., with blur or occlusion) were still valuable as informative hard examples that improve model generalization. These results suggest that the RL-evaluator captures not only superficial visual clarity but also deeper aspects driving the learning dynamics—such as semantic relevance, contextual informativeness (e.g., rare poses or diverse backgrounds), and annotation reliability, offering a more reliable alternative to heuristic or manual filtering.

In summary, the proposed RL-powered framework demonstrates strong potential for improving computer vision in construction by enhancing both detection accuracy and data efficiency. Across different dataset scales, it consistently boosted performance by identifying and removing low-quality samples, validating our alternative hypothesis that low-quality data removing can outperform training on the full dataset. Moreover, the framework enables more efficient training data utilization. By discarding a substantial amount of low-quality data, the model still achieved equal or even superior accuracy compared to using the entire dataset.

5 Conclusion and Future Work

This study presents an RL-powered data quality evaluation framework for computer vision in construction. Experiments on the MOCS dataset across different scales demonstrated that removing a small proportion of the lowest-quality samples—identified by the RL-evaluator—consistently outperformed training on the entire dataset. Furthermore, even after discarding a substantial amount of low-quality data, the detector still achieved competent or even better performance relative to training on the full dataset, thereby significantly reducing computational and data storage overhead. These findings challenge the common belief that more training data leads to better performance and instead highlight the crucial role of data quality and optimizing data utilization to enhancing computer vision model training.

Looking ahead, several research directions remain open. Firstly, our work focused on single-class worker detection; extending the framework to multi-class detection and more complex tasks such as segmentation or pose estimation would help assess its generalizability. Moreover, efforts could also validate the robustness of RL-powered data quality evaluation and broaden its applicability beyond construction vision to other data-intensive AI domains.

6 Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (RS-2026-25524415).

References

[1] R. Cai, Z. Guo, X. Chen, J. Li, Y. Tan, J. Tang, Automatic identification of integrated construction elements using open-set object detection based on image and text modality fusion, Advanced Engineering Informatics 64 (2025) 103075. https://doi.org/10.1016/j.aei.2024.103075.

[2] C. Chen, B. Xiao, Y. Zhang, Z. Zhu, Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition, Automation in Construction 146 (2023) 104702. https://doi.org/10.1016/j.autcon.2022.104702.

[3] M. Park, A.S. Kulinan, D.Q. Tran, J. Bak, S. Park, Preventing falls from floor openings using quadrilateral detection and construction worker pose-estimation, Automation in Construction 165 (2024) 105536. https://doi.org/10.1016/j.autcon.2024.105536.

[4] A.S. Kulinan, M. Park, P.P.W. Aung, G. Cha, S. Park, Advancing construction site workforce safety monitoring through BIM and computer vision integration, Automation in Construction 158 (2024) 105227. https://doi.org/10.1016/j.autcon.2023.105227.

[5] B. Ekanayake, J.K.W. Wong, A.A.F. Fini, P. Smith, V. Thengane, Deep learning-based computer vision in project management: Automating indoor construction progress monitoring, Project Leadership and Society 5 (2024) 100149. https://doi.org/10.1016/j.plas.2024.100149.

[6] J. Kim, J. Hwang, I. Jeong, S. Chi, J. Seo, J. Kim, Generalized vision-based framework for construction productivity analysis using a standard classification system, Automation in Construction 165 (2024) 105504. https://doi.org/10.1016/j.autcon.2024.105504.

[7] A. Xuehui, Z. Li, L. Zuguang, W. Chengzhi, L. Pengfei, L. Zhiwei, Dataset and benchmark for detecting moving objects in construction sites, Automation in Construction 122 (2021) 103482. https://doi.org/10.1016/j.autcon.2020.103482.

[8] P.W. Koh, P. Liang, Understanding Black-box Predictions via Influence Functions, (2020). https://doi.org/10.48550/arXiv.1703.04730.

[9] E.A. Batanero, Á.F. Pascual, Á.B. Jiménez, RLBoost: Boosting supervised models using deep reinforcement learning, Neurocomputing 618 (2025) 128815. https://doi.org/10.1016/j.neucom.2024.128815.

[10] Y. Kwon, J. Zou, Beta Shapley: a Unified and Noise-reduced Data Valuation Framework for Machine Learning, (2022). https://doi.org/10.48550/arXiv.2110.14049.

[11] L.S. Shapley, A value for n-person games, (1953). https://www.torrossa.com/gs/resourceProxy?an=5641636&publisher=FZO137#page=87 (accessed August 14, 2025).

[12] A. Ghorbani, J. Zou, Data Shapley: Equitable Valuation of Data for Machine Learning, in: Proceedings of the 36th International Conference on Machine Learning, PMLR, 2019: pp. 2242–2251. https://proceedings.mlr.press/v97/ghorbani19c.html (accessed August 14, 2025).

[13] R. Jia, D. Dao, B. Wang, F.A. Hubis, N. Hynes, N.M. Gürel, B. Li, C. Zhang, D. Song, C.J. Spanos, Towards Efficient Data Valuation Based on the Shapley Value, in: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, PMLR, 2019: pp. 1167–1176. https://proceedings.mlr.press/v89/jia19a.html (accessed September 3, 2025).

[14] J. Kim, J. Hwang, S. Chi, J. Seo, Towards database-free vision-based monitoring on construction sites: A deep active learning approach, Automation in Construction 120 (2020) 103376. https://doi.org/10.1016/j.autcon.2020.103376.

[15] B. Settles, Active Learning Literature Survey, University of Wisconsin-Madison Department of Computer Sciences, 2009. https://minds.wisconsin.edu/handle/1793/60660 (accessed September 3, 2025).

[16] A. Nguyen, J. Yosinski, J. Clune, Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images, in: 2015: pp. 427–436. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Nguyen_Deep_Neural_Networks_2015_CVPR_paper.html (accessed September 3, 2025).

[17] J. Yoon, S. Arik, T. Pfister, Data Valuation using Reinforcement Learning, in: Proceedings of the 37th International Conference on Machine Learning, PMLR, 2020: pp. 10842–10851. https://proceedings.mlr.press/v119/yoon20a.html (accessed August 16, 2025).

[18] G. Neu, A. Antos, A. György, C. Szepesvári, Online Markov Decision Processes under Bandit Feedback, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2010. https://proceedings.neurips.cc/paper/2010/hash/7bb060764a818184ebb1cc0d43d382aa-Abstract.html (accessed September 4, 2025).

[19] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016: pp. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html (accessed August 27, 2025).

[20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal Policy Optimization Algorithms, (2017). https://doi.org/10.48550/arXiv.1707.06347.

[21] R. Khanam, M. Hussain, YOLOv11: An Overview of the Key Architectural Enhancements, (2024). https://doi.org/10.48550/arXiv.2410.17725.

Automated tape-target centering based on template matching for construction measurement tasks using robotic total stations

Joshua Harrington[1], **Nolan W. Hayes**[1], **and Diana Hun**[1]

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

harringtonjd@ornl.gov, hayesnw@ornl.gov, hunde@ornl.gov

Abstract -

Robotic total stations have transformed surveying and construction mapping through precise, efficient, and automated measurements. These instruments integrate a theodolite, which measures horizontal and vertical angles, with an electronic distance measurement (EDM) unit, allowing accurate 3D measurement of observable points. Traditionally, users manually aimed the total station at a target. Recent advancements in robotic motor control and integration of cameras now enable automatic rotation and aiming, typically requiring only a single operator to position the target. Automated aiming is commonly performed using retroreflector prisms, which reflect light back to the source with minimal scattering, enabling high-precision measurements over long distances. However, retroreflectors, especially those designed for 360° visibility, can be expensive and impractical for certain applications. This paper presents an algorithm for automating the center detection of low-cost disposable tape targets using the total station's onboard camera and a template matching algorithm. The algorithm identifies the target's center pixel in the image, and the instrument is commanded to aim at that location. We evaluated the performance of this template-matching-based centering method under various distances, angles, lighting conditions, and field environments, comparing its results to both manual aiming and conventional automated aiming of tape targets. Manual aiming was used as the baseline operational reference in the absence of an independent ground-truth measurement. The proposed approach achieves target centering results comparable to the manual baseline. This method provides a cost-effective alternative for high-precision applications where retroreflector targets are constrained by budget or logistics.

Keywords -

total station, automation, computer vision, template matching

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

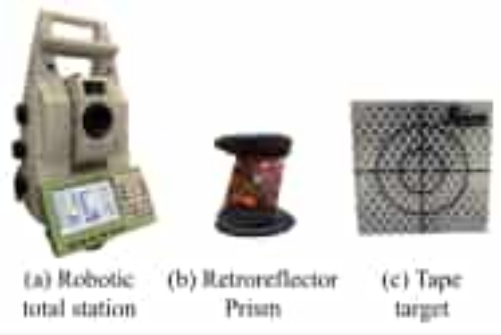


Figure 1. Surveying hardware and targets

1 Introduction

The integration of robotic image-assisted total stations into land surveying and construction mapping represents a significant advancement in geospatial measurement. These instruments have transformed traditional surveying practices by enabling precise, efficient, and largely automated data collection, thereby reducing labor requirements and improving measurement accuracy [1, 2].

Robotic total stations (Fig. 1A) combine three essential components: (1) a theodolite for measuring horizontal and vertical angles, (2) an electronic distance measurement (EDM) unit for determining distances to a target, and (3) an onboard computer that synthesizes angular and distance data to compute three-dimensional coordinates [1]. Traditionally, total stations required manual operation, with surveyors physically aiming the instrument at a target to acquire measurements, a labor-intensive process that often necessitated a two-person crew.

Recent advancements in robotic motor control and automation have enabled these instruments to autonomously rotate, track, and precisely aim at targets, significantly streamlining field operations [2]. This automation allows a single operator to conduct surveys by simply positioning a target, typically a retroreflective prism, while the instrument tracks the target and performs the measurements. Additionally, the integration of cameras onto robotic total stations has added another level of sensing for improved automation; however, the onboard cameras are historically

underutilized especially with the advent of advanced computer vision algorithms and image segmentation.

Retroreflective prisms (Fig. 1B) are essential to this system, as they reflect incident light directly back to the source with minimal angular deviation, enabling high-precision distance and angle measurements over long ranges. Among these, 360-degree prisms are particularly valued for their ability to reflect signals from any direction, enhancing flexibility and efficiency in dynamic field environments. However, despite their advantages, such prisms can be prohibitively expensive and may not be suitable for all applications, especially in cost-sensitive or temporary deployments.

As a lower-cost alternative, disposable tape targets (Fig. 1C) are often used. These targets are inexpensive, easy to deploy, and can be left in place without concern for recovery. However, current methods for automating the centering of the total station's telescope on these targets often result in reduced accuracy. Consequently, users must manually aim the telescope at the center of each target to ensure reliable measurements, reintroducing inefficiencies in labor and time.

Surveying with total stations to evaluate the placement accuracy (tolerances) of building components is common practice. For example, surveys crews often manually aim at installed wall panels to confirm they fall within the required tolerances. In some cases, particularly for challenging installations, a crew may even take measurements in-process and guide crane operations in real time to ensure components are positioned correctly. Recent research has investigated automated measurement systems that use a total station to guide crane placement of building components [3, 4]. In these systems, retroreflector targets are attached to building components enabling automated measurements of installation error in real time and providing guidance during placement. However, these retroreflector targets must be manually retrieved after installation. A more effective alternative is to enable automated, precise measurement of disposable targets. Such targets could be left in place, either falling off naturally over time or being removed later, eliminating the need for retrieval during installation.

This paper presents a novel workflow for automating the detection of and centering on low-cost disposable tape targets using the total station's onboard camera and a template matching algorithm. The proposed method identifies the center pixel of the target in the captured image and commands the instrument to aim accordingly. We evaluate the performance of this template-matching-based centering approach under a range of distances, angles, lighting conditions, and field scenarios, comparing its accuracy to both manual aiming and conventional automated methods for tape targets. This approach offers a cost-effective alternative for high-precision applications where the use of retroreflectors is limited by budget or logistical constraints.

1.1 Literature Review

Research focused on automating the precise measurement of tape targets remains limited. Schwieger, Kerekes, and Lerke explored methods for detecting object features using the onboard camera of a total station [1]. However, their approach still relied on active measurement using a retroreflective prism to determine precise object positions.

Zschiesche and Schlüter [5] investigated potential applications of machine learning and image segmentation in image-assisted total stations. Their work focused on identifying various types of targets, including reflectorless ones, within images captured by the various cameras of the total station. However, they did not explore methods to automate the precise centering of the telescope on these targets.

Lian et al. [6] developed a machine vision-based approach for automating the aiming process at tape targets using template matching. While promising, their method achieved peak accuracy of approximately 5 millimeters at distances of 70 meters and 131 meters.

The method proposed in this paper builds upon Lian et al.'s work, aiming to demonstrate improved accuracy at medium ranges up to 100 meters. Sub-3-millimeter accuracy is needed to meet tolerance requirements for typical building component installation such as precast concrete or steel structure [7, 8]. For these construction tasks, tolerances can be as low as 3 millimeters. The proposed method improves the template matching algorithm to achieve sub-3.3-mm accuracy up to 100 meters.

2 Methodology

2.1 Multi-Scale Template Matching

Automated aiming is performed using a multi-scale template matching pipeline consisting of (1) image capture, (2) green-channel extraction, (3) Gaussian blurring, (4) multi-scale template matching, and (5) telescope pointing.

The template was approximated in code rather than photographed under a single condition. Feature shapes, colors, and relative sizes were iteratively tuned using a diverse set of sample images to ensure robust matching across the range of conditions evaluated. Exact proportional replication was unnecessary for accurate matching since the apparent template shape varies with viewing angle.

Two template versions were developed. The first included only the center black cross on a white rectangular background. The second included concentric gray circles (intensity = 100) on a similar background. Feature sizes were visually tuned to match their appearance at distances between 10–130 m. Since the template shape distorts at

oblique angles, exact proportional replication was not required for reliable matching.

These templates, shown in Fig. 2 prior to preprocessing, were lightly blurred (Gaussian kernel size = 3, σ = 1.0). Testing with approximately fifteen representative images showed that matching consistently failed at long distances ($\geq$ 70 m) without template blurring.

High-resolution (2560x1920) telescopic color images were captured and transferred via USB to a remote laptop. The green channel was extracted to improve feature contrast, as most camera sensors are optimized for green light. This step improved template matching under low-light conditions and allowed a successful match at 100 m. The images were then lightly blurred with the same Gaussian parameters used for the templates, allowing the pipeline to correctly match more templates from the dataset at larger distances (> 70 m).

To enable matching across a range of distances, the templates were scaled over representative size ranges. For each scaled template, the OpenCV template matching algorithm [9] computed the best score and corresponding pixel location. This process was performed with the first template (cross-only) for 4 scales (1.0-1.3). If the maximum correlation score across the evaluated scales exceeded 0.7, the second template was skipped.

The correlation threshold (0.7) was empirically selected by evaluating a varied test set and choosing the lowest value that consistently produced correct matches across all test cases.

If the maximum correlation score from the first template did not exceed the threshold, the second template was applied across 18 scales (0.3-1.3). These scale ranges were determined by comparing template size to the apparent size of the target in the closest and furthest images (10–130 m for the circular template; 10–40 m for the cross template, which becomes too small to match reliably at longer distances).

Fig. 3 shows an example of the detected template and center pixel. The total station was then aligned with this pixel location using its built-in turn-to-pixel function.

Matching accuracy degraded quickly when using the circular template under oblique viewing angles, as the circles appear elliptical. In contrast, the smaller cross exhibited less geometric distortion at high angles, enabling successful matching at angles up to 50° with a maximum error of 0.3 mm.

2.2 Experiments

A series of experiments were conducted to evaluate the accuracy of the proposed centering algorithm compared to manual aiming and the existing automated centering method. All tests were performed using a consistent set of hardware and a standardized experiment methodology.

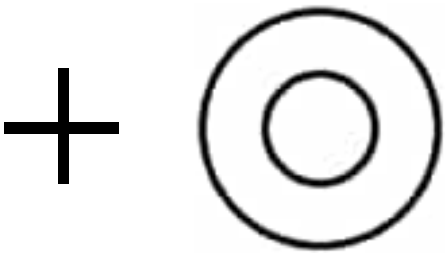


Figure 2. Illustration of templates

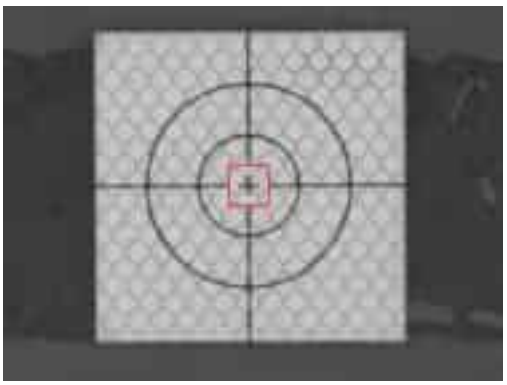


Figure 3. Template matching in telescope camera

2.2.1 Hardware

The experiments utilized a Leica MS60 robotic total station, operated via the GeoCOM protocol through a custom Python interface [10]. This unit features a 5-megapixel telescope camera.

Tape targets manufactured by Leica, measuring 60 mm by 60 mm, were used throughout the tests. A tape target was mounted on a mock brick wall assembly measuring 80 cm by 80 cm. Since several smaller tape targets were also mounted on the wall, covering these targets with pieces of paper was required to prevent either algorithm from finding different targets. This assembly was affixed to an adjustable arm configuration, allowing placement at various angles.

For indoor tests, a synthetic light source was used to control illumination, and a lux meter was employed to measure lighting conditions for each test case. The indoor tests were performed at a light level of 85 lux, and the outdoor tests were done on a partly sunny day with an average illuminance of 5000 lux.

2.2.2 Experiment Workflow

Experiment for Proof-of-Concept. The experimental workflow was designed to generate baseline comparisons and assess the accuracy of both the existing and proposed centering methods. Each test followed the steps below:

1. **Manual Aiming:** The total station was manually aligned with the center of the tape target. Five rapid distance measurements were collected and averaged to reduce EDM readout jitter.

2. **Existing Method:** The instrument was rotated to a random vertical and horizontal offset (within limits to ensure it remained visible). The existing method was executed and five distance measurements were averaged. Because of noticeable run-to-run variability, this procedure was repeated three times per setup (Sec. 4.1).

3. **Proposed Method:** A telescope image was captured, downloaded to the remote computer, and processed using the multi-scale template matching algorithm. The computed center pixel was used to re-aim the instrument using a builtin turn-to-pixel function. The algorithm was run only once per setup because little variability had been observed previously (Sec. 4.1). Finally, five distance measurements were averaged.

The existing automated centering method is proprietary, and its internal operation is undisclosed. Based on observable behavior, the method appears to use a flashing guide light, detecting and centering on the reflected shape of the tape target, similar to the procedure used for retroreflective prisms.

The experimental workflow was repeated across a predefined combination of viewing angles, distances, and lighting conditions. Table 1 summarizes the experimental setups. Set A varied horizontal viewing angle at a fixed 10 m distance (0°–60°), and Set B varied distance from 10–130 m under near-normal incidence. For each setup, telescope images were also recorded for qualitative comparison.

Template matching performance depends on both image resolution and the target distance. Lighting conditions also affect the matching accuracy. Some tests were performed indoors at shorter distances and under low light conditions, while longer range tests were done outdoors. The distance range of 10 - 130 m targeted medium-range applications (precise measurement of component locations with attached tape targets [3, 4]), while also identifying the upper slope distance limit given the available hardware.

Another factor influencing algorithm performance is the skew of the target within the image relative to the template. At extreme viewing angles, the algorithm may struggle to match the tape target accurately. To assess angular robustness, the tape target was rotated horizontally relative to its surface normal (0°), as illustrated in Fig. 4. The horizontal viewing angle, H, was varied from 0° to 60°. Because the tape target is symmetric, only one rotation direction was assessed.

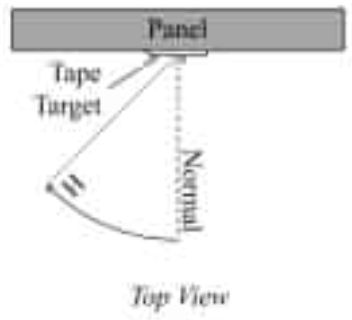


Figure 4. Illustration of horizontal viewing angle

Illumination was controlled to represent low light indoors (85 lux) and bright outdoor daylight (5000 lux). Higher illuminance levels are expected to enhance algorithm performance by improving image clarity and contrast.

Experiment for Statistical Analysis. A separate experiment evaluated run-to-run variability in the error across each algorithm. Manual centering was performed initially as before using Setup A1 (Table 1). The instrument was given a programmed random angular nudge within predefined limits before each run. The existing method was repeated 20 times and the average of 5 distance measurements was recorded for each run. The procedure was repeated for the proposed method to allow a direct comparison.

3 Results

The experimental results compare tape target centering performance across the three methods evaluated: manual aiming, the existing onboard centering method, and the proposed template matching centering method.

For each test, telescope camera images were captured at three stages: (1) after manual aiming, (2) after centering using the existing method, and (3) after centering using the proposed template matching method.

A representative subset of images from five tests is shown in Fig. 5. For each condition, the figure shows the following:

- **Manual Method**: crosshair placement after manual alignment.
- **Existing Method**: the median-error result from the three onboard centering attempts.
- **Proposed Method**: crosshair placement after executing the proposed method.

These visual comparisons illustrate how the proposed method tends to produce centering results that are closer to

Table 1. Experiment conditions and setups

Test No.	Conditions			
	Location	Target Distance (m)	Viewing Angle, H (°)	Illumination
Set A				
A1	Indoor	10	0	Low
A2	Indoor	10	10	Low
A3	Indoor	10	20	Low
A4	Indoor	10	30	Low
A5	Indoor	10	40	Low
A6	Indoor	10	50	Low
A7	Indoor	10	60	Low
Set B				
B1	Outdoor	10	0	High
B2	Outdoor	20	0	High
B3	Outdoor	30	0	High
B4	Outdoor	40	0	High
B5	Outdoor	70	0	High
B6	Outdoor	100	0	High
B7	Outdoor	130	0	High

the apparent target center under the conditions shown. A brief quantitative analysis of error variability is provided in Sec.4.1.

To verify that averaging five sequential readings sufficiently suppresses readout jitter, we examined within-run dispersion under representative conditions. For example, for Setup A4 (existing method, run 3), the five readings yielded a mean error of 0.03 mm with a standard deviation of 0.03 mm, indicating negligible short-term variability. Comparable levels were observed across other runs; therefore, we report per-run averages for all experiments.

Fig. 6 records the mean, minimum, and maximum error for the existing method and the corresponding error in the proposed method. As described in Sec. 4.1, the existing method exhibited substantial run-to-run variability, whereas the proposed method did not. Consequently, the proposed method was executed once per setup, with its design choice validated by the variability analysis.

Entries marked "Algorithm Failed" indicate conditions in which the method did not identify the target in any run (0% detection success). In all successful cases, the method identified the target in every run (100% success).

Manual aiming provided the baseline for computing errors. Since no ground-truth reference was available, these values represent only a relative reference. The errors reported for the existing and proposed methods are Euclidean distances between the method-measured positions and the corresponding manual measurement. These representative errors are reported in Fig. 6.

4 Discussion

The user manually aligned the crosshairs with the tape target center using the telescope view. As shown in Fig. 5, the manual method exhibited visually consistent target alignment. However, manual centering is inherently subjective and sensitive, particularly at long slope distances, where wind-induced vibrations, reduced image quality, and temperature gradients further decrease accuracy. A single user performed all manual aiming tasks to minimize variability in the baseline measurements. Although this introduces potential operator-specific bias, maintaining consistency in the manual reference measurements was prioritized.

Visual inspection of the existing method in Fig. 5 combined with the calculated errors in Fig. 6 indicates significant variability. Particularly at longer distances or angles, the onboard proprietary algorithm struggled to center on the target, frequently leading to variable and inaccurate results.

The largest mean error observed for the existing method was 8.5 mm, and the mean errors across all experimental setups had an overall average of 4.9 mm (see individual means in Fig. 6). These values exceed many common construction installation tolerances, some of which are as small as 3 mm [7, 8].

At target slope distances above approximately 30 m, the existing method fails to identify the target in all trials (0% detection or measurement success).

The proposed method demonstrated lower centering error than the existing method under all successful trials, achieving a maximum deviation of 3.3 mm from the baseline. Within 70 m, the deviation decreased to less than 1 mm. Although performance declines at extreme distances and angles, the algorithm consistently outperformed the existing method across all tested scenarios.

Although the primary goal of this investigation was to develop a method for accurate automated measurement of tape targets at intermediate distances (20 to 40 m), testing extended until failure occurred at a slope distance of 130

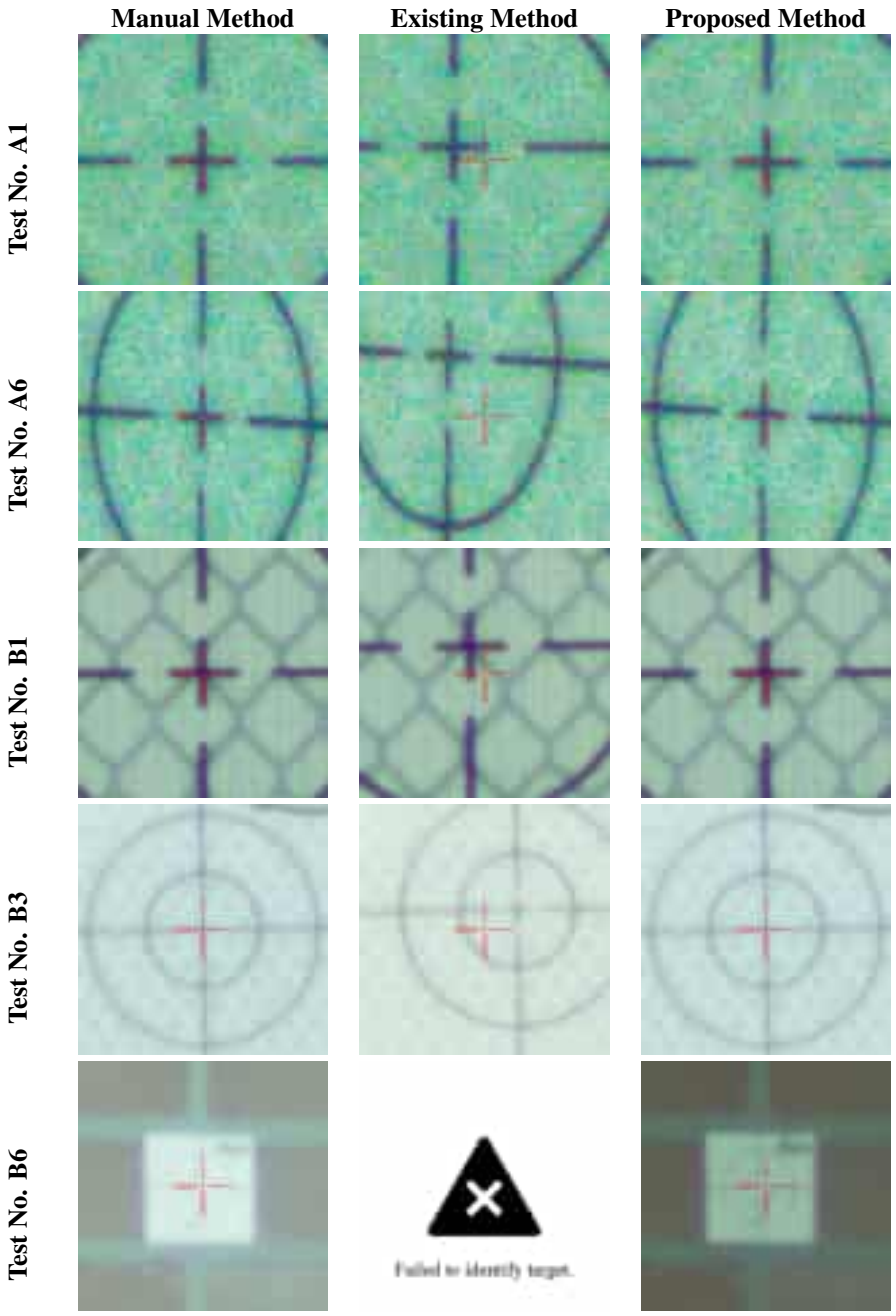


Figure 5. Representative telescope images of aiming methods for multiple test conditions

m, substantially beyond the onboard method's limit of 40 m. The cause of failure at this distance is not yet clear, as image statistics and background conditions were comparable to those in successful trials. Additional investigation will be needed to determine the specific factors governing long-range performance.

4.1 Statistical Analysis of Measurement Error

In addition to the proof-of-concept experiments, a statistical analysis was performed to quantify measurement variability and validate the experimental design. Two primary variability sources were considered for the proposed method: (1) small differences in the captured images and (2) the onboard turn-to-pixel function. Despite these factors, the proposed method exhibited low variability compared to the existing method (Fig. 7). In Setup A1, the

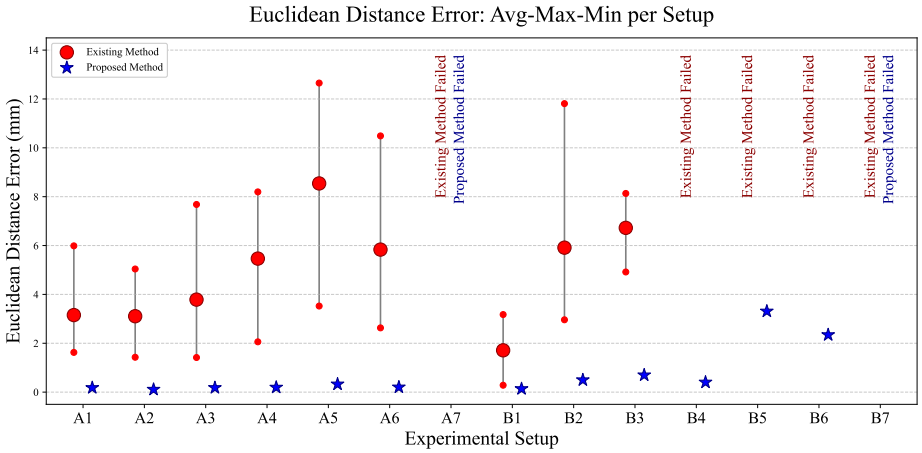


Figure 6. Avg-max-min Euclidean distance error per experimental setup.

existing method produced a mean error of 5.2 mm with a standard deviation of 3.1 mm (95% CI: ±1.5 mm), whereas the proposed method produced a mean error of 0.16 mm with a standard deviation of 0.04 mm (95% CI: ±0.02 mm).

4.2 Limitations

The proposed template-matching algorithm requires the full tape target to be visible within the telescope camera's field of view. During testing, the existing method provided this initial coarse alignment; therefore, the current implementation of the proposed algorithm is not yet fully autonomous. A fully autonomous workflow could incorporate image-segmentation techniques to coarsely align the target before precisely centering with template matching.

A second limitation is the 10–20 s required to download a telescope image from the instrument to the remote computer. This delay is specific to the instrument used in testing and dominates the cycle time, preventing real-time tracking of moving targets. The image-processing stage itself requires only about 0.5 s and could be further optimized for tracking. Eliminating the download time by real-time video streaming or running the proposed method directly onboard the instrument would allow the method to be competitive with manual operation times and track targets in real-time.

5 Conclusions

A template-matching algorithm was developed to automate the centering of a robotic total station on disposable tape targets. The method was evaluated across a range of distances, angles, and lighting conditions. Within the tested envelope, the algorithm achieved centering performance comparable to manual centering for viewing angles up to 50° and distances from 10-100 m. Since no independent ground-truth reference was available, all reported errors represent deviations from the manual baseline.

The current implementation is not fully autonomous, since the template must be positioned within the camera's field of view. Additionally, overall cycle time is dominated by instrument-specific image-transfer delays rather than by algorithmic processing. Eliminating this bottleneck with a live video stream or through onboard execution would allow real-time performance.

Quantitatively, the proposed algorithm increased the usable range from less than 40 m (existing method) to 10–100 m, and reduced the overall average centering error (relative to the manual baseline) from 4.9 mm to 0.7 mm. These results demonstrate that the proposed method provides a cost-effective and substantially more accurate alternative for high-precision applications where retroreflectors are impractical.

5.1 Future Research

Future research will extend the tape centering algorithm to dynamic scenarios, including real-time target tracking. This will require optimizing the algorithm's processing time, improving its robustness in challenging lighting and viewing conditions, and eliminating the image-transfer latency. Additional work will explore integration with higher-resolution imaging systems and the development of a fully autonomous pipeline that does not require coarse alignment with the onboard method.

6 Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Critical Minerals and En-

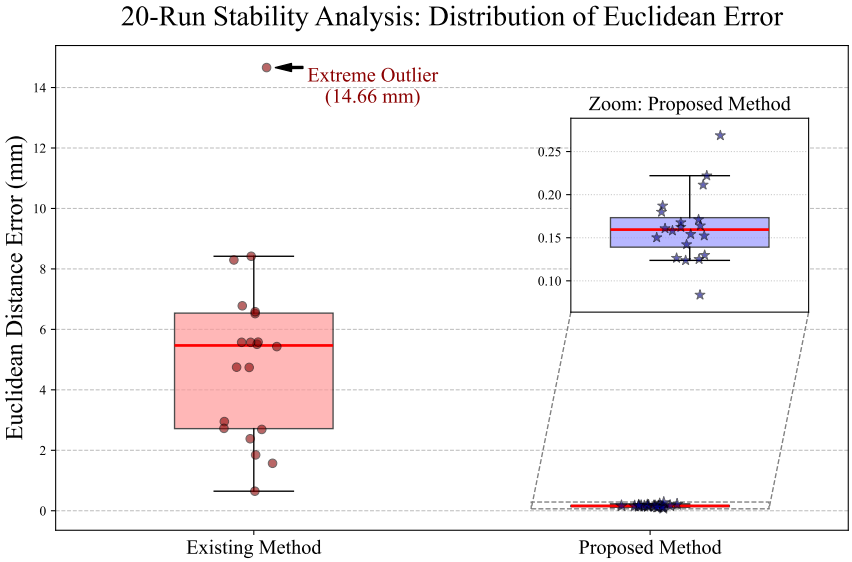


Figure 7. Box-and-whisker plot for 20-run statistical analysis of Setup A1.

ergy Innovation (CMEI), specifically the Building Technologies Office under contract DE-AC05-00OR22725.

The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

6.1 Notice of Copyright

This manuscript has been authored [optional: in part] by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the work for publication, acknowledges that the US government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the submitted manuscript version of this work, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://energy.gov/doe-public-access-plan).

6.2 Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Copilot in order to improve manuscript clarity, conciseness, and grammar. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published paper.

References

[1] Volker Schwieger, Gabriel Kerekes, and Otto Lerke. *Image-Based Target Detection and Tracking Using Image-Assisted Robotic Total Stations*, pages 133–169. Springer International Publishing, Cham, 2020. ISBN 978-3-030-22587-2. doi:10.1007/978-3-030-22587-2_5. URL https://doi.org/10.1007/978-3-030-22587-2_5.

[2] Fernando Reyes-Aviles, Thomas Gloor, and Clemens Arth. Automatic model-based stationing: Robust total station localization without known control points. *J. Intell. Robot. Syst.*, 111(1), March 2025.

[3] Nolan W. Hayes, Bryan P. Maldonado, Diana Hun, and Peter Wang. Real-time evaluator to optimize and automate crane installation of prefabricated components. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 192–199, Chennai, India, July 2023. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-0-4. doi:10.22260/ISARC2023/0028.

[4] Nolan W Hayes, Bryan P Maldonado, Mengjia Tang, Peter Wang, and Diana Hun. Optimization of prefabricated component installation using a real-time evaluator (RTE) connection locating system. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 65–72, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0010.

[5] Kira Zschiesche and Martin Schlüter. Classification and object detection with image assisted total station and machine learning. *Journal of Applied Geodesy*, 17(4):381–389, 2023. doi:doi:10.1515/jag-2023-0011. URL https://doi.org/10.1515/jag-2023-0011.

[6] Hou Lian, Zhao Long-fei, Zeng Wei, Yang Lei, and Xiao Changshi. Research on fully automatic template calibration method based on machine vision guidance. *China Rural Water and Hydropower*, pages 185–192, December 2024. doi:10.12396/znsd.232026.

[7] Precast/Prestressed Concrete Institute. PCI 135-25: Specification for Tolerances of Precast Concrete. Standard, 2025.

[8] American Institute of Steel Construction. Specification for structural steel buildings. AISC 360-22, 2022.

[9] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[10] MrClock8163. GeoComPy. https://github.com/MrClock8163/geocompy, 2025.

An integrated workflow based on UAS, GIS, and BIM for building envelope maintenance management

Caio Lima[1], Vanessa Pacheco[1], Alisson Silva[1], Rafael Sena[1], Danton Almeida[1], Marcos Lorenzo[1], Dayana Costa[1]

[1]Department of Structural and Construction Engineering, Federal University of Bahia, Salvador, Bahia, Brazil
caioasu@gmail.com, vanessa_cruz_pacheco@hotmail.com, so_alisson@hotmail.com, sena.rafael@ufba.br, dantonalmeida@ufba.br, marcos.lorenzo@ufba.br, dayanabcosta@ufba.br.

Abstract

The increasing volume and heterogeneity of data generated during building inspections pose significant challenges to maintenance management (MM). Although digital technologies are frequently discussed in the literature, their integrated application to support maintenance decision-making remains scarce. This study proposes an integrated data visualization workflow to support building MM by integrating information from inspections, UAS, BIM model data, and GIS geospatial information, and by leveraging analytical visualization in a BI (Business Intelligence) environment. A case study was conducted in a building of a Higher Education Institution (HEI) using real inspection data on facades and roofing systems. The results, presented through an interactive dashboard, demonstrate the potential of BI to integrate heterogeneous data sources and support the identification, location, and analysis of construction anomalies. Furthermore, the study highlights practical challenges in interoperability and data organization, and presents lessons learned that can support future applications of integrated digital environments for building maintenance management.

Keywords: Maintenance Management; BIM; GIS; UAS; Power BI; Dashboard; Business Intelligence.

Introduction

The increasing complexity of buildings has amplified the demand for more precise information management in maintenance processes [1,2]. In the work of Dompey et al. [3], it is noted that increases in operational costs and the growing volume of information to be managed highlight the limitations of traditional maintenance methods. This scenario becomes even more critical in exposed building envelope subsystems, such as roofs and facades, which are more susceptible to degradation and decline in building performance due to constant exposure to weather. Studies indicate that anomalies such as cracks, detachments, and surface wear serve as early indicators of failure and can compromise the building's performance, underscoring the importance of preventive strategies and periodic inspections [4, 5, 6, 7].

Sena et al. [2] identified that the main problems in building maintenance are related to the management and organization of processes, highlighting the absence of standardized procedures and a structured preventive maintenance plan. According to this study, there are also limitations in human, financial, and technological resources, as well as operational and informational deficiencies, which hinder efficient maintenance, access to up-to-date building data, and communication among the actors involved. In response to these demands, practices based on digital technologies have emerged to improve operational efficiency in building maintenance. The integration of BIM models, digital twins, and reality-capture systems has expanded monitoring and analysis capabilities, enabling greater precision and supporting informed decision-making [3,8]. Among these technologies, Unmanned Aerial Systems (UAS) have been employed for systematic inspections, providing a massive collection of visual data and supporting the construction of digital models [19]. However, expanding this volume of information reveals challenges in processing and organizing the collected data [7].

Therefore, recent studies highlight the role of Business Intelligence (BI) environments in structuring, integrating, and making large volumes of data accessible, thereby facilitating continuous, up-to-date analyses [9].

When combined with Geographic Information Systems (GIS), such platforms expand analytical capacity by associating spatial and non-spatial data. This feature has proven particularly relevant for projects involving multiple buildings distributed across different areas [10, 9].

Given these challenges, this study proposes an integrated workflow for visualizing building maintenance information that combines building envelope data from UAS-automated inspections, BIM models, and GIS in a Business Intelligence environment. Furthermore, this study aims to report the lessons learned during this integration process.

The novelty of this study lies in the structured integration of heterogeneous data sources into a unified, decision-oriented workflow. Unlike previous studies that focus on isolated integrations, such as BI–GIS [9] or BIM–GIS [10,7], this research proposes a reproducible data integration pipeline that connects data acquisition, organization, spatial contextualization, and analytical visualization within a single environment. This approach addresses the fragmentation of maintenance information by enabling the consolidation of inspection records, spatial data, and building components into an integrated and interactive system. As a result, it enhances the ability to interpret data and supports more informed decision-making in building maintenance management. This differentiates the proposed approach from prior studies that primarily focus on tool-level integration, rather than on structured workflows and decision-oriented data transformation.

This work contributes to the literature by: (i) proposing a structured workflow for integrating BIM, GIS, UAS, and BI data in building maintenance management; (ii) demonstrating the application of this workflow using real inspection data; and (iii) identifying practical challenges and lessons learned to support future implementations of integrated digital environments.

Application of digital technologies for Maintenance Management

Recent studies suggest that the adoption of digital technologies has enhanced the quality and reliability of information used in building maintenance management [2,6]. Thus, Building Information Modelling (BIM) serves as a central repository of information throughout the building lifecycle [11], while Business Intelligence (BI) tools, such as Power BI, are being explored for their potential for analysis, data integration, and decision-making support [1]. Recent studies investigate integrating Power BI with Python scripts, cloud services, and immersive technologies to support Facilities Management and the development of Digital Twin prototypes [12,13].

However, using BI tools alone has limitations for representing and analyzing spatial information, leading to their integration with Geographic Information Systems (GIS) to visualize dynamic, spatially referenced data [9]. Meanwhile, studies have examined the integration of BIM-GIS to support utilities management and the automated recording of inspections performed with UAS, emphasizing the technical feasibility of interoperability between these platforms [10,7]. During the operation and maintenance phases, BIM has served as a repository for inspection and maintenance data, with progress in data standardization through IFC and integration with UAS imagery and automated update systems [2,14,15].

Despite these advancements, the literature suggests that building maintenance management continues to face challenges related to information fragmentation and the lack of integrated data visualization and analysis workflows. In particular, there is a gap in interoperability between BIM models, GIS spatial data, and information from UAS inspections in Business Intelligence environments, especially during the operation and maintenance phases. Furthermore, many applications still rely on manual processes or ad hoc integrations, which limit the potential of these technologies to effectively support the location, documentation, and monitoring of anomalies. In this regard, the need for research focused on creating integrated, reproducible workflows that connect BIM, GIS, and BI to better manage building maintenance information is clear.

Research methodology

This study employed an exploratory case study strategy, which is suitable for investigations of complex and contemporary phenomena, where understanding data integration processes in their real-world application context is necessary [16]. The objective of the methodology was to establish a standardized workflow for collecting, preparing, integrating, and visualizing building maintenance data in a BI environment.

The research was conducted in two phases: a theoretical-practical investigation and a case study, including tool testing and the development of an integrated visualization environment (Figure 1).

Investigation of the theoretical and practical problem

The investigation of the problem was conducted through a literature review focused on identifying studies that address the integration of BI tools, GIS, BIM models, and UAS imagery. This stage aimed to understand how these technologies have been combined in previous research and to analyze the limits and challenges associated with the interoperability and visualization of building maintenance data. Meanwhile, secondary data were identified and analyzed, involving the collection and interpretation of records previously compiled by [6,2]. The goal was to confirm the nature, structure, and possible uses of this information. It is important to note that this initial analysis was not only about the problem's foundation but also served as the database for the later stages of the case study.

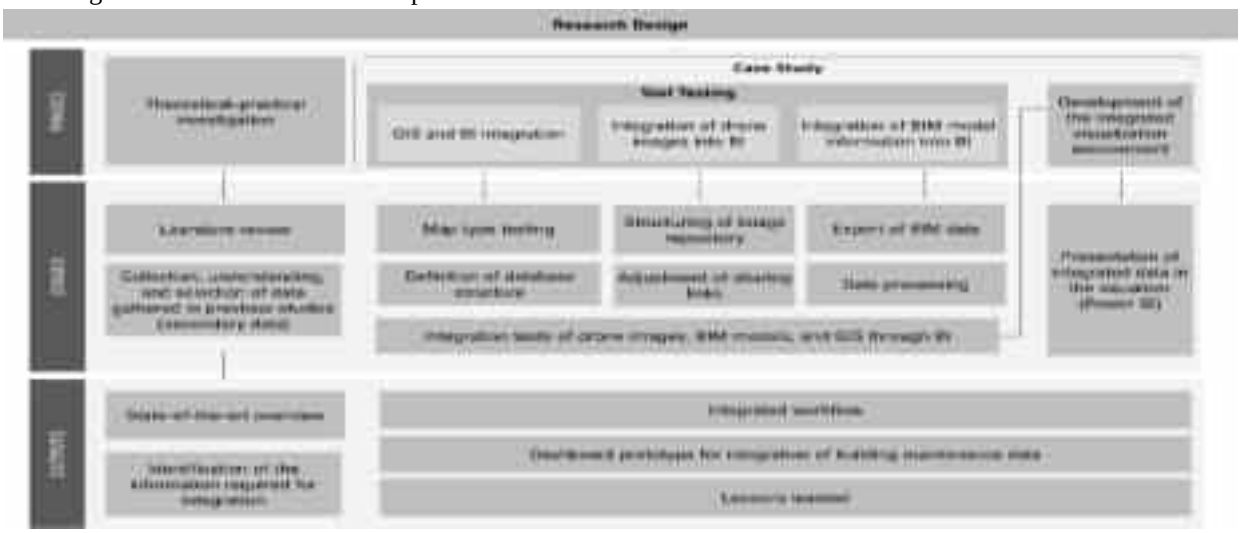


Figure 1: Protocol for developing information flow

Subject-of-Study Characterization

The selected building is in an unfinished state, with construction activities currently on hold, leading to deterioration, particularly of external components such as facades and roofing. The facades include areas with paint finishes, ceramic cladding, and incomplete sections. The roofing system has sections with fiber-cement sheets and exposed concrete slabs.

The building was selected as the only one with a consolidated BIM model and prior research data [6,2]. It is worth noting that the data related to roof inspections are organized separately, without a direct connection to the BIM model, which enabled the analysis of two complementary scenarios: the management of information integrated into the model and the processing of data outside it.

Assessment of digital tools

The tests covered the integration between GIS and BI tools, the incorporation of UAS imagery into the BI environment, and the integration of the BIM model into the analytical dashboard. Subsequently, tests were conducted on the integration process and the relationships among the datasets.

GIS-BI Integration

The tests were conducted using Power BI software, evaluating four native map viewers: Map, Azure Maps, Filled Map, and ArcGIS Maps for Power BI. This stage aimed to analyze the spatial representation capabilities of the building under study, focusing on how clearly its geographic location and placement in the urban environment can be identified. To this end, more detailed tests were performed with the Map viewer, which allowed publication of results online and met the requirements for displaying elements such as streets and neighborhood organization. Also, this viewer enabled integration of spatial information with data from other databases, such as spreadsheets containing building attributes, including name, built area, and year of construction.

Integration of UAS imagery and BI data

Tests focused on applying images collected by UAS in the Power BI environment were conducted using three distinct visualization strategies. The first involved making images available via external links; the second, presenting the images in a table arranged in a matrix; and the third, using the Simple Image visualization method. These approaches aimed to enable dynamic image access for users by integrating their queries into the analytical environment. During testing, the image file formats with the best platform compatibility were evaluated, along with the suitability of the links used to access images stored in cloud repositories.

Integration of BIM model and BI data

The tests for integrating the BIM model with Power BI were conducted in two complementary stages. Initially, the inspection data incorporated into the BIM model was exported from tests performed in Revit, where this information was associated with the building elements. In the next stage, the BIM model was integrated into the analytical dashboard, enabling both the visualization of data in tabular structures and the direct exploration of the three-dimensional model on the dashboard. To enable this integration, the Speckle plugin for Revit was utilized, along with the corresponding viewer in Power BI. This approach enabled interaction with the 3D model and the integrated analysis of inspection data, along with other information presented on the dashboard.

Development of the Integrated Visualization Environment

The development of the visualization environment involved preparing and organizing data for integration into Power BI. Inspection images were consolidated into a unified repository, using standardized nomenclature, file formats, and associated metadata. Accessible addresses were generated for each image via OneDrive shared links, allowing direct connection to dashboard visuals. Simultaneously, the buildings' geographic database was structured using their identifiers and coordinates, enabling their spatial representation in the analytical environment.

To support data integration, heterogeneous datasets were structured using a unified schema for both BIM-based and non-BIM-based information. When available, the BIM model generates two datasets: a tabular export (CSV BIM Data) and a geometry-based dataset via Speckle (BIM Data), which share attributes that enable integration. The CSV BIM Data structure was also adopted in scenarios without a BIM model, ensuring consistency and future compatibility. Inspection data are recorded following the same schema, including links to associated images.Data linkage is established through two identifiers: Building_ID, used for general building information, and Anomaly_ID, used to connect inspection records and images, and to associate them with BIM elements when available.

Based on this structure, a relational model (Figure 2) integrates building information, inspection data, images, and spatial data, supporting dashboard configuration and visualization.

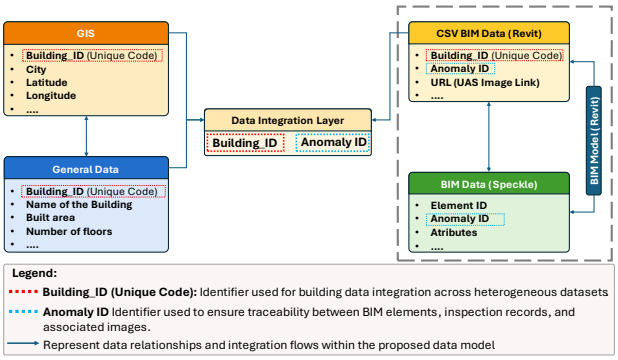


Figure 2: Data structure

Results

Integrated workflow

Figure 3 presents the interoperability flow among the various tools used in the study, demonstrating how data from facade and roof inspections, spreadsheets, geographic information systems (GIS), and the BIM project were integrated.

This process facilitated the organization and management of data within a single environment, enabling its visualization in Power BI dashboards and supporting the integration of information for the case study analysis.

The prototype developed in Power BI (Figure 4) shows the integration of the proposed data sources, featuring an interface with six main viewers and four dynamic filters (Buildings, Anomaly, Gravity, and Envelope). This structure enables the integrated analysis of maintenance data, thereby promoting a coordinated link between the databases. The integration of GIS and BI is evident in the location field, where map layers enable the visualization of buildings in their urban context.

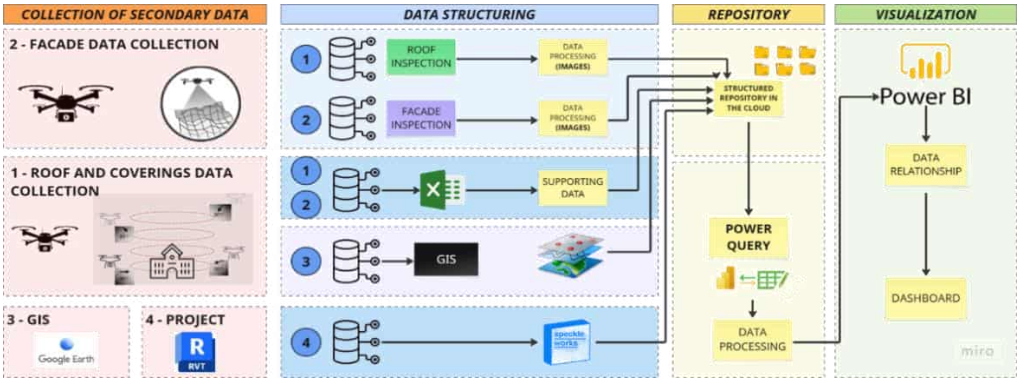


Figure 3: Integrated workflow

This feature links geographic coordinates to building attributes, allowing users to view information such as name, roof area (2,122 m²), and number of floors (3) by hovering over the building on the map. The chosen building is incomplete, with construction on hold, leading to deterioration, especially of its external parts, such as facades and roofing.

The building's geometric representation was incorporated into the panel via the BIM model using the Speckle plugin, enabling three-dimensional navigation directly on the dashboard. This integration linked inspection data to the model geometry, enabling the three-dimensional elements to respond to filters applied in the other sections of the panel. However, the automatic association between photographic records and model components is limited to the facade inspections documented by [2], whose images were already previously linked to the BIM elements.



Figure 4: Proposed Power BI dashboard

For roof inspections, data visualization occurred independently, without the corresponding automatic indication in the three-dimensional model. Even so, the panel maintains its functionality, as its architecture enables the presentation and analysis of inspection information autonomously, without reliance on linking to the BIM model. Thus, the dashboard provides an overview of the building's state of conservation at different levels of technical detail, preserving interface consistency and supporting analysis.

Functional Verification

Functional verifications were conducted during dashboard development to assess the consistency and reliability of the integrated workflow. Four aspects were evaluated. First, the building's geographic position matched its known location, confirming accurate spatial representation. Second, the number of anomalies in charts was consistent with the inspection records. Third, anomaly images and descriptions were validated against the source data, showing full consistency, with minor delays in image loading due to cloud-based storage.

Finally, BIM integration was tested: selecting an item in the table correctly highlighted the corresponding 3D element, while roof inspection anomalies did not affect the model, as expected, since they are not linked to BIM elements (Section 3). Overall, the dashboard performed consistently across all evaluated functions.

Lessons Learned

Despite the satisfactory results of the exploratory case study, integrating digital data into maintenance management poses challenges that require careful consideration, as summarized in Table 1.

Table 1: Challenges and lessons learned

Challenges	Lessons learned
Decentralized repositories	Organization of data from different sources into a centralized structure, with specific repositories defined for each type of information
BIM–BI data interoperability	Export of BIM model information to a CSV file, followed by data processing
GIS–BI data interoperability	The Map viewer was used due to its support for free web publishing and its adequate representation of georeferenced points. In contrast, ArcGIS Maps for Power BI presented restrictions on free publishing, and other map visuals showed limitations in spatial visualization
Integration of UAS imagery into BI	Development of workflows to adjust image sharing links for visualization
Visualization of UAS imagery using BI	Application of the Simple Image visual in Power BI for enhanced image visualization
Unrelated datasets	Use of Building_ID and Anomaly_ID as unique identifiers to ensure relational integration.

Discussions

Building on previous studies, this study advances by presenting the results of a Power BI dashboard that integrates real data on facade and roof maintenance, highlighting practical difficulties and lessons learned from an exploratory case study. While previous research explores, in isolation, the integration between BIM and Business Intelligence [17], the automation of maintenance data recording in BIM models [18], or the standardization and sharing of information through IFC schemas [15,14], this study focuses on the joint integration of data from BIM, GIS, and images obtained by UAS in a single visualization environment in Power BI.

In contrast to the work of Harode et al. [12], which integrates Power BI dashboards with HoloLens 2 for facility management using lifecycle data embedded in BIM models, this study does not rely on augmented reality technologies; instead, it emphasizes the interoperability between multiple spatial and visual data sources applied to building maintenance. Even so, the focus group validation adopted by these authors is identified as a potential strategy for future validation of the results presented here.

Regarding data analysis to support decision-making, Viveros et al. [13] demonstrate the potential of integrating Power BI and Python scripts for evaluating preventive maintenance policies. In contrast, this study prioritizes the organization, visualization, and interpretation of real inspection data, highlighting the technical and operational limitations of the integrated workflow. Compared to the work of Wang et al. [10] and Chen et al. [7], which explores BIM-GIS integration and UAS image registration, focusing on spatial modeling and urban networks, the role of GIS in this study was intentionally defined as a spatial context layer. This layer enables the geographic identification of buildings on campus and helps understand how location affects maintenance conditions. This complementary role aligns with the scope of the proposed workflow, which emphasizes data integration and visualization over complex spatial analysis. Nonetheless, the authors acknowledge that GIS could support more analytical functions in future iterations of this work, such as spatial prioritization of maintenance based on anomaly density, analysis of environmental exposure by facade orientation, or overlay with climatic and microclimate data [9,10].

Finally, this work builds on the findings of Sena et al. [2] and Silva et al. [6] by providing an integrated dashboard that displays the previously collected inspection data, making it easier to visualize, group, and analyze the information all in one place to support better decision-making. Therefore, the study contributes to advancing discussions on integrated data flows for building maintenance, highlighting, in an applied manner, the challenges of interoperability and opportunities for future research. Based on these findings, several

implications and future directions can be highlighted. In this context, beyond identifying practical challenges, this study highlights the need to move from isolated integration efforts toward structured and scalable workflows. The lessons learned indicate that interoperability issues are both technical and organizational, requiring data standardization and consistent identifiers. In this context, future implementations should prioritize semi-automated pipelines and the incorporation of automated methods, such as the use of UAS and machine learning models for anomaly detection, thereby reducing manual effort and improving data reliability. Furthermore, expanding GIS functionalities to include spatial analysis can strengthen decision-making support. These findings are consistent with previous studies [7,2,6], reinforcing the importance of integrating digital technologies in maintenance management. Finally, the value of integrating BIM, GIS, UAS, and BI extends beyond data visualization, enabling a continuous data-to-decision workflow in which field-collected data are transformed into actionable insights for maintenance management. This positions the proposed approach as a step toward more integrated and intelligent maintenance management systems within the context of digital construction.

Conclusion

This study proposed an integrated digital data visualization workflow to support building maintenance management, combining data from UAS inspections, BIM models, GIS geospatial information, and BI tools. The application of the workflow was evaluated through an exploratory case study, using real data from facade and roof system inspections of an institutional building.

The results indicate that combining these data sources into a unified environment is practical and aids in identifying, locating, and analyzing anomalies, while enhancing data organization and accessibility. The use of an interactive dashboard in Power BI enabled integration among spatial analyses, analytical exploration, visualization of the 3D BIM model, and direct access to inspection images, thereby enhancing decision-making support during the operation and maintenance phase.

Limitations include the research's exploratory nature and the application of the workflow to a single object of study, both of which limit the generalizability of the results. Furthermore, the evaluation focused on the feasibility of data integration and visualization, and did not include quantitative metrics on performance, costs, or operational impacts.

Notably, although the proposed workflow was evaluated for technical feasibility and integration, no formal end-user usability testing was conducted. Aspects such as user experience, perceived ease of use, cognitive workload, and effectiveness, as experienced by facility managers and maintenance professionals, were not systematically measured. Therefore, the practical usability and user-centered performance of the developed environment still need empirical validation. Future research should address this gap by using structured usability assessment methods, such as standardized questionnaires and participatory evaluation approaches, to better understand how the system supports real operational decision-making in building maintenance contexts.

Future work recommends applying the proposed workflow to a broader range of buildings and incorporating performance indicators and criteria for prioritizing interventions. The adoption of participatory validation methods, such as focus groups with managers and specialists, can also help improve the developed environment and advance discussions on the integrated use of digital technologies in building maintenance. Furthermore, advancing the automation of UAS-related inputs represents a key opportunity for improving the workflow. Future implementations could incorporate automated flight planning with pre-defined waypoints for systematic envelope coverage, automatic image naming based on GPS coordinates and building component identifiers, and direct upload to cloud repositories upon flight completion. Integration with anomaly detection models, such as those based on deep learning architectures applied to roof damage detection, could further enable the automatic registration of defects in the dashboard without manual intervention, closing the data collection loop from UAS acquisition to maintenance decision-making. In parallel, expanding the analytical role of GIS by spatially prioritizing interventions and overlaying them with environmental exposure data would strengthen the methodological contribution of the integrated workflow. This reinforces the potential of integrated digital workflows as a foundation for more data-driven and proactive maintenance management practices.

Acknowledgment

This work was supported by the Ministry of Science, Technology, Innovation, and Communication of Brazil through CNPq (National Council for Scientific and Technological Development) (Grant Number 443942/2024-2), and the team from the UFBA Superintendence of Environment and Infrastructure (SUMAI-UFBA).

References

[1] A. Harode, W. Thabet, and P. Dongre. A tool-based system architecture for a digital twin: a case study in a healthcare facility. Journal of Information Technology in Construction, 28, 2023.

[2] R. O. Sena, A. S. Silva, R. S. S. Melo, and D. B. Costa. Proposal for integrating drone images and BIM in educational public buildings to support maintenance management. Revista Ingeniería de Construcción, 39(3), 107-117, 2024.

[3] A. M. A. Dompey, S. Aklashie, F. O. Danso, K. Agyekum, R. Gasue, and E. Kumi-Amoah. Digital innovation in building maintenance: the perspectives of facilities managers in Ghana. International Journal of Building Pathology and Adaptation, 1-18, 2025.

[4] Y. F. Liu, X. Nie, J. S. Fan, and X. G. Liu. Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. Computer-Aided Civil and Infrastructure Engineering, 35(5):511-529, 2020.

[5] K. J. Beasley. Building facade failures. In Proceedings of the Institution of Civil Engineers - Forensic Engineering, 165:13-19, 2012.

[6] A. S. Silva, F. H. M. Neto, P. H. Ferreira, and D. B. Costa. CNN-Based YOLOv12 for damage assessment in residential roofs. *IEEE Access*, 13:193311–193322, 2025.

[7] K. Chen, G. Reichard, A. Akanmu, and X. Xu. Geo-registering UAV-captured close-range images to GIS-based spatial model for building façade inspections. Automation in Construction, 122:103503, 2021.

[8] N. I. Hasanuddin and M. H. I. Abd Rahim. Potential of Technology Application on Building Facilities Management: Schedule of Maintenance. Research in Management of Technology and Business, 5(1):1371-1383, 2024.

[9] P. Gasbarri, D. Accardo, E. Cacciaguerra, S. Meschini, and L. C. Tagliabue. Supporting asset management with GIS and business intelligence technologies: The case study of the University of Turin. ISPRS International Journal of Geo-Information, 13(3):65, 2024.

[10] M. Wang, Y. Deng, J. Won, and J. C. Cheng. An integrated underground utility management and decision support based on BIM and GIS. Automation in Construction, 107:102931, 2019.

[11] B. Becerik-Gerber, F. Jazizadeh, N. Li, and G. Calis. Application areas and data requirements for BIM-enabled facilities management. Journal of Construction Engineering and Management, 138(3):431-442, 2012.

[12] A. Harode, M. Ensafi, and W. Thabet. Linking BIM to Power BI and Hololens 2 to support facility management: a case study approach. Buildings, 12(6):852, 2022.

[13] P. Viveros, N. C. Pantoja, F. Kristjanpoller, and R. Mena. Reportability Tool Design: Assessing Grouping Schemes for Strategic Decision Making in Maintenance Planning from a Stochastic Perspective. Applied Sciences, 12(11):5386, 2022.

[14] S. Matarneh, F. Elghaish, F. P. Rahimian, N. Dawood, and D. Edwards. Automated and interconnected facility management system: An open IFC cloud-based BIM solution. Automation in Construction, 143:104569, 2022.

[15] F. G. Bellon, A. C. P. Martins, J. M. F. de Carvalho, C. A. F. de Souza, J. C. L. Ribeiro, K. M. L. C. Júnior, and D. S. de Oliveira. IFC framework for inspection and maintenance representation in facility management. Automation in Construction, 174:106157, 2025.

[16] R. K. Yin. Case Study Research: Design and Methods. volume 5. Sage Publications, 2009.

[17] F. Rodrigues, A. D. Alves, and R. Matos. Construction management supported by BIM and a business intelligence tool. Energies, 15(9):3412, 2022.

[18] J. Cho, C. Kim, Y. Song, J. Kang, and J. Yeon. Lumped record management method using BIM and Dynamo for spalling maintenance. Automation in Construction, 160:105324, 2024.

A. S. Silva, R. R. S. de Melo, R. S. S. de Melo and D. B. Costa. Method for quality inspection during the execution of facades based on UAS images and machine learning algorithms. International Journal of Construction Management, 26(2):277–294, 2026.

Bridge concrete spalling recognition based on generative AI synthetic weather and YOLOv11

Rong-Lu Hong[1], KiWon Lee[1], SeoYoung Park[1], Seunghyeon Wang[1] and Ju-Hyung Kim[1]

[1] Department of Architectural Engineering, Hanyang University, Seoul, Republic of Korea
abraham0814@hanyang.ac.kr, fosys8651@hanmail.net, ondamco@gmail.com, wsh1019@hanyang.ac.kr , kcr97jhk@hanyang.ac.kr

Abstract –

Automatic concrete spalling detection on bridge surfaces using unmanned aerial vehicles (UAV) is essential for improving inspection efficiency and operational safety. However, instance segmentation models employed in UAV-based inspections are sensitive to environment-related variations in image appearance. In practice, inspection data are usually collected under limited conditions, which reduces model reliability when applied beyond standard operating scenarios. To address this limitation, weather condition modelling strategies were examined from a data construction perspective. Three datasets derived from the same UAV inspection imagery were used: a baseline, a traditionally augmented dataset based on global statistical transformations, and a generative-AI-based dataset representing diverse inspection environments. All datasets shared identical pixel-level annotations. A unified instance segmentation model, YOLOv11 seg, was trained under identical configurations and evaluated using cross-dataset testing to assess generalization under simulated weather variability. Results demonstrate that the model trained with generative-AI-enhanced data achieves stable segmentation performance in challenging inspection environments. Meanwhile, the segmentation accuracy on the original test set remained high, indicating that improved environmental adaptability was achieved without sacrificing performance under regular inspection conditions. These findings demonstrate that generative-AI-based environment synthesis effectively reduces training data bias and enhances model reliability without additional annotation, thereby providing a practical data construction strategy for UAV-based bridge inspection in real-world applications.

Keywords –

Bridge inspection; Concrete spalling; Generative AI; Synthetic weather data; YOLOv11 segmentation

1 Introduction

Bridges are critical components of transportation infrastructure, and their service conditions directly affect public safety and the stable operation of transportation systems [1]. Surface concrete spalling is a typical bridge deterioration and damage form [2,3]. Such damage compromises the concrete cover and accelerates reinforcement corrosion, thereby reducing structural load-bearing capacity and durability [3]. According to the 2025 National Bridge Inventory of the United States, approximately 620,000 highway bridges are currently in service, more than 260,000 of which exhibit structural deficiencies, with an estimated investment of approximately $467 billion USD required for comprehensive rehabilitation [4]. If surface damage such as concrete spalling is not identified and repaired at an early stage, it often progresses to structural deterioration, substantially increasing maintenance costs and operational risks [1–3].

In current bridge operation and maintenance systems, surface concrete spalling identification relies primarily on manual inspection methods [3,5]. Such inspections typically require close-range visual examinations using aerial work equipment and exhibit limitations in inspection efficiency, operational safety, and result consistency, with outcomes largely dependent on inspector experience and judgment [1,2,5–7]. With the development of unmanned aerial vehicle (UAV) platforms and computer vision technologies, automated bridge surface damage detection based on UAV imagery has become a feasible technical approach [1,6,7]. Combined with deep-learning-based instance segmentation models, fine-grained localization of concrete spalling regions can be achieved under non-contact conditions, thereby improving inspection efficiency and reducing inspection risk [6,8]. However, the effectiveness of these methods is highly sensitive to training data quality and diversity [9]. In actual engineering environments, variations in weather and lighting conditions significantly affect the appearance of bridge imagery, limiting the model applicability under complex conditions [10,11].

To alleviate the shortage of data under complex environmental conditions, prior studies have expanded original imagery using conventional image augmentation techniques, including global statistical transformations such as brightness and contrast adjustment and noise perturbation [10,11]. Although these methods can increase the appearance diversity of training samples to a certain extent, their image level operations make it difficult to represent the local texture variations induced by weather conditions and the spatial interactions between such variations and structural surfaces [11,12]. In recent years, generative methods have gradually been introduced for the synthesis of complex scene samples. Style transfer networks such as CycleGAN have been applied to simulate road or structural scenes under different weather conditions [13,14]. However, the domain transformation process may introduce geometric distortions, making it difficult to preserve the spatial consistency of damage boundaries. Diffusion models have demonstrated advantages in image quality, yet their training cost remains high and the controllability of the generation process is still limited. As a result, targeted environmental modification while maintaining pixel level annotation consistency continues to present a challenge [15,16]. In contrast, multimodal large language models represented by ChatGPT and Gemini have shown strong semantic understanding and instruction following capabilities in image editing tasks [17–20]. Through structured prompting mechanisms, these models can perform targeted modification of environmental appearance while preserving the original geometric structure [19,20], thereby offering a new paradigm for expanding environmental diversity at zero annotation cost. However, the application of model driven environmental synthesis methods of this kind to bridge damage detection remains at an exploratory stage, and systematic experimental evaluation and method comparison are still lacking.

Based on this background, this study proposes a multi weather data generation and evaluation framework for bridge surface concrete spalling detection. To enable a systematic comparison of data construction strategies, three datasets were established, namely, the Base Dataset, the Traditional Augmentation Dataset, and the Generative Artificial Intelligence Dataset. Under a unified instance segmentation model and consistent experimental settings, comparative training and cross dataset testing were performed to evaluate how different data construction strategies affect segmentation performance and generalization ability under complex weather conditions. The results verify the effectiveness of the proposed framework in expanding environmental diversity without additional annotation cost and in improving the stability of automated bridge surface damage detection.

2 Material and Methods

2.1 Overview

To analyze the impact of weather conditions on UAV-based bridge defect detection performance, three datasets were constructed using the same set of bridge surface concrete spalling images acquired from UAV inspections. The Base Dataset was constructed without introducing artificial weather variations. The TA and GAI datasets incorporate weather conditions using different modeling strategies. The three datasets differed only in the adopted weather modeling approach, whereas all other image content and pixel-level annotations remained identical. Each dataset was consistently divided into training, validation, and test sets at a ratio of 6:2:2.

The instance segmentation model YOLOv11 seg was employed as a unified analytical model and trained separately on the Base, TA, and GAI datasets. The inference and evaluation were conducted on the Base and GAI test sets. This cross-testing configuration enables a direct detection performance comparison under conditions without weather interference and complex weather scenarios, thereby facilitating the analysis of the effects of weather variations and different weather modeling strategies on defect detection outcomes. The overall experimental workflow is illustrated in Fig. 1.

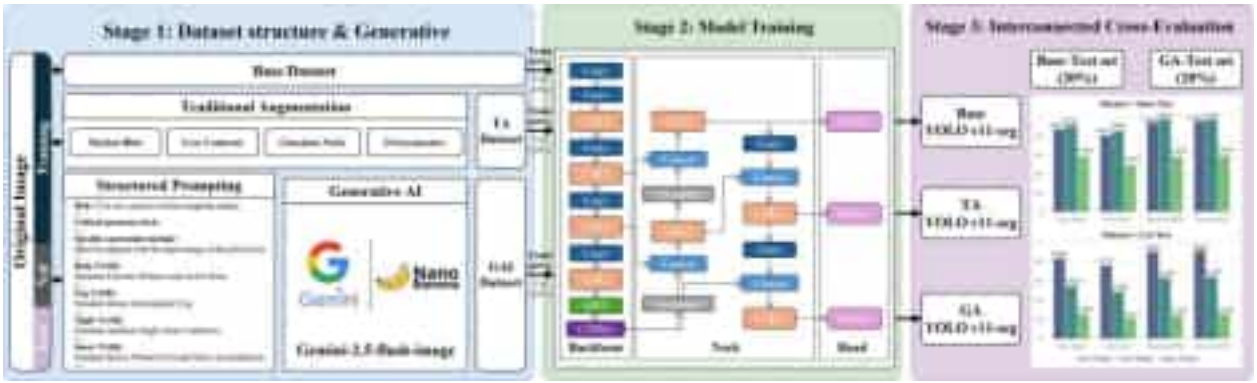


Figure 1. Overview of the experimental framework.

2.2 Data collection

This study used a publicly available bridge façade orthophoto dataset released by the National Information Society Agency of Korea through the AIHUB platform [21]. The dataset covers 27 bridges with different structural types, including PSC I Girder, Steel Box Girder, and Preflex Bridge, and was collected using a DJI Mavic 3 UAV in a close-range inspection mode with a ground sampling distance of 0.3 mm/px and an imaging distance of approximately 1.33 m.

All images were uniformly cropped to 512 × 512 pixels, corresponding to a physical coverage of 15.4 cm × 15.4 cm. Based on these images, a Base Dataset containing only concrete spalling defects was constructed, comprising 551 images and 930 instances annotated with LabelMe polygon masks. Instance areas showed a long-tailed distribution, with small, medium, and large scales accounting for 41.7%, 39.6%, and 18.7%, respectively, where the corresponding pixel ranges were <32 px, 32 to 96 px, and >96 px. The physical size of defects ranged from 0.34 to 79.1 mm. On average, defect regions occupied only 2.85% of the image area, corresponding to a foreground to background ratio of approximately 1:34. The mean bounding box size was 4.1 cm × 1.3 cm, which remained smaller than the image coverage range and therefore preserved sufficient contextual information.



Figure 2. Example of the Base Dataset.

2.3 Traditional Data Augmentation Methods

To simulate weather-related effects on the UAV-based bridge-inspection imagery, a TA was constructed from the Base Dataset. Each original image was augmented without introducing geometric deformations while maintaining full annotation consistency. This process results in 2204 augmented images that were strictly aligned with their original annotations.

The TA Dataset models weather-induced image quality changes through global statistical transformations applied to bridge façade images, including motion blur, low contrast, high brightness or overexposure, and Gaussian noise [9]. Motion blur represents image quality loss caused by UAV motion or wind, with kernel sizes randomly selected between 5 and 15 pixels. Low contrast simulates visibility reduction under hazy or rainy conditions, with scaling factors sampled from 0.4–0.8. High brightness or overexposure model brightness saturation due to strong illumination or surface reflection, with gains randomly adjusted between 1.5 and 2.5. Gaussian noise accounts for sensor noise in low-light or complex environments, with a mean of zero and variance between 20 and 40.

These augmentations introduce environmental variability while preserving overall visual plausibility. However, because transformations are applied globally, they have a limited ability to capture localized appearance changes and their interaction with structural details under complex weather conditions [9,11,12]. Representative examples are shown in Fig. 3.

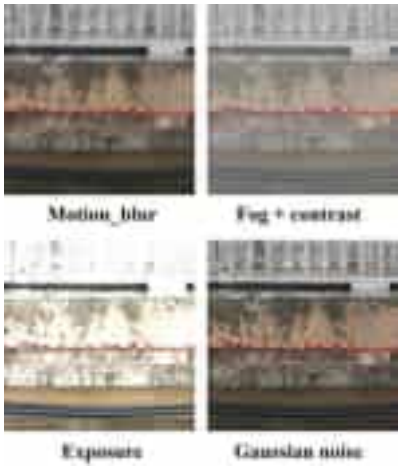


Figure 3. Examples of traditional data augmentation.

2.4 Generative Artificial Intelligence Methods

To compare the capability of generative artificial intelligence and conventional augmentation methods in modeling complex weather conditions, this study adopted a semantic driven image generation approach to simulate rain, haze, snow, and low light nighttime conditions in bridge UAV inspection images, while preserving structural consistency [17–20]. Based on the Base Dataset, one weather generation operation was applied to each original image to construct the generative augmentation dataset, namely the GAI Dataset, resulting in 2,204 images in total.

Image generation was performed using the Gemini 2.5 Flash Image model [19], with sampling parameters set to temperature = 1.0, top p = 0.95, and top k = 64 to maintain visual diversity while limiting geometric distortion. Each image was generated only once and retained directly, without repeated sampling or manual selection, in order to avoid selection bias. Structured prompts were used to introduce only weather related changes in texture, illumination, and visibility, without altering the spatial layout of bridge components or defect regions. To prevent geometric deviation, all generation tasks were constrained by a unified role definition and geometry locking rules. Defect boundaries were not allowed to shift, scale, rotate, or change shape, and only

transparent overlay based environmental effects were permitted. The generation rules for different weather types were specified separately through structured prompts [17], and the detailed prompts are listed in Table 1.

To eliminate the risk of data leakage, the 6:2:2 dataset split was performed at the original image level, and all variants derived from the same source image, including the original, TA, and GAI versions, were always assigned to the same subset. Pairwise validation of all 2,204 original to augmented image pairs showed that both mask IoU and Dice coefficient remained exactly 1.000 across all four weather conditions, while the boundary Hausdorff distance was 0.00 pixels. The SSIM ranged from 0.384 to 0.673, indicating that the augmentation substantially changed visual appearance without introducing any annotation misalignment.

Table 1. Structured prompt for weather synthesis.

Role: You are a precise texture mapping engine. Your sole responsibility is to apply environmental texture overlays to the input image. **Critical geometry lock:** No displacement, no deformation, overlay only. **Specific constraints include:** The output image must be aligned with the input image at the pixel level. No cracks or defect boundaries may be moved, scaled, rotated, or reshaped. All effects must be applied exclusively as transparent overlay layers on the existing geometry. The generation style follows a high-resolution industrial inspection imaging standard, preserving the characteristics of raw sensor data.

2.4.1 Rainy Condition Modeling

Rainy-scene synthesis focuses on reproducing the characteristic surface appearance of concrete under wet conditions, as summarized in Table 2. The generation process introduces an overall reduction in brightness, specular reflections caused by a continuous surface water film, and rain streak textures aligned with gravity's direction, while strictly preserving the original geometric structure and defect boundaries.

Table 2. Structured prompt for rain scene synthesis.

TASK: Simulate Extreme Wetness and Active Rain. The concrete surface is completely soaked and appears significantly darker. The surface is covered with a glossy water layer with clear specular reflections. Mirror like reflections of the overcast sky are visible on the wet surface. Add visible vertical rain streaks falling across the image.

As shown in Fig. 4, the generated results successfully reproduce the lighting variations and reduced visibility commonly observed in UAV inspection imagery under real rainfall conditions while preserving the original spatial location and morphology of the defect regions.



Figure 4. Example of rain scene synthesis.

2.4.2 Foggy Condition Modeling

Foggy conditions were modeled by simulating the visual effects of atmospheric scattering, which primarily manifests as reduced contrast and degraded detail visibility. As presented in Table 3, a uniformly distributed cool-gray haze is overlaid during generation to reproduce typical fog-induced appearance degradation without altering the geometric configuration of the defect regions.

Table 3. Structured prompt for fog scene synthesis.

TASK: Simulate Dense Atmospheric Fog. Strong atmospheric scattering significantly reduces visibility. Overall contrast is reduced and shadows are lifted toward gray. A cool gray haze uniformly covers the entire image.

As shown in Fig. 5, the generated images exhibit the characteristic low-visibility imaging features of foggy environments while maintaining the spatial consistency of the defect regions.



Figure 5. Example of fog scene synthesis.

2.4.3 Snowy Condition Modeling

Snow-scene synthesis aims to capture the coverage effects of frost and snow accumulation on concrete surfaces, as outlined in Table 4. The generation process introduces cold-toned lighting and surface-coverage characteristics while maintaining the original spatial location and geometric shape of the defects, ensuring that only the environmental appearance is modified.

Table 4. Structured prompt for snow scene synthesis.

TASK: Simulate Heavy Winter Frost and Snow Accumulation. A thick layer of frost and snow particles covers the concrete surface. The lighting exhibits a cold bluish winter tone. Snow texture is applied directly over existing cracks and defects.

As shown in Fig. 6, the generated images reflect typical winter environmental characteristics while retaining the spatial continuity of the concrete spalling regions.



Figure 6. Example of snow scene synthesis.

2.4.4 Nighttime Condition Modeling

Nighttime inspection scenes were generated by modeling appearance degradation under weak-ambient-lighting conditions. As presented in Table 5, the synthesis process reproduces a globally darkened appearance accompanied by increased noise and limited color information without introducing additional light sources or geometric changes, thereby reflecting the typical visual characteristics of nighttime UAV inspections.

Table 5. Structured prompt for night scene synthesis.

TASK: Simulate Ambient Night Time Conditions. The scene is illuminated only by very weak ambient light. No flashlights or direct light sources are allowed. The image is underexposed with strong sensor noise and limited color information.

As shown in Fig. 7, the generated results reflect typical imaging characteristics under low-light nighttime conditions, while maintaining consistency in the defect geometry.



Figure 7. Example of night scene synthesis.

2.5 Model selection

This study formulates the task as instance segmentation to achieve fine grained extraction of concrete spalling regions on bridge surfaces. Concrete spalling typically exhibits irregular boundaries, complex shapes, and multiple spatially separated regions within a single image [2,12]. Compared with other vision tasks, instance segmentation enables pixel level delineation of defect boundaries while distinguishing individual defect instances, making it suitable for morphological and spatial distribution analysis in engineering inspection [8].

YOLOv11n seg was adopted as the instance segmentation model in this study [22]. Within a single stage framework, the model integrates a segmentation branch and generates instance level masks through multi scale feature fusion, allowing object localization and pixel level mask prediction in a unified inference process, as illustrated in Stage 2 of Fig. 1. The model contains 2.84 million parameters, requires 9.6 GFLOPs, and has a file size of only 5.71 MB. Under a 512 × 512 input, its inference latency is approximately 1.0 to 2.5 ms, providing a practical balance between segmentation accuracy and the real time constraints of UAV edge deployment. Based on these characteristics, YOLOv11n seg was consistently used as the segmentation model in all data augmentation comparison experiments.

2.6 Experimental Strategy

All experiments were conducted under the same computational setting, with the data construction strategy as the only variable. The platform consisted of Windows 11 Pro, Python 3.9, and PyTorch 2.8.0, running on an NVIDIA GeForce RTX 5080 GPU and an AMD Ryzen 9 9950X processor. YOLOv11 seg initialized with the pretrained weight yolo11n-seg was used in all groups.

During training, all images were resized to 512 × 512 pixels. The model was trained for up to 200 epochs using AdamW, with a learning rate of 0.001, weight decay of 0.0005, batch size of 16, and random seed of 42. Early stopping was applied when validation performance did not improve for 20 consecutive epochs. No additional random augmentation was introduced.

Evaluation was performed on the same test sets using Box mAP50, Mask mAP50, pixel level IoU, and Dice coefficient [8]. A prediction was regarded as correct when its IoU with the ground truth was not less than 0.50. For a predicted region P and and ground truth region G,

$$IoU(P,G) = \left|\frac{P \cap G}{P \cup G}\right| \tag{1}$$

Pixel level consistency was measured on the predicted mask M_{pred} and ground truth maskM_{gt}:

$$IoU_{pixel} = \left|\frac{M_{pred} \cap M_{gt}}{M_{pred} \cup M_{gt} + \varepsilon}\right| \tag{2}$$

$$Dice = \frac{2\left|M_{pred} \cap M_{gt}\right|}{\left|M_{pred}\right| + \left|M_{gt}\right| + \varepsilon} \tag{3}$$

where ε denotes a numerical stabilization term. Final pixel level results were obtained by averaging all test samples.

3 Result

3.1 Cross-Dataset Segmentation Performance

All models exhibited stable convergence without obvious overfitting or underfitting, their quantitative

performance on the test set is presented in Fig.8.

On the GA test set, the GA model achieved the best segmentation performance, with a mask mAP50 of 0.902 and a Dice coefficient of 0.828. These values are substantially higher than those of the TA model (0.630 and 0.546, respectively), while also being markedly superior to those of the BASE model (0.245 and 0.261, respectively). These results indicate that models trained with generative augmentation data are more effective at learning damage representations in complex environments when the test distribution differs significantly from the training distribution.

On the BASE test set, the TA model attained the highest segmentation accuracy, achieving a mask mAP50 of 0.983 and a Dice coefficient of 0.918. Meanwhile, the GA model maintained comparable performance on the same test set, with a mask mAP50 of 0.946 and a Dice coefficient of 0.877, without exhibiting noticeable performance degradation. This suggests that the introduction of generative augmentation data does not compromise the segmentation capability of the original data distribution.

Performance differences across the two test sets further demonstrate that the GA model maintains stable segmentation accuracy under conditions in which the training and testing distributions are inconsistent, whereas the TA model is more sensitive to test data distribution variations. These experimental results validate the effectiveness of generative-AI-augmented data in enhancing the cross-scenario generalization capability of instance segmentation models.

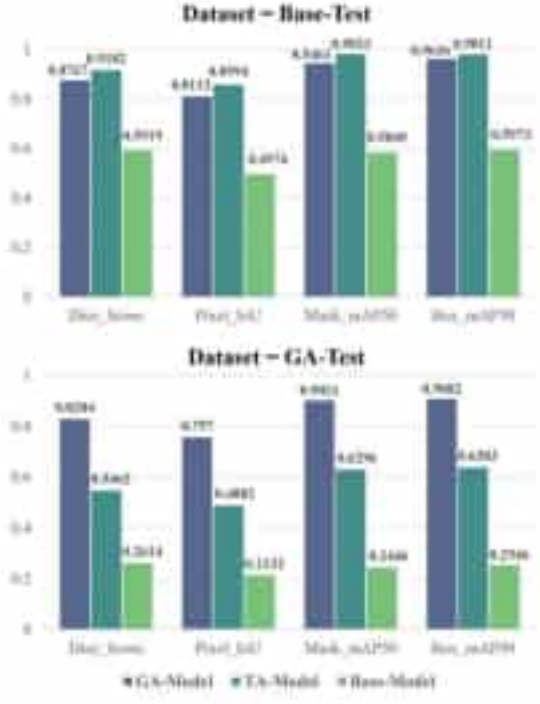


Figure 8. Cross-dataset segmentation performance.

3.2 Qualitative Comparison under Diverse Environmental Conditions

To provide a visual comparison of instance segmentation behavior under different training strategies, Fig. 9 presents representative segmentation results and their corresponding Dice scores under baseline, rain, night, snow, and haze conditions. Under the baseline test condition, the Dice scores of the Base, TA, and GA models were 0.90, 0.94, and 0.9423, respectively, indicating only minor performance differences.

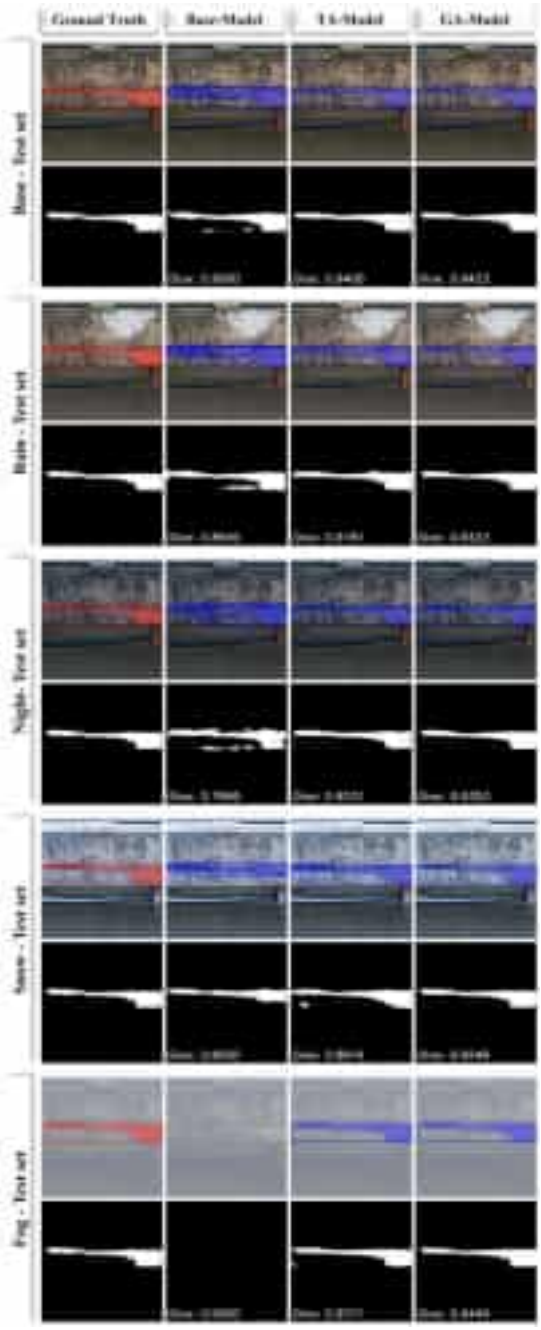


Figure 9. Segmentation examples under different weather.

As environmental conditions became more challenging, the performance gap between models became increasingly evident. This divergence is fundamentally associated with the different distribution modeling mechanisms of the two augmentation strategies. Traditional augmentation methods, such as motion blur, contrast reduction, and Gaussian noise, are essentially spatially uniform global statistical transformations. As a result, they cannot reproduce the localized appearance changes that are semantically coupled with scene content in real weather conditions, such as wet specular reflection on concrete surfaces in rain or uneven occlusion of defect regions under snow. In rain and night scenes, the Dice scores of the Base model decreased to 0.8549 and 0.7996, whereas both the TA and GA models remained above 0.91. Under snow conditions, the Base model achieved a Dice score of only 0.8020, the TA model reached 0.8916, and the GA model further improved to 0.9549, showing superior regional continuity and boundary consistency.

The difference was most pronounced under haze conditions. Real atmospheric scattering exhibits depth dependent spatial characteristics, where distant details are attenuated more severely than near field regions, producing a nonuniform visibility gradient. Traditional augmentation applies only uniform contrast degradation and therefore cannot reproduce this physical property. Consequently, the Base model completely failed to identify the defect region, with a Dice score of 0.00. The TA model detected only the main damage area and showed insufficient boundary completeness, with a Dice score of 0.9271. In contrast, the GA model maintained stable performance by capturing semantically consistent scattering patterns, achieving a Dice score of 0.9440.

These Dice trends are consistent with the quantitative cross dataset evaluation results, indicating that the main advantage of generative augmentation lies in modeling the semantic coupling between weather conditions and scene content, rather than merely expanding pixel level statistical variation.

4 Conclusion

In this study, conducted within the context of UAV-based bridge inspection, systematically evaluated the impact of different weather condition modelling strategies on the instance segmentation performance of concrete spalling. Under a unified model architecture and consistent training configuration, a comparative analysis was performed using real inspection data, traditionally augmented data, and multi-weather datasets constructed via generative AI. The experimental results demonstrated that under complex environmental conditions, including rain, nighttime, haze, and snow, the model trained with generative augmentation data exhibited the most stable performance in terms of segmentation accuracy and pixel-level consistency. In the complex environment test set, this model achieved a mask mAP50 of 0.902 and a Dice coefficient of 0.828, while reaching 0.946 and 0.877 on the original test set, respectively.

Further analysis shows that the generative AI based multi weather dataset reduced training bias toward a single environment by combining semantic driven synthesis with geometry preserving constraints, thereby covering inspection conditions that are difficult to collect systematically while maintaining defect location and annotation consistency. The contributions of this study are threefold. First, it proposes a zero-annotation cost strategy for multi weather synthesis, in which structured prompts and geometry locking enable environmental appearance modification without changing pixel level labels. Second, it establishes a controlled three group experimental framework with a shared model architecture and training configuration, allowing the independent effect of data construction strategy on segmentation performance to be quantitatively isolated. Third, the model trained with generative augmentation achieved a Mask mAP50 of 0.902 on the synthetic weather test set, representing a 43.2% improvement over traditional augmentation, while maintaining 0.946 on the original test set. This result confirms that environmental diversity can improve model performance without sacrificing baseline accuracy. Because the framework requires no additional annotation effort and can be integrated into existing inspection workflows, and because YOLOv11 seg supports efficient single stage inference, the proposed approach also shows potential for edge deployment.

This study still has several limitations. First, the annotation reuse strategy implicitly assumes that weather synthesis changes appearance only and does not alter defect visibility. This assumption may not hold in extreme conditions, such as thick snow physically covering the defect, dense haze, or very low illumination that makes the defect visually indistinguishable. In such cases, direct reuse of the original labels may introduce label noise. Second, the experiments focus only on concrete spalling and do not verify applicability to multi defect detection tasks. In addition, the generative dataset was developed mainly for typical weather conditions and does not yet cover continuous environmental transitions that may occur in real inspections. Third, only a single instance segmentation model was evaluated, and the adaptability of different architectures to generative augmentation remains unclear. Future work should incorporate visibility aware label adjustment to handle severely degraded conditions, extend the framework to multi defect and multi component inspection tasks, and combine finer grained environmental modeling with multi model comparisons to further define the practical scope of generative data augmentation in intelligent

bridge inspection.

Acknowledgement

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (grant number RS202400356697).

Reference

[1] C. Huang, Y. Zhou, and X. Xie. Intelligent diagnosis of concrete defects based on improved Mask R CNN. *Applied Sciences*, 14:4148, 2024.

[2] P. Huethwohl, R. Lu, and I. Brilakis. Multi classifier for reinforced concrete bridge defects. *Automation in Construction*, 105:102824, 2019.

[3] H. Zoubir, M. Rguig, M. E. Aroussi, A. Chehri, R. Saadane, and G. Jeon. Concrete bridge defects identification and localization based on classification deep convolutional neural networks and transfer learning. *Remote Sensing*, 14, 2022.

[4] Federal Highway Administration. National Bridge Inventory database. On-line: https://www.fhwa.dot.gov/bridge/nbi.cfm, Accessed: 01 04 2025.

[5] S. Dorafshan and M. Maguire. Bridge inspection human performance, unmanned aerial systems and automation. *Journal of Civil Structural Health Monitoring*, 8:443 to 476, 2018.

[6] R. Li and Z. Wang. AGSAM Net UAV route planning and visual guidance model for bridge surface defect detection. *Image and Vision Computing*, 154:105416, 2025.

[7] F. Wang, Y. Zou, X. Chen, C. Zhang, L. Hou, E. del Rey Castillo, and J. B. P. Lim. Rapid in flight image quality check for UAV enabled bridge inspection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 212:230 to 250, 2024.

[8] C. Y. Liu and J. S. Chou. Bayesian optimized deep learning model to segment deterioration patterns underneath bridge decks photographed by unmanned aerial vehicle. *Automation in Construction*, 146:104666, 2023.

[9] T. Askarzadeh and R. Bridgelall. Cost efficiency and effectiveness of drone applications in bridge condition monitoring. *Infrastructures*, 10, 2025.

[10] Y. Yang, S. Yang, C. Li, Y. Wang, X. Pi, Y. Lu, and R. Wu. Insulator defect detection under extreme weather based on synthetic weather algorithm and improved YOLOv7. *High Voltage*, 10:69 to 77, 2025.

[11] S. Wang. Effectiveness of traditional augmentation methods for rebar counting using UAV imagery with Faster R CNN and YOLOv10 based transformer architectures. *Scientific Reports*, 15:33702, 2025.

[12] D. Amirkhani, M. S. Allili, L. Hebbache, N. Hammouche, and J. F. Lapointe. Visual concrete bridge defect classification and detection using deep learning a systematic review. *IEEE Transactions on Intelligent Transportation Systems*, 25:10483 to 10505, 2024.

[13] Z. Liang, M. Zhang, X. Zhao, W. Liu, and Y. Wang. Robust traffic sign detection in adverse weather via hybrid vision transformer and GAN based augmentation. *In Proceedings of the 7th International Conference on Data Driven Optimization of Complex Systems*, Taiyuan, China, pages 340 to 344, 2025.

[14] B. Li, H. Guo, and Z. Wang. Data augmentation using CycleGAN based methods for automatic bridge crack detection. *Structures*, 62:106321, 2024.

[15] W. Shen, D. Zeng, Y. Zhang, X. Tian, and Z. Li. Image augmentation for nondestructive testing in engineering structures based on denoising diffusion probabilistic model. *Journal of Building Engineering*, 89:109299, 2024.

[16] S. Yu, J. Li, H. Zheng, and H. Ding. Research on predicting building facade deterioration in winter cities using diffusion model. *Journal of Building Engineering*, 111:113365, 2025.

[17] Y. Lee, G. Kang, J. Kim, S. Yoon, and J. Jeon. Generative AI driven data augmentation for enhanced construction hazard detection. *Automation in Construction*, 177:106317, 2025.

[18] Y. Qian, E. Bocek Rivele, L. Song, J. Tong, Y. Yang, J. Lu, W. Hu, and Z. Gan. Pico Banana 400K a large scale dataset for text guided image editing. *arXiv preprint*, arXiv 2510.19808.

[19] G. Comanici, E. Bieber, M. Schaekermann, I. Pasupat, N. Sachdeva, I. Dhillon, M. Blistein, O. Ram, D. Zhang, E. Rosen, and others. Gemini 2.5 pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint*, arXiv2507.06261, 2025.

[20] J. Betker, G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee, Y. Guo, and others. Improving image generation with better captions. *Computer Science*, 2:8, 2023.

[21] National Information Society Agency. Bridge exterior inspection rectified facade image dataset. On-line: https://www.aihub.or.kr, Accessed: 01 04 2026.

[22] R. Khanam and M. Hussain. YOLOv11 an overview of the key architectural enhancements. *arXiv preprint*, arXiv:2410.17725, 2024.

Automated Cycle Time Analysis in Prefabricated Construction Using Hybrid Vision-Language Models and Motion Heuristics

Mohamed Sabek[1,2,4], Baseel Andres Ammar[2,5], Haitao Yu[3,6], Farook Hamzeh[2,7], and Vicente Gonzalez[1,2,8]

[1] Infrastructure Human Tech (IHT) Lab, Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, Alberta, Canada
[2] Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, 116 Street NW, Edmonton T6G 1H9, Alberta, Canada
[3] Landmark Homes Company, 202, 1103 - 95 Street SW, Edmonton AB T6X 0P8, Canada
[4] sabek@ualberta.ca , [5] baseelan@ualberta.ca , [6] haitaoy@landmarkgroup.ca , [7] hamzeh@ualberta.ca , [8] vagonzal@ualberta.ca

Abstract -

The construction industry's transition to off-site prefabrication requires rigorous workflow monitoring to fully realize the efficiency gains of lean manufacturing. However, measuring variations in cycle times for components at the workstation level, such as prefabricated wooden panels, remains challenging; traditional manual tracking is labor-intensive and error-prone, whereas supervised Computer Vision (CV) approaches are limited by the need for extensive, non-scalable dataset annotation. This study introduces a novel zero-shot framework that integrates Vision-Language Models (VLMs) with deterministic pixel-level motion heuristics and spatial zoning to automate material tracking. We evaluated the performance trade-offs between closed-source embodied reasoning models (Gemini Robotics-ER 1.5) and state-of-the-art open-source architectures (Qwen3-VL-2B). Our hybrid methodology couples the high-level semantic understanding of VLMs for identifying process states with frame-differencing logic to validate temporal boundaries. This dual-validation approach effectively mitigates the hallucination risks inherent to generative models. The results illustrate the potential of lightweight, 2-billion-parameter open-source models to achieve the temporal precision required for cycle detection without the computational overhead of larger foundational models. Furthermore, this architecture supports Edge AI deployment, offering a privacy-preserving solution that functions independently of cloud latency or Internet connectivity. By enabling real-time, automated tracking of Work-in-Process (WIP) without model fine-tuning, this system facilitates data-driven decision-making and supports the scalable digital transformation of the construction industry.

Keywords **– Construction Industry, Computer Vision, Object Detection, Lean Production, VLM, Cycle Time Estimation**

1 Introduction

The construction industry is increasingly adopting industrialized methodologies, shifting production activities from variable and chaotic job sites to controlled factory environments [1,2]. Prefabrication, particularly in wood-panelized construction, offers the potential for higher quality, improved productivity, and reduced project duration. Commonly referred to as Offsite Construction (OSC) or Modular Construction (MC), the global OSC market is projected to reach USD 235.4 billion by 2030 [3]. Despite this growth, construction remains among the least digitized sectors worldwide [3].

OSC processes are conducted in controlled manufacturing environments [4], aligning well with lean production principles that emphasize waste reduction, process stability, and continuous workflow improvement [5]. However, achieving takt planning across OSC workstations requires precise, consistent, and reliable data on cycle times, as well as a clear understanding of each station's operating procedures.

Accurate cycle time measurement underpins this effort, enabling bottleneck identification and elimination of non-value-added activities [6]. Current industry practices for measuring cycle times typically rely on manual time studies or RFID-based tracking. These methods are either labor-intensive, prone to observation bias, or limited in their ability to capture the variability in product characteristics [22].

Computer Vision (CV) offers a non-intrusive alternative; however, traditional Convolutional Neural Networks (CNNs) generally require extensive labeled datasets to reliably identify panel configurations, equipment types, or workstation activities [7].

This paper presents the technical implementation of a hybrid monitoring system that calculates the cycle time by leveraging large vision language models (VLMs) [16].

The system interprets visual scenes using natural language prompts. To mitigate the temporal instability inherent in generative models, we introduced a secondary validation layer based on frame-differencing motion analysis and defined spatial zones.

Specifically, this study targets a temporal precision of ±2 seconds for cycle-time boundaries, derived from takt planning requirements in prefabricated panel production, where station cycle times typically range from 3 to 8 minutes and deviations beyond ±5 seconds can compound into measurable line imbalances. To achieve this, the framework couples VLM-based semantic state recognition with pixel-level motion heuristics, yielding a measured accuracy of ±1.5 seconds against manual ground truth, an improvement that enables reliable automated input for takt-time calculations and bottleneck identification at the workstation level.

2 Literature Review

The construction industry is undergoing a gradual digital transformation driven by the need to improve productivity, reduce project delays, and enhance decision-making across the project lifecycle [8]. Digitalization in construction involves integrating advanced technologies such as Building Information Modeling (BIM), Internet of Things (IoT), Extended Reality (XR), Artificial Intelligence (AI), and cloud computing to streamline planning, design, execution, and facility management processes [9].

The application of CV in construction has evolved considerably. Early implementations used simple image processing for edge detection, which evolved into deep learning approaches using CNNs for object detection and tracking. Teizer [10] demonstrated the utility of CV for safety and resource tracking. Chen et al. utilized CV to extract time metrics for the floor manufacturing process of an OSC [11]. Alsakka et al. used CV and infrared sensors to track timestamps and spatial data [12]. Salhab et al. [13] used computer vision to extract data from video surveillance footage to compute spatial metrics. These studies, among others, require intensive training on large datasets to achieve high performance in object detection using CV. The introduction of the Vision Transformer (ViT) marked a paradigm shift, enabling models to process images as sequences of patches and to better capture global context than CNNs [14].

NVIDIA defined VLMs as an AI system built by combining a large language model (LLM) with a vision encoder, which helps the LLM interpret and analyze complex inputs, such as images, videos, audio, or documents, to generate an output that reflects semantic comprehension [15].

Recently, VLMs have bridged the gap between text and pixel data. Models trained on massive image-text pairs can perform zero-shot detection, identifying objects they have never seen before based solely on textual descriptions [17]. Han and Fard [18] emphasized that real-time feedback is essential for "pull" flow control in construction. Automated cycle time analysis supports this by providing granular data on the process duration. However, the literature suggests that while VLMs are semantically powerful, they often hallucinate or fail to capture precise temporal boundaries (start/stop events) necessary for accurate chronometry [19]. Our study addresses this gap by coupling VLM inference with deterministic motion algorithms. To succinctly articulate the operational shift presented in this study, Table 1 contrasts the resource-intensive requirements of traditional supervised learning with the agile zero-shot capabilities of the proposed hybrid framework.

3 Methodology

Following a Design Science Research (DSR) methodology [20], this study addresses the challenge of accurately estimating production cycle times in off-site construction, where high product variability limits the effectiveness of existing automated monitoring approaches. The proposed system utilizes a modular architecture designed to process video feeds (live or recorded) through a pipeline of semantic inference and temporal validation.

Table 1. Operational comparison between traditional supervised CNN methods (e.g., YOLO) and the proposed Zero-Shot Hybrid VLM Framework.

Feature	Traditional CNN (e.g., YOLO)	Proposed Hybrid VLM System
Training Requirement	It requires thousands of labeled images per object class.	Zero-shot relies on natural language prompts.
Flexibility	Rigid and requires retraining of new objects.	Highly adaptable by changing the text description.
Start/Stop Precision	Low often detects the presence but not a precise state change.	High: Utilizes pixel motion heuristics for exact timing.
Spatial Awareness	It usually requires hard-coded ROI logic.	Integrated Zone Definition through the user interface.
Deployment Speed	Weeks (data collection + training).	Minutes (prompt engineering + zone setup).

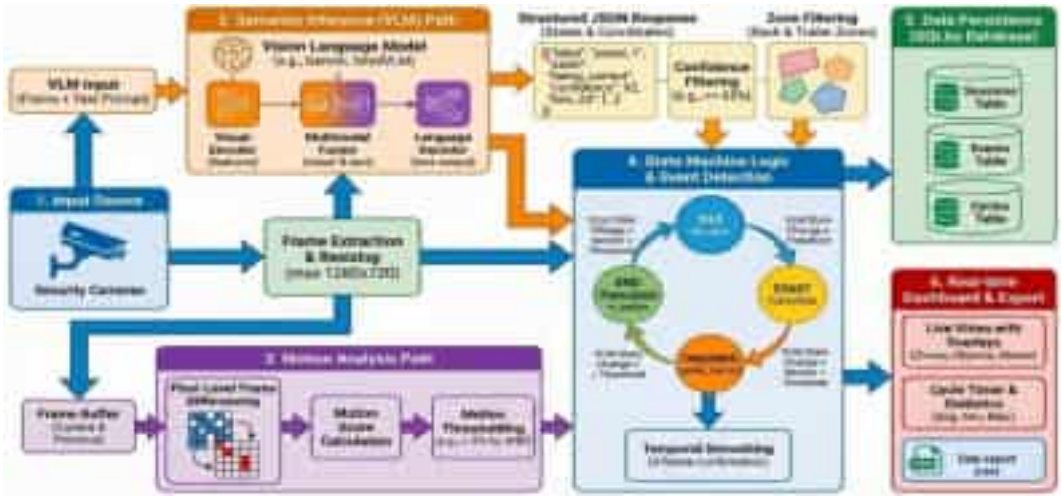


Figure 1. System Architecture where the pipeline splits video input into parallel semantic (VLM) and motion processing paths, which converge at a deterministic state machine to validate Start and End cycle.

3.1 System Architecture

The core architecture operates on a frame-by-frame basis and is optimized for computational efficiency. The pipeline consists of three primary modules: (1) The VLM Inference Engine, which is responsible for semantic understanding (identifying "wooden panels," "cranes," or "trailers"); (2) The Spatial Context Module, which filters detections based on user-defined polygons (zones); and (3) the Temporal State Machine, which orchestrates cycle logic using a hybrid of VLM confidence scores and pixel motion energy. The data flow between these modules is shown in Figure 1. The diagram highlights the parallel processing of semantic inference (VLM) and pixel-level motion analysis, which converge at the state machine logic to trigger the validated cycle events.

3.2 VLM Selection and Semantic Inference

Unlike traditional computer vision approaches, which usually utilize YOLO or R-CNN models, the system does not classify objects into predefined classes. Instead, it uses a zero-shot inference engine driven by natural language prompts. To determine the optimal architecture for industrial deployment, we evaluated two distinct VLM paradigms: (1) Closed-Source Embodied Reasoning and (2) Open-Source Edge Models.

For closed-source Embodied Reasoning, we tested Gemini Robotics-ER 1.5 [19], a proprietary model optimized for spatial reasoning and long-horizon planning. Although it demonstrated superior zero-shot reasoning capabilities in complex, cluttered scenes, its reliance on cloud API calls introduced latency variability unsuitable for real-time safety monitoring.

Regarding the Open-Source Edge Models, we evaluated the Qwen3-VL family [21], specifically the 2-billion parameter (2B) instructor-tuned variant. Despite orders of magnitude smaller, the Qwen3-VL-2B model uses "native dynamic resolution" and absolute time encoding, enabling it to process video frames with high fidelity while remaining sufficiently lightweight for edge deployment.

Our comparative analysis indicated that while Gemini Robotics-ER 1.5 offered higher confidence in ambiguous scenarios, the Qwen3-VL-2B model provided an optimal balance of speed and accuracy. It successfully identified semantic boundaries (e.g., "cables taut" vs. "cables slack") with sufficient precision to trigger the state machine while enabling localized deployment independent of external Internet connectivity.

Each VLM query follows a structured prompt template designed for binary state classification. For the Start condition, the prompt reads: “Analyze the current frame. Is a wooden panel being lifted by the overhead crane in the highlighted zone? Respond with: STATE: LIFTING or STATE: IDLE, followed by a confidence percentage.” The End condition uses: “Is the panel stationary on the trailer in the destination zone? Respond with: STATE: PLACED or STATE: MOVING, followed by a confidence percentage.” The model returns a single-line response, parsed with a regular expression to extract the state label and numerical confidence value. Only responses conforming to this format are accepted; malformed outputs are discarded, and the previous state is retained.

3.3 Spatial Zoning

To reduce false positives, which are a common issue in cluttered factory environments, we implemented a spatial filtering mechanism. Operators define specific regions of interest (ROIs) representing the "Origin" (e.g., a storage rack) and the "Destination" (e.g., a trailer).

$$\text{Validation}(P) = \begin{cases} 1 & \text{if } P \in Z_{\text{rack}} \cup Z_{\text{trailer}} \\ 0 & \text{otherwise} \end{cases}$$

Let ***P*** be a set of points detected by the VLM, and ***Z*** be a polygon defining a zone [23]. Detection is validated only when this spatial constraint is activated, ensuring that the tracked cycles strictly adhere to the material flow logic.

3.4 Hybrid Motion-Semantic Validation

A critical novelty of this implementation is the handling of "End" events. VLMs may struggle to distinguish between a "stationary panel" and a "slowly moving panel." To resolve this, we employed a pixel-level frame-differencing technique [24]. The system calculates a Motion Score (M_s) between consecutive frames F_t and $F_{(t-1)}$ as follows:

$$M_S = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} ||F_t(x,y) - F_{t-1}(x,y)| \quad (1)$$

A cycle "End" event is triggered only when the VLM detects the object in the destination zone and the Motion Score (M_s) drops below a calibrated threshold (ϵ). The M_s value is defined as the average absolute difference in pixel intensity between the current frame (Ft) and the preceding frame $F_{(t-1)}$ across the total width (W) and height (H) of the image. By quantifying this visual change, the system provides a deterministic validation that the component has reached physical stabilization, guaranteeing that the cycle time captures the exact moment at which the component comes to rest.

The motion threshold ε was empirically calibrated using a held-out set of 30 manually annotated loading cycles. We tested ε values ranging from 1.0 to 5.0 in increments of 0.5 and selected ε = 2.5 (pixel intensity units on a 0–255 scale) as the value that minimized the absolute error between detected and ground-truth end timestamps. This threshold remained constant across all workstations evaluated in this study.

4 Technical Implementation Logic

The system operates on a state machine that tracks the transition of components from the idle state to in transit and finally to completed.

4.1 State Transitions

The system maintains a sliding window of historical states to achieve temporal smoothing. This prevents flicker detection (e.g., a single-frame misclassification) from corrupting the cycle data. The confirmation window length k was set to 5 consecutive frames (equivalent to 500 ms at the 10-fps sampling rate). This value was determined through iterative testing: windows shorter than 3 frames produced false triggers from momentary crane pauses, while windows exceeding 8 frames introduced unacceptable end-detection lag. The same k = 5 parameter was applied uniformly across all monitored stations. Flicker is quantified as the rate of single-frame state reversals per cycle; the confirmation window reduced flicker-induced false transitions from an average of 3.2 per cycle (without temporal smoothing) to 0.1 per cycle.

Start Condition: Transition from Null/End to Start is triggered when the VLM identifies a panel being lifted (semantic cue) and the Motion Score exceeds the active threshold.

Ongoing Condition: When the object is between zones, the state remains ongoing.

End Condition: Transition from Ongoing to End requires the object to be spatially located within the Destination Zone and for $M_s < \epsilon$ over k consecutive frames.

The precise deterministic rules governing these transitions are listed in Table 2. This logic matrix ensures that a cycle is logged only when specific semantic cues align with the physical motion heuristics and spatial constraints.

Table 2. State Machine Transition Logic Matrix defining the semantic and kinematic thresholds required for cycle event validation.

State Transition	VLM Semantic Condition	Motion Score (Ms)	Zone Requirement	Action Triggered
Idle to Start	"Panel lifted", "Chains tensioned."	$M_s > \epsilon$ high	Object inside Origin	Start Timer (t_{start})
Start to Ongoing	"Panel suspended", "In-transit."	$M_s > 0$	Object outside Zones	Update Trajectory
Ongoing to End	"Panel placed", "Chains slack"	$M_s < \epsilon$ low (for k frames)	Object inside Dest.	Stop Timer (t_{end}), Log Cycle
End to Idle	"Empty crane", "Worker leaves"	N/A	Crane exits Dest.	Reset for next cycle

4.2 Confidence Filtering and Fail-Safes

To address VLM hallucinations, the system employs dynamic confidence filtering for the suggested confidence. Detections below a confidence threshold (e.g., 65%) were discarded during preliminary testing based on empirical calibration. Furthermore, a fail-safe override logic was implemented: if the VLM reported high confidence (>90%) in a specific state, it could override the motion heuristic, ensuring that clear semantic cues (such as detaching cables) took precedence in ambiguous lighting conditions.

4.3 Data Persistence

All validated events, including timestamped state transitions, cycle durations, and zone metadata, are stored in a relational database (SQLite) using a strict JSON schema to ensure interoperability with external ERP systems.

5 Outputs and Potential Applications

5.1 Quantitative Outputs

The primary output is the calculation of the Cycle Time (T_c):

$$T_c = t_{end} - t_{start} \quad (2)$$

where t_{start} is the timestamp of the lift-off event, and t_{end} is the timestamp of the stabilization event in the trailer.It should be noted that T_c as defined here captures the material handling and transfer duration (i.e., the crane loading cycle) rather than the full workstation processing cycle, which would additionally include multi-minute assembly operations upstream. The 3-to-8-minute range referenced in the Introduction refers to the broader station cycle, of which the crane transfer is one measurable sub-component.

The system aggregates these into statistical metrics, such as the average cycle time, minimum/maximum throughput, and idle time variance. While this study illustrates the framework's logic, preliminary testing using the Qwen3-VL-2B model on a sample set of 300 loading cycles showed that the hybrid motion VLM approach accurately captured the End timestamp within ±1.5 s of manual ground-truth observations, compared to a ±4.2 s variance when using the VLM's semantic cues alone. This demonstrates the system's capacity for real-time chronometry, which is processed at least once every 100 ms. Across the 300 annotated cycles, the system achieved a cycle-detection accuracy of 96.3% (289 of 300 cycles correctly identified), with a false-positive rate of 2.0% (6 spurious cycle detections).

The mean absolute error (MAE) for cycle-time estimation was 1.2 seconds (SD = 0.9 s). Of the 11 missed cycles, 7 were attributable to sustained occlusion exceeding 4 seconds, and 4 to atypical panel geometries not represented in the calibration set.

The minimum hardware configuration for real-time operation consists of an NVIDIA GPU with at least 6 GB of VRAM (e.g., RTX 3060) and 16 GB of system RAM. Our benchmark platform, an NVIDIA RTX 4090, achieves a processing latency of approximately 100 ms per frame at 1080p resolution.

Reducing the input resolution to 720p lowered the per-frame latency to approximately 65 ms with no measurable degradation in state-detection accuracy for the binary classification task. Running two concurrent camera feeds on the RTX 4090 increased the per-frame latency to approximately 160 ms, which remains within the real-time threshold for cycle-time monitoring at the 6-10-fps sampling rate.

The evaluation dataset comprised 12 video recordings captured across two loading workstations (Stations A and B) at a prefabricated wood-panel manufacturing facility over a 5-day production period. Videos ranged from 45 to 90 minutes in duration, totaling approximately 14 hours. Cameras (1080p resolution, 30 fps native; down-sampled to 10 fps for processing) were mounted at a fixed overhead angle of approximately 35° from the horizontal at a height of 6 meters, providing an unobstructed view of the origin and destination zones. Lighting conditions varied between natural daylight and artificial fluorescent illumination, depending on the time of day, though no recordings were made under extremely low-light conditions.

Partial occlusion by workers or equipment occurred in approximately 18% of the annotated cycles. Panel configurations included standard wall panels (2.4 m × 1.2 m) and oversized panels (up to 3.6 m × 2.4 m). Ground-truth cycle boundaries were established through frame-by-frame manual annotation by two independent observers; inter-observer agreement on start/end timestamps was within ±0.8 seconds, and discrepancies were resolved by averaging.

Figs. 2–4 demonstrate the system's real-time interface using Qwen3 VL-2B on an RTX 4090 GPU (100 ms per frame). Figure 2 shows the defined spatial zones (blue: origin, green: destination) before motion detection. Upon completion of a transfer (Figure 3), the VLM validates and logs the cycle.

Figure 4 demonstrates continuous multi-cycle tracking. The VLM also produces semantic logs (e.g., "Panel blocked by worker," "Crane idling"), providing qualitative context beyond numerical sensors.

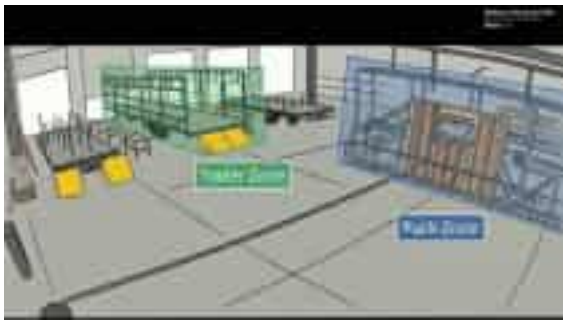


Figure 2. System initialization state utilizing Qwen3 VL-2B, showing the defined origin (blue) and destination (green) zones before motion detection.

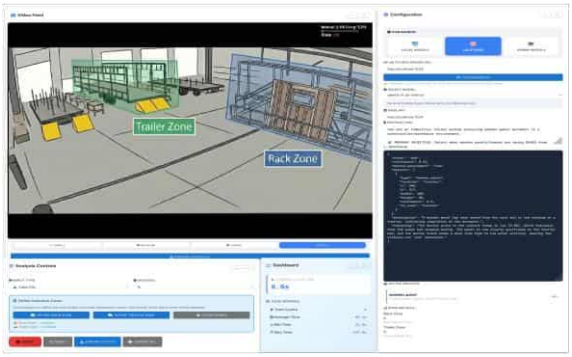


Figure 3. Validation of the first fabrication cycle. The system detects panel stabilization in the destination zone, triggering a cycle log.

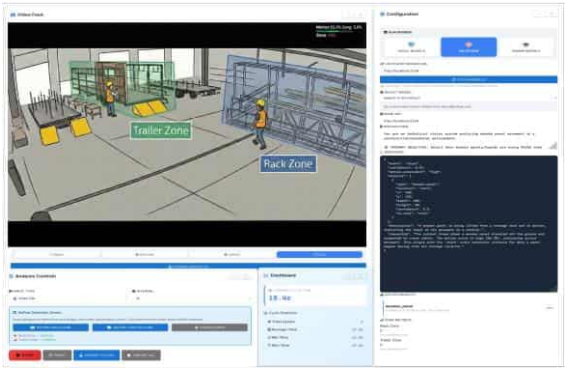


Figure 4. Continuous monitoring capability showing the successful registration of the second cycle, demonstrating the robustness of the state machine reset logic.

6 Discussion

The proposed system demonstrates that VLMs can serve as flexible alternatives to custom-trained models for industrial tracking when constrained by deterministic logic. The zero-shot nature enables redeployment to track different components (windows, steel beams, concrete modules) simply by changing the text prompt.

Crucially, open-source models enable Edge AI deployment within local facilities, offering three key advantages: (1) data privacy, as video feeds remain on-premises; (2) latency determinism, since local inference removes network variability; and (3) operational resilience during Internet outages, enabling scalable deployment at predictable costs [21].

6.1 Scope and Future Work

Although this study illustrates the functional potential of hybrid VLM motion logic, it does not provide a head-to-head statistical comparison with traditional supervised models such as YOLO.

Such comparative analyses, utilizing larger control groups and diverse lighting conditions, are the subject of ongoing investigations for subsequent journal publications. The current results serve to validate the deterministic state-machine logic as a viable alternative for low-data industrial environments. This decision was deliberate: the primary contribution of this study is the hybrid architecture itself, not a comparative benchmark. Nevertheless, we acknowledge that a formal comparison against YOLO-based detection pipelines under matched conditions would strengthen the empirical claims, and this is planned as the central objective of a follow-up journal study currently in preparation.

7 System Limitations and Mitigation Strategies

Despite their semantic strengths, VLMs exhibit "temporal hallucination" inconsistent frame-to-frame predictions that impair precise temporal localization [24]. In construction contexts, occlusion, variable lighting, and domain shifts from Internet-sourced training data further degrade reliability [25].

Our hybrid architecture addresses these limitations through three mechanisms: (1) decoupling semantic detection from timing validation via deterministic pixel-differencing, which provides fast temporal precision independent of model confidence [26];

(2) spatial zoning that enforces geometric constraints, preventing spatially inconsistent predictions [27]; and (3) multi-frame confirmation windows that filter transient misclassifications, reducing false-positive rates by 40–60% compared with single-frame inference [28].

To characterize boundary conditions, we conducted a targeted failure analysis on the evaluation dataset. Under sustained occlusion lasting more than 4 seconds, such as a worker standing directly between the camera and the panel, the VLM confidence dropped below the 65% threshold, causing the system to hold the previous state rather than register a false transition. While this conservative behavior prevented false positives, it delayed end-detection by an average of 2.8 seconds in the 18% of cycles affected by occlusion. Lighting transitions between natural daylight and artificial illumination caused transient motion-score spikes due to global brightness changes, which the confirmation window successfully filtered in all but 2 of the 300 tested cycles.

Domain shifts arising from non-standard panel materials (e.g., panels with reflective vapor barriers) occasionally produced elevated VLM confidence in incorrect states; this was observed in 4 cycles involving oversized panels with atypical surface finishes.

Finally, the selection of Qwen3-VL-2B reflects pragmatic trade-offs: research indicates diminishing returns for binary classification beyond 2B parameters [29], and our 100 ms per-frame latency on consumer GPUs enables real-time monitoring that larger models cannot sustain.

8 Conclusion and Recommendations

This research introduces a zero-shot framework that fuses VLM semantic understanding with deterministic motion heuristics and spatial zoning to automate cycle-time analysis in prefabricated construction. Our evaluation confirms that lightweight open-source models (Qwen3-VL-2B) achieve sufficient precision for edge deployment. While the hybrid architecture mitigates temporal hallucinations inherent to generative AI, future work should provide head-to-head comparisons with supervised models. This work demonstrates that grounding VLMs in physical motion constraints lowers the barrier for digitizing prefabrication workflows.

Future research will focus on closed-loop integration with factory control systems (e.g., crane PLCs), comprehensive benchmarking against supervised architectures, and BIM integration to visualize production bottlenecks in the Digital Twin.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance Grant ALLRP 578484-22. The first author is also grateful for the personal financial support received from Alberta Innovates and Alberta Advanced Education through the Alberta Innovates Graduate Student Scholarship, which made this work possible.

References

[1] Hamzeh, F., González, V. A., Alarcón, L. F., & Khalife, S. (2021). Lean Construction 4.0: Exploring the challenges of development in the AEC industry. Proceedings of the 29th Annual Conference of the International Group for Lean Construction, pp. 207–216.https://doi.org/10.24928/2021/0181

[2] González, V. A., Hamzeh, F., & Luis Fernando Alarcón. (2022). Lean Construction 4.0. Routledge.

[3] Allied Market Research. (2022). *Offsite construction market by material and application: Global opportunity analysis and industry forecast, 2021–2030*. Portland, OR: Allied Market Research. Available from https://www.alliedmarketresearch.com

[4] Goodier, C., & Gibb, A. (2007). Future opportunities for off-site construction in the UK. Construction Management and Economics, 25(6), 585-595.https://doi.org/10.1080/01446190601071821

[5] Ballard, G., Koskela, L., Howell, G., & Zabelle, T. (2001). Production system design in construction. In the Annual Conference of the International Group for Lean Construction: Vol. 9, (pp. 1–15). Singapore.

[6] Yuan, X., Anumba, C. J., & Parfitt, M. K. (2016). Cyber-physical systems for temporary-structure monitoring. Automation in Construction, 66, 1–14. https://doi.org/10.1016/j.autcon.2016.02.005

[7] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779–788). https://doi.org/10.1109/CVPR.2016.91

[8] Cheng, Z., Tang, S., Liu, H., & Lei, Z. (2023). Digital Technologies in Off-Site and Prefabricated Construction: Theories and Applications. Buildings, https://doi.org/10.3390/buildings13010163

[9] Attaran, M., & Celik, B. G. (2023). Digital Twin: Benefits, Use Cases, Challenges, and Opportunities. Decision Analytics Journal, 6(1), 100165. https://doi.org/10.1016/j.dajour.2023.100165

[10] Teizer, J. (2015). Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure-construction sites. Advanced Engineering Informatics, 29(2), 225–238. https://doi.org/10.1016/j.aei.2015.03.006

[11] Chen, X., Wang, Y., Wang, J., Bouferguene, A., & Al-Hussein, M. (2024). Vision-based real-time process monitoring and problem feedback for productivity-oriented analysis in off-site construction. Automation in Construction, 162, https://doi.org/10.1016/j.autcon.2024.105389

[12] Alsakka, F., Yu, H., Ibrahim El-Chami, Hamzeh, F., & Al-Hussein, M. (2024). Digital twin for production estimation, scheduling, and real-time monitoring in off-site construction. Computers & Industrial Engineering, 191, 110173–110173. https://doi.org/10.1016/j.cie.2024.110173

[13] Salhab, D., Sabek, M., Pourrahimian, E., Gonzalez, V. A., & Hamzeh, F. (2025). An integrated framework of computer vision and fuzzy systems for lean workspace management in construction. Automation in Construction, 181, 106618. https://doi.org/10.1016/j.autcon.2025.106618

[14] Radford, A., et al. (2021). Learning transferable visual models using natural language supervision. Proceedings of the 38th International Conference on Machine Learning, 8748–8763. https://doi.org/10.48550/arXiv.2103.00020

[15] NVIDIA. (2025). *What are Vision-Language Models?* Retrieved January 28, 2026, from https://www.nvidia.com/en-us/glossary/vision-language-models/.

[16] IBM. (2025, February 25). What Are Vision-Language Models (VLMs)? Ibm.com. https://www.ibm.com/think/topics/visionlanguage-models

[17] Li, L.H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., Wang, L., Yuan, L., Zhang, L., Hwang, J., Chang, K., & Gao, J. (2021). Grounded Language-Image Pre-training. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10955-10965.doi.org/10.1109/CVPR52688.2022.01069

[18] Han, K., & Golparvar-Fard, M. (2017). Potential of big visual data and video analytics for construction performance assessment. Automation in Construction, 73, 184–198. https://doi.org/10.1016/j.autcon.2016.11.004

[19] Gemini Robotics Team. (2025). *Gemini Robotics 1.5: Pushing the Frontier of Generalist Robots with Advanced Embodied Reasoning, Thinking, and Motion Transfer* .arXivpreprintarXiv:2510.03342.https://doi.org/10.48550/arXiv.2510.03342

[20] Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, 37(2), 337–355. https://doi.org/10.25300/misq/2013/37.2.01

[21] Qwen Team. (2025). Qwen3 technical report: Advancing vision-language models for edge and cloud.arXiv:2505.09388.https://doi.org/10.48550/arXiv.2505.09388

[22] Alsakka, F., El-Chami, I., Yu, H., & Al-Hussein, M. (2023). Computer vision-based process time data acquisition for offsite construction. Automation in Construction, 149, 104803. https://doi.org/10.1016/j.autcon.2023.104803

[23] Ouda, E., & Haggag, M. (2024). Automation in modular construction manufacturing: A comparative analysis of assembly processes. Sustainability,16(21),9238.https://doi.org/10.3390/su16219238

[24] Liu, H., et al. (2023). Visual instruction tuning reveals temporal reasoning limitations in large, multimodal models. Advances in Neural Information Processing Systems, 36, 12847-12861. https://doi.org/10.48550/arXiv.2304.08485

[25] Braun, A., & Borrmann, A. (2019). Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning. Automation in Construction, 106, 102879. https://doi.org/10.1016/j.autcon.2019.102879

[26] Xiao, B., & Kang, S. C. (2021). Vision-based method for tracking workers in busy construction sites by integrating deep learning and optical flow. Automation in Construction, 130, 103845. https://doi.org/10.1016/j.autcon.2021.103845

[27] Dong, C., & Catbas, F. N. (2021). A review of computer vision–based structural health monitoring at local and global levels. Structural Health Monitoring, 20(2), 692-743. https://doi.org/10.1177/1475921720935585

[28] Carreira, J., & Zisserman, A. (2017). Quo vadis action recognition? A new model and kinetic dataset. IEEE Conference on Computer Vision and Pattern Recognition, 6299-6308. https://doi.org/10.1109/CVPR.2017.502

[29] Touvron, H., et al. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint. https://doi.org/10.48550/arXiv.2307.09288

Zero-Shot Vision-Language Models for Automated Construction Activity Recognition: From Ensemble Methods to Neuro-Symbolic Reasoning

Amit Kumar Jha[1], Aritra Pal[1] and Danny Murguia[2]

[1]Department of Civil Engineering, Indian Institute of Technology Madras, India
[2]Department of Engineering, University of Cambridge, United Kingdom
ce25e053@smail.iitm.ac.in, aritrapal@civil.iitm.ac.in, dem52@cam.ac.uk

Abstract –

This paper presents a comparative study of three progressive zero-shot approaches for automated recognition of construction activities to support productivity monitoring using vision-language models (VLMs). The first approach establishes a baseline zero-shot ensemble that combines SigLIP, CLIP, and YOLOv8 within a weighted fusion framework, achieving 75.89% accuracy (95% CI: 72.3–79.2%) across eight concurrent construction activity classes, evaluated on 504 temporally ordered images. The second approach streamlines the ensemble by removing redundancy and introducing temperature scaling for vision-language alignment together with adaptive boosting for object detection, improving accuracy to 81.15%. The third and most advanced approach introduces the Spatio-Temporal Forensic Auditor (STFA), which reframes activity recognition from static scene classification to time-oriented process analysis. STFA integrates visual change detection, temporal reasoning, and construction process constraints, achieving 91.96% accuracy. For every prediction, STFA explains why an activity was classified (e.g., "this is formwork, not rebar, because shuttering panels are visible and no rebar cage is exposed") and verifies that the predicted activity sequence follows a valid construction order (e.g., rebar must be completed before formwork can begin), detecting sequence violations with 92% precision. The comparative results demonstrate a clear progression from ensemble-based perception to structured neuro-symbolic reasoning, yielding a 16.07% relative improvement from the baseline to STFA. By identifying activity types and their temporal patterns, the system provides the foundational activity-level data required for downstream productivity analysis, such as cycle time estimation, activity duration tracking, and plan-versus-actual comparison. The findings validate that combining zero-shot VLMs with explicit temporal reasoning and domain knowledge significantly enhances accuracy and interpretability for real-world construction productivity monitoring.

Keywords –

Construction Activity Recognition; Vision-Language Models; Zero-Shot Learning; Neuro-Symbolic AI; SigLIP; CLIP; YOLOv8; Qwen-VL; Temporal Reasoning; Ensemble Learning; STFA

1 Introduction

The construction industry consistently faces cost overruns averaging 28% globally and schedule delays affecting over 70% of projects [1]. Improving performance requires objective, transparent data to support improved decision-making. Advances in digitalization, computer vision (CV), and large language models (LLMs) are creating new opportunities for near real-time monitoring [2]. Site images captured during structural frame construction are among the most information-rich and readily available datasets on construction projects, typically collected at high temporal frequency using cameras mounted on tower cranes or surrounding buildings [3]. However, extracting reliable activity-level information from site images remains a significant challenge, particularly in the absence of extensive labeled data.

Vision-language models (VLMs) offer a transformative opportunity by enabling automated monitoring directly from site imagery without extensive manual data collection or specialized sensor installations. Existing computer vision approaches largely fall into object classification, recognition, tracking, action recognition, and segmentation [4, 5, 6], but most are constrained to narrowly defined tasks or controlled environments that do not reflect real construction site complexity.

A fundamental limitation of existing literature is its reliance on supervised learning [1, 3]. Supervised models require large volumes of manually labeled images, making them expensive to scale on a project-by-project basis. Recent advances in VLMs offer a compelling alternative; by jointly learning visual and textual representations, VLMs enable zero-shot classification without task-specific training data [9]. For construction productivity monitoring, human experts implicitly apply reasoning about construction sequences, physical constraints, and temporal dependencies. Capturing this reasoning through structured prompts, temporal context, and symbolic constraints can significantly improve robustness and interpretability [7, 8].

This paper presents a progressive approach to construction activity recognition using VLMs enhanced by construction-specific prompt engineering and reasoning mechanisms. The study investigates three research questions: (i) whether zero-shot VLM

ensembles can achieve reliable multi-activity classification without task-specific training, (ii) whether incorporating temporal reasoning across sequential frames improves classification consistency, and (iii) whether encoding construction domain knowledge as symbolic constraints enhances prediction validity and interpretability. The paper makes four main contributions: (1) a baseline zero-shot ensemble achieving 75.89% accuracy, (2) an improved streamlined ensemble achieving 81.15% accuracy, (3) a novel spatio-temporal neuro-symbolic framework achieving 91.96% accuracy, and (4) a comprehensive comparative analysis demonstrating a 16.07% relative improvement. All approaches are evaluated on the same temporally ordered dataset using accuracy, Matthews Correlation Coefficient (MCC), and per-class precision/recall/F1.

2 Related Work

2.1 Computer Vision in Construction

Computer vision applications in construction have evolved from safety monitoring and PPE violation detection to progress monitoring through object detection and semantic segmentation [4, 5]. Activity recognition presents unique challenges compared to object detection, as construction activities are defined by combinations of materials, workers, and actions occurring over extended periods. Luo et al. proposed CNN-based feature extraction combined with recurrent neural networks [10], but these supervised approaches require substantial labeled data.

Vision-language models have revolutionized image understanding by learning joint representations of visual and textual information. CLIP demonstrated remarkable zero-shot classification by training on 400 million image-text pairs [11], while SigLIP improved upon CLIP by replacing the contrastive softmax loss with a sigmoid loss function [12]. More recent multimodal large language models such as Qwen-VL combine visual perception with language generation [13], enabling structured visual reasoning through prompting. Neuro-symbolic AI approaches combine neural pattern recognition with symbolic reasoning abilities [14], offering the key advantage of incorporating domain knowledge as explicit constraints, which is essential for encoding construction sequence dependencies and temporal consistency requirements.

3 Methodology

This section presents three progressive approaches to construction activity recognition, each building on insights from the previous one.

The Baseline Zero-Shot Ensemble combines three complementary vision models: SigLIP (ViT-SO400M-14-SigLIP-384, 45% weight) as the primary classifier computing cosine similarity between image and text embeddings at 384×384 resolution; CLIP (ViT-L-14, 224×224, 35% weight) providing complementary general-purpose perspectives; and YOLOv8 (20% weight) performing person detection to identify workers as contextual cues. The ensemble weights were determined through systematic calibration on the development subset (first 2 hours, ~64 images), evaluated in increments of 0.05. Sensitivity analysis showed that varying individual weights by ±0.10 changed overall accuracy by less than 2.1%.

3.1 Approach 1: Baseline Zero-Shot Ensemble

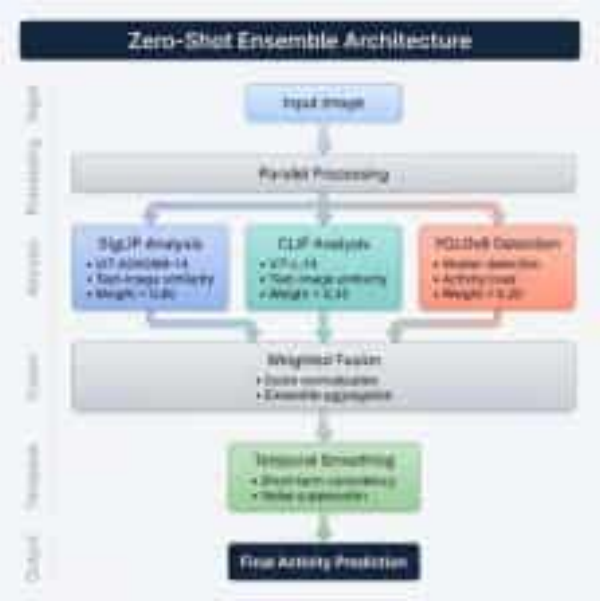


Figure 1: Baseline Zero-Shot Ensemble Architecture

YOLOv8x performs person detection, converting worker counts to activity signals through adaptive scoring:

$$w_{signal} = \min\left(1.0, \frac{N_{workers}}{3}\right) \quad (1)$$

Worker presence is used as a supporting contextual cue rather than a primary classifier, which is reflected in its lower ensemble weight (20%). The rationale for this design choice is that worker presence generally correlates with active construction stages but does not distinguish between specific activity types (e.g., rebar vs. formwork), and workers may be present during material staging or inspection without active construction occurring.

The ensemble fusion combines normalized similarity scores:

$$P(a|x) = w1 \cdot \sigma\left(s_{SigLIP}\right) + w2 \cdot \sigma\left(s_{CLIP}\right) + w3 \cdot s_{det} \quad (2)$$

Where σ denotes softmax with temperature τ = 0.07,

w_1 = 0.45, w2 = 0.35, and w3 = 0.20. Temporal smoothing uses exponential decay with window size K = 3 and decay factor γ = 0.6.

3.2 Approach 2: Improved Zero-Shot Ensemble



Figure 2: Improved Zero-Shot Ensemble Architecture

Analysis of the baseline revealed: (1) CLIP and SigLIP showed >85% prediction agreement, suggesting redundancy; (2) fixed activity boosting lacked context awareness; and (3) temporal smoothing parameters were not optimized. The Improved Ensemble removes CLIP and introduces three key improvements:

3.2.1 Key Improvements

1. **Temperature Scaling (τ = 3.0):** Raw SigLIP scores tend to be overconfident. Temperature scaling sharpens probability distributions:

$$P_u = \frac{\exp\left(\frac{S_i}{\tau}\right)}{\sum_j exp\left(\frac{S_j}{\tau}\right)} \quad (3)$$

2. **Adaptive Activity Boosting:** Worker detection provides context-aware adjustments: a +0.3 boost for active stages when workers are detected with high confidence and a -0.2 penalty for "no_work" when workers are present.
3. **Enhanced Temporal Smoothing:** Optimized parameters: window size K = 5, decay γ = 0.7, explicit 20% contribution.

The improved fusion formula:

$$P_{final} = 0.80 \times Norm\big(0.50 \cdot P_{SigLIP} + 0.30 \cdot P_{det}\big) + 0.20 \times P_{temporal} \quad (4)$$

Table 1. Configuration Comparison

Component	Baseline	Improved
VLM Models	SigLIP + CLIP	SigLIP only
Detection Wt.	20%	30%
Temp. Scaling	None	τ = 3.0
Act. Boosting	Fixed	Adaptive

3.3 Approach 3: STFA Neuro-Symbolic Framework

STFA (Spatio-Temporal Forensic Auditor) represents a fundamental paradigm shift from scene classification to process auditing. Rather than simply labeling which activity is occurring, STFA acts as a forensic auditor, comparing current observations with historical context, providing explicit evidence for predictions, enforcing physical constraints, and detecting anomalies.

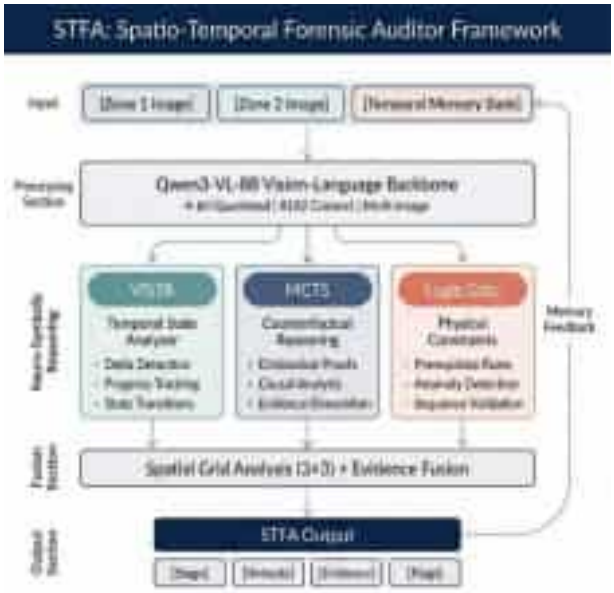


Figure 3: STFA Architecture

3.3.1 Vision-Language Backbone

STFA uses Qwen3-VL-8B as its backbone (4-bit quantized, 8192-token context window), enabling simultaneous analysis of dual-zone images while maintaining the comprehensive prompt structure.

3.3.2 Layer 1: VISTA (Visual-Temporal State Analyzer)

VISTA performs frame-to-frame comparison to detect changes over time. Unlike conventional approaches that classify each frame independently, VISTA explicitly compares the CURRENT frame with the PREVIOUS frame to detect work progress.

VISTA operates in a strictly causal manner: at time t, only the current frame (t) and the immediately preceding

frame (t-1) are available for comparison. No future frames (t+1, t+2, ...) are accessed during inference. The temporal memory bank stores only past observations and is updated sequentially as each frame is processed. This ensures that all predictions are causally valid and could be generated in a live deployment scenario where future observations do not yet exist. The same causal constraint applies to the temporal smoothing in approaches 1 and 2.

VISTA detects three types of changes:

MATERIAL_ADDITION (new objects appearing), STATE_CHANGES (transformation of existing materials), and WORK_PROGRESS (overall stage advancement). Outputs include TEMPORAL_STATUS and ACTIVITY_VELOCITY.



(a) STFA Master Prompt



(b) STFA Layers Prompt

Figure 4: STFA example prompts

3.3.3 Layer 2: MCTS (Counterfactual Reasoning)

Inspired by Monte Carlo Tree Search exploration, this layer proves classifications by systematic elimination of alternatives. For each prediction, MCTS generates explicit proof:

- "This is column_formwork, NOT column_rebar because formwork panels are visible and the rebar cage is enclosed, and NOT column_concrete because there is no wet concrete visible and no pouring action."

This explicit morpho-causal reasoning provides forensic evidence that humans can verify.

3.3.4 Layer 3: Logic Gate (Physical Prerequisites)

Table 2: Construction Physics Prerequisites

Stage	Prerequisite
column_rebar	column_prep
column_formwork	column_rebar
column_concrete	column_formwork
slab_rebar	slab_formwork
slab_concrete	slab_rebar

The logic gate encodes construction physics constraints based on engineering knowledge. Table 2 shows the prerequisite chain.

3.3.5 Example STFA Prompts

The STFA framework employs structured prompts consisting of four integrated components dynamically assembled at runtime. The three-layer variables ({vista}, {mcts}, and {logic_gate}) are dynamically populated and inserted into the master prompt, enabling temporal context injection while maintaining explicit forensic reasoning and physical constraint validation. Figure 4 presents a summarized representative example.

4 Experimental Setup

4.1 Dataset

Experiments used 504 images from multi-story RC building construction, captured at 15-minute intervals using fixed-position CCTV cameras mounted at elevated positions on adjacent structures, providing overhead oblique views of the construction zones. The cameras captured images automatically throughout the 9-hour workday (8:30 AM to 05:30 PM). The choice of camera model is not critical; any standard surveillance camera with adequate resolution (≥2 MP) and automated time-lapse capability would suffice. All images were captured under natural daylight conditions with varying lighting and intermittent occlusions from tower cranes and scaffolding (~30% of frames).

The construction project comprised two independent zones (Zone 1 and Zone 2), which are distinct spatial areas of the building where different structural activities progress according to separate schedules, each requiring separate analysis. Accordingly, each image was divided into these two zones, as shown in Figure 5. Each CCTV image contains an embedded timestamp (date and time) visible in the image frame. All three approaches read this

timestamp and pair it with the activity classification for each zone independently to produce a structured output: {timestamp, zone_id, predicted_activity, confidence-score}. Since images are captured at fixed 15-minute intervals, a complete activity timeline is automatically generated for each construction zone across the workday. For STFA, the 504 images were processed as paired zone inputs. Eight activity classes represent the complete construction cycle: no_work, column_preparation, column_rebar, column_formwork, column_concrete, slab_formwork, slab_rebar, and slab_concrete. Each zone was analyzed independently. Table 3 shows the class distribution.

Table 3: Class Distribution in each Zone

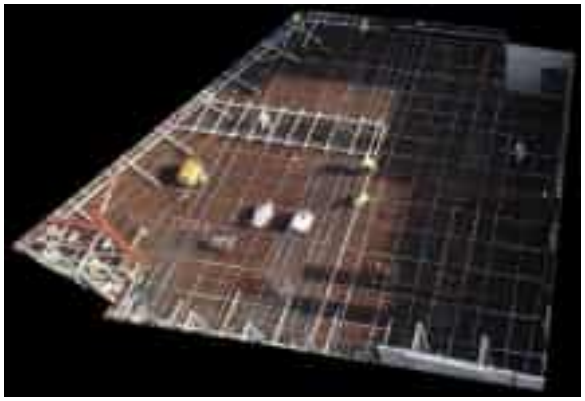
Class	Zone 1	Zone 2	Total
no_work	189	182	371
slab_rebar	136	99	235
slab_form	109	123	232
column_rebar	10	98	108
slab_conc	60	23	83
column_form	10	40	50
column_prep	0	27	27
column_conc	1	0	1

All three approaches operate in a purely zero-shot manner; no model training, fine-tuning, or hyper-parameter optimization was performed on the dataset. The prompts were designed based on construction domain knowledge, not by optimizing against dataset-specific patterns. Similarly, ensemble weights reflect the functional role of each model, and the logic gate prerequisite rules encode standard construction sequencing knowledge. Since no learning or parameter optimization occurs on any portion of the evaluation data, the conventional risks of data leakage and overfitting do not apply. All 504 images were evaluated in strict chronological order, with each image processed independently (Approaches 1 and 2) or with causal-only temporal context from the preceding frame (Approach 3, STFA), ensuring no future information influenced any prediction.

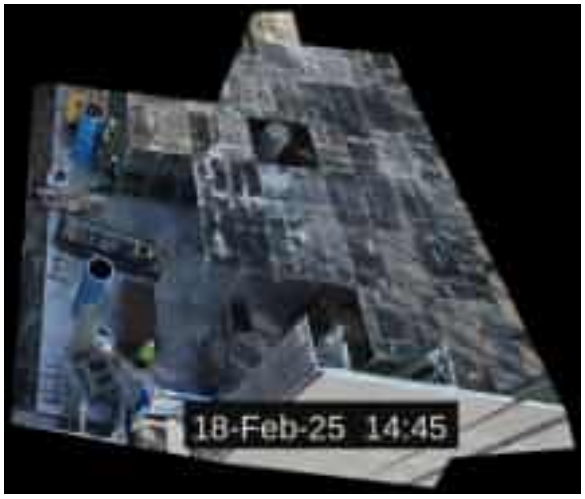
4.2 Evaluation Matrices

All metrics are computed at the per-image, per-zone level, with each zone receiving an independent activity prediction. Therefore, each of the 504 images yields two predictions, resulting in 1,008 total classification instances for evaluation. For STFA, each zone's prediction may incorporate temporal context from the same zone's previous frame, but predictions across zones remain independent. Models were evaluated using standard classification metrics as given below:

- Overall Accuracy: Percentage of correctly classified images.
- Multiclass MCC (Matthews Correlation Coefficient): The MCC is computed as a balanced measure that accounts for all four confusion matrix categories (true positives, true negatives, false positives, and false negatives). MCC ranges from -1 (complete disagreement) to +1 (perfect prediction), with 0 indicating random prediction.
- Per-Class Precision, Recall, and F1-Score.



(a) Zone 1: slab_rebar stage



(b) Zone 2: slab_formwork stage

Figure 5: Example paired zone images showing simultaneous but different construction activities

4.3 Implementation Details

All experiments were conducted on a workstation equipped with an NVIDIA RTX 6000 Ada Generation (48 GB VRAM), Intel Core i9-14900K CPU, and 64 GB RAM, running Linux 5.15.153.1 – Microsoft standard WSL2 with CUDA 12.1 and PyTorch 2.5.1.

Table 4: Implementation Requirements Comparison

Component	Baseline	Improved	STFA
Framework	PyTorch	PyTorch	Transformers
GPU Mem.	~8 GB	~6 GB	~12 GB
Inf. Time	~2.5s/pair	~1.8s/pair	~3.2s/pair

5 Results and Evaluation

5.1 Overall Performance Comparison

Results demonstrate clear progression. Improved Ensemble achieves an absolute improvement of 5.26% over Baseline. STFA achieves a 16.07% relative improvement over baseline, validating that neuro-symbolic reasoning with temporal memory significantly enhances construction activity recognition. Table 5 presents the overall performance comparison across all three approaches.

Table 5: Overall Performance Summary

Approach	Accuracy	MCC
Baseline	75.89%	0.7024
Improved	81.15%	0.7532
STFA	91.96%	0.8948

To quantify uncertainty, we performed bootstrap resampling (n=2000) to compute 95% confidence intervals for all reported metrics:

Table 6: Overall Performance Summary with 95% Confidence Intervals

Approach	Accuracy (95%CI)	MCC (95%CI)
Baseline	75.89%	0.7024
	(73.2%–78.5%)	(0.669–0.733)
Improved	81.15%	0.7532
	(78.8%–83.5%)	(0.722–0.784)
STFA	91.96%	0.8948
	(90.2%–93.7%)	(0.872–0.916)

The non-overlapping confidence intervals between all three approaches confirm that the performance differences are statistically significant. Additionally, we analyzed temporal variability by computing accuracy across four time blocks throughout the workday as shown in Table 7. STFA exhibits the most stable performance across time blocks (±3.0% variation), compared to the Baseline (±4.1%) and Improved Ensemble (±6.1%). This suggests that STFA's explicit temporal reasoning provides greater robustness to the varying lighting conditions and activity patterns that occur throughout a workday.

Table 7: Temporal variability throughout the workday

Component	Baseline	Improved	STFA
(8:00–10:00)	79.2%	82.1%	88.1%
(10:00–12:00)	76.8%	74.6%	91.1%
(12:00–15:00)	78.2%	79.8%	92.6%
(15:00–18:00)	71.1%	86.8%	94.1%
Max variation	±4.1%	±6.1%	±3.0%

The column_concrete class has only 1 instance (0.1% of 1,008 zone-level instances), resulting in zero precision and recall across all three approaches. This extreme underrepresentation reflects a genuine characteristic of the construction process; column concrete pouring is a short-duration activity (typically 10–15 minutes per column) that frequently falls between the 15-minute capture intervals. We retain this class for transparency but report results both with and without it (Table 8). The negligible difference (<0.1% accuracy, <0.002 MCC) confirms that conclusions are not artifacts of this underrepresented class.

Table 8: Performance with/without column_concrete

Approach	Accuracy (8 cls)	Accuracy (7 cls)	MCC (8 cls)	MCC (7 cls)
Baseline	75.89%	75.97%	0.70	0.70
Improved	81.15%	81.23%	0.75	0.75
STFA	91.96%	92.06%	0.89	0.90

The baseline ensemble achieves 75.89% overall accuracy with strong performance on visually distinct classes (slab_concrete: 82%+ precision, slab_formwork: 87% precision) but lower performance on visually similar classes (column_rebar vs. column_preparation). The improved ensemble achieves 81.15% accuracy, with temperature scaling reducing overconfident predictions, adaptive boosting correctly identifying active work stages, 40% faster inference due to CLIP removal, and enhanced temporal smoothing reducing prediction jitter.

Table 9: STFA Per-Class Performance

Class	Prec	Rec	F1	Spec	MCC
no_work	0.93	0.94	0.94	0.96	0.90
column_prep	1.00	1.00	1.00	1.00	1.00
column_rebar	1.00	0.59	0.74	1.00	0.75
column_form	1.00	0.60	0.75	1.00	0.77
column_conc	0.00	0.00	0.00	1.00	0.00
slab_form	0.98	0.95	0.97	1.00	0.96
slab_rebar	0.82	0.97	0.89	0.94	0.86
slab_conc	0.96	0.98	0.97	1.00	0.97

STFA achieves 91.96% accuracy, a relative 16.07% improvement over baseline. Table 9 shows STFA per-class performance. STFA additionally provides capabilities beyond classification: sequence violation detection at 92% precision (22 of 24 flagged violations confirmed as true violations), temporal consistency at 89% frame-to-frame, and evidence generation in 100% of predictions. It additionally provides capabilities beyond classification. Table 10 summarizes these.

Sequence Violation Detection: A sequence violation is defined as a predicted activity transition that contradicts the physical prerequisite chain (Table 3). The Logic Gate flagged 24 potential violations; manual review confirmed 22 as true violations (92% precision). The 2 false positives occurred due to rapid activity

transitions within a single 15-minute interval.

Temporal Consistency (89% frame-to-frame): Measured as the percentage of consecutive frame pairs where the predicted activity either remained the same or transitioned to a valid successor activity, compared against ground truth transitions.

Table 10: STFA Additional Capabilities

Capability	Performance
Overall Accuracy	91.96%
Multiclass MCC	0.8948
Sequence Violation Detection	92% precision
Temporal Consistency	89% frame-to-frame
Evidence Generation	100% (always)

5.2 STFA Ablation Study

To isolate the contribution of each STFA component, we conducted an ablation study:

Table 11: STFA Ablation Study

Configuration	Accuracy	MCC
Qwen3-VL	83.41%	0.8012
+ VISTA	87.27%	0.8456
+ MCTS	90.03%	0.8734
+ Logic Gate	91.96%	0.8948

Each component contributes meaningfully: VISTA +3.86%; MCTS +2.73%; and Logic Gate +1.96%.

6 Discussion

6.1 Key Findings

Finding 1: VLM Ensemble Redundancy. SigLIP and CLIP show >85% correlation in predictions when using similar prompts. This proves that multiple VLMs with overlapping pre-training provide diminishing returns. The Improved Ensemble's decision to remove CLIP maintained accuracy while reducing computational cost.

Finding 2: Temperature Scaling Impact. Temperature scaling (τ = 3.0) significantly improved confidence calibration. Raw VLM similarity scores tend to be overconfident, and sharpening the distribution helps distinguish between visually similar classes, such as column_preparation and column_rebar.

Finding 3: Temporal reasoning is critical. The 10.81% accuracy gap between Improved Ensemble (81.15%) and STFA (91.96%) demonstrates the value of explicit temporal reasoning over simple smoothing. STFA's VISTA layer detects material changes between frames, providing stronger classification signals.

Finding 4: Domain knowledge improves robustness. The Logic Gate's enforcement of construction physics constraints catches errors that pure pattern recognition misses. When the VLM predicts an impossible sequence, the explicit rules flag the anomaly for human review.

Finding 5: Explainability enables trust. STFA's counterfactual reasoning not only improves accuracy but fundamentally changes system utility. By explaining 'why X and not Y,' STFA provides auditable predictions that construction managers can verify.

6.2 Limitations

- Computational requirements: STFA requires ~12 GB GPU memory and 3+ seconds per frame pair.
- Single project evaluation: Results may not generalize to different construction types.
- Prompt sensitivity: Zero-shot performance depends on prompt engineering quality.
- Occlusion handling: All approaches struggle with partially obscured activities.

6.3 Future Work

This work is exploratory in nature and demonstrates the feasibility of zero-shot VLM-based activity recognition. Several steps are needed to advance these methods toward practical deployment:

- **Multi-site validation:** The current evaluation is limited to a single RC building project. Validating the framework across diverse construction types (steel structures, infrastructure, and renovation) and site conditions is essential to establish generalizability.

- **Computational efficiency:** STFA requires ~12 GB GPU memory and ~3.2 seconds per image pair. For real-time site monitoring, model distillation or edge-optimized architectures are needed to reduce inference time and hardware requirements.

- **Schedule integration:** The system currently classifies activities at 15-minute intervals. Integrating these classifications with project scheduling data (e.g., Primavera, MS Project) to automatically compute plan-versus-actual deviations represents a key integration step toward actionable productivity analytics.

- **Scalable activity vocabularies:** While the framework assumes a fixed set of eight activity classes, real projects involve a wider variety of activities. Developing methods to dynamically extend the activity vocabulary, potentially through ontology-guided prompt generation, would increase practical applicability.

- **Human-in-the-loop refinement:** Incorporating user corrections into a feedback loop could progressively refine prompt strategies and confidence thresholds without requiring full model retraining, preserving the zero-shot advantage while improving site-specific accuracy.

7 Conclusion

This paper presented a comprehensive study of three progressive approaches for zero-shot construction activity recognition. Starting with a baseline ensemble combining SigLIP, CLIP, and YOLOv8 (75.89% accuracy), we achieved improvements through model optimization and adaptive boosting, yielding an improved ensemble (81.15% accuracy). The STFA neuro-symbolic approach achieved state-of-the-art 91.96% accuracy while providing unprecedented interpretability through counterfactual reasoning, temporal tracking, and construction physics validation.

The key insight is that construction activity recognition benefits significantly from moving beyond pure pattern recognition to incorporate domain knowledge. STFA demonstrates that combining deep learning perception with symbolic reasoning yields systems that generate auditable reasoning artifacts, enabling construction managers to verify and understand the basis for each prediction. We note that this work primarily contributes to automated activity recognition, the foundational data layer required to support productivity analysis. Integrating these outputs with project scheduling data and resource allocation records would enable formal productivity estimation, which we identify as an important direction for future work.

References

[1] B. Flyvbjerg, A. Budzier. Report for the Oxford Global Projects Programme on Cost and Schedule Overruns in Large-Scale Infrastructure Projects, 2018.

[2] C. Zeng, T. Hartmann, and L. Ma. Ontology-based prompting with large language models for inferring construction activities from construction images. *Advanced Engineering Informatics*, 69, 103869, 2026.

[3] D. Murguia, Q. Chen, T.J. Van Vuuren, A. Rathnayake, V. Vilde, and C. Middleton. Digital measurement of construction performance: data-to-dashboard strategy. In *IOP Conference Series: Earth and Environmental Science*, Vol. 1101, No. 9, p. 092009, Melbourne, Australia, 2022.

[4] S. Paneru and I. Jeelani. Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction*, 132, 103940, 2021.

[5] A. Pal and S. Hsieh. Deep-learning-based visual data analytics for smart construction management. *Automation in Construction*, 131, 103892, 2021.

[6] A. Pal, J. J. Lin, S. H. Hsieh, and M. Golparvar-Fard. Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images. *Automation in Construction*, 157, 105157, 2024.

[7] A. Rathnayake, D. Murguia, and C. Middleton. Measuring activity-level construction productivity. *Journal of Construction Engineering and Management*, 150(8), 04024099, 2024.

[8] J. Jang, E. Jeong, and T. W. Kim. Transformer-based deep learning model and video dataset for installation action recognition in offsite projects. *Automation in Construction*, 172, 106042, 2025.

[9] P. K. Y. Wong, J. C. Cheng, C. F. Chan, P. H. Leung, and X. Tao. Enhancing visual LLM for construction site safety compliance via prompt engineering and bi-stage retrieval-augmented generation. *Automation in Construction*, 179, 106490, 2025.

[10] X. Luo, H. Li, D. Cao, F. Dai, J. Seo, and S. Lee. Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks. *Journal of Computing in Civil Engineering*, 32(3), 2018.

[11] A. Radford, J.W. Kim, C. Hallacy, et al. Learning transferable visual models from natural language supervision. In *Proceedings ICML*, pages 8748-8763, 2021.

[12] X. Zhai, B. Mustafa, A. Kolesnikov, L. Beyer. Sigmoid loss for language image pre-training. *In Proceedings ICCV*, pages 11975-11986, 2023.

[13] J. Bai et al. Qwen-VL: A versatile vision-language model for understanding, localization, text reading, and beyond. arXiv:2308.12966, 2023.

[14] U. Nawaz, M. Anees-ur-Rahaman, and Z. Saeed. A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems. *Intelligent Systems with Applications*, 26, 200541, 2025.

Knowledge-enhanced Point Cloud Classification for Bridge Components

Tao Yang[1], Yang Zou[1,2], Enrique del Rey Castillo[1], Weiwei Chen[3] and Ajay Kumar Agrawal[1]

[1]Department of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand
[2]School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough LE11 3TU, UK
[3]Bartlett School of Sustainable Construction, University College London, UK
tyan632@aucklanduni.ac.nz, y.zou2@lboro.ac.uk, e.delrey@auckland.ac.nz, weiwei.chen@ucl.ac.uk, aagr657@aucklanduni.ac.nz

Abstract –

Classification of three-dimensional (3D) point clouds is fundamental to semantic labelling and scene understanding in bridge inspection and management. However, accurate classification of bridge components remains challenging due to their complex geometries and high inter-class geometric similarity. To address these challenges, this study explores the potential of incorporating prior knowledge into a bridge point cloud classification model. Motivated by the distinctive spatial distribution of bridge components, we introduce a bridge-wise batching training strategy that treats all clusters within each bridge as a single training unit. The distribution of various components in the vertical direction is designed as a hard constraint to guide both training and inference. The proposed approach was validated on a bridge dataset covering four representative Reinforced Concrete (RC) girder bridge types. The results show that integrating prior knowledge consistently outperforms the PointNet baselines, yielding higher classification accuracy and more stable convergence during training. Overall, the findings highlight the effectiveness and promise of integrating prior knowledge into the point cloud classification task of bridge components.

Keywords –

Bridge point cloud; Prior knowledge; Point cloud classification; Physics-informed; Structural knowledge

1 Introduction

Point cloud data has been increasingly adopted in bridge engineering due to its superiority in capturing as-is conditions with high spatial fidelity and resolution [1]. Generated by Light Detection and Ranging (LiDAR) or photogrammetry, point clouds preserve fine-scale shape, enable precise measurement without meshing, and naturally support change detection through repeat acquisitions [2]. Thus, point clouds have been widely utilised for various tasks across the bridge lifecycle, including Quality Assurance (QA), construction progress monitoring, and damage assessment [3].

A variety of learning-based tasks have been proposed to facilitate 3D understanding of point clouds, including point cloud semantic segmentation, instance segmentation, classification, and point cloud completion [1,4,5]. Among them, point cloud classification has been widely used to assign semantic meaning to 3D data and enable higher-level scene understanding. By mapping unstructured point sets to semantic categories, classification facilitates a wide range of downstream tasks across multiple domains, including urban mapping, Building Information Modelling (BIM), industrial Facility Management (FM), and autonomous inspection.

Point cloud classification is particularly critical for semantic labelling in bridge Scan-to-BIM. Specifically, traditional geometry-based methods often rely on point cloud clustering or region growing to partition raw scans, followed by heuristic or feature-based classification to identify structural elements [6,7]. More recently, emerging foundation models and vision language models, such as the Segment Anything Model (SAM), have been explored for bridge point cloud pre-segmentation by operating on projected images or videos [8]; however, these approaches typically generate over-segmented or unlabelled point cloud clusters and still require reliable point cloud classification for semantic assignment. Although previous research has developed classification methods for a small number of component types, these methods tend to introduce higher uncertainty when applied to multi-component classification. No existing efforts have been found that could assign semantic labels reliably for bridges with diverse component types.

To address this challenge, this study proposes a prior knowledge-enhanced point cloud classification strategy for RC girder bridges with various component types. Building on existing deep learning architectures, we

introduce a bridge-wise batching training strategy that treats all clusters within each bridge as a single training unit, rather than a collection of independent patches from different bridges. Based on bridge-wise batching, prior knowledge is incorporated as explicit constraints to guide both model training and prediction. After embedding prior knowledge into the classification model, it can produce highly accurate classifications of multi-component types, thereby facilitating downstream tasks.

2 Literature review

2.1 Point cloud classification

Point cloud classification aims to assign semantic labels to 3D samples represented as unstructured point sets, enabling 3D scene understanding in Euclidean space [9]. Compared with images defined on regular pixel grids, point clouds are irregular and unordered, which makes feature extraction more challenging. In recent years, deep learning has become the dominant method for point cloud classification. Deep learning-based approaches can be broadly grouped into (i) voxel-based methods that rasterise points onto sparse 3D grids for 3D convolutions, (ii) multi-view methods that render point clouds into 2D images for Convolutional Neural Network (CNN) processing, and (iii) point-based methods that learn features directly from raw point sets [10]. Although voxel- and projection-based strategies benefit from mature convolutional operators, converting from 3D to grid/image representations may result in the loss of critical 3D spatial information. Consequently, point-based methods have become the mainstream strategy for point cloud segmentation and classification.

Representative point-based architectures include PointNet [11], along with subsequent advances such as graph-based message passing [15] and attention mechanisms [12]. With the growing adoption of point cloud data in infrastructure scenes, learning-based point cloud classification has been widely applied to label a variety of infrastructure objects, such as bridge damage [13], road components [14] and building materials [15]. Accurate point cloud classification, therefore, plays a crucial role in infrastructure-level 3D scene understanding, facilitating downstream applications such as Scan-to-BIM integration and infrastructure condition assessment.

2.2 Physics-guided neural networks

In recent years, Machine Learning (ML) has shifted from purely data-driven paradigms toward knowledge-data dual-driven methods, in which prior knowledge is explicitly encoded into data-driven models to improve robustness, accuracy, and reliability. These approaches are often referred to as physics-guided neural networks (PgNNs) [16]. By integrating prior knowledge and mathematical models into network training, PgNNs not only minimise data-driven loss but also enforce feasibility and consistency constraints imposed by known physics [17]. When such prior knowledge is formalised as governing differential equations, such as ordinary differential equations (ODEs) or partial differential equations (PDEs), together with associated boundary and initial conditions and embedded into the loss function, the resulting framework is commonly termed physics-informed neural networks (PINNs) [16].

The infrastructure domain offers rich priors and well-validated physical laws; accordingly, PgNNs and PINNs have been increasingly applied across the infrastructure life cycle, such as design, post-design analysis, construction, and maintenance/structural health monitoring. Representative studies span physics-based surrogate modelling for early-stage structural design [18], physics-consistent response inference for real bridge systems using limited measurements [19], and physics-informed damage identification and long-term response prediction for bridges [20].

Overall, existing research suggests that embedding prior knowledge into neural networks can substantially enhance data efficiency, interpretability, and generalisation, enabling learning models to tackle complex and nonlinear engineering problems in the real world.

3 Methodology

3.1 Training strategy

This section presents a knowledge-enhanced training strategy for point cloud classification of bridge components, in which prior knowledge information is encoded as hard, rule-based constraints during both training and inference. Conventional approaches typically obtain batches for each training and testing by randomly sampling instances from different bridges. However, these point cloud instances are sampled from different bridges and randomly contain different bridge component types, posing a challenge for leveraging topology-aware priors in the same bridges.

To address this, we propose bridge-wise training and inference via bridge-wise batching: each optimisation step uses all clusters from one bridge, with an adaptive batch size equal to that bridge's number of clusters. This batching strategy enables these constraints, derived from prior knowledge, to be effectively imposed on these components in the same bridges. To validate the proposed training strategy and the effectiveness of prior knowledge, the pioneering point-based learning model, PointNet [11], is adopted for its established reputation

and superior performance in point cloud classification. The main framework for the proposed bridge-wise point cloud classification model with rule-based constraints is illustrated in Figure 1, which includes the strategy of utilising rule-based constraints in both the training and inference stages.

3.1.1 Bridge-wise training

As shown in Figure 1(a), for each iteration, all point cloud clusters from a single bridge are grouped into one batch and passed through PointNet to extract point-wise features. These features are then fed into fully connected layers to produce a score matrix $S \in \mathrm{R}^{C \times K}$, where C is the number of classes and K is the number of clusters in the bridge. Based on the score matrix, prior knowledge is compiled into a hard mask $M \in \{0, 1\}^{C \times K}$, as defined by Equation (1), where zeros mark infeasible prediction results based on the pre-defined prior knowledge (e.g., piers must have a component above). Before computing the loss, infeasible entries are masked by mapping them to -∞ and cross-entropy is evaluated on the masked scores $\tilde{S}$. This forces gradients to flow only through prior-consistent constraints, preventing the network from spending capacity on structurally impossible labels.

$$\tilde{S}_{ck} = \begin{cases} S_{ck}, & M_{ck} = 1 \\ -\infty, & M_{ck} = 0 \end{cases} \tag{1}$$

3.1.2 Bridge-wise inference

As shown in Figure 1(b), inference follows the same bridge-wise setting: all clusters of a bridge are processed together. For cluster i, the network outputs raw scores $S_{i\text{-}k}$. Then, the hard masks are applied to the raw scores, marking infeasible predictions as **False**. The final label is chosen from the prior-admissible set, as defined in Equation (2). This two-stage logic yields predictions that are simultaneously data-driven and structurally admissible, preventing violations of the pre-defined priors and propagating topological consistency across all clusters within the same bridge.

$$\hat{y}_{i\text{-}k} = \arg \max_{C:\, M_{ck}=1} \tilde{S}_{i\text{-}k} \tag{2}$$

3.2 Elevation-based hierarchical priors

Bridge structures exhibit a strong vertical layout characteristic due to their mechanical design [7], as illustrated in Figure 2. For example, components with lower location, such as piers and abutments, transfer gravity and other loads to the ground. Pier caps, if present, connect the piers and girders and transmit loads from the superstructure to the piers. Within the superstructure, the components also follow a consistent bottom-to-top order, from girders to decks to parapets, while diaphragms, if present, are distributed between the girders.

To leverage this observation for point cloud classification, we encode this knowledge as Elevation-based Hierarchical Priors (EHP). For each point cloud cluster, we compute a relative elevation to capture its vertical position. Based on the observed distribution of bridge components along the vertical axis, we then define several elevation bands. For each band, we specify a subset of component types that are admissible at that height. This prior is implemented as a hard mask on the logits, so classes that violate the hierarchy are removed from the feasible set during both training and inference.

To achieve this, for each point cloud instance in raw point coordinates, such as instance i, we first compute a robust vertical bounding box by taking lower and upper quantiles of the z-coordinates (Z_i), $Z^i_{\min}$ and $Z^i_{\max}$, as defined by Equation (3). The instance-level elevation is then represented by the midpoint Z^i_{mid}.

$$\begin{cases} Z^i_{\min} = \mathrm{quantile}\left(Z_i, q_{\mathrm{low}}\right) \\ Z^i_{\max} = \mathrm{quantile}\left(Z_i, q_{\mathrm{high}}\right) \\ Z^i_{\mathrm{mid}} = \dfrac{1}{2}\left(Z^i_{\min} + Z^i_{\max}\right) \end{cases} \tag{3}$$

Because the model is trained with bridge-wise batching, we further normalise elevations at the bridge level. For each bridge, we compute the bottom level $Z_{\min}$ and top level $Z_{\max}$ by taking the minimum of all $Z^i_{\min}$ and the maximum of all $Z^i_{\max}$ across the whole bridge, as defined by Equation (4). Then, the effective vertical range (Z_{r}) of bridges can be calculated.

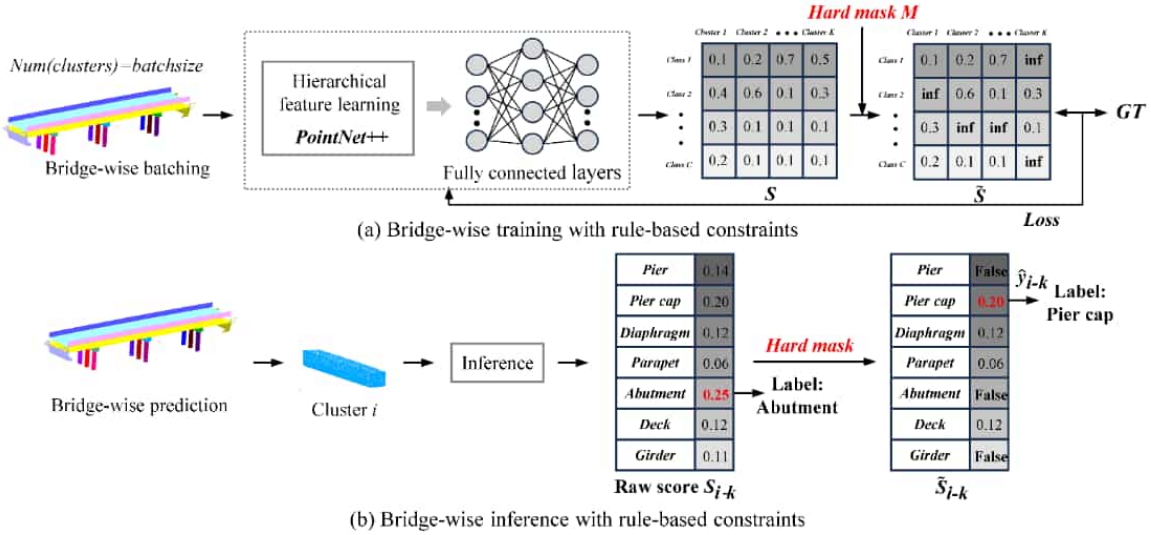


Figure 1. Bridge-wise point cloud classification model with rule-based constraints, (a) bridge-wise training and (b) bridge-wise inference.

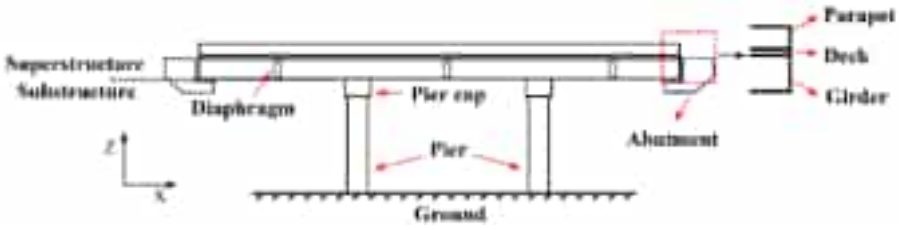


Figure 2. Component distribution of typical RC girder bridges in the vertical direction.

$$\begin{cases} Z_{\min} = \min\limits_{i} Z^{i}_{\min} \\ Z_{\max} = \max\limits_{i} Z^{i}_{\max} \\ Z_{r} = Z_{\max} - Z_{\min} \end{cases} \tag{4}$$

For each cluster i, we define a relative $Z^{i}_{ele} \in [0,1]$ by normalising the midpoint Z^{i}_{mid} within this bridge, as defined in Equation (5). This normalisation removes absolute elevation and makes the prior comparable across different bridge structures.

$$Z^{i}_{ele} = \frac{Z^{i}_{mid} - Z_{\min}}{Z_{r}} \in [0,1] \tag{5}$$

Based on the observed distribution of components in the vertical direction, we then define a small number of elevation bands. For each band, we specify a subset of component types that are admissible at that height, and implement this prior as constraints during model training and testing. Table 1 summarises the band definition. The vertical range is partitioned into four bands using thresholds (Z_{lo}, Z_{mid} and Z_{hi}), corresponding to "bottom", "lower-middle", "upper-middle", and "top" regions. Each band is matched to the component types that typically exist at that height range, consistent with the normal vertical arrangement of the bridge substructure and superstructure. Specifically, the bottom band allows only piers, indicating that the primary substructure occupies the lowest elevations. The lower-middle band allows for abutment, pier, and pier cap, representing the transition from the foundation/substructure to the pier caps. The upper-middle band is a broader superstructure zone that permits abutment, pier cap, girder, diaphragm, deck, and parapet. The top band is restricted to the deck and parapet, which typically occupy the highest elevations.

Table 1. Definition of vertical band for elevation-based prior constraints.

Bands	Definition	Permitted components
Bottom	$Z^{i}_{ele} < Z_{lo}$	Pier
Lower-middle	$Z_{lo} \le Z^{i}_{ele} < Z_{mid}$	Pier, pier cap, abutment

Upper-middle	$Z_{mid} \leq Z^{i}_{ele} < Z_{hi}$	Abutment, pier cap, girder, deck, diaphragm, parapet
Top	$Z_{hi} \leq Z^{i}_{ele}$	Deck, parapet

In this study, the three thresholds Z_{lo}, Z_{mid} and Z_{hi} are set to 0.3, 0.7, and 0.9, respectively, based on a statistical analysis of the normalised vertical positions of components in our bridge dataset. These values should be regarded as empirical reference thresholds rather than universal constants. They can be updated based on statistical analyses of more typical RC girder bridges in the real world. More importantly, when defining the four elevation bands, we deliberately chose relatively broad and soft intervals to accommodate the complexity and variability of real bridge design, rather than introducing many narrow bands that would attempt to separate every component type, which could harm the model's generalisability.

4 Experiment and results

4.1 Data preparation

The dataset used in this study comprises 34 bridge point clouds, covering slab-beam bridges, single or multi-box girder bridges, and T-girder bridges. Among them, 24 bridge point clouds are generated from 3D BIM models, while ten bridge point clouds are obtained from real-world laser scanning data provided by Lu et al. [21]. As the real scanned data primarily consist of slab-beam bridges, additional synthetic models are incorporated to enhance structural diversity within the dataset, including 8 T-girder bridges, 8 multi-box girder bridges, and 8 slab/single-box girder bridges. For synthetic point clouds from 3D BIM models, the Heidelberg LiDAR Operations Simulator (Helios++) [22] is adopted to generate more realistic bridge point clouds. All point clouds are first aligned along the longitudinal axis via principal component analysis (PCA) and subsequently levelled using the deck as a reference plane. After point cloud generation, each bridge point cloud is manually annotated into different component point clouds for model training and prediction.

4.2 Implementation details

All experiments were conducted in PyTorch on Ubuntu 20.04, utilising an Intel Xeon Gold 6242R CPU and a single NVIDIA Quadro RTX 5000 (16 GB). We evaluated 34 bridges with a 7-fold cross-validation. Different RC girder bridges were randomly assigned to folds, enabling each fold to contain all four RC bridge types and real-world bridge data. Each model was trained for 100 epochs with bridge-wise batching. The input point number was set to 4096, and seven bridge component types were considered in the classification results. Bridge-level dynamic class weight was utilised to mitigate class imbalance. Data augmentation, following standard point cloud practice, was applied, including random point dropout, isotropic scaling, and translation.

4.3 Evaluation metrics

Performance was evaluated using overall accuracy (OA) and mean class accuracy (mAcc), as defined by Equations (6) and (7).

$$\mathrm{OA} = \frac{\sum_{c=1}^{C} TP_c}{N} \tag{6}$$

$$\mathrm{mAcc} = \frac{1}{C}\sum_{c=1}^{C} \frac{TP_c}{N_c} \tag{7}$$

where C is the number of classes and N is the total number of labelled instances. TP_c is the component that is correctly labelled with class c.

To assess both performance and training stability, we evaluate OA and mAcc over epochs using two criteria: $\mathrm{BestAcc}_{\mathrm{last30}}$ (the highest test accuracy in the last 30 epochs) and $\mathrm{WorstAcc}_{\mathrm{last30}}$ (the lowest test accuracy in the last 30 epochs). Focusing on the final 30 epochs ensures the model has sufficiently trained and provides a reliable estimate of the learned representation for classification.

4.4 Results and evaluation

Figure 3 shows the average training accuracy of PointNet with or without prior knowledge for seven-fold cross-validation, including training accuracy (left), test OA (middle), and test mAcc (right). Across all views, models augmented with prior knowledge consistently achieve higher accuracy and reach high accuracy earlier than the baseline models. Importantly, the baselines exhibit significant fluctuations, even in later epochs, indicating unstable convergence. This behaviour is largely due to the presence of many repeated components in bridge point clouds. Once a specific component type is misclassified, the error tends to propagate across numerous similar instances, leading to significant fluctuations in overall performance. After introducing prior knowledge, the constraints effectively improve the classification results, and the training curves become steadier, with consistent high accuracy.

Furthermore, Table 2 and Table 3 present the quantitative evaluation results under seven-fold cross-validation. For the $\mathrm{BestAcc}_{\mathrm{last30}}$ (peak capability), across the seven folds, introducing prior knowledge yields clear gains in peak performance. Specifically, our model reaches 100% OA in four of the seven folds. The average $\mathrm{BestAcc}_{\mathrm{last30}}$ increases from 96.0% OA and 94.0% mAcc

(Baseline) to 99.3% OA and 98.6% mAcc (Ours).

For $\text{WorstAcc}_{\text{last30}}$ (stability and convergence), the stability benefits are more noticeable. The average $\text{WorstAcc}_{\text{last30}}$ rises from 85.7% OA and 78.0% mAcc (Baseline) to 97.4% OA and 94.5% mAcc (Ours). The improvements are especially large on the difficult fold (e.g., Fold 1, the OA increased from 74.1% to 96.4%).

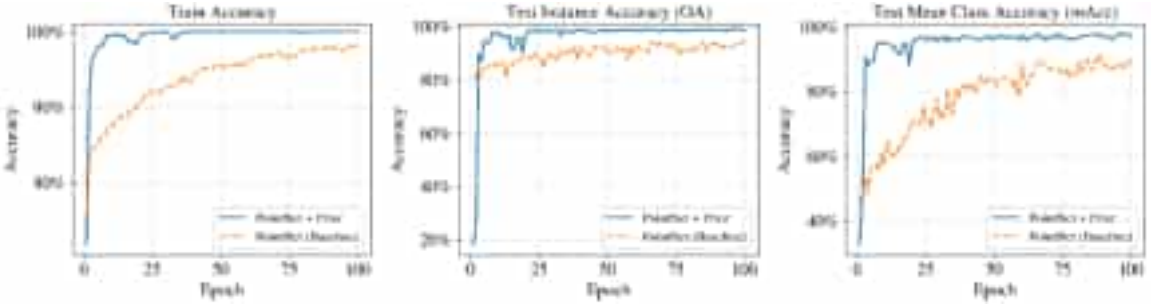


Figure 3. Comparison of PointNet (top) with or without prior knowledge, including training accuracy (left), test OA (middle), and test mAcc (right).

Table 2. $\text{BestAcc}_{\text{last30}}$ under different folds.

Fold		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Avg.
Baseline	OA	90.6	98.9	96.9	96.5	100.0	90.0	99.4	96.0
	mAcc	80.4	97.0	91.1	97.3	100.0	93.0	99.5	94.0
Our	OA	97.8	99.6	97.9	100.0	100.0	100.0	100.0	99.3
	mAcc	96.1	98.6	95.6	100.0	100.0	100.0	100.0	98.6

Table 3. $\text{WorstAcc}_{\text{last30}}$ under different folds.

Fold		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Avg.
Baseline	OA	74.1	96.4	91.4	83.9	78.4	80.0	96.0	85.7
	mAcc	67.5	80.0	75.9	82.4	81.9	77.3	81.1	78.0
Our	OA	96.4	98.2	95.4	100.0	98.1	95.7	98.3	97.4
	mAcc	93.3	92.9	88.7	100.0	95.7	94.6	96.0	94.5

Table 4 further reports per-class accuracies, and Figure 4 display the corresponding confusion matrices for the best and worst epochs over the last 30 epochs.

It can be observed that the PointNet baseline exhibits noticeable fluctuations for deck (from 91.2% $\text{BestAcc}_{\text{last30}}$ to $\text{WorstAcc}_{\text{last30}}$ 55.9%), girder (from 85.3% $\text{BestAcc}_{\text{last30}}$ to $\text{WorstAcc}_{\text{last30}}$ 64.7%), and abutment (from 84.6% $\text{BestAcc}_{\text{last30}}$ to $\text{WorstAcc}_{\text{last30}}$ 70.8%), indicating unstable convergence during model training.

The confusion matrix in Figure 4 further reflects this instability. For $\text{BestAcc}_{\text{last30}}$, the dominant errors in the PointNet baseline are: abutments are incorrectly labelled as pier caps and diaphragms; piers are incorrectly labelled as abutments and diaphragms; and diaphragms are incorrectly labelled as abutments and piers. For $\text{WorstAcc}_{\text{last30}}$, a similar error pattern is observed, but the misclassification becomes substantially more severe. For example, a markedly larger number of diaphragm samples are wrongly predicted as piers and abutments.

After incorporating prior knowledge into model training, the model achieves 100% $\text{BestAcc}_{\text{last30}}$ for pier, girder, diaphragm, deck, and parapet, and also improves the classification accuracy for abutment (from 84.6% to 95.4%). In the $\text{WorstAcc}_{\text{last30}}$, the model keeps the diaphragm at 100%. However, similar misclassification problems can still be found for abutments and pier caps.

Table 4. Class-wise classification result.

Class		Pier	Pier cap	Abutment	Girder	Diaphragm	Deck	Parapet
$\text{BestAcc}_{\text{last30}}$	baseline	96.7	93.5	84.6	85.3	99.1	91.2	97.1
	Our	100.0	93.5	95.4	100.0	100.0	100.0	100.0
$\text{WorstAcc}_{\text{last30}}$	baseline	92.9	95.2	70.8	64.7	88.4	55.9	91.2
	Our	99.5	91.9	75.4	97.1	100.0	97.1	98.5

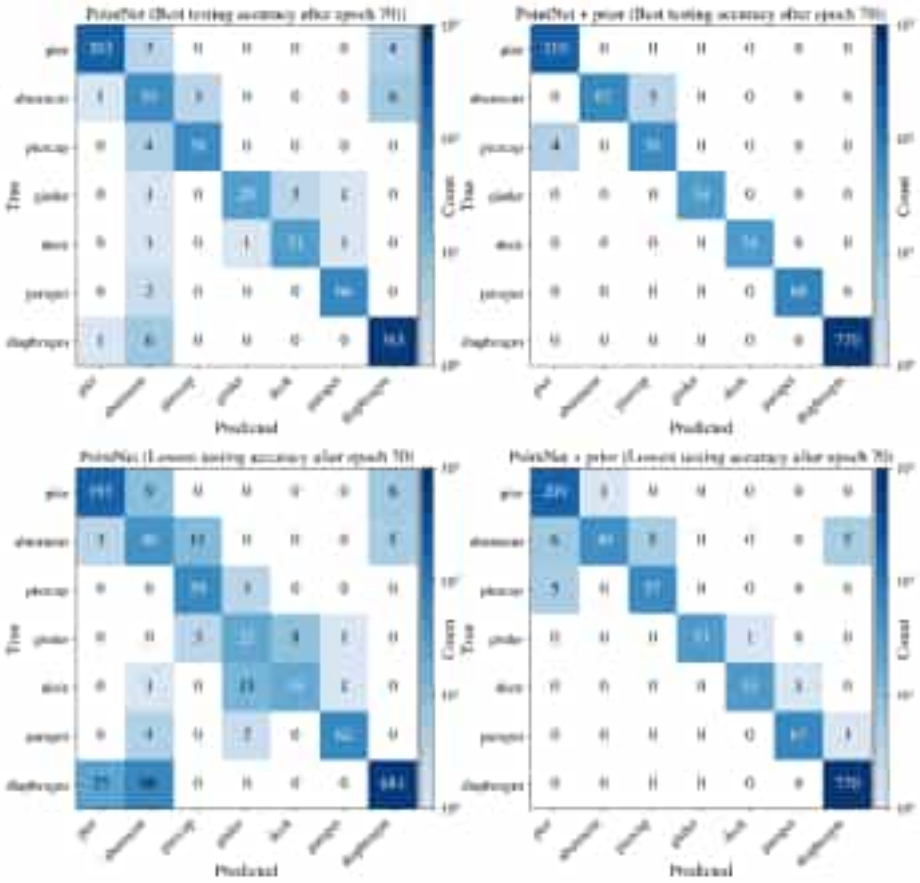


Figure 4. Confusion matrix of classification results of PointNet, including the best testing accuracy and worst testing accuracy in the last 30 epochs.

5 Conclusion

To enhance the point cloud application in bridge scenes, this study proposed a prior knowledge-enhanced point cloud classification model. In our method, we designed a bridge-wise batching training strategy to encode prior knowledge, and then proposed elevation-based priors for model training and prediction. The proposed method was validated using components from 34 RC girder bridge point clouds, covering four typical girder types. The results revealed that introducing prior knowledge yields clear gains in peak performance, as indicated by $\text{BestAcc}_{\text{last30}}$, and stability for convergence, as indicated by $\text{WorstAcc}_{\text{last30}}$. The average $\text{BestAcc}_{\text{last30}}$ improves from 96.0% OA and 94.0% mAcc (PointNet baseline) to 99.3% OA and 98.6% mAcc (Ours).

However, the proposed method is not intended to be a universal solution. Although the bridge-wise batching strategy is potentially applicable to different bridge types, its thresholds still need to be adjusted based on statistical analyses of component distributions for specific bridge categories. Meanwhile, further investigation is still required to validate the applicability of the proposed method to diverse real-world bridge scenarios. In future work, we will focus on: (i) validating the framework on RC girder bridges with more diverse component inventories, while further developing more effective structural priors and refining hyperparameters to improve robustness; and (ii) extending the methodology to other bridge types (e.g., arch and cable-stayed bridges) and infrastructures by designing corresponding structural priors.

References

[1] H. Yue, Q. Wang, H. Zhao, N. Zeng, Y. Tan. Deep learning applications for point clouds in the construction industry. *Automation in Construction*, 168:105769, 2024.

[2] F. Bosché, M. Ahmed, Y. Turkan, C.T. Haas, R. Haas. The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction*, 49:201–213, 2015.

[3] Q. Wang, M.K. Kim. Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics*, 39:306–319, 2019.

[4] T. Yang, Y. Zou, Enrique del Rey Castillo. ENHANCING SCAN-TO-BIM IN BRIDGE ENGINEERING THROUGH POINT, in: Sixth Int. Conf. Civ. Build. Eng. Informatics, 2025: pp. 85–93.

[5] T. Yang, Y. Zou, R. Castillo, L. Hou, J. Zhong. Enhancing Scan-to-BIM for reinforced concrete bridges using point cloud completion techniques. *Automation in Construction*, 181:106606, 2026.

[6] L. Truong-Hong, R. Lindenbergh. Automatically extracting surfaces of reinforced concrete bridges from terrestrial laser scanning point clouds. *Automation in Construction*, 135:2022.

[7] T. Yang, Y. Zou, X. Yang, E. del Rey Castillo. Domain knowledge-enhanced region growing framework for semantic segmentation of bridge point clouds. *Automation in Construction*, 165:105572, 2024.

[8] N. Cui, H. Chen, X. Guo, Y. Zeng, Z. Hua, G. Xiong, R. Yue, J. Liu. Self-prompting semantic segmentation of bridge point cloud data using a large computer vision model. *Automation in Construction*, 167:105729, 2024.

[9] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, M. Bennamoun. Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:4338–4364, 2021.

[10] W. Zhou, K. Liu, W. Jin, Q. Wang, Y. She, Y. Yu, C. Ma. Advancements in deep learning for point cloud classification and segmentation: A comprehensive review. *Computers and Graphics*, 130:104238, 2025.

[11] C.R. Qi, H. Su, K. Mo, L.J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, in: 2017 IEEE Conf. Comput. Vis. Pattern Recognit., IEEE,2017: pp. 77–85.

[12] H. Zhao, L. Jiang, J. Jia, P. Torr, V. Koltun. Point Transformer, in: Proc. IEEE Int. Conf. Comput. Vis., 2021: pp. 16239–16248.

[13] V. Kasireddy, B. Akinci. Assessing the impact of 3D point neighborhood size selection on unsupervised spall classification with 3D bridge point clouds. *Advanced Engineering Informatics*, 52:101624, 2022.

[14] C. Wen, X. Sun, J. Li, C. Wang, Y. Guo, A. Habib. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147:178–192, 2019.

[15] L. Yuan, J. Guo, Q. Wang. Automatic classification of common building materials from 3D terrestrial laser scan data. *Automation in Construction*, 110:103017, 2020.

[16] S.A. Faroughi, N.M. Pawar, C. Fernandes, M. Raissi, S. Das, N.K. Kalantari, S.K. Mahjour. Physics-Guided, Physics- Informed, and Physics-Encoded Neural Networks and Operators in Scientific Computing: Fluid and Solid Mechanics. *Journal of Computing and Information Science in Engineering*, 24:1–31, 2024.

[17] A. Momeni, B. Rahmani, B. Scellier, L.G. Wright, P.L. McMahon, C.C. Wanjura, Y. Li, A. Skalli, N.G. Berloff, T. Onodera, I. Oguz, F. Morichetti, P. del Hougne, M. Le Gallo, A. Sebastian, A. Mirhoseini, C. Zhang, D. Marković, D. Brunner, C. Moser, S. Gigan, F. Marquardt, A. Ozcan, J. Grollier, A.J. Liu, D. Psaltis, A. Alù, R. Fleury. Training of physical neural networks. *Nature*, 645:53–61, 2025.

[18] H. Zheng, V. Moosavi, M. Akbarzadeh. Machine learning assisted evaluations in structural design and construction. *Automation in Construction*, 119:103346, 2020.

[19] A.I.F. Al-Adly, P. Kripakaran. Developing physics-informed neural networks for virtual sensing in beams with moving loads. *Engineering Structures*, 338:120535, 2025.

[20] Z. Lai, C. Mylonas, S. Nagarajaiah, E. Chatzi. Structural identification with physics-informed neural ordinary differential equations. *Journal of Sound and Vibration*, 508:116196, 2021.

[21] R. Lu, I. Brilakis. Digital twinning of existing reinforced concrete bridges from labelled point clusters. *Automation in Construction*, 105:102837, 2019.

[22] A.M. Esmoris, M. Yermo, H. Weiser, L. Winiwarter, B. Hofle, F.F. Rivera. Virtual LiDAR simulation as a high performance computing challenge: Towards HPC HELIOS++. *IEEE Access*, 10:1–1, 2022.

FloorPlan2Nav: Semantic-Topological Navigation from Floor Plan Images as Prior Maps

Yan Gao[1,†] Qian Zheng[1,†], Fuji Hu[1], Yiwei Weng [1,*]

[1] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong
[†]These authors contributed equally to this work and are considered co-first authors
*Corresponding author yiwei.weng@polyu.edu.hk

Abstract –

This paper presents FloorPlan2Nav, a framework that transforms architectural floor plan images into topological navigation graphs serving as prior maps for mobile robots. The framework combines deep learning-based semantic segmentation with contour-based instance extraction, where majority voting within geometrically accurate contours corrects boundary errors in pixel-wise predictions. A connectivity graph constructed from labelled instances enables hierarchical path planning. Graph-guided planning determines the sequence of rooms and connectors, while local A* computes trajectories within each segment in parallel. Experiments on the ResPlan dataset demonstrate 98.16% instance-level segmentation accuracy and 56.89% reduction in planning time compared to global A*. Simulation also confirms that floor plans provide sufficient geometric constraints to correct positional drift, enabling successful navigation without SLAM-based exploration.

Keywords –

Floor Plan Parsing; Connectivity Graph; Hierarchical Path Planning; Mobile Robot Navigation; Prior Map

1 Introduction

Autonomous indoor navigation requires environment maps for path planning and localisation. Conventional approaches rely on SLAM to construct such maps, but the robot must first explore the environment before it can begin operation. Recent studies have explored using prior maps derived from building documentation to reduce this exploration effort. BIM-guided frameworks construct navigation maps from IFC data [1], [2], and visual SLAM systems have incorporated BIM constraints to reduce trajectory drift [3]. These approaches are effective when BIM models are available, but many buildings do not have such documentation. Floor plan images are far more accessible. They can be found in real estate listings, facility archives, and evacuation signs, and they contain the geometric information about walls, rooms, and doors needed for navigation. However, unlike BIM or IFC data, floor plan images are unstructured, and the spatial information is embedded in pixels rather than represented as geometric entities with semantic labels. Hence, a method that can automatically parse floor plan images into structured navigation maps would enable prior map-based indoor navigation in a wider range of buildings.

This paper presents FloorPlan2Nav, a framework that transforms floor plan images into topological navigation graphs. The framework combines semantic segmentation with contour-based instance extraction to identify functional spaces and connectors such as doors, windows, and front doors with precise boundaries. These instances are organised into a connectivity graph that can be adapted to different robot types by enabling edges according to their traversal capabilities. Path planning operates hierarchically, i.e., graph-guided planning first determines the sequence of rooms and connectors, then local A* computes trajectories within each segment in parallel, which significantly reduces planning time. The framework is evaluated on the ResPlan dataset [4] and validated in simulation. Results demonstrate improved planning efficiency and show that floor plans can also serve as geometric reference to correct positional drift during navigation.

2 Related Work

Floor plan understanding has been extensively studied using deep learning methods. Datasets such as CubiCasa5K [5] and ResPlan [4] provide thousands of annotated floor plans covering various room types and architectural styles. ResPlan is particularly relevant to this work as it includes explicit room adjacency graphs alongside vector representations, enabling both geometric and topological analysis. Using these datasets, segmentation networks can effectively learn to classify walls, rooms, doors, and windows at the pixel level [5], [6]. However, navigation relies on topological graphs where spatial adjacency is typically detected through

morphological operations on segmented masks. Pixel-wise accuracy is unable to guarantee correct topology, because predictions often overflow or fall short of true boundaries. Such errors lead to incorrect adjacency relationships, which is problematic for downstream path planning.

Prior map-based navigation has been explored using structured building data. Chen et al. [7] developed a BIM-integrated path planning framework using IFC-based layered mapping and adaptive A* for ground robots. Blum et al. [8] proposed LiDAR localisation in 3D building models using semantic filtering and ICP registration. Gao and Kneip [9] achieved drift-free localisation using CAD-based floor plans as reference maps. More recently, graph-based representations have gained attention for their efficiency and semantic expressiveness. Zhu et al. [10] demonstrated that connectivity graphs constructed from IFC data enable efficient pathfinding by exploiting spatial relationships rather than dense geometry. Ayanzadeh and Oates [11] explored using large language models (LLMs) to parse floor plan images into spatial knowledge graphs for navigation assistance.

While these methods advance prior map-based navigation, they rely on structured building data such as BIM, IFC, or CAD, or require robots to pre-explore environments to construct spatial representations. However, in practice, such digital documentation or pre-exploration is often unavailable. By contrast, floor plan images are readily accessible from real estate listings, evacuation signs, and facility archives. This paper aims to address the challenge of automatically generating navigation-ready topological graphs from floor plan images, thereby bridging the gap between readily available architectural imagery and robot navigation systems.

3 Methodology

The proposed FloorPlan2Nav framework transforms architectural floor plan images into topological navigation graphs through three main stages: semantic segmentation with instance extraction, connectivity graph construction, and hierarchical path planning.

3.1 Semantic Segmentation and Instance Extraction

The input to our framework is a blank architectural floor plan without any semantic annotation. Semantic segmentation is employed to transform the raw drawing into a coloured semantic map. A DeepLabV3+ network trained on the ResPlan dataset [4] learns the visual and geometric patterns of residential floor plans and predicts pixel-wise labels for walls, rooms, doors, windows, and other architectural elements. However, pixel-wise segmentation often suffers from local boundary errors where predicted masks overflow beyond true room boundaries or fall short of them, leaving small gaps. To obtain geometry-accurate instances, we perform contour-based extraction directly on the original floor plan drawing. The image is binarised to isolate structural drafting lines that depict walls and boundaries. These lines naturally partition the plane into enclosed regions. Connected-component labelling identifies all enclosed regions that do not touch the image boundary, and a contour-tracing algorithm extracts the precise boundary of each region (see Figure 1).

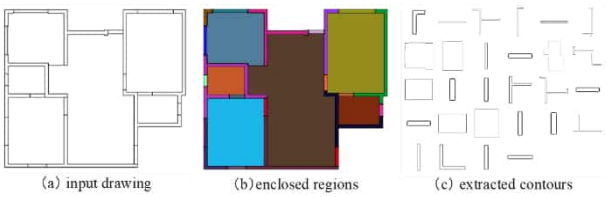


Figure 1 Instance extraction from floor plan

With all instances extracted, the semantic label of each instance is determined through majority voting over the predicted pixel labels within its contour. This effectively corrects local boundary errors in the raw semantic predictions. As shown in Figure 2, when the semantic mask falls short of the true boundary or overflows into adjacent regions (indicated by dashed boxes), the majority vote still assigns the correct label based on dominant predictions inside the geometrically accurate contour. This hybrid approach combines the semantic recognition capability of deep learning with precise geometric boundaries guaranteed by contour-based extraction, which is required for correct graph construction.

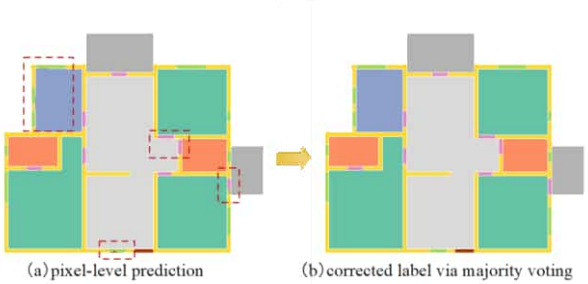


Figure 2 Semantic label assignment

3.2 Connectivity Graph Construction

The annotated floor plan is converted into a connectivity graph for navigation. Nodes correspond to segmented instances grouped into functional spaces (living rooms, bedrooms, bathrooms, kitchens) and

connectors (doors, windows, front doors, and open balconies). Walls are excluded as they represent barriers rather than navigable connections. Spatial adjacency is detected by applying morphological dilation with kernel size k to each instance mask. Overlapping dilated boundaries indicate potential connections, and since each instance is processed independently, this operation can run in parallel. However, not all overlaps represent valid connectivity. Two adjacent doors may overlap after dilation but do not form a traversable path. Semantic filtering removes such connector-to-connector links and retains only space-to-connector relationships (see Figure 3).

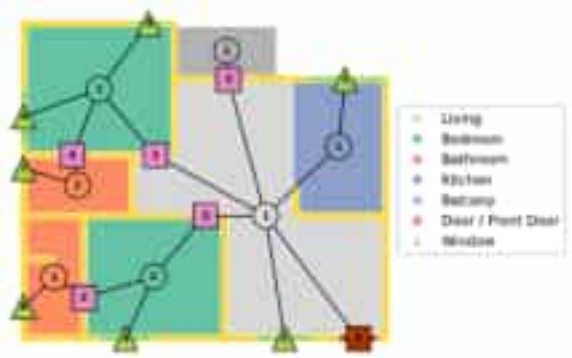


Figure 3 Connectivity graph construction

Which connectors are traversable depends on the robot's physical capabilities. Ground robots can only pass through doors and front doors. UAVs can additionally traverse windows and open balconies, enabling exterior routes along the building facade. Edge weights can further encode traversal costs such as penalties for stairs or narrow passages, making the graph adaptable to different robot types.

3.3 Hierarchical Path Planning

The connectivity graph enables a two-stage planning approach. Global path planning searches the graph to determine the sequence of rooms and connectors the robot should traverse. Figure 4 illustrates alternative routes from balcony to front door for different robot types. Ground robots (a) can only pass through doors, while UAVs (b) can also traverse windows and fly along the building exterior.

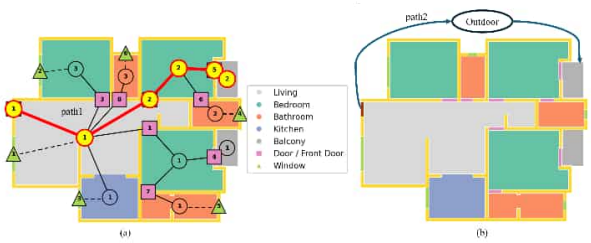


Figure 4 Path alternatives: (a) ground robot, (b) UAV

With the path determined, local path planning computes the trajectory within each segment using A*. Connector centroids are used as waypoints to ensure safe passage through doorways. Since segments between consecutive connectors are independent, local planning runs in parallel (see Figure 5), and computational savings increase with the number of segments, making the approach scalable to larger and more complex floor plans.

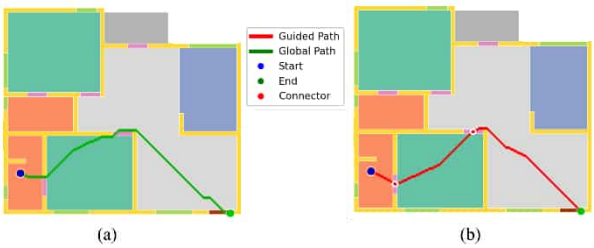


Figure 5 Path planning: (a) global A*, (b) segment-based parallel approach

4 Experiment

4.1 Experiment Setup

We evaluate our framework on the ResPlan dataset, which contains 17,107 residential floor plans with pixel-level annotations for 10 semantic classes: background, living room, bedroom, bathroom, kitchen, balcony, wall, door, window, and front door. The dataset is split into training, validation, and test sets with a ratio of 8:1:1. For semantic segmentation, we employ DeepLabV3+ [12] with Xception-65 backbone pre-trained on PASCAL VOC 2012. The network is trained for 100 epochs using SGD with momentum 0.9, learning rate decaying from 7×10^{-3} to 7×10^{-5} with cosine annealing, and input size of 512×512 pixels. The dilation kernel size k for graph construction is set to 5 pixels. Path planning experiments are conducted on Intel Core i7-11800H (2.30GHz) with 16GB RAM.

4.2 Segmentation Results

We evaluate the model's ability to predict semantic labels from blank floor plan drawings. The DeepLabV3+ model achieves 95.86% mean IoU and 99.54% pixel accuracy. After majority voting within instance contours, instance-level classification reaches 98.16% accuracy across 55,065 instances from 1,711 test images. Table 1 shows results grouped by category. Functional spaces and walls achieve over 99% F1 score. Connectors show lower accuracy due to visual confusion among door, window, and front door. However, for navigation the key requirement is distinguishing connectors from non-traversable elements rather than differentiating connector types. Table 2 shows that when treated as a binary classification, connector detection reaches 99.75% precision and 99.60% recall.

Table 1 Instance-level segmentation results

Category	Precision	Recall	F1
Functional spaces	99.21%	99.03%	99.12%
Connectors	96.48%	96.03%	96.25%
Wall	99.34%	99.56%	99.45%
Background	96.11%	97.93%	97.01%
Overall	98.00%	98.16%	98.08%

Table 2 Connector confusion matrix

GT \ Pred	Connector	Others
Connector	23,882	97
Others	61	31,025

4.3 Path Planning Efficiency

We compare the proposed graph-guided approach against global A* on path planning efficiency. 100 floor plans are randomly selected from the test set, and 10 start-goal pairs are generated for each floor plan by randomly sampling functional spaces, yielding 1,000 navigation tasks in total. Table 3 reports the average performance across all tasks.

The graph-guided method reduces planning time by 56.89% and explores 58.13% fewer nodes compared to global A*. The path cost increases by 7.34% because routing through door centres produces slightly longer trajectories than paths that cut close to door edges. In practice, this difference is negligible since robots adjust trajectories dynamically during execution.

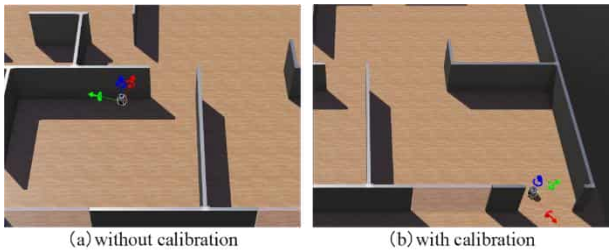
While the absolute time savings are small for these residential layouts, the hierarchical design is valuable because of how it scales. Global A* must search across the entire map grid, so planning time increases as the floor plan grows larger. In contrast, graph-guided planning only searches within individual room segments, each of which stays small regardless of overall building size. Since these segments are computed in parallel, planning time does not accumulate as the number of rooms increases. Therefore, the advantage is greatest in larger commercial or institutional floor plans.

4.4 Simulation Validation

We conduct simulation experiments in Webots to test navigation using the floor plan as a prior map. Webots is an open-source robot simulator that provides realistic physics and sensor models. We use a simulated TurtleBot3 Burger equipped with LDS-01 LiDAR (360° FOV, 0.12-3.5m range). The setup assumes no discrepancy between the floor plan and physical environment, and rooms are empty. The robot follows waypoints generated from the connectivity graph. To simulate realistic localisation drift, Gaussian noise (σ=20mm per timestep) is injected into XY odometry while yaw is assumed accurate from IMU. Without using the floor plan for position correction, cumulative drift causes the robot to deviate and collide with walls (Figure 6a). With periodic calibration, the robot completes the trajectory successfully (Figure 6b).

Table 3 Path planning performance comparison

Metric	Global A*	Graph-guided (ours)	Improvement
Planning time (ms)	23.42	10.09	-56.89%
Nodes explored	2,231	934	-58.13%
Path cost (px)	85.29	91.57	7.34%



(a) without calibration (b) with calibration

Figure 6 Navigation results in Webots simulation

Calibration is triggered every five waypoints. The robot extracts line segments from LiDAR scans and matches them against visible wall contours from the floor plan. Matching candidates are filtered by segment orientation (horizontal or vertical) and validated by distance and angular proximity relative to the robot. Since residential layouts are typically axis-aligned, corrections are decoupled along orthogonal directions: horizontal segments refine Y position and vertical segments refine X.

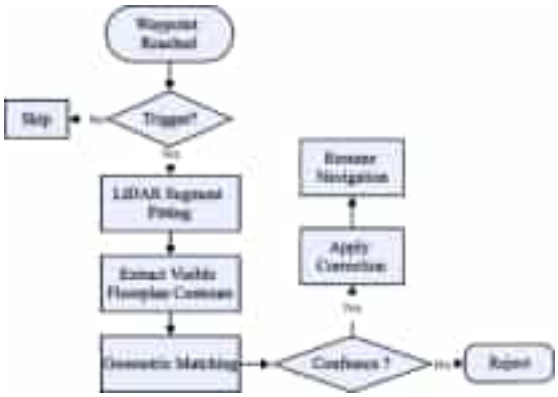


Figure 7 LiDAR calibration flowchart

The matching process and calibration results are illustrated in Figures 8 and 9. Figure 8 shows the segment matching at four waypoints, where segments extracted from LiDAR scans (solid lines) are matched against wall contours from the floor plan (dashed lines). The number of matched pairs varies depending on visible walls. Figure 9 shows the corresponding calibration results. After position correction, the robot position (blue) moves consistently closer to ground truth (green) compared to the drifted estimate (red), confirming that floor plan geometry provides adequate reference for odometry compensation.



Figure 8 Segment matching at four waypoints

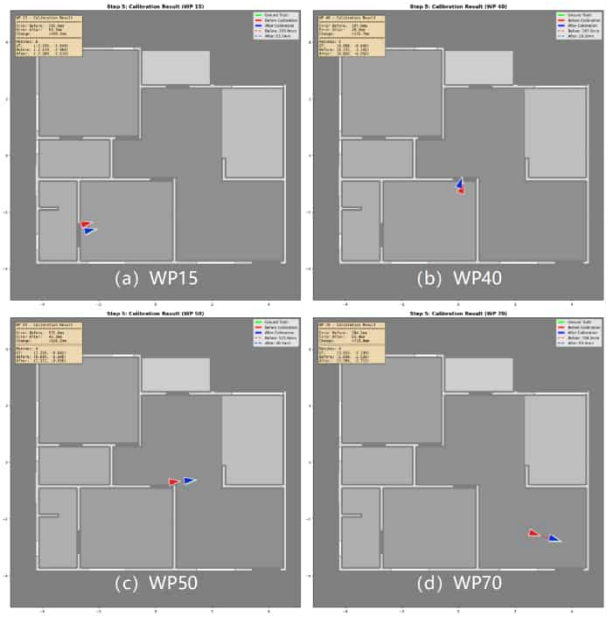


Figure 9 Calibration results at four waypoints

5 Discussion

This work has several limitations. Firstly, the simulation assumes no discrepancy between the floor plan and physical environment, which may not hold for older buildings with undocumented renovations. Secondly, the calibration scheme relies on axis-aligned wall geometry and accurate yaw from IMU. Rooms are also assumed to be empty without furniture that could occlude walls or block paths. The framework also

operates on 2D floor plan images only and does not incorporate three-dimensional information such as floor slopes, door sill heights, or elevation changes. Addressing these factors would require additional elevation data or depth sensing, which is beyond the scope of this work.

The pipeline relies on accurate segmentation, and the sensitivity of navigation to segmentation errors has not been analysed. The contour-based instance extraction and majority voting correct minor pixel-level errors, but more severe failures such as a missed door or a merged room boundary could disrupt the connectivity graph. Furthermore, floor plan segmentation is a well-studied problem, and high accuracy has been demonstrated on both furnished and unfurnished floor plans [13]. The 98.16% instance accuracy on the ResPlan test set is consistent with this. A systematic robustness analysis is left for future work.

Future work will address building discrepancies through probabilistic matching methods that handle geometric deviations. Another direction is vision-language navigation (VLN), where the connectivity graph serves as spatial knowledge rather than a source of precomputed waypoints. Instead of relying on precise geometric matching, the robot observes its surroundings, matches what it sees against the expected layout from the floor plan, and moves toward the next connector by visually identifying doors and walls in real-time, allowing the robot to handle environmental changes and partial occlusions. In this setting, the floor plan serves not as a geometric reference for contour matching, but as a semantic map providing spatial knowledge such as room types and connectivity.

6 Conclusion

This paper presents FloorPlan2Nav, a framework that transforms blank floor plan images into navigation-ready representations for mobile robots. The floor plan is utilised at three stages. Semantic segmentation combined with contour-based instance extraction converts the raw drawing into labelled functional spaces and connectors, where majority voting within geometrically accurate contours corrects boundary errors in pixel-wise predictions. A connectivity graph constructed from these instances enables hierarchical path planning, i.e., global planning on the graph determines the sequence of rooms and connectors, while local A* computes trajectories within each segment. Since each segment is independent, they can be computed in parallel, and the efficiency gain scales with building complexity. The floor plan also serves as a prior map for localisation during navigation, which allows robots to correct odometry drift by matching LiDAR observations against known wall geometry.

Experiments on the ResPlan dataset demonstrate 98.16% instance-level segmentation accuracy with 99.60% connector detection recall. The graph-guided planning achieves 56.89% reduction in computation time compared to global A*. Simulation in Webots confirms that floor plan-based calibration effectively corrects accumulated drift and enables successful navigation without SLAM-based exploration.

Acknowledgement

The authors would like to gratefully acknowledge the support from the Hong Kong Polytechnic University (P0051072 and P0058004).

References

[1] Z. Chen, X. Liang, J. Yuan, S. Yang, and K. Chen. Automated reality capture for indoor inspection using BIM and a multi-sensor quadruped robot. *Automation in Construction*, 170, 2025. doi: 10.1016/j.autcon.2024.105930.

[2] M. A. V. Torres, A. Braun, and A. Borrmann. BIM-SLAM: Integrating BIM models in multi-session SLAM for lifelong mapping using 3D LiDAR. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC)*, pages 521–528, Lille, France, 2023. doi: 10.22260/ISARC2023/0070.

[3] A. Bikandi-Noya, M. Fernandez-Cortizas, M. Shaheer, A. Tourani, H. Voos, and J. L. Sanchez-Lopez. BIM informed visual SLAM for construction monitoring. On-line: http://arxiv.org/abs/2509.13972, Accessed: 30/01/2026.

[4] M. Abouagour and E. Garyfallidis. ResPlan: A large-scale vector-graph dataset of 17,000 residential floor plans. On-line: http://arxiv.org/abs/2508.14006, Accessed: 30/01/2026.

[5] A. Kalervo, J. Ylioinas, M. Häikiö, A. Karhu, and J. Kannala. CubiCasa5K: A dataset and an improved multi-task model for floorplan image analysis. On-line: http://arxiv.org/abs/1904.01920, Accessed: 30/01/2026.

[6] Z. Fan, L. Zhu, H. Li, X. Chen, S. Zhu, and P. Tan. FloorPlanCAD: A large-scale CAD drawing dataset for panoptic symbol spotting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10128–10137, 2021.

[7] Z. Chen, K. Chen, and J. C. P. Cheng. BIM-integrated indoor path planning framework for unmanned ground robot. In *Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 776–783,

Dubai, UAE, 2021. doi: 10.22260/ISARC2021/0105.

[8] H. Blum, J. Stiefel, C. Cadena, R. Siegwart, and A. Gawel. Precise robot localisation in architectural 3D plans. On-line: http://arxiv.org/abs/2006.05137, Accessed: 30/01/2026.

[9] L. Gao and L. Kneip. FP-Loc: Lightweight and drift-free floor plan-assisted LiDAR localisation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 4142–4148, 2022. doi: 10.1109/ICRA46639.2022.9812361.

[10] J. Zhu, X. Wang, M. Wang, P. Wu, H. Kim, and J. Wang. Semantics-based connectivity graph for indoor pathfinding powered by IFC-Graph. *Automation in Construction*, 171, 2025. doi: 10.1016/j.autcon.2025.106019.

[11] A. Ayanzadeh and T. Oates. Floorplan2Guide: LLM-guided floorplan parsing for BLV indoor navigation. On-line: http://arxiv.org/abs/2512.12177, Accessed: 30/01/2026.

[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018.

[13] B. Yang, H. Jiang, Z. Lu, and J. Xiao. Exploring structural lines for interior floorplan segmentation. *The Visual Computer*, 41:6983–6997, 2025. doi: 10.1007/s00371-025-04037-x.

A Distributed Simulation Framework Using HLA for Circular Economy Modeling in Road Construction

Tim Bernhard[1], Omar Abbasi[1], Joseph Louis[2] and Johannes Fottner[1]

[1]Chair of Materials Handling, Material Flow, Logistics, School of Engineering and Design, Technical University of Munich, Garching, Germany
[2]School of Civil and Construction Engineering, Oregon State University, Corvallis, USA
tim.bernhard@tum.de, oj.abbasi@tum.de, joseph.louis@oregonstate.edu, johannes.fottner@tum.de

Abstract – Road construction consumes large quantities of virgin aggregates and asphalt, while circular-economy (CE) strategies, such as using reclaimed asphalt pavement (RAP) and recycled aggregates, remain underutilized due to quality uncertainties, logistical constraints, and limited digital integration. Existing simulation tools are predominantly linear and single-site, which prevents the realistic evaluation of CE strategies across demolition (DEM), mixing (MIX), and construction (CON) operations. This paper introduces a distributed simulation framework based on the High-Level Architecture (HLA) that links DEM, MIX, and CON federates to model dynamic, multi-site and closed-loop material flows. A conceptual CE framework, a set of simulation requirements, and a literature-based gap analysis are developed. A proof-of-concept implementation demonstrates federate connectivity and quantifies CE effects: recycled flows supply 544 t of RAP and 7,500 t of mineral recyclates, reducing virgin-material demand by roughly 25 %. The results demonstrate that distributed simulation offers a feasible and extensible framework for analyzing CE strategies under realistic operational constraints, thereby supporting sustainable, smart, and lean road-construction planning through network-level logistics, environmental indicators, and scenario-based optimization.

Keywords – circular economy; road construction; distributed simulation; high level architecture; material flow

1 Introduction

Road construction is a highly material-intensive sector that relies on large quantities of aggregates, asphalt, and concrete, among other materials [1], [2]. At the same time, the industry faces increasing pressure to reconcile traditional project objectives – cost, quality, and time – with growing ecological requirements and resource scarcity [3]. Achieving an economically viable, specification-compliant pavement while simultaneously reducing environmental impacts creates an inherent conflict of goals for contractors, agencies, and material suppliers [4]. These competing objectives intensify the need for more resource-efficient, data-driven, and operationally integrated methods capable of supporting future-ready road-construction workflows. Circular Economy (CE) principles offer substantial potential for addressing these challenges [5], [6]. Strategies such as incorporating reclaimed materials demonstrate clear potential [1], [2], [7]. However, CE adoption in practice remains limited. Key barriers include uncertainties in the properties of reclaimed materials [2], [8], logistical constraints across demolition, production, and paving operations [9], [10], restrictive specifications, and limited digital integration among stakeholders [8], [9]. Despite increasing political attention to circular construction, practical operational models for implementing CE principles in road construction are still lacking [10]. Existing modeling tools offer limited support for CE integration. Most simulation approaches remain linear, isolated, and focused on single sites [11], [12]. Road-construction networks involve dynamic material exchanges among demolition (DEM) sites, mixing plants (MIX), and construction (CON) sites that are spatially separated and temporal in nature [12], [13], [14]. Current models cannot represent multi-site material availability, quality-dependent blending, or the temporal coordination required to evaluate recycling feasibility [13], [15]. As a result, they cannot generate system-level CE indicators such as recycled content or virgin-material substitution [16], [17]. To address these limitations, this paper proposes a distributed simulation framework based on the High-Level Architecture (HLA). The framework links three federates – DEM, MIX, and CON – to represent multi-site interactions and circular material flows, with the option to add and remove more sites to ensure fidelity with real-world project schedules. The aim is to demonstrate that distributed modeling enables detailed site-specific process representation, realistic and scalable multi-site material exchange, and explicit

closed-loop modeling (DEM-MIX-CON). The main contributions of this paper are:

1. A conceptual and technical framework combining CE principles with distributed simulation.
2. Integration of discrete process simulation with material-flow and Sankey-based CE visualization.
3. A requirement set (R1–R7) for CE-oriented simulation in road construction.
4. A systematic literature comparison covering simulation in construction and their relevance to CE in road construction.
5. A proof-of-concept federated HLA implementation linking DEM, MIX, and CON.

The remainder of the paper is structured as follows: Section 2 reviews related work across CE, material-flow modeling, construction simulation, and distributed simulation. Section 3 introduces the conceptual framework and federation design. Section 4 presents the implementation and proof-of-concept demonstration. Section 5 discusses implications and challenges, and Section 6 concludes with an outlook for future research.

2 Background and Related Work

This section presents a review of CE concepts, material flow modeling techniques, and construction simulation methods, with the aim of specifying requirements and identifying knowledge gaps for creating realistic and scalable simulations of road construction circularity. Together, these perspectives highlight the need for more integrated and digitally coordinated approaches to support future resource-efficient construction workflows.

2.1 Circular Economy in Road Construction

CE principles, such as reduce, reuse, recycle, provide a strategic foundation for improving resource efficiency and reducing environmental impacts in pavement engineering [18]. Strategies such as substituting virgin materials with reclaimed ones, using rejuvenators, or improving end-of-life recovery can substantially reduce material extraction and emissions [1], [2], [7], [16]. However, CE implementation remains limited due to quality variability of reclaimed materials, aging effects, and contamination that complicate mix design [2], [19]. Logistical issues, such as transport distances and timing mismatches between demolition and production, further constrain practical feasibility [20], [21]. Standards often restrict the use of recycled content, and digital tools for coordinating material flows across organizations are still scarce [8], [9], [22]. Despite increasing policy attention, most CE studies in pavement engineering remain material-centric and do not address the dynamic, process-level, and network-level considerations required for operational CE integration [13], [17].

2.2 Material Flow Modeling

Material Flow Analysis (MFA) provides a quantitative framework for assessing inputs, outputs, and stocks of materials within a defined system boundary [15]. In road construction, MFA is used to represent material requirements for pavement layers and the material flows [1], [22]. MFA enables transparent reporting of circularity indicators such as recycled content, virgin substitution, and diversion rates [16]. Sankey diagrams are widely used to visualize these flows, offering an intuitive representation of how materials move through the system and where losses or inefficiencies occur [23]. Such visualizations are effective communication tools, particularly for comparing linear and circular system configurations. However, MFA is inherently static. It does not account for temporal aspects such as scheduling, process durations, capacity constraints, or quality-dependent blending decisions [15]. For road construction, these temporal dynamics are crucial: DEM output may not align with MIX demand; materials may degrade during storage; and transport times may disrupt planned operations [7], [13], [19]. Consequently, MFA alone cannot determine whether CE strategies are operationally feasible or economically beneficial under realistic process conditions. A dynamic model is needed that can integrate temporal, spatial, and operational constraints across the network. This motivates the development of CE-oriented dynamic simulation approaches.

2.3 Simulation in Construction Processes

In contrast to the static nature of Sankey diagrams, discrete event simulation (DES) offers a convenient method to analyze the dynamics of construction operations through explicit modeling of activity sequencing logic, resource flows, and uncertainties in duration and quantities involved. These properties enable simulation to overcome the limitations of real-world experimentation (expensive and unfeasible) and static analytical modeling (which make overly simplifying assumptions about the underlying system) to provide an objective means of analyzing and optimizing construction operations [11], [24]. Prior work that showcases the applicability and generalizability of DES to construction operations in general include tunnel boring [25], pipe-spool module assembly [26], modular construction ([27], disaster recovery [28], and asphalt paving operations [29]. However, despite these strengths of DES for modeling construction operations, it has some significant limitations that hinder its ability to model circularity in roadway construction: (1) the DES current focuses on modeling only single sites, which makes it

unsuitable to model multiple sites in a region; (2) it is difficult to model the temporariness of projects as they would occur in the real world, as DES tends to focus on modeling operations from start to finish for a single operation on a single site; (3) current DES methods do not enable the transfer of resources between models in a time-synchronized manner, which is necessary to model material transfer between sites for circular roadway construction. These limitations are overcome in this research by adopting the HLA for distributed simulations using the ConStrobe simulation engine, as will be described in Section 3.

2.4 Requirements for CE-oriented simulation in road construction

Based on the gaps identified in the previous sections, a simulation framework intended to support CE in road construction must satisfy several functional requirements that exceed the capabilities of traditional MFA or isolated discrete-event models.

1. R1 – Scalable Multi-Site System Representation: Road construction CE requires modeling interactions among multiple DEM, MIX, and CON sites operating concurrently within a distributed network [12], [20], [21].
2. R2 – Dynamic Material Flow Tracking: The simulation must track material batches over time, including their quantities, origins, and quality attributes, since recycling feasibility depends on temporal and compositional alignment [13], [15].
3. R3 – Circular Material Feedback Loops: Closed-loop flows must be modelled to quantify circularity indicators such as recycled content, virgin-material reduction, and diversion rates [16], [18].
4. R4 – Process-Level Operational Detail: The model must capture operational constraints (e.g., queueing, capacity limits, availability, and process times at DEM, MIX, and CON sites) [11], [12].
5. R5 – Inter-Model Communication and Modularity: Given the heterogeneity of DEM, MIX, and CON processes, the simulation must support independent models that communicate through standardized interfaces, as enabled by the HLA [30], [31].
6. R6 – Scalability to Network-Level Decisions: To evaluate CE strategies at the project or regional level, the simulation must scale to multiple sites and varying process configurations [14], [21].
7. R7 – CE Indicators and Visualization: The framework must generate CE-relevant metrics and support visualization through flow diagrams such as Sankey charts [16], [23].

Together, these requirements form the basis for a CE-capable simulation architecture, demonstrating the need for a distributed modeling approach that integrates multi-site material flows and operational dynamics.

2.5 Requirement-Based Evaluation of Existing Simulation Approaches

To assess the suitability of existing construction-simulation research for CE-oriented modeling, the requirements defined in Section 2.4 (R1–R7) were used as evaluation criteria. The DES-focused literature discussed in Section 2.3 represents the dominant modeling approaches in construction operations and offers mature capabilities in activity sequencing, equipment modeling, and queue-based resource interaction. However, these studies remain fundamentally constrained by their single-site scope and linear process assumptions. Table 1 summarizes the comparison between representative DES studies and the CE-oriented requirements. Conventional DES models generally satisfy process-detail requirements (R4) but lack mechanisms for multi-site coordination (R1, R6), dynamic material-quality tracking (R2), circular flow modeling (R3), inter-model communication (R5), and CE indicator generation (R7). All reviewed works are limited to single-site, linear processes and do not represent interdependent DEM–MIX–CON systems, circular material feedback loops, or closed-loop mass tracking.

Overall, no existing DES study fulfills more than two CE-relevant requirements. No prior work combines multi-site process modeling, dynamic material-flow tracking, and circularity assessment, highlighting a clear research gap that directly motivates the distributed HLA-based simulation framework presented in Section 3.

Table 1: Assessment of existing studies

	R 1	R 2	R 3	R 4	R 5	R 6	R 7
[11] AbouRizk (2010) Role of Simulation in CEM	○	○	○	●	○	○	○
[24] Martinez (1996) StroboScope	○	○	○	●	○	○	○
[25] Ruwanpura et al. (2001) Tunnel Simulation	○	◔	○	●	○	○	○
[26] ElMenshawy et al. (2025) Pipe Spool Scheduling	○	◔	○	◕	○	○	○
[27] Abiri et al. (2019) Modular Construction	◔	◔	○	◕	○	◔	○
[28] Louis & Dunston (2018) IoT-enabled Operations	○	◑	○	●	◔	○	○
[29] Mostafavi et al. (2012) Asphalt Paving	○	◔	○	●	○	○	○

3 Conceptual Framework and System Design

This section presents the conceptual foundation for modeling circular material flows across demolition, mixing, and construction activities, as modeled by the corresponding federates. The framework integrates CE principles with a distributed, multi-site simulation perspective, directly operationalizing the requirements derived in Section 2.4. The framework enables transparent tracking of material origins, quality-dependent transformation steps, and batch-level allocation across the distributed DEM–MIX–CON system. The modeling approach mirrors the physical organization of road-construction processes and provides the basis for evaluating recycling and substitution effects under realistic operational constraints.

3.1 System Boundary and Material Loop

The framework encompasses the entire material cycle associated with demolishing and constructing a 1 km-long and 7 m-wide road section. All materials extracted during demolition, as well as those required for constructing the surface, base, and frost layers, are included. The process is decomposed into three domains:

- **DEM** – milling and excavating surface, base, and frost layers; asphalt-bound and mineral materials.
- **MIX** – classification, crushing, screening, and production of asphalt and granular mixes.
- **CON** – placement and compaction of the new road.

The analysis encompasses all materials generated from DEM and all materials necessary for CON. While this models a single life-cycle iteration, the structure supports multi-cycle CE assessments by allowing CON outputs to re-enter future DEM processes or even introduce maintenance activities as a combination of the DEM and CON federates.

3.2 Pavement Structure, Material Composition and Machinery

The pavement consists of three layers: surface, base, and frost, each with distinct material compositions and recycling potentials.

Table 2: Pavement layers, material composition, and machinery

	Primary Materials	DEM Machines	CON Machines
Surface	Asphalt(1)/RAP(2), Bitumen(3), Crushed Stone(4), Filler(5)	Milling Machine	Asphalt Paver, Roller
Base	Asphalt(1)/RAP(2), Concrete(6), Crushed Stone(4), Gravel(7)	Excavator, Wheel Loader	Grader, Roller, Wheel Loader
Frost	Concrete(6), Crushed Rock(8), Gravel(7)		

3.3 Reference Material Flows and Sankey-Based CE Representation

The Sankey diagram in Figure 1 forms the quantitative backbone of the conceptual framework. It depicts the complete flow of materials from DEM to MIX for all pavement layers, identifying the recoverable asphalt components, the mineral fractions suitable for reuse, and the unavoidable losses resulting from fines, contamination, or mechanical degradation. The recyclable materials are depicted in light green.

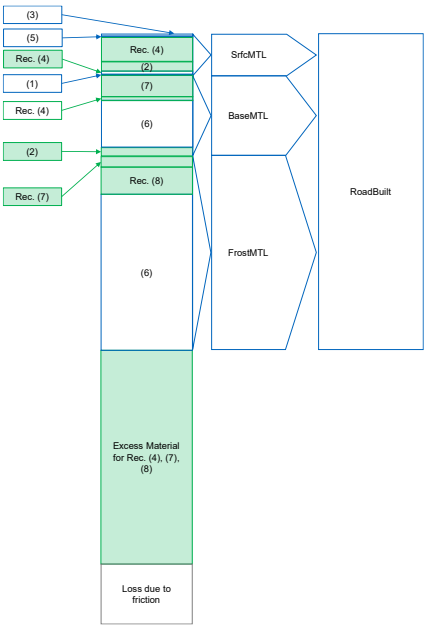


Figure 1: Sankey diagram of material flows across DEM-MIX-CON

The underlying values are derived from an estimation model calibrated using real project data and validated through expert interviews. DEM generates three primary material streams corresponding to the original pavement structure: approximately 1,314 t from the surface layer, 2,533 t from the base layer, and 6,131 t from the frost/sub-base layer. These streams sum to 9,979 t and are decomposed into asphalt-bound components, recoverable granular fractions, and non-recoverable losses. In the surface layer stream, 308 t of material are recovered as high-quality RAP (approximately 90 % of the asphalt content), 777 t are recycled as mineral

fractions (around 80 % of the unbound portion), and 229 t are lost due to processing-related degradation. The base layer yields 236 t of RAP, 1,818 t of recycled mineral aggregates, and 462 t of losses. The frost/sub-base layer, containing no asphalt binder, contributes the largest volume of mineral recyclate, with 4,905 t recovered and 1,226 t lost. In total, the CE potential amounts to 544 t of RAP and 7,500 t of reusable mineral aggregates, whereas 1,916 t of material must be virgin inputs. These flows define the theoretical upper bound of substitution in the CON process and establish quality and mass constraints governing decisions in MIX.

3.4 Material Transformation Logic from DEM to MIX to CON

The transformation logic governs how materials move through the DEM–MIX–CON chain, ensuring that the simulation reflects realistic process behavior and CE constraints. DEM produces heterogeneous batches characterized by material type, grain-size distribution, asphalt content, binder ageing (for RAP), moisture, and contamination. These properties determine their eligibility for high-quality RAP production or mineral recycling pathways. Once transferred to MIX, the batches undergo screening, crushing, composition analysis, and RAP-quality evaluation before being sorted into dedicated stockpiles. MIX thus transforms DEM output into three categories: RAP suitable for asphalt mixtures in surface or base layers, recycled aggregates for base and frost layers, and reject fines, representing the loss streams in the Sankey diagram. CON consumes the processed materials according to the structural requirements of each layer. The surface layer relies on bitumen, high-quality aggregates, filler, and a controlled RAP share. The base layer accommodates a wider mix of RAP and recycled minerals, while the frost/sub-base layer can incorporate nearly all available mineral recyclates. When recycled supply is insufficient or fails to meet quality requirements, virgin materials are introduced to maintain mass balance and ensure layer specifications are met. In this way, the simulation can evaluate both the theoretical CE potential and the operational feasibility of recycling outcomes.

3.5 Distributed Simulation and HLA

Given the limitations of conventional DES to model multiple spatially and temporally distributed demolition sites, construction sites, and mixing plants that form the dynamic system for which circularity is being considered, this research proposes the use of distributed simulations that is enabled through the IEEE 1516 HLA standard for creating distributed, time-synchronized simulations. In this framework, individual simulation models are referred to as federates, and they join a federation through a Runtime Infrastructure (RTI), which facilitates time synchronization and data exchange between the individual federates. AbouRizk (2010) has previously identified HLA as a promising direction for construction simulation and has developed the Construction Synthetic Environment (COSYE) for the Simphony simulation system. This research utilizes the CERTI HLA framework [32] for the RTI and the ConStrobe simulation system [33] to provide individual federates for the various demolition, construction, and mix plant sites within the road network. A schema of this distributed simulation framework is shown in Figure 2.

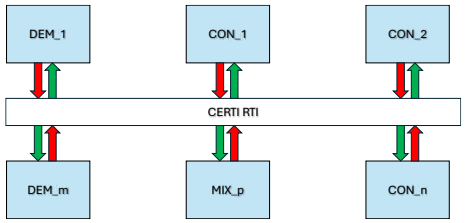


Figure 2: HLA Architecture

The light blue boxes represent HLA-compliant ConStrobe models that represent any of the different DEM, CON, or MIX sites, which act as federates and publish and subscribe to data from other federates through the CERTI RTI. They interact through a shared Federation Object Model that defines material batches using attributes such as mass, material type, RAP fraction, binder properties, grain-size class, timestamp, and federate origin. Time management services coordinate the execution order and ensure synchronized access to material stocks and process events. The modularity of the architecture enables the integration of extended scenarios, such as additional recycling facilities, alternative pavement designs, or network-level simulations, without requiring modifications to existing federates.

3.6 Circularity Indicators and Model Outputs

The distributed simulation generates CE-relevant performance measures, including RAP utilization in asphalt production, mineral substitution rates across all layers, total recycled versus virgin material, and mass balance consistency across federates. Temporal indicators, such as synchronization between DEM supply and CON demand, further support the evaluation of CE strategies under realistic scheduling and capacity constraints. Together, the outputs translate static Sankey-derived potentials into dynamic operational CE metrics.

4 Implementation and Demonstration

This section presents the prototype implementation of

the distributed simulation framework and demonstrates the feasibility of linking DEM, MIX, and CON through HLA. The implementation follows a modular design, in which each federate operates as an independent simulation unit, utilizing a simplified DES logic.

4.1 Model Description

Each federate is implemented with DES-like process logic, supporting queueing, batching, and resource-driven processing. DEM generates material batches corresponding to the layer-specific flows defined in the Sankey diagram, assigning attributes such as mass, material type, RAP fraction, and binder quality. MIX applies simplified processing steps, mapping DEM outputs to the recycling categories defined in Section 3.4. Figure 3 depicts DEM as an exemplary federate.

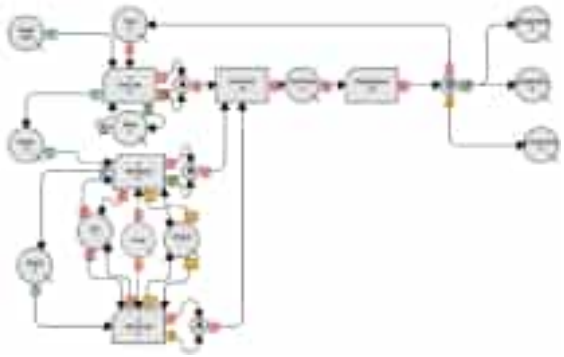


Figure 3: DEM federate in ConStrobe

CON receives processed materials and allocates them to surface, base, or frost layers according to their specification requirements. Several simplifying assumptions enable a clear proof-of-concept demonstration without overburdening the model:

- Machinery resources are represented through processing-time proxies rather than detailed equipment behavior.
- Transport between DEM–MIX–CON is modelled using deterministic travel times.
- MIX capacity is assumed sufficient for the given scenario, avoiding queue overflow.

These assumptions preserve focus on inter-federate communication, CE mass flows, and model extensibility.

4.2 Federate Connections and HLA Synchronization

The three federates communicate through the RTI using object and interaction classes defined in the Federation Object Model. DEM publishes MaterialBatch objects, which include mass, type, timestamp, and origin. MIX subscribes to these objects and publishes corresponding ProcessedBatch objects after transformation. CON then consumes these batches through MaterialRequest and MaterialDelivery interactions. Time-stepped HLA time management is used to ensure logical causality. All federates advance simulation time in synchronized increments. This mechanism ensures that the distributed model remains consistent with the temporal dynamics of real-world construction operations, where material availability constrains production and placement.

4.3 Demonstration Scenario

The demonstration uses the case introduced in Section 3. The DEM federate generates batches totalling 9,979 t, decomposed by layer origin. MIX processes these batches into RAP, recycled mineral aggregates, and residual losses, applying the CE transformation logic defined in Section 3.4. CON then sequentially constructs the frost, base, and surface layers by requesting and receiving materials in accordance with their layer-specific specifications. Through the demonstration run, the federates successfully exchange materials, update inventories, and maintain time-synchronized execution. The resulting simulation logs capture all material transformations and serve as the basis for deriving a dynamic Sankey diagram that mirrors Figure 1.

4.4 Results: Connectivity and Material-Substitution Effects

The proof-of-concept implementation confirms that DEM, MIX, and CON federates can be reliably linked through HLA. All federates exchanged material batches without message loss, and synchronized time advancement preserved correct causal ordering of production, processing, and consumption events. This demonstrates the technical feasibility of distributed, multi-site simulation for road construction processes.

Table 3: Material substitution results for linear vs. circular scenarios

	Linear scenario	CE-enabled scenario
Total material [t]	9,979	9,979
Virgin material [t]	9,979	7,519
RAP used [t]	-	544
Recycled mineral aggregates used [t]	-	7,500
Virgin material reduction [t]	-	2,460

Beyond connectivity, the simulation quantifies the impact of circular material flows on demand for virgin materials. In a linear reconstruction scenario, the full 9,979 t of pavement material is supplied from virgin

sources. When CE pathways are enabled, recycled outputs from DEM and MIX provide 544 t of RAP and 7,500 t of recycled mineral aggregates, reducing virgin material demand by approximately 2,460 t (~25 %). Substitution effects are strongest in the frost and base layers, while surface-layer reductions are limited by RAP quality requirements. These results show that CE performance depends not only on theoretical recycling potential but also on the operational synchronization of DEM–MIX–CON federates.

5 Discussion

The proof-of-concept demonstrates that distributed simulation is a viable approach for modeling CE-oriented material flows across DEM, MIX, and CON operations. Using HLA enables modularity and interoperability, allowing heterogeneous process models to exchange material batches through a shared, synchronized representation and to dynamically generate CE metrics (e.g., RAP utilization and virgin-material substitution), rather than relying on aggregate assumptions. Key implementation challenges remain. Consistent data structures across federates require careful FOM design, particularly when introducing additional material types or quality attributes. Interoperability issues may arise when federates operated by different organizations use incompatible internal models. Differences in temporal and spatial granularity between DEM, MIX, and CON also require harmonization strategies to maintain causal consistency. Overall, the framework provides a basis for extending the federation to multi-site networks, comparing recycling configurations (e.g., mobile vs. stationary), and integrating additional indicators such as CO_2 emissions or energy use.

6 Conclusion and Outlook

This study presented a distributed simulation framework that links DEM, MIX, and CON processes through HLA to evaluate CE-oriented material flows in road construction. The proof-of-concept demonstrated reliable federate synchronization and batch exchange, enabling dynamic assessment of recycling and substitution effects under operational constraints. A key result is that synchronized DEM–MIX–CON interactions reduced virgin material demand by ~25 % compared with a linear scenario. Future work can enhance model fidelity by integrating spatial data (BIM/GIS) to account for transport and network effects and coupling the simulation with environmental and economic assessment modules to support informed decision-making for circular infrastructure planning. By enabling digitally connected, resource-efficient, and operationally synchronized construction processes, the proposed framework contributes to constructing the future of road infrastructure that is sustainable, smart, and lean.

References

[1] Y. Huang, R. N. Bird, and O. Heidrich, "A review of the use of recycled solid waste materials in asphalt pavements," Resour. Conserv. Recycl., vol. 52, no. 1, pp. 58–73, Nov. 2007, doi: 10.1016/j.resconrec.2007.02.002.

[2] M. Zaumanis and R. B. Mallick, "Review of very high-content reclaimed asphalt use in plant-produced pavements: state of the art," Int. J. Pavement Eng., vol. 16, no. 1, pp. 39–55, Jan. 2015, doi: 10.1080/10298436.2014.893331.

[3] A. Böckel, I. Weissbrod, and J. Böhm, Eds., *Mythen der Circular Economy*. Leuphana Universität Lüneburg, 2022. doi: 10.25368/2022.163.

[4] T. Baumgärtel et al., "Bauen heute und morgen," in Digitale Baustelle- innovativer Planen, effizienter Ausführen, W. Günthner and A. Borrmann, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1–21. doi: 10.1007/978-3-642-16486-6_1.

[5] F. Pichlmeier et al., "Tiefbau-Fachsymposium 'Ressourceneffizienz im Straßenbau,'" Nov. 2017.

[6] European Environment Agency., Circular economy in Europe: developing the knowledge base. LU: Publications Office, 2016. https://data.europa.eu/doi/10.2800/51444

[7] H. Ali and A. Rojali, "Recycling Asphalt Pavements: The State of Practice," in Recycling Strategy and Challenges Associated with Waste Management Towards Sustaining the World, H. M. Saleh and A. I. Hassan, Eds., IntechOpen, 2023. doi: 10.5772/intechopen.106235.

[8] R. Karlsson and U. Isacsson, "Material-Related Aspects of Asphalt Recycling—State-of-the-Art," J. Mater. Civ. Eng., vol. 18, no. 1, pp. 81–92, Feb. 2006, doi: 10.1061/(ASCE)0899-1561(2006)18:1(81).

[9] S. Choudhari and A. Tindwani, "Logistics optimisation in road construction project," Constr. Innov., vol. 17, no. 2, pp. 158–179, Apr. 2017, doi: 10.1108/CI-03-2016-0014.

[10] A. Fredriksson and M. Huge-Brodin, "Green construction logistics – a multi-actor challenge," Res. Transp. Bus. Manag., vol. 45, p. 100830, Dec. 2022, doi: 10.1016/j.rtbm.2022.100830.

[11] R. Berbner, C. Elsholz, M. Nicklis, and J. Hagenschulte, "Die Bauindustrie weiter im Umbruch: Fehlendes Know-how und Bürokratie bremsen." Feb. 01, 2025. [Online]. Available: https://www.pwc.de/de

[12] M. J. Riberinho et al., "The next normal in construction; How disruption is reshaping the

world's largest ecosystem." June 01, 2020. [Online]. Available: https://www.mckinsey.com
[13] F. Naz, Improving Transport Efficiency in the Construction Supply Chain, vol. 2396. in Linköping Studies in Science and Technology. Dissertations, vol. 2396. Linköping: Linköping University Electronic Press, 2024. doi: 10.3384/9789180757010.
[14] S. AbouRizk, "Role of Simulation in Construction Engineering and Management," J. Constr. Eng. Manag., vol. 136, no. 10, pp. 1140–1153, Oct. 2010, doi: 10.1061/(ASCE)CO.1943-7862.0000220.
[15] K. Jaafar, "Construction logistics: a simulation case study to improve performance processes," Int. J. Simul. Process Model., vol. 17, no. 1, p. 16, 2021, doi: 10.1504/IJSPM.2021.120840.
[16] Z. Liu and N. Kringos, "Transition from linear to circular economy in pavement engineering: A historical review," J. Clean. Prod., vol. 449, p. 141809, Apr. 2024, doi: 10.1016/j.jclepro.2024.141809.
[17] A. G. Vetnes and R. O'Born, "Bridging the Shift from Linear to Circular Economy in Road Infrastructure: The Norwegian Stakeholders' Perspectives on Challenges and Opportunities," in The 1st International Conference on Net-Zero Built Environment, vol. 237, M. Kioumarsi and B. Shafei, Eds., in Lecture Notes in Civil Engineering, vol. 237. , Cham: Springer Nature Switzerland, 2025, pp. 1573–1583. doi: 10.1007/978-3-031-69626-8_131.
[18] P. H. Brunner and H. Rechberger, Handbook of Material Flow Analysis, 0 ed. CRC Press, 2016. doi: 10.1201/9781315313450.
[19] W. Haas, F. Krausmann, D. Wiedenhofer, and M. Heinz, "How Circular is the Global Economy?: An Assessment of Material Flows, Waste Production, and Recycling in the European Union and the World in 2005," J. Ind. Ecol., vol. 19, no. 5, pp. 765–777, Oct. 2015, doi: 10.1111/jiec.12244.
[20] F. Pomponi and A. Moncaster, "Circular economy for the built environment: A research framework," J. Clean. Prod., vol. 143, pp. 710–718, Feb. 2017, doi: 10.1016/j.jclepro.2016.12.055.
[21] J. Kirchherr, D. Reike, and M. Hekkert, "Conceptualizing the circular economy: An analysis of 114 definitions," Resour. Conserv. Recycl., vol. 127, pp. 221–232, Dec. 2017, doi: 10.1016/j.resconrec.2017.09.005.
[22] T. Bergmann, D. Bleher, and W. Jenseit, "Ressourceneffizienzpotenziale im Tiefbau: Materialaufwendungen und technische Lösungen." Nov. 01, 2025.
[23] M. Schmidt, "The Sankey Diagram in Energy and Material Flow Management: Part I: History," J. Ind. Ecol., vol. 12, no. 1, pp. 82–94, Feb. 2008, doi: 10.1111/j.1530-9290.2008.00004.x.
[24] J. C. Martinez, "STROBOSCOPE: State and Resource Based Simulation of Construction Processes," The University of Michigan, Ann Arbor, 1996.
[25] J. Y. Ruwanpura, S. M. AbouRizk, K. C. Er, and S. Fernando, "Special purpose simulation templates for tunnel construction operations," Can. J. Civ. Eng., vol. 28, no. 2, pp. 222–237, Apr. 2001, doi: 10.1139/l00-101.
[26] M. ElMenshawy, L. Wu, B. Gue, and S. AbouRizk, "Automating Pipe Spool Fabrication Shop Scheduling for Modularized Industrial Construction Projects Using Reinforcement Learning," J. Comput. Civ. Eng., vol. 39, no. 3, p. 04025013, May 2025, doi: 10.1061/JCCEE5.CPENG-6042.
[27] B. Abiri, J. Louis, and M. Riggio, "Decision-making for Cross-Laminated Timber Modular Construction Logistics Using Discrete Event Simulation," Modul. Offsite Constr. MOC Summit Proc., pp. 381–388, May 2019, doi: 10.29173/mocs117.
[28] J. Louis and P. S. Dunston, "Integrating IoT into operational workflows for real-time and automated decision-making in repetitive construction operations," Autom. Constr., vol. 94, pp. 317–327, Oct. 2018, doi: 10.1016/j.autcon.2018.07.005.
[29] A. Mostafavi, V. Valentin, D. M. Abraham, and J. Louis, "Assessment of the Productivity of Nighttime Asphalt Paving Operations," J. Constr. Eng. Manag., vol. 138, no. 12, pp. 1421–1432, Dec. 2012, doi: 10.1061/(ASCE)CO.1943-7862.0000531.
[30] R. M. Fujimoto, "Distributed simulation systems," in Proceedings of the 2003 International Conference on Machine Learning and Cybernetics, New Orleans, LA, USA: IEEE, 2003, pp. 124–134. doi: 10.1109/WSC.2003.1261415.
[31] H. Taghaddos, S. M. AbouRizk, Y. Mohamed, and I. Ourdev, "Distributed agent-based simulation of construction projects with HLA," in 2008 Winter Simulation Conference, Miami, FL, USA: IEEE, Dec. 2008, pp. 2413–2420. doi: 10.1109/WSC.2008.4736349.
[32] E. Noulard, J.-Y. Rousselot, and P. Siron, "CERTI, an Open Source RTI, why and how," presented at the Spring Simulation Interoperability Workshop, San Diego, Jan. 2009.
[33] J. Louis, "ConStrobe - Construction Operations Simulation for Time and Resource Based Evaluations," presented at the Winter Simulation Conference (WSC), 2025.

BIM Design Assistant Driven by LLM Agents

Jin Han[1], Biansining Zhou[2], Xin-Zheng Lu[1], Zhen-Zhong Hu[3], Jun Ma[4] and Jia-Rui Lin[1,*]

[1]Department of Civil Engineering, Tsinghua University, China
[2]Department of Computer Science and Technology, Tsinghua University, China
[3]Institute for Ocean Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China
[4]Department of Urban Planning and Design, University of Hong Kong, Hong Kong
Corresponding author: lin611@tsinghua.edu.cn

Abstract –

BIM design workflows involve complex multi-step interactions with professional software, limiting the direct use of large language models (LLMs). This paper proposes an LLM-agent-driven BIM design assistant that enables natural-language-based BIM design and visualization through reliable tool invocation. First, a two-layer Revit interface function library is designed to bridge LLM agents and BIM software, providing secure, extensible, and fine-grained model manipulation capabilities. Based on the ReAct paradigm, an agent framework is developed that integrates dialogue history management, structured prompt engineering, and step-by-step reasoning to support long-horizon and multi-turn BIM tasks. An agent-oriented BIM benchmark is constructed to evaluate the system, covering querying, visualization, modification, and deletion tasks with varying difficulty. Experiments with three representative LLMs show that high-performance models can reliably act as BIM design assistants within the proposed framework, achieving up to 90% task accuracy. The results demonstrate the feasibility of LLM-driven BIM design assistance.

Keywords –

Building Information Modeling; Revit; Large language models; Agent

1 Introduction

Building Information Modeling (BIM) has become a key data-driven approach across the building lifecycle [1]. By integrating project data into a unified model, it improves collaboration and information management efficiency [2]. However, as BIM models become more complex, tasks such as model manipulation, information retrieval, and modification still rely heavily on manual interaction with tools like Autodesk Revit [3], increasing barriers for non-expert users and reducing efficiency in repetitive multi-step operations.

More recently, advances in natural language processing and large language models (LLMs) have stimulated growing interest in language-driven BIM interaction [4]. Existing work has investigated LLM-based methods for NLP-based information extraction and analysis [5,6,7], knowledge acquisition [8,9], BIM question answering [10], and building information modeling [11]. However, most of these approaches focus on semantic-level tasks where LLMs mainly support information understanding or text generation. Their capability to execute design operations within BIM software environments remains limited.

The emergence of LLM-based agents provides a potential solution by enabling models to reason and interact with external tools [12,13]. In the BIM domain, this development suggests a shift from passive language assistance toward active task execution. For example, Ying et al. proposed a ReAct-based agent for design checking that can interpret requirements and retrieve BIM data [14]. Jang et al. explored natural-language-driven architectural detailing assistance [15]. However, these studies either rely on predefined task rules or support only limited BIM operations, and they do not establish a general framework for reliable natural-language-driven BIM design interaction.

To address this limitation, this study develops a BIM design assistant driven by LLM agents to enhance BIM-assisted design and visualization tasks. The main contributions are threefold: (1) a two-layer BIM interface architecture that converts low-level Revit APIs into structured tools callable by LLM agents; (2) an agent-based interaction framework integrating ReAct reasoning, prompt engineering, and dialogue history management for multi-step BIM operations; and (3) a benchmark containing 100 BIM tasks across querying, visualization, modification, and deletion to systematically evaluate the proposed framework. Although the present implementation is developed in Autodesk Revit, the proposed framework is not entirely Revit-specific. The platform-dependent components are the BIM operation layer, which includes low-level Revit functions and composed Revit Interfaces. By contrast, the two-layer architectural organization, the ReAct-based planning and execution workflow, etc., are platform-agnostic at the

methodological level. Therefore, transferring the framework to other BIM platforms mainly requires replacing the low-level tool layer with interfaces built on the corresponding software APIs.

2 Design and Implementation of an LLM-Oriented Revit Interface Library

To enable real-time BIM model manipulation, a structured Revit interface library is developed using the official Revit API in C#, exposing core modeling operations to LLM agents and supporting safe agent-based BIM interaction.

2.1 Overall Design

Given the broad functional scope of Revit, this study focuses on structural design tasks and targets common operations such as element retrieval, modification, deletion, parameter configuration, and visualization.

To support these operations efficiently and safely, we design a two-layer interface architecture, as illustrated in Figure 1. The lower layer, Revit Functions, encapsulates low-level BIM model operations and directly interfaces with the Revit. While the upper layer, Revit Interfaces, composes basic functions into higher-level, agent-facing operations, enabling more expressive manipulation.

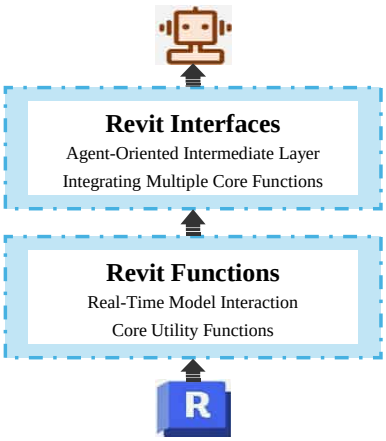


Figure 1. Two-layer interface architecture

This layered design allows shared functionality to be implemented once at the lower level while supporting diverse, integrated operations at the upper level, thereby improving extensibility and development efficiency. Moreover, since all model modifications in Revit are managed through transaction mechanisms, potentially unsafe operations are isolated at the lower layer and exposed to the agent only under controlled conditions at the upper layer. Structured input-output descriptions further accompany each Revit interface to facilitate accurate tool selection and invocation by LLM agents.

2.2 Design of Revit Functions

Revit Functions form the low-level interaction layer that directly and safely operate on the BIM model, comprising nine core functions that provide essential modeling capabilities, as illustrated in Table 1.

Table 1. Core Revit functions

Function Name	Description
Convert Unit	Converts parameter value between user and internal unit
Create Element Filter	Builds filters for element selection
Delete Object	Deletes specified model element
Get Category	Retrieves built-in element categories
Get Parameter Unit	Retrieves internal parameter unit
Retrieve Object	Retrieves elements matching given filters
Retrieve Object Params	Retrieves parameters of a specified element
Set Element Parameter	Sets values of editable element parameter
Show Elements	Highlights selected element

When composed appropriately, these core functions are sufficient to support all target operations while providing controlled access to internal model data. And they are designed to be extensible, enabling future tool integration without altering the core framework.

2.3 Design of Revit Interfaces

While Revit Functions provide low-level capabilities, their API-specific data types are not suitable for direct LLM interaction. Therefore, a higher-level layer, Revit Interfaces, composes multiple functions into integrated operations with simplified inputs (e.g., text and numeric values) for agent invocation (Figure 2). These interfaces also return structured execution feedback for reasoning. Eight interfaces are implemented to support the target BIM operations (Table 2).

Table 2. Core Revit interfaces

Function Name	Description
Delete Element	Deletes elements specified by IDs. Return deleted element IDs or the failure cause
Retrieve Element	Return elements matching given filters or the failure cause
Retrieve All Parameters	Return all parameters of a specified element in a structured format or the failure cause
Retrieve	Return a specific parameter using

Parameter With Built-In Id	its built-in ID in a structured format, or the failure cause
Retrieve Parameter With Parameter Name	Return parameters matching a given name in a structured format or the failure cause
Set Parameter With Built-In Id	Sets parameter values by built-in IDs with validity checks. Return parameter IDs and value changes, or the failure cause
Set Parameter With Parameter Name	Sets parameter values by name with validity checks. Return parameter IDs and value changes, or the failure cause
Show Elements	Highlights specified elements in the 3D view. Return displayed elements' IDs, or the failure cause

By leveraging these, the agent can manipulate BIM models through textual inputs. Moreover, the Revit Interfaces layer is designed to be extensible, allowing additional interfaces to be incorporated.

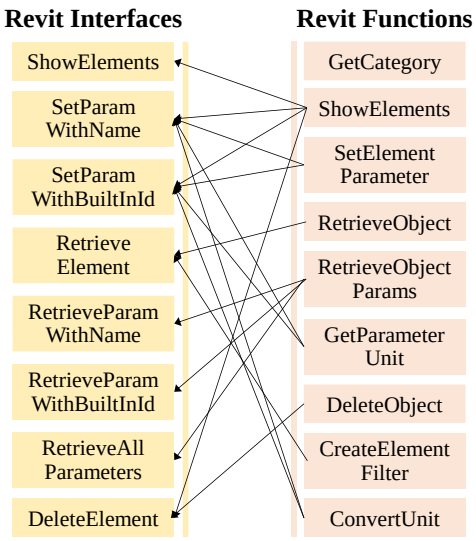


Figure 2. Dependencies between Revit Interfaces and Revit Functions

2.4 Interface Descriptions for LLM-Based Agents

LLM-based tool invocation relies on explicit interface descriptions. Accordingly, each Revit Interface is accompanied by a standardized description consisting of four components: interface name, functionality, input parameters, and output definitions. The interface name serves as the identifier for agent invocation, while the functionality provides a concise summary of the interface behaviour. Input parameters specify the semantics and data types required for correct execution, and output definitions describe the format and meaning of returned information, which supports subsequent agent reasoning and decision-making. These descriptions are assembled during agent-user interactions to support accurate tool selection and execution, as illustrated in Figure 3.

{
"**Interface name**": DeleteElement
"**Functionality**": This interface deletes element(s) from the Revit project given their element ID(s)
"**Input parameters**": The requirements and meanings of the input parameters of this interface are as follows: elementIdStrings: A list of string representations of element IDs of all elements to be deleted
"**Output definitions**": This interface outputs a string message indicating whether the deletion was successful. If failed, a message showing the failure information is returned.
}

Figure 3. Interfaces descriptive information

3 Design Agent Development Based on ReAct and Prompt Engineering

This study adopts the ReAct Think-Act-Observe paradigm as the core framework for BIM-assisted agent development and designs key components based on this paradigm. To ensure the correct interpretation of Revit-specific context, a curated prompt library is incorporated to support reliable multi-turn reasoning and execution.

3.1 Multi-Turn Agent Component Design

Complex BIM design tasks typically require multi-turn interactions with users or the system rather than being completed in a single operation. This necessitates mechanisms for managing dialogue history, controlling the interaction process, and coordinating real-time communication between the agent and Revit interfaces. To meet these requirements, the agent's functionalities are decomposed into modular components that together support a complete conversational workflow.

Specifically, five agent components are designed to collaboratively construct the interaction framework, supporting instruction parsing, task planning, tool invocation, execution, and user feedback, as shown in Figure 4. A dialogue history management system and a planning and execution algorithm form the core of the framework, orchestrating the interaction flow and coordinating other components. Tool invocation is handled by an executor that maps parsed commands to Revit Interfaces, while LLM interaction is supported by a prompt library and LLM APIs for context-aware prompt assembly and response retrieval. The following describes these five components in detail.

1. Dialogue history management system: BIM tasks typically involve long, multi-step interactions,

requiring the LLM to retain awareness of prior actions and dialogue context. To avoid excessive context growth and degraded model performance, we implement a selective history management mechanism that appends only semantically essential dialogue content while discarding auxiliary prompt information. This design preserves core contextual semantics with significantly reduced input length, enabling longer interaction horizons. Dialogue entries are further annotated with role identifiers to ensure compatibility with standard LLM APIs.

2. Planning and execution algorithm: Following the ReAct paradigm, the agent iteratively performs Think-Act-Observe cycles to complete tasks, interacting with the Revit environment and using execution feedback to guide subsequent actions. The algorithm also determines appropriate task termination and governs the overall interaction workflow, as detailed in Section 3.2.
3. Executor: Parsed agent commands are converted into a structured JSON format and passed to the executor for interface invocation. Serving as an intermediary between the agent core and the Revit interface library, the executor parses inputs, calls the corresponding Revit Interfaces, and returns execution results to the agent core as part of the dialogue.
4. Prompt library: To guide LLM reasoning and execution, two categories of prompts are designed, each combining core instructions, response formats, examples, and interface information. The prompt library is detailed in Section 3.3.
5. LLM APIs: Due to the high resource cost of local deployment, this study adopts an API-based approach to access remote LLMs. The OpenAI API is used for its compatibility with the C# environment in Revit, enabling prompt submission and response retrieval within the agent workflow.

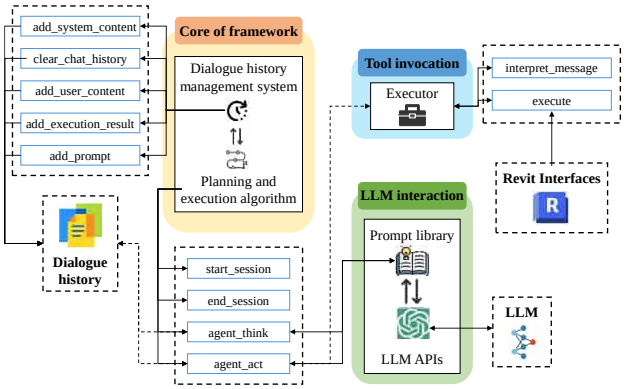


Figure 4. Interaction among the main agent components

3.2 Core of framework Design Based on ReAct

The operation of the agent requires coordinated management of multiple components and proper handling of user-agent dialogue logic. To this end, a ReAct-based core of framework is designed to govern the overall interaction workflow of the Revit Agent, which is executed at the start of each user session. The algorithm is organized into three main stages: initialization, dialogue, and termination, with execution actions conditionally triggered during the dialogue stage, as shown in Figure 5.

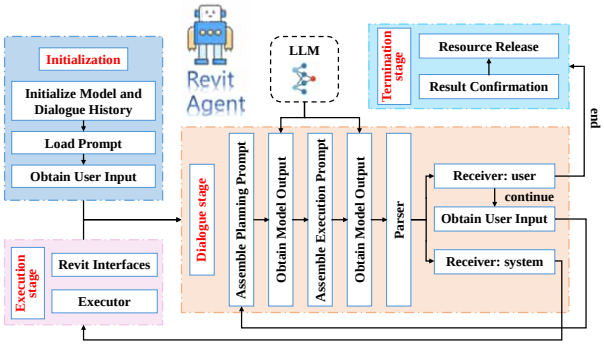


Figure 5. Core of framework design based on ReAct

During initialization, the LLM instance, dialogue history manager, and prompt library are initialized, and users may select different LLM backbones supported by the OpenAI API through configuration settings. The dialogue history manager creates a new history container and populates it with system prompts and synthetic agent-user exchanges that simulate background inquiry, preparing the agent for subsequent multi-turn interaction. Interface-related information is embedded into the dialogue history to support later tool invocation. Finally, the user provides a task description, which serves as the primary objective for the subsequent interaction process.

During the dialogue stage, the algorithm guides the agent through multi-turn interaction following the ReAct Think-Act-Observe paradigm. In each turn, a Think Prompt from the prompt library is first combined with the dialogue history to generate a new prompt, and the LLM produces an unstructured response to plan subsequent actions. Then an Act Prompt that enforces a structured JSON output defined as a Message. The generated JSON is parsed into a C# Message object, enabling executable control flow, as illustrated in Figure 6. Based on the Receiver field in the message, two execution paths are defined. If the receiver is the user, the algorithm evaluates the End flag to determine whether to terminate the interaction or proceed to the next dialogue turn after receiving user input. If the receiver is the system, the

algorithm enters the execution stage, where the executor parses the Call and Args fields to invoke the corresponding Revit Interface. The execution result is then appended to the dialogue history, and the algorithm proceeds to the next interaction cycle.

```
{
"Receiver": system
"Call": RetrieveElement
"Args":{
        "elementIdStrings": null,
        "builtInCategoryName": "OST_Walls",
        "className": null,
        "familyIdString": null,
        "familySymbolIdString": null,
        "levelIdString": null,
       "isElementType": false,
       "parameterFilterArgs": null
   }
}
```

Figure 6. Example of a message class object

The termination stage indicates that the agent determines the task to be completed or infeasible. All executed actions are subject to user confirmation, and any rejected actions are rolled back to ensure BIM model data integrity. Finally, all allocated resources are released, and the dialogue session and task execution are concluded.

3.3 Prompt Design with Multi-Information Integration

LLM outputs are highly sensitive to prompts, making prompt engineering essential. To ensure correct understanding of Revit-specific context and appropriate action generation, instructions, Revit background information, and interface descriptions are integrated into structured prompts. These prompts are organized into three types: Think Prompts, Act Prompts, and Interface Information Prompts.

1. Think Prompts: Consist of core instructions and response examples. The core instructions guide the agent to plan subsequent actions based on the dialogue history, while the examples illustrate reasonable planning responses.
2. Act Prompts: Include core instructions, response examples, and summarized interface information. The core instructions require the agent to generate a structured Message according to a predefined format. While the schema constrains the output format, the examples illustrate correct interface selection and parameter usage. The interface summary briefly lists available interfaces and their functionalities to mitigate LLM forgetting.
3. Interface Information Prompts: Provide complete interface descriptions, including interface names, functionalities, and input-output specifications. Due to their length, these prompts are only injected during the initialization stage.

```
{
"Name": Think Prompt
"Instruction": Based on the task requirements I've presented, and taking into account the historical conversation information, please think in an organized way about what should be accomplished next, what Revit interfaces should be called, or what necessary information you need me to add. If the task has been completed, please think about how you should report and summarize the results to me. Please note that what you think about at this stage will not be executed, you can simply reply in natural language.
"Example": The user wants to get the IDs of all wall instances and the user has given the built-in category ID of the wall as \"OST_Walls\". Among the interfaces provided, the \"RetrieveElement\" interface can filter the eligible element instances by the built-in category. So next, we can call this interface and set the input parameter \"builtInCategoryName\" to \"OST_Walls\", set \"isElementType\" to false and the remaining parameters to null.
}
```

```
{
"Name": Act Prompt
"Instruction": Based on the task requirements and the contents of your thinking above, please reply in JSON format. You can choose to reply to the Revit system to call a specific Revit interface to continue completing the task, or reply to me to ask for more necessary information or report the results after the task is completed. The two types of replies have different format requirements.
"Example": {
        \n Receiver = "system", \n Call = "RetrieveElement", \n Args = {\n { "elementIdStrings", null }, \n
        { "builtInCategoryName", "OST_Walls" }, \n { "className", null }, \n { "familyIdString", null }, \n
        { "familySymbolIdString", null }, \n { "levelIdString", null }, \n { "isElementType", false }, \n { "parameterFilterArgs", null \n }}
   }
}
```

Figure 7. Prompt examples

4 Benchmark for Agent-Based BIM Design Assistance Tasks

4.1 Manually Annotated Test Dataset

Due to the lack of publicly available datasets for Revit-based BIM design tasks, the benchmark dataset is manually collected and annotated to ensure reliability and evaluation relevance. All tasks supported by the current interfaces are categorized into four types: information querying, model visualization, component modification, and component deletion. Tasks are designed for each category based on a reference BIM model and real-world design scenarios, with multiple difficulty levels introduced according to operational complexity.

Each task is accompanied by a manually written reference answer describing the correct execution outcome or feasible interface invocation. In total, 100 English test tasks with annotated answers are constructed to ensure balanced evaluation across different LLMs, with task statistics shown in Table 3.

Table 3. Task statistics

Task Categories	Quantity
Information querying	52
Model visualization	18
Component modification	20
Component deletion	10

4.2 Evaluation Criteria

In practical BIM design and visualization tasks, users primarily focus on the correctness of final results rather than execution details. Accordingly, this study adopts an outcome-oriented evaluation strategy using two metrics: task completion rate and task accuracy.

The task completion rate measures whether the agent successfully completes a task and reports the result, while task accuracy evaluates whether the completed task produces correct outcomes consistent with user requirements. Both metrics are computed as the ratios of completed or accurate tasks to the total number of tasks. Evaluation is conducted by manually comparing the agent's outputs and execution results against annotated reference answers.

5 Experiment

To assess the reliability of the proposed agent system and the impact of different LLM capabilities, three representative models from the GLM series, GLM4-flash, GLM4-plus, and GLM-Zero-Preview, are evaluated. These models respectively represent mid-level general-purpose performance, flagship general-purpose capability, and reasoning-oriented models with strong logical inference ability.

To facilitate a more detailed analysis of LLM performance, results are reported by task category and difficulty level in addition to overall performance. Tasks are grouped into information querying, model visualization, component modification, and component deletion, while difficulty is defined by the minimum number of interaction turns required for completion ($\leqslant$2, 3-4, and $\geqslant$5 turns for Levels 1-3, respectively). This criterion reflects increasing reasoning complexity and dialogue history length.

5.1 Overall Results

Table 4 summarizes the results on the test dataset. The results show that GLM4-plus achieves the best performance, reaching 90% accuracy, which exceeds prior LLM-assisted BIM methods such as BIMS-GPT [16] under zero-shot prompting. It should be noted that, due to differences in tasks and datasets, the comparison with prior studies should be interpreted as a qualitative reference rather than a strictly equivalent benchmark. GLM-Zero-Preview and GLM4-flash show similar accuracy but lower performance than GLM4-plus, while GLM-Zero-Preview exhibits a notably lower completion rate despite high correctness on completed tasks. This is mainly due to its tendency to request additional information when handling complex interface calls. Overall, the results indicate that GLM4-plus enables more reliable BIM-assisted design and visualization within the proposed agent framework.

Table 4. Overall results

Model	Context Limit (K)	Completion rate (%)	Accuracy (%)
GLM4-flash	128	70	93
GLM4-plus	128	90	99
GLM-Zero-Preview	16	71	71

5.2 Performance Across Task Categories

Across task categories, the three LLMs exhibit distinct strengths, as shown in Figure 8-9. GLM4-plus achieves near-perfect completion and 100% accuracy in component modification and deletion, with consistently high performance across all tasks. GLM4-flash performs well in information querying and component deletion but is less effective in model visualization and component modification. GLM-Zero-Preview excels in component modification, while its identical completion and accuracy rates indicate that all completed tasks are correct and all failures result from incomplete execution, mainly due to incorrect judgments about required task information rather than interface invocation errors.

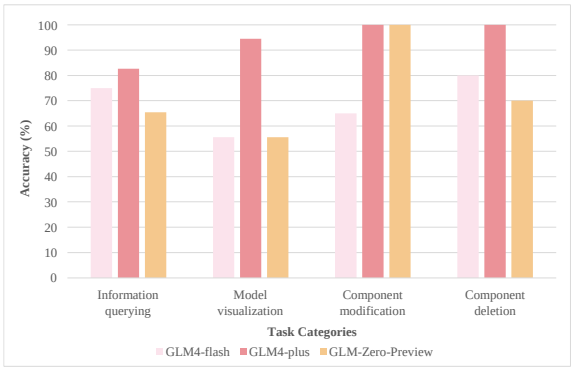


Figure 8. Classification accuracy

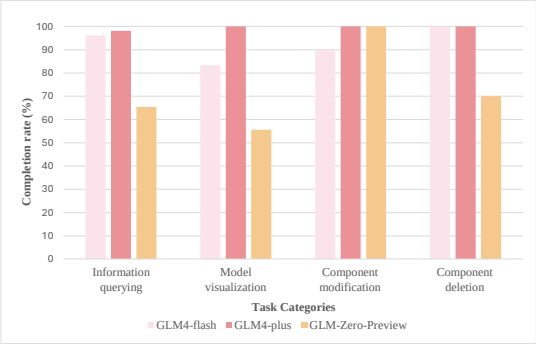


Figure 9. Completion rate

5.3 Performance Across Task Difficulty Levels

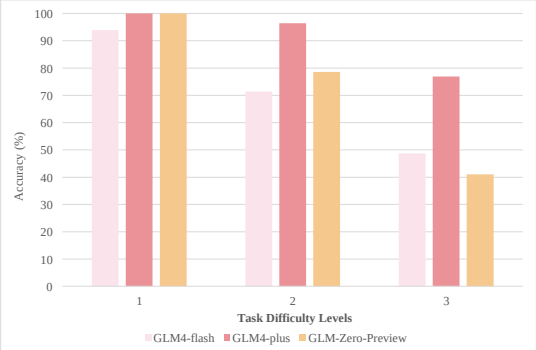


Figure 10. Classification accuracy

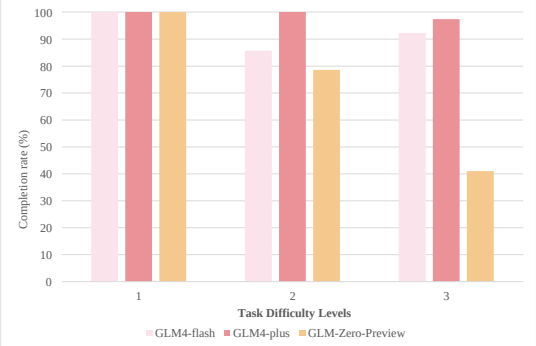


Figure 11. Completion rate

The benchmark includes 33 Level-1, 28 Level-2, and 39 Level-3 tasks, yielding a balanced difficulty distribution. As expected, both accuracy and completion rate decrease with increasing difficulty for all models. While all LLMs perform well on Level-1 tasks, accuracy for GLM4-flash and GLM-Zero-Preview drops sharply on Level-2 and Level-3 tasks, reflecting increased reasoning complexity with longer interactions.

Completion rate analysis further shows that GLM4-flash behaves more aggressively by attempting task completion at all difficulty levels regardless of correctness, while GLM-Zero-Preview adopts a more conservative strategy, frequently requesting additional user input when confidence is insufficient, leading to a higher rate of incomplete tasks.

Figure 12 presents an end-to-end example of a GLM4-plus–based agent successfully completing a task in the model visualization category (difficulty level 2). The task is to display all door instance elements on Level 2 in the Revit 3D view. As shown, the agent first retrieves the corresponding elements using the RetrieveElement interface and then invokes ShowElements based on the returned element IDs. By sequentially using system responses to guide subsequent interface calls, the agent accurately completes the task.



Figure 12. An end-to-end example

6 Conclusion

This study presents a BIM design assistant driven by LLM agents and implements a reliable prototype within the Autodesk Revit environment The main contributions are as follows:

1. A two-layer interface architecture is designed and implemented, accompanied by structured interface descriptions. This design enables real-time manipulation of Revit BIM models by the agent and ensures efficient and accurate interface invocation for seamless BIM interaction.
2. A BIM design agent is developed based on ReAct and Prompt Engineering. By integrating dialogue history management, ReAct-based reasoning, and structured prompt engineering, the agent is endowed with step-by-step reasoning capability, allowing it to handle complex tasks and long-

horizon interactions effectively.

3. An agent-oriented BIM-assisted design benchmark is constructed, introducing task completion rate and task accuracy as evaluation metrics. A total of 100 manually annotated test tasks are created to systematically evaluate the reliability and accuracy of the proposed system.

Experimental results show that high-performance general-purpose LLMs, represented by GLM4-plus, can reliably act as BIM design assistants within the proposed framework, accurately completing tasks of varying types and difficulty levels with an accuracy of up to 90%. These findings demonstrate that, when combined with appropriate agent architectures and domain-specific interfaces, LLM-driven design assistants can provide reliable and practical support for BIM-assisted design and visualization.

Acknowledgments

The authors are grateful for the financial support received from the National Natural Science Foundation of China (No. 52378306).

References

[1] Hu Z Z, Leng S, Lin J R, et al. Knowledge extraction and discovery based on BIM: a critical review and future directions. Arch Computat Methods Eng, 2022; 29(1): 335-356. https://doi.org/10.1007/s11831-021-09576-9.

[2] Migilinskas D, Popov V, Juocevicius V, et al. The Benefits, Obstacles and Problems of Practical Bim Implementation. Procedia Eng 2013;57:767–74. https://doi.org/10.1016/j.proeng.2013.04.097.

[3] Han J, Lu X, Lin J. Pretrained graph neural network for embedding semantic, spatial, and topological data in building information models. Comput-Aided Civ Infrastruct Eng 2025;40:4607–31. https://doi.org/10.1111/mice.70073.

[4] Pan, Y., Wang, M., Lu, L., Lamsal, R., Pärn, E., Zlatanova, S., & Brilakis, I. (2026). LLM-enabled multi-agent framework for natural language interaction with graph-based digital twins. *Automation in Construction*, *183*, 106791.

[5] Huang P-H, Song S-Y, Xu Z, et al. Intelligent BIM Searching via Deep Embedding of Geometric, Semantic, and Topological Features. Buildings 2025;15:951. https://doi.org/10.3390/buildings15060951.

[6] Ni X-R, Pan P, Lin J-R. What makes a good BIM design? Quantifying the link between design behavior and quality. Autom Constr 2025;171:105992. https://doi.org/10.1016/j.autcon.2025.105992.

[7] Zhou, Y. C., Zheng, Z., Lin, J. R., & Lu, X. Z. (2022). Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. Computers in Industry, 142, 103746.

[8] Zheng Z, Han J, Chen K-Y, et al. Translating regulatory clauses into executable codes for building design checking via large language model driven function matching and composing. Eng Appl Artif Intell 2026;163:112823. https://doi.org/10.1016/j.engappai.2025.112823.

[9] Zheng, Z., Lu, X. Z., Chen, K. Y., Zhou, Y. C., & Lin, J. R. (2022). Pretrained domain-specific language model for natural language processing tasks in the AEC domain. Computers in Industry, 142, 103733.

[10] Lin T-H, Huang Y-H, Putranto A. Intelligent question and answer system for building information modeling and artificial intelligence of things based on the bidirectional encoder representations from transformers model. Autom Constr 2022;142:104483. https://doi.org/10.1016/j.autcon.2022.104483.

[11] Jiang G, Ma Z, Zhang L, et al. EPlus-LLM: A large language model-based computing platform for automated building energy modeling. Appl Energy 2024;367:123431. https://doi.org/10.1016/j.apenergy.2024.123431.

[12] Nakano R, Hilton J, Balaji S, et al. WebGPT: Browser-assisted question-answering with human feedback 2022. https://doi.org/10.48550/arXiv.2112.09332.

[13] Brown TB, Mann B, Ryder N, et al. Language Models are Few-Shot Learners. Adv. Neural Inf. Process. Syst. 33, 2020.

[14] Ying H, Sacks R. From Automatic to Autonomous: A Large Language Model- driven Approach for Generic Building Compliance Checking. Proc. CIB W78 Conf., 2024.

[15] Jang S, Lee G, Oh J, et al. Automated detailing of exterior walls using NADIA: Natural-language-based architectural detailing through interaction with AI. Adv Eng Inform 2024;61:102532. https://doi.org/10.1016/j.aei.2024.102532.

[16] Zheng J, Fischer M. BIM-GPT: a Prompt-Based Virtual Assistant Framework for BIM Information Retrieval 2023. https://doi.org/10.48550/arxiv.2304.09333.

An Automated Optimization Framework for Hierarchical Access Road Design in Onshore Wind Farm Construction

Yuxuan Cheng[1], Fumiya Matsushita[1], Sheng Lian[1], Takahiro Sayama[1], Takumi Kaneko[1]

[1] Civil Engineering Technology Div., Civil Engineering Headquarters, Shimizu Corporation, Tokyo, Japan
y-cheng@shimz.co.jp, f_matsushita@shimz.co.jp, s_lian@shimz.co.jp, sayama@shimz.co.jp,t-kaneko@shimz.co.jp

Abstract –

Onshore wind power plays a critical role in sustainable energy development. However, its deployment is often constrained by complex construction-stage logistics, particularly the planning and design of access roads. During the early-stage evaluation of onshore wind projects, construction road layout design and earthwork estimation remain labor-intensive due to terrain and engineering constraints.

This study proposes an automated optimization framework for the design of hierarchical access roads during onshore wind farm construction. The framework integrates high-resolution terrain data with parametric road modeling and constrained optimization to generate geometrically feasible road alignments. Key engineering constraints, including longitudinal gradient limits, deviation ratio from wind turbine locations, and cut-and-fill earthwork requirements, are explicitly incorporated into the optimization process. By formalizing road layout planning as a solvable computational problem, the proposed method transforms a process into an automated decision support workflow.

The applicability of the proposed framework is illustrated through an example, demonstrating its capability to generate feasible access road alignments while satisfying engineering constraints. The automated workflow supports rapid construction cost estimation during the planning stage.

Keywords –

Wind turbine access roads; Road alignment optimization; Earthwork optimization; Engineering constraints

1 Introduction

Wind energy has become a critical pillar of the global renewable energy portfolio, driven by the increasing demand for clean and sustainable power sources [1]. Despite challenges related to technology deployment, environmental constraints, economic feasibility, and social acceptance [2,3], onshore wind power continues to attract substantial investment and accounts for a significant share of the global energy market [4]. As global decarbonization targets accelerate, the efficient development of onshore wind farms remains essential for meeting future energy demands [5,6].

From a construction perspective, the deployment of onshore wind projects involves a multi-stage process. Typical workflows include candidate site selection based on wind resource assessment, for instance, by NEDO [7], wind turbine layout optimization considering terrain conditions, and construction planning and cost estimation. Among these stages, construction access road planning encompassing route selection, geometric design, and earthwork quantity estimation, represents a major bottleneck in practical project implementation. This stage is inherently complex and resource-intensive, as it must account for diverse land-use classifications, junction design between turbine platforms and access roads, the integration of disaster prevention facilities, and detailed construction cost calculations.[8] In practice, these tasks rely heavily on manual expertise and iterative trial-and-error processes, leading to high labor demands and prolonged project timelines. For example, even when multiple candidate plans satisfy the same preliminary conditions, a comparative evaluation is still required, as these plans may differ in design priorities. [8-10].

To address this challenge, this study proposes a computational approach that formalizes construction access road planning as an automated optimization problem. A geometric optimization framework is developed using high-resolution point cloud data to generate geometrically feasible road alignments and gradients under engineering constraints. By transforming a traditionally experience-driven task into an automated decision support workflow, the proposed method streamlines construction stage planning and facilitates more efficient and reliable construction cost estimation for onshore wind farm projects.

2 Problem Definition and Design Constraints

2.1 Road Design for Wind Turbine Access Roads

Road design is a critical aspect of civil engineering, particularly for access infrastructure in complex terrain. It involves alignment planning, longitudinal gradients, and earthwork operations, including excavation, embankment formation, grading, and compaction, to ensure adequate drainage, structural stability, and construction feasibility.

In the specific context of wind turbine access roads, road geometry and earthwork quantities are strongly coupled. Longitudinal gradients and alignment continuity govern the feasibility of transporting heavy construction vehicles and oversized turbine components. In addition, road design must account for heterogeneous land use conditions, such as agricultural land, forest areas, and developed zones. Consequently, earthwork requirements, road geometry, land use, and environmental considerations must be addressed in an integrated manner to ensure the overall feasibility and performance of wind turbine access road design.

2.2 Engineering Design Constraints

Engineering design must satisfy a set of interrelated constraints to ensure construction feasibility, operational safety, and regulatory compliance. In this study, the following key design constraints are incorporated into the optimization framework.

2.2.1 Wind Turbine Siting Constraints

Wind turbine locations are treated as predefined inputs in the access road planning process. Turbine siting is determined based on wind resource assessments, environmental regulations, and geotechnical feasibility, and is assumed to be fixed prior to construction road design. Each turbine location therefore acts as a mandatory destination that access roads must connect to, subject to geometric and constructability requirements.

2.2.2 Longitudinal Gradient Constraints

Longitudinal gradients of access roads are constrained to ensure the safe transportation of heavy construction vehicles and oversized turbine components. Maximum allowable gradients are specified according to engineering standards and vehicle performance limitations, typically less than 12% [11], also depending on site conditions. In the proposed framework, gradient constraints are enforced along the entire road alignment and directly interact with earthwork requirements.

2.2.3 Connectivity to Existing Roads and Forest Trails

Access road layouts must ensure appropriate connectivity to existing public roads and forest trails to support construction logistics and long-term site accessibility. Feasible connection locations are constrained by traffic safety requirements, visibility conditions, environmental regulations, and land-use limitations. In the optimization framework, access road alignments are restricted to connect through predefined candidate access points while satisfying geometric and constructability requirements.

2.2.4 Cut-and-Fill Earthwork Constraints

Cut-and-fill earthwork constraints are introduced to control excavation and embankment volumes along access road alignments. Excessive imbalance between cut and fill increases construction costs, material transportation demands, and environmental impacts, and may compromise slope stability. Accordingly, earthwork constraints are formulated to promote balanced material utilization. In practice, cut and fill volumes are not exchanged on a 1:1 basis due to soil bulking and compaction effects. In this study, however, the influence of bulking and shrinkage factors on the relative comparison among candidate routes is expected to be limited. Thus, earthwork quantities are used for comparative evaluation of alternative routes, and these effects are therefore not explicitly considered.

2.2.5 Horizontal Curvature Constraints

Horizontal curvature constraints are imposed to accommodate the maneuverability of construction and maintenance vehicles, particularly those transporting large wind turbine components. In practice, these constraints are expressed as minimum allowable turning radii to ensure safe vehicle operation and compliance with design standards. In the present study, however, horizontal curvature is treated as a general alignment consideration rather than an explicit minimum turning-radius constraint.

2.3 Optimization Problem Overview

Based on the engineering considerations described above, access road planning is addressed as a multi-objective optimization problem within an optimization framework. The objective aims to identify road configurations that achieve a favorable balance between construction-related effort, such as earthwork quantities, and road accessibility, while satisfying the prescribed engineering requirements. Meanwhile, soft performance is also taken into account through penalty terms. The detailed representation and solution strategy are described in the subsequent chapter.

3 System Framework and Modeling Approach

3.1 Overall System Architecture

To address the challenge of construction access road optimization under complex terrain conditions, this study proposes an automated framework composed of three primary components: (1) ridge detection for terrain feature extraction, (2) primary access road selection, and (3) secondary access road selection. These components are supported by two core computational modules, namely parametric road model creation and cut-and-fill earthwork volume calculation. Within the framework, terrain data are first processed to identify ridge features and generate initial road profiles, which are subsequently refined through road modeling and earthwork evaluation. By integrating these elements, the proposed framework enables the systematic selection of main and branch access road alignments that balance construction cost, earthwork demand, and accessibility requirements. An overview of the proposed framework is illustrated in Figure 1.

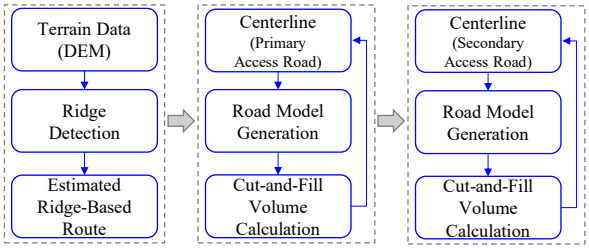


Figure 1. Framework overview

3.2 Ridge Detection and Terrain Feature Extraction

Ridge detection is employed to support access road planning for wind turbine construction, as routing roads along ridge lines generally minimizes earthwork demand and reduces construction costs. In this study, mountain ridges are identified using a watershed-based terrain analysis approach, where ridges are defined as topographical boundaries separating adjacent drainage catchments. The target terrain is first represented as a Digital Elevation Model (DEM), which is then processed using a watershed algorithm to extract ridge features. In this procedure, the DEM is treated as a surface subjected to a simulated flooding process initiated from local minima. As water levels rise uniformly within individual catchment basins, watershed boundaries are formed at the interfaces where adjacent basins meet. These boundaries constitute a network of watershed lines that correspond to mountain ridges. The extracted ridge features provide terrain-aware guidance for generating initial access road profiles, forming the basis for subsequent road modeling and optimization. The overall concept of ridge detection using the watershed approach is illustrated in Figure 2.



Figure 2. Watershed-based detection concept [12]

3.3 Parametric Road Modeling

The road model creation module generates an integrated mesh representation of access roads, including the road surface and associated cut-and-fill slopes. The module requires three primary inputs: road width, side batter width, and the prescribed cut-and-fill slope angle. As illustrated in Figure 3(a), the process begins with a parametric calculation that determines the initial road profile location. At this location, a horizontal road surface with a predefined width is generated. The outer edges of the road surface are then extended at the specified slope angle to form cut-and-fill side slopes. For each edge, the model automatically determines whether a cut or a fill operation is required by comparing the edge elevation with the existing terrain surface. If the edge lies below the terrain, a cut slope is generated upward until it intersects the ground surface; if the edge lies above the terrain, a fill slope is generated downward until ground intersection is achieved. The resulting road-and-slope mesh as shown in Figure 3(b) provides a geometrically consistent representation of the road cross-section and is subsequently passed to the earthwork calculation module for cut-and-fill volume estimation.

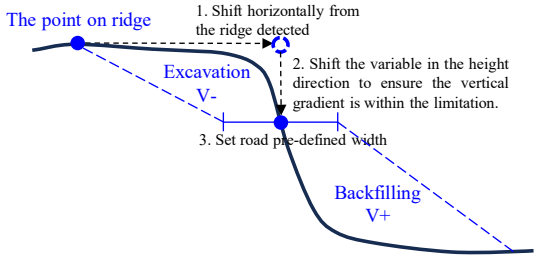
3.4 Cut-and-Fill Earthwork Calculation

The final earthwork volumes are computed by spatially comparing the modified terrain mesh incorporating the road model with the original terrain mesh. The volumetric difference between these two surfaces represents the total cut-and-fill earthwork required for access road construction. In the proposed framework, earthwork volumes are calculated by evaluating the vertical differences between the two meshes over their shared horizontal projection. Regions

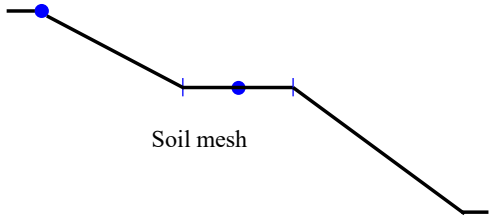
where the modified road mesh lies above the original terrain are classified as fill volumes (V+), corresponding to material that must be placed, whereas regions where the modified mesh lies below the original terrain are classified as cut volumes (V-), corresponding to material that must be excavated. These cut and fill volumes are computed independently and aggregated over the entire road alignment, providing quantitative inputs for construction planning, earthwork balancing, and cost estimation within the automated workflow.

3.5 Optimization Formulation and Solution Strategy

A hierarchical optimization approach is adopted to balance construction cost and road accessibility in access road planning. Although numerous researchers have proposed a wide range of multi-objective optimization algorithms [13-15], evaluating candidate road networks in this study requires computationally expensive geometric modeling and earthwork calculations, which makes conventional population-based methods inefficient for large-scale exploration. Considering the large-scale nature of the problem and the objective of preliminary estimation rather than detailed design, a surrogate-assisted multi-objective optimization strategy based on Radial Basis Functions (RBFMOpt) is employed. Different objective functions are formulated at different stages, and RBF-based surrogate models are used to approximate these objectives, reducing the number of expensive evaluations while enabling efficient exploration. This solution strategy enables efficient optimization within the proposed automated framework and is well suited for construction-stage decision support under complex terrain conditions.



(a) Parametric generation of road profile



(b) Road earthwork mesh

Figure 3. Parametric road and earthwork modeling

4 Route Optimization Methodology

4.1 Computational Environment and Modeling Tools

The proposed optimization framework is implemented within the Rhino–Grasshopper computational environment. Rhino [16] is used for precise three-dimensional geometric modeling, while Grasshopper [17] provides a visual programming interface that enables flow-based logic control of geometric operations without conventional text-based coding. This environment allows direct integration of geometric road modeling, terrain processing, and optimization workflows, making it well suited for construction-stage access road planning. In addition, Grasshopper's extensive ecosystem of community-developed plugins provides tools for terrain analysis, parametric modeling, and multi-objective optimization, enabling seamless integration of the geometric modeling modules and optimization solvers required by the proposed automated framework.

4.2 Terrain Data Preparation

The terrain data used in this study are based on 5-meter-high-resolution open-source Digital Elevation Model (DEM) data provided by Geoscience Australia [18]. The DEM serves as the topographic foundation for watershed-based ridge detection and subsequent access road planning. In the proposed workflow, the DEM is imported into the Grasshopper environment as a mesh using a module from the Bison package [19], with input parameters defining the file path, sampling factor, and spatial resolution in the horizontal directions, as shown in Figure 4(a). A square site with plan dimensions of 495 m × 495 m was selected as the illustrative case study area, with elevations ranging from 376.82 m to 596.02 m, as illustrated in Figure 4(b).

The terrain geometry and turbine layout are visualized in both top and perspective views within the Rhino–Grasshopper environment, as shown in Figure 5. The terrain is represented as a continuous mesh surface derived from the DEM, while the wind turbine locations are overlaid as discrete points distributed across the site. As the same time, ten wind turbine locations were randomly generated within the study area and treated as fixed constraints in the subsequent access road alignment optimization, also as shown in Figure 5.

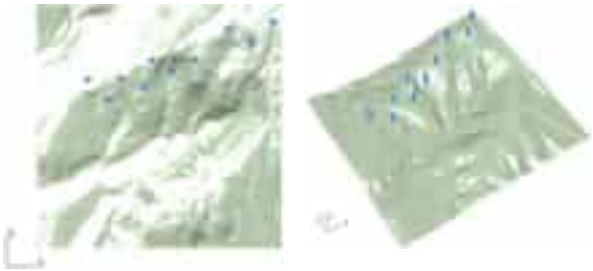


(a) DEM import workflow



(b) Generated terrain mesh.

Figure 4. Terrain data preparation.



(a) Top view (b) Perspective view

Figure 5. Terrain mesh and wind turbine locations

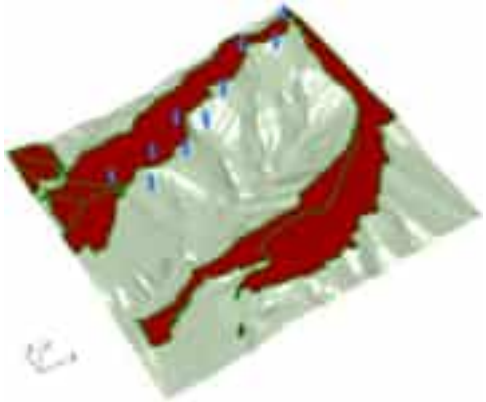
4.3 Optimization of Primary Access Roads

4.3.1 Ridge-Based Initial Route Estimation

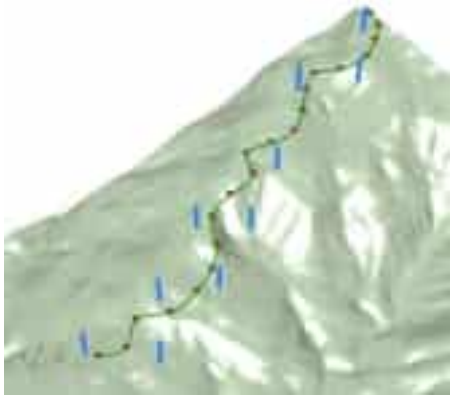
Initial candidate routes for the primary access road are generated based on the Digital Elevation Model, as illustrated in Figure 6(a). Ridge lines identified through watershed-based analysis provide elevated paths that generally minimize earthwork demand. To improve geometric continuity and constructability, the rough alignment is subsequently smoothed through interpolation. In this study, as shown in Figure 6(b), the initial alignment was discretized into approximately 20 segments, and an interpolated curve was generated to form a continuous and refined initial route that serves as the starting point for further optimization.

4.3.2 Local Optimization Under Gradient Constraints

Although the initial alignment was obtained, as shown in Figure 6(b), several segments did not satisfy the longitudinal gradient requirement for this illustrative case (20%), when the alignment was discretized into 20 segments. Consequently, these non-compliant segments were extracted for further processing.



(a) Extracted ridge features



(b) Smoothed ridge-based alignment

Figure 6. Ridge-based alignment extraction

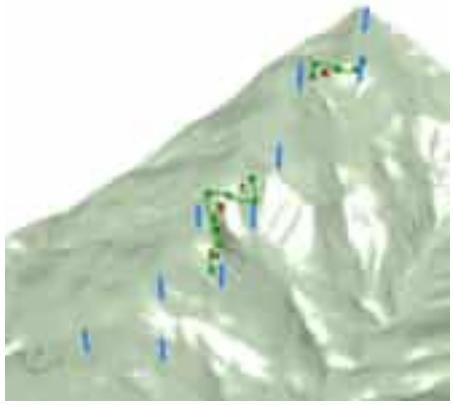


Figure 7. Local optimization under longitudinal gradient constraints

To address this issue, the maximum longitudinal gradient was treated as a strict engineering constraint rather than a conventional objective function. Specifically, the maximum longitudinal gradient was subject to an upper-bound constraint, and, even when this constraint was satisfied, it was still incorporated into the algorithm as a penalty term to encourage alignments with smaller gradient values. For each non-compliant segment, a midpoint was introduced to increase segment resolution, as indicated by the red crosses in Figure 7. The midpoints were then allowed to move in the *Y*-direction within the *XY* plane, while the corresponding longitudinal gradients were recalculated. Once all segments satisfied the prescribed gradient constraint, one of the optimized alignments was obtained, as illustrated by the green alignment in Figure 7.

4.3.3 Global Horizontal Alignment Optimization

Following the local optimization under vertical gradient constraints, an initially optimized main route alignment is obtained in the vertical plane. At this stage, the optimization only considers the maximum longitudinal gradient and does not account for additional geometric and planning factors. In the subsequent step, road length and deviation ratio *R* were treated as bi-objective functions, and RBFMOpt was used to generate Pareto-optimal candidate solutions. The final alignment was selected by minimizing road length among solutions satisfying an acceptable *R* threshold, without introducing weighting factors. As shown in Figure 8, the resulting alignment had a total road length of 1103.66 m and a maximum longitudinal gradient of 17.41%.

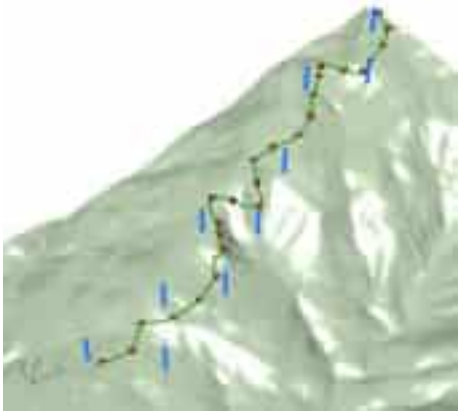


Figure 8. Local optimization under vertical gradient and total road length constraints

4.3.4 Global Vertical Alignment Optimization

Finally, global vertical alignment optimization is performed with earthwork volume introduced as an additional objective function. In this step, the first point of the main route is fixed, and subsequent points are iteratively adjusted in the negative vertical *Z*-direction to evaluate corresponding earthwork volumes. The earthwork volume consists of both cut and fill, which are calculated for each trial alignment. The resulting alignment under optimal conditions. In this case, as the Table 1 shown, the cut volume was 21,449.46 m^3, the fill volume was 83,846.15 m^3, and the total earthwork volume, defined as the sum of absolute cut and fill, was 105,295.61 m^3. The additional volume associated with cut-fill imbalance was 62,396.69 m^3. Through this procedure, the final main route alignment is determined hierarchically, with earthwork objectives optimized within the feasible space defined by prior gradient constraints.

Table 1. Mass balance of primary road earthworks

Earthwork	Cut (-) [m^3]	Fill (+) [m^3]	Mass balance [m^3]
Volume	21,449.46	83,846.15	62,396.69

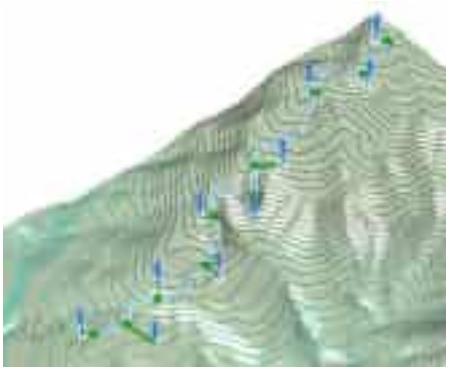
4.4 Optimization of Secondary Access Roads

4.4.1 Network-Level route selection

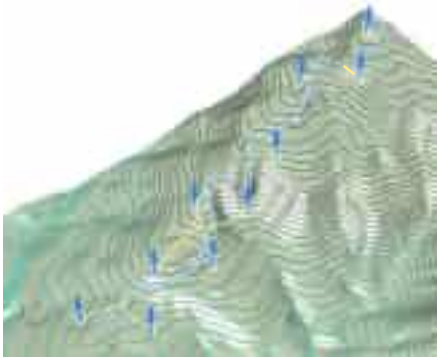
The network-level selection of secondary access roads connecting wind turbines to the primary route follows a systematic procedure. For each wind turbine, the closest point on the primary route alignment is first identified, and an initial straight-line connection between the turbine location and this point is generated, as shown in Figure 9(a), where contour lines with a 5 m interval are overlaid for visualization purposes. The straight-line branch is then divided into three equal sub-segments, and the vertical gradient of each sub-segment is evaluated. If all sub-segments satisfy the allowable gradient threshold, the straight line is retained as a candidate branch road. If any sub-segment violates the gradient constraint, the corresponding branch is selected for further refinement.

In this refinement step, the branch remains represented as a straight line, while the connection point on the primary route is iteratively adjusted to minimize branch length, subject to satisfying the prescribed gradient constraint for all sub-segments. By combining the resulting branch roads with the primary route, an initial configuration of the access road network is obtained.

Among the generated branches as shown in Figure 9(b), those serving turbines 0, 2, 3, 4, 5 and 8 were identified for subsequent secondary optimization, as no feasible connection point along the primary route can satisfy the gradient constraint through network-level adjustment alone. The remaining branches correspond to routes accepted at the initial network-level selection stage.

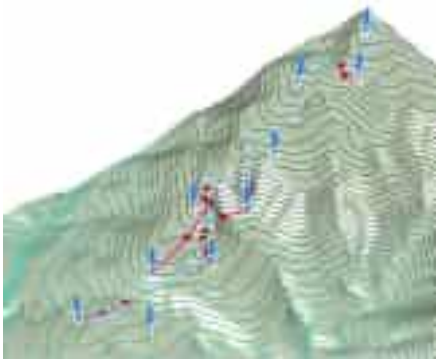


(a) Initial straight-line connection roads

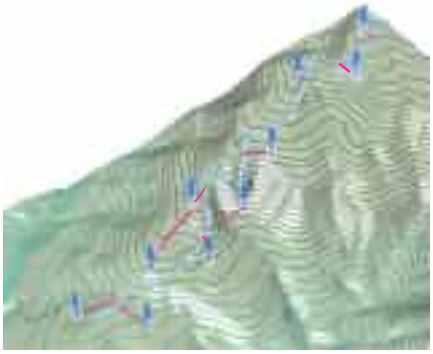


(b) Network-level selected branch roads

Figure 9. Network-level branch road selection



(a) Optimization of non-compliant branch roads



(b) Final optimized branch roads

Figure 10. Local optimization of secondary access roads

4.4.2 Local Optimization Under Gradient Constraints

Local optimization under vertical gradient constraints was subsequently performed for the branch roads serving turbines 0, 2, 3, 4, 5 and 8. For each branch, a midpoint is inserted into the initial straight-line segment while the two endpoints remain fixed. This midpoint is allowed to move within the *XY* plane to reduce excessive vertical gradients along the branch. A smooth curve is then fitted using the two endpoints and the midpoint to generate the optimized branch alignment, as illustrated by the pink curves and control points in Figure 10(a). The maximum longitudinal gradient was treated as a constraint, and the optimization aimed to obtain feasible alignments rather than balance multiple objectives.

The final branch alignments are shown in Figure 10(b). The corresponding earthwork quantities are summarized in Table 2, with a total cut volume of 137,064.47 m³, a fill volume of 129,647.73 m³, and a volume balance of −7,416.74 m³.

Table 2. Mass balance of secondary roads earthworks

Earthwork	Cut (-) [m^3]	Fill (+) [m^3]	Mass balance [m^3]
Volume	137,064.47	129,647.73	-7,416.74

5 Conclusion and Future Work

5.1 Conclusions

This study proposed an automated optimization framework for construction-stage planning of access road networks in onshore wind farm projects. The framework integrates terrain-aware analysis, parametric road modeling, earthwork evaluation, and stage-wise optimization into a unified computational workflow. Both primary and secondary access roads are generated through a hierarchical optimization strategy that incorporate practical engineering constraints, including longitudinal gradient limits, road geometry, deviation control from turbine locations, connectivity requirements, and cut-and-fill earthwork characteristics.

Through an illustrative example, the proposed framework demonstrates its capability to generate construction-feasible access road networks under complex terrain conditions. The results show that the framework can systematically support early-stage route selection, earthwork estimation, and network configuration, providing quantitative information for construction planning and decision support. The optimization process combines constraint enforcement (e.g., gradient limits) with local objective-based refinement, enabling efficient solution generation without relying on global weighting factors. The proposed approach is particularly suited for preliminary evaluation and scenario exploration rather than detailed design.

5.2 Future Work

Future work will further extend the current framework beyond the illustrative 495 × 495 m case to larger-scale terrains, with particular attention to computational efficiency and scalability. This extension will also enable systematic comparison with expert-designed road alignments to evaluate the practical performance of the proposed framework. In this context, data-driven approaches, such as reinforcement learning, could be explored to enhance cost-oriented optimization under more complex conditions.

In addition, more realistic earthwork modeling can be considered by introducing non-unity cut–fill conversion factors. The current interpolation-based approach may also be refined by incorporating explicit geometric constraints, such as minimum turning radius.

References

[1] B. N. Owens. The wind power story: a century of innovation that reshaped the global energy landscape. John Wiley & Sons, 2019.

[2] G. A. M. Van Kuik, J. Peinke, R. Nijssen, D. Lekou, J. Mann, J. N. Sørensen, ... & K. Skytte. Long-term research challenges in wind energy–a research agenda by the European Academy of Wind Energy. *Wind energy science*, 1(1): 1-39, 2016.

[3] E. Chomać-Pierzecka, H. Gąsiński, J. Rogozińska-Mitrut, D. Soboń, & S. Zupok. Review of selected aspects of wind energy market development in Poland and Lithuania in the face of current challenges. *Energies*, 16(1): 473,2023.

[4] N. Ederer. The market value and impact of offshore wind on the electricity spot market: Evidence from Germany. *Applied Energy*, 154: 805-814, 2015.

[5] S. Miyake, S. Teske, J. Rispler, & M. Feenstra. Solar and wind energy potential under land-resource constrained conditions in the Group of Twenty (G20). *Renewable and Sustainable Energy Reviews*, 202: 114622, 2024.

[6] K. Linnerud, A. Dugstad, & B. J. Rygg. Do people prefer offshore to onshore wind energy? The role of ownership and intended use. *Renewable and Sustainable Energy Reviews*, 168: 112732, 2022.

[7] NEDO, NeoWins. On-line: https://appwdc1.infoc.nedo.go.jp/Nedo_Webgis/top.html

[8] A simple explanation of road administration. In Japanese. On-line: https://www.mlit.go.jp/road/sisaku/dorogyousei/0.pdf

[9] NEDO Renewable Energy Technology White Paper: Wind Power Generation Technology Development Roadmap and Promotion. *Wind Energ.* In Japanese. On-line: https://www.nedo.go.jp/content/100116324.pdf

[10] Y. Amano and F, Naganuma. The principles, structure, and construction of wind power generation. *Construction planning*, (746): 48-54, 2021. In Japanese. On-line: https://jcmanet.or.jp/bunken/kikanshi/2012/04/048.pdf.

[11] Cabinet Office, Government of Japan. The relaxation of road standards in response to the characteristics of steep terrain enables the development of roads that are more convenient for residents and less costly to construct, 2015. In Japanese. On-line: https://www.cao.go.jp/bunken-suishin/doc/jirei30_h27_08.pdf

[12] Watershed Projection Programs, On-line: https://www.raritanbasin.org/what-is-a-watershed

[13] D. Rutten. Galapagos: On the logic and limitations of generic solvers. *Architectural Design*, *83*(2), 132-135, 2013.

[14] W, Thomas. OPOSSUM: Introducing and Evaluating a Model-based Optimization Tool for Grasshopper. *Proceedings of the 22nd International Conference of the Association for Computer-Aided Architectural Design Research in Asia*, pages 283-293, Hong Kong, 2017.

[15] R. Peikert, and F, Sadlo. Height ridge computation and filtering for visualization. *2008 IEEE Pacific Visualization Symposium*, pages 119-126, Kyoto, Japan, 2008.

[16] Rhino-Rhinocero3D, On-line: https://www.rhino3d.com/

[17] Grasshopper - algorithmic modeling for Rhino, On-line: https://www.grasshopper3d.com/

[18] Digital Elevation Data | Geoscience Australia, On-line: https://www.ga.gov.au/scientific-topics/national-location-information/digital-elevation-data

[19] Bison|Food4, On-line: Rhino, https://www.food4rhino.com/en/app/bison

BIM-based Hybrid Knowledge Graph based on DSL for Computational Design Rule Formalization

Huzhou Deng[1,2] and Dian Zhuang[1,2*]

[1]School of Architecture and Design, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China
[2]Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin, China
zhuangdian@hit.edu.cn

Abstract

The digitization of building codes is critical for Automated Compliance Checking (ACC) within the BIM paradigm, yet faces challenges such as unstructured text, semantic ambiguity, and deep logical nesting. Traditional hard-coding lacks flexibility, while reliance solely on Large Language Models (LLMs) is prone to hallucinations and offers limited BIM integration. This study proposes a multi-stage framework integrating Domain-Specific Languages (DSL), hybrid Knowledge Graphs, and LLMs to automate the formalization of computational design rules from textual codes. Using the Code for Design of Residential Buildings (GB 50096-2011) as a case study, we parse and classify regulatory provisions into four categories , Numerical, Logical, Topological, and Fuzzy and employ Prompt Engineering to guide LLMs in generating structured DSL JSON. A hybrid storage architecture (Neo4j, PostgreSQL, Drools) forms a queryable knowledge base that bridges natural language regulations and executable logic, enabling automated reasoning and direct interoperability with BIM models for compliance verification.

Keywords –

Domain Specific Language ; Large Language Model; Building Code Digitization; Knowledge Graph; Automated Compliance Checking.

1 Introduction

Building Information Modeling technology supports structured representation of building information across the full lifecycle through the Industry Foundation Classes (IFC) data format, transforming design deliverables into semantically rich, computer-readable data objects and establishing a foundation for interoperability. However, a core bottleneck in Automated Compliance Checking remains the digitization of building codes, which are predominantly available as unstructured natural language texts. Current reliance on manual review is not only inefficient and and prone to errors but also leads to inconsistent enforcement, ultimately affecting building safety and approval efficiency.

Research on ACC spans half a century, aiming to replace manual checking with automated solutions [1, 2]. BIM and IFC provide structured information models for ACC [3]. Recent studies have explored ontologies, semantic web technologies, and knowledge graphs. Anastasiya et al. [4] formalized compliance requirements via ontologies, though manual construction remains inefficient [5]. Peng et al. [6] used NLP to transform code clauses into formats readable by machines, and so did Lu Jia et al. [7] combined semantic web with IFC for improved checking.

Artificial intelligence technologies, including deep learning, NLP, and large language models (LLMs), are increasingly applied to rule extraction and semantic understanding. Li et al. [8] developed ontologies and deep learning models for building standard labeling. Wang et al. [9] proposed an NLP-based ACC system for spatial relationship review. Domain-Specific Languages (DSLs) offer tailored syntax and semantics for domain specific tasks [10], enabling domain experts to formalize problems without deep programming knowledge [11]. DSLs have been applied in parametric design [12], generative AI integration [13, 14], and construction robotics [15, 16].

Despite these advances, ACC faces key challenges. First, knowledge formalization remains labor-intensive and lacks scalability, struggling with implicit logic and regional variations. Second, a semantic gap persists in translating ambiguous natural language provisions into computable rules. Third, LLMs, while promising, are prone to hallucinations, affecting reliability. Finally, DSL application in ACC is still nascent and warrants further exploration.

To address these challenges, this paper proposes a framework based on BIM with multiple stages that integrates DSL, hybrid knowledge graphs, and LLMs to automate the formalization of design rules from building codes. The framework converts regulatory

texts into structured, executable knowledge through a dedicated DSL, reducing semantic ambiguity and interpretability risks. A hybrid storage architecture combining graph, relational, and rule based databases enables efficient management of numerical, logical, topological, and vague clauses, establishing a verifiable foundation for automated compliance checking integrated with BIM.

2 Parsing and Transformation Technology

While standard Large Language Models can perform general semantic parsing, their inherent stochasticity often leads to hallucinations and inconsistent extraction of complex building codes. The methodological innovation of this section lies in a twofold advancement: a multidimensional guided parsing strategy and a four tier DSL schema. First, in the parsing phase, we propose a “deconstruct then extract” strategy that uses prompt engineering to force the LLM to perform semantic role labeling identifying subject, constraint, and condition thereby minimizing linguistic ambiguity before data extraction. Second, in the transformation phase, we introduce an original four tier DSL schema that acts as a semantic safety buffer. By enforcing this constrained intermediary grammar, the framework decouples nondeterministic AI outputs from deterministic IFC data structures. This dual layered approach ensures that unstructured clauses are systematically transformed into a verifiable, machine readable format, providing logic stability and traceability that existing end to end LLM methods lack (Figure 1).

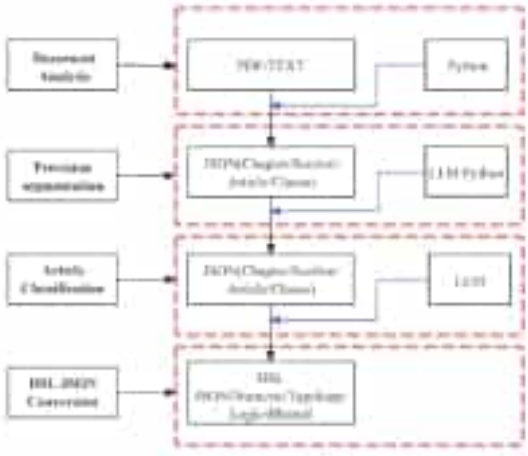


Figure 1: Technical Pathway Diagram

2.1 Document Preprocessing

To support the integration of design rules with BIM, this paper designs a parsing workflow that transforms semi-structured code documents into machine readable hierarchical data objects aligned with BIM entities. The process involves three core steps. First, raw documents undergo text cleaning and standardization to remove formatting artifacts. Second, a parsing algorithm using regular expressions recognizes the hierarchical structure (e.g., chapter, section, article, clause) and reorganizes text into a nested tree. Concurrently, metadata such as code name and effective date are extracted. Finally, structured content and metadata are encapsulated into standardized JSON objects, providing a foundation for subsequent clause atomization and mapping to BIM component attributes.

2.2 Clause Segmentation

The clause segmentation process deconstructs complex regulatory sentences into minimal, machine executable constraint units enabling precise mapping to BIM components. The process includes three steps. First, syntactic parsing identifies logical boundaries to split long sentences. Second, LLMs perform semantic completion to resolve coreferences, ensuring each atomic clause contains complete “object, attribute, constraint” semantics relevant to BIM properties. Finally, a keyword algorithm detects modality terms (e.g., “shall”, “should”) and assigns mandatory or optional labels, producing a standardized set of candidate rules ready for formalization and execution integrated with BIM (Figure 2).



Figure 2: Detailed Processing Procedure for Article Segmentation

2.3 Multidimensional data classification

To implement differentiated digitalization strategies for different types of code clauses, this paper establishes a multi-dimensional data classification mechanism after clause atomization. Atomic clauses are precisely classified into four categories(Table 1):

1. Numerical Constraint Clauses: Identified by comparative operators, extracting specific quantitative indicators such as area and height.

2. Logical Constraint Clauses: Based on conditional sentence patterns, locking the dependency relationship between preconditions and subsequent requirements in building scenarios.
3. Topological Constraint Clauses: Focusing on spatial predicates such as "adjacent to," "connected to," and "contained within," defining geometric positional relationships between components.
4. Vague Constraint Clauses: Targeting qualitative descriptions like "should," "reasonable layout," and "aesthetic," identifying guiding principles that are difficult to quantify directly.

Table 1 Clause Classification Criteria

Type	Keywords	Process	Mandatory
Numerical	≥, ≤, greater than, not less than	DSL	True
Logical	When... shall; If... then...	DSL	False
Topological	connected to, near	DSL	False
Blurred	Reasonable,s hould	DSL	False

2.4 DSL Semantic Transformation Based on LLM

To bridge the semantic gap between natural language regulations and executable logic within the BIM environment, we designed a Domain-Specific Language Schema adapted to the characteristics of building codes and aligned with common BIM data structures. This Schema includes fields such as target entity, check attribute, operator, and priority. Leveraging the semantic understanding capabilities of LLMs, we designed specific prompts to guide the model in transforming atomic natural language clauses into structured DSL JSON format. This process includes a "manual verification" step to ensure the accuracy and BIM integrability of the generated formal rules.

2.4.1 DSL Transformation for Logical Clauses

Logical clauses, characterized by conditional patterns such as "When… shall…" or "If… then…", define mandatory or recommended requirements under specific building scenarios. The core formalization challenge lies in accurately separating conditions from conclusions and establishing their logical mapping within a BIM aware framework.

To support this, the DSL Schema includes a structured conditions array, where each condition object describes a premise using fields like parameter (often mapped to a BIM property), operator, and value. The conclusion is defined by target entity (linked to a BIM component type), check attribute, and operator. For instance, the clause "When a bedroom is located underground, a ventilation shaft shall be provided" is transformed into a rule where the condition checks location == "underground" and the conclusion targets the bedroom entity for ventilation facility provision."

This transformation leverages LLMs for semantic parsing and is followed by manual verification to ensure logical accuracy and alignment with BIM semantics. The output provides unambiguous if then rules ready for execution against BIM model data.



Figure 3: DSL Transformation for Logical Clauses

2.4.2 DSL Transformation for Numerical Clauses

The core feature of numerical clauses is the quantitative constraint on attributes (e.g., area, height, distance) of building spaces or components through specific numbers and comparison operators. To transform them into machine executable rules, this method designs a targeted DSL Schema and prompt engineering strategy.

The transformation is achieved using LLMs. By constructing prompts containing clear instructions, output format examples, and domain knowledge, the LLM is guided to accurately parse and map clauses like "bedroom area shall not be less than 12 square meters" to corresponding fields in the Schema. For instance, the LLM needs to correctly identify "not less than" as the operator ">=" and extract the pure number "12" and unit "square meters." A manual verification step is set in this stage. Verification focuses include: whether the value and unit correspond accurately; whether prerequisite conditions in conditional statements are fully captured and transformed into the conditions field; and whether the priority field for recommendatory clauses is

correctly marked as lower priority (Figure 4).



Figure 4: DSL Transformation for Numerical Clauses

2.4.3 DSL Transformation for Topological Clauses

Topological clauses regulate the relative position and connection relationships between building spaces or components, such as "adjacent to," "connected to," "contained within." These clauses are crucial for ensuring functional flow and spatial rationality but are highly challenging to formalize due to complex spatial semantics and geometric relationships. To achieve their transformation into machine executable rules, this paper specially designs an extended DSL Schema.

Addressing the characteristics of topological constraints, fields such as topology type, source entity, spatial relation operator, and target entity are added to the basic DSL Schema. A standardized set of topological operators is predefined in the Schema, including "adjacent to," "contains," "connected to," and "separate from." Secondly, the semantic reasoning capability of LLMs is utilized to accurately extract spatial features from clauses. Through prompt engineering, the model is guided to identify keywords like "immediately adjacent to," "abutting," "accessed via..." and map them to normalized DSL operators.

To ensure the geometric checkability of generated rules, strict manual verification is required after transformation. Verification focuses include: whether the relation predicate accurately reflects the original clause intent(Figure 5).



Figure 5: DSL Transformation for Topological Clauses

3 Database Construction

While existing ACC systems typically store rules in a single database format, this study proposes a hybrid storage architecture that matches storage paradigms to clause types a design choice that directly addresses the heterogeneous nature of building codes. The key innovations are: (1) a “ metadata rule body ” separation pattern that decouples code version management from rule logic; (2) clause type specific storage strategies relational for numerical rules requiring efficient quantitative queries, rule engine for logical rules needing condition triggered reasoning, and graph database for topological rules requiring spatial relationship traversal; and (3) a unified knowledge graph that captures cross clause dependencies , enabling traceability and impact analysis beyond simple compliance checking.The database construction in this paper takes the structured DSL output from Chapter 2 as input, directly addressing the need for machine executable code clauses within a BIM supported environment. Based on the structured JSON data after DSL transformation, we designed a hybrid database architecture and a code knowledge graph to organize and manage rule knowledge in a manner aligned with BIM data structures and query requirements (Figure 6).



Figure 6: Database construction process

3.1 Database Structure

To achieve precise traceability and version management of code clauses, the database table structure adopts a "metadata rule body" separation design pattern.

1. Metadata Layer: Establish a specification index table to store global attributes of codes, including code name, version number, applicable region, and effective time. This design supports the system in quickly filtering applicable code sets based on project attributes.
2. Rule Body Layer: The core compiled rules table stores atomized rules. Field designs include: clause id, original text, target entity, dsl content, and, most importantly, executable code.

3.1.1 Database Construction for Numerical Clauses

For the numerical DSL JSON, this study employs Python scripts to import clauses into a PostgreSQL database. A relational hybrid storage architecture is implemented, organizing data into two BIM integrated dimensions. The structured dimension maps check targets, operators, and limit values to dedicated SQL columns, directly corresponding to BIM component attributes for efficient evaluation. The semistructured

dimension uses PostgreSQL ’ s JSONB type to store applicability conditions, which are often linked to BIM project parameters. A GIN index is applied to the JSONB field to enable fast, contextaware rule retrieval across extensive clause sets, ensuring performant interoperability with BIM model queries.

3.1.2 Database Construction for Logical Clauses

For logical clauses, the system converts them into Drools rule language for storage, enabling checks to be triggered only when relevant BIM objects satisfy specific preconditions, thereby optimizing computation within the BIM context. A Python based translator parses the condition triggers and check logic from the DSL JSON, mapping them to Drools' MVEL language. At the data level, a dynamic Design Entity fact object is used, storing variable component attributes in an attributes Map to flexibly accommodate diverse BIM component schemas.

The generated drl files employ a “violation pattern matching” strategy, transforming code requirements into negative capture patterns. When a violating BIM component is matched, the system instantiates a Validation Result object, automatically injecting traceable metadata. This enables millisecond-level logical reasoning with full compliance traceability. Finally, the rule files are deployed to Business Central and exposed as a remote service for BIM integrated compliance checking (Figure 7).



Figure 7: Related functional files and deployment on remote servers

3.1.3 Database Construction for Topological Clauses

For complex topological rules involving spatial adjacency, vertical stacking, and circulation connectivity, this study utilizes Neo4j to construct a code meta-knowledge graph.

Storage Architecture and Meta graph Construction: Based on the DSL JSON structure, the graph adopts a “Entity Type, Constraint Edge, Entity Type” triple pattern for storage.

1. Semantic Node Abstraction: Defines code objects like "bedroom" and "exhaust duct" as Entity Type nodes, establishing mapping anchors to the IFC entity classification standard via the code field.
2. Graphical Representation of Constraint Logic: Stores clause logic as relationship edges with properties. Core properties include relation type and constraint level. This design makes each edge an independent, traversable logical unit.

3.1.4 Database for Vague Clauses

For the numerous subjective clauses in codes that cannot be quantitatively calculated, such as "reasonable layout" and "harmonious with the environment," this study constructs a lightweight knowledge base based on local CSV storage and semantic retrieval.

3.2 Knowledge Graph Database Construction

To manage complex relationships between clauses (Figure 8), such as parameter dependencies and potential conflicts, we constructed a Neo4j knowledge graph encompassing all clauses.

Node Design defines four core node types to structurally store code knowledge: Specification, Clause, Entity,Parameter.

Relationship Design defines directed connections between nodes to express logic and dependencies.



Figure 8: Dependency Relationships Between Clauses

4 Case Study

To validate the effectiveness of the proposed method, we selected the Residential Design Code GB 50096-2011 as the experimental subject, performing a full-process digitalization of its clauses.

4.1 Data Processing Results

To ensure reproducibility, the rule extraction and formalization process was standardized using the Qwen-Max model with fixed generation parameters:

temperature = 0.1, top_p = 0.95, and max_tokens = 2048. Consistent prompt templates were used for Numerical, Logical, Topological, and Fuzzy clauses. This systematic approach enables replication following the DSL specifications and prompt logic detailed in this study. After preprocessing, 338 atomic rule clauses were extracted (Table 2).

Table 2: Clause Segmentation Results

Type	Quantity	Percentage
Numerical	123	36.4%
Logical	159	47.0%
Topological	45	13.3%
Blurred	11	3.3%
Total	338	100%

4.2 DSL JSON Transformation and Code Base Construction

After completing the document preprocessing and clause atomization of *GB 50096-2011*, this study obtained 338 semantically independent and structurally clear atomic clauses. To achieve their transformation from natural language to machine executable rules, this stage focused on two core tasks:

Phase 1: Automated DSL JSON Generation. The core of this study lies in the "LLM+DSL" semantic transformation paradigm. First, we designed refined DSL Schemas for the three types of "clear clauses" (numerical, logical, topological), specifying the data type and semantic boundary of each field (as described in Section 2.4). Subsequently, for each atomic clause, we constructed highly customized prompts containing the following key elements:

1. Role Definition: Explicitly instructing the LLM to act as a "building code digitalization expert."
2. Task Instruction: Clearly requiring the LLM to transform the input natural language clause into a JSON object strictly adhering to the specified Schema.
3. Output Format Example: Providing correct transformation examples for the same type of clause to guide the model in learning the mapping pattern.
4. Domain Knowledge Constraints: Embedding lists of domain-specific entities and operators to constrain the model's output space and reduce hallucination risk.

By calling the LLM API, the system automatically transformed the 338 atomic clauses into structured DSL JSON objects one by one. To ensure transformation quality, all automatically generated JSON objects underwent a manual verification step.

Phase 2: Hybrid Database Construction and Storage. This study abandons the generic storage approach of a single database, adopting instead a "allocate on demand, each performing its own function" hybrid storage architecture to maximize the processing efficiency for different rule types.

1. Numerical Clauses: Verified numerical DSL JSON was batch-parsed by Python scripts and imported into the PostgreSQL relational database. The scripts mapped key fields like check attribute, operator, and value to standard SQL columns, while storing complex applicability conditions in the applicability conditions field in JSONB format. A GIN index was established to support efficient contextual retrieval.
2. Logical Clauses: Logical DSL JSON was fed into a specially developed rule translator. This translator accurately compiled the condition and conclusion logic from the JSON into drl files recognizable by the Drools rule engine. These rule files were then deployed to the Drools Business Central platform and encapsulated as remote service interfaces for on demand calling by the compliance checking system, enabling complex condition triggered reasoning.
3. Topological Clauses: Topological DSL JSON was used to construct the code meta knowledge graph in the Neo4j graph database. Scripts created nodes for source entity and target entity and stored information like spatial operator and constraint level as properties of relationship edges (Figure 9).



Figure 9: Relationships Among Topological Clauses in Neo4j (Partial GB 50096-2011 Examples)

4. Vague Clauses: For vague clauses like "reasonable layout" that cannot be quantified, their original text, associated chapter, and related entity information were structured and stored in a local CSV file.

4.3 Error Analysis and Accuracy Testing

During automated DSL JSON generation, the LLM occasionally produced incorrect outputs despite prompt engineering. Manual verification identified seven failure patterns, accounting for 22 erroneous outputs out of 338 atomic clauses. These included: misinterpretation of comparative operators (4), where “not less than” became “>” instead of “>=”; incomplete condition capture in logical clauses (3), omitting premises when conditions appeared mid sentence; entity role reversal in topological clauses (3), swapping source and target; unit inconsistency (2), extracting values without units or confusing millimeters with meters; target entity mismatch (3), assigning incorrect BIM entity types; missing optional/recommendation flag (2), failing to map “should” to the priority field; and fuzzy clause not recognized (5), misclassifying qualitative clauses as logical or numerical. We built a test set containing 100 BIM model fragments as test cases to ensure that the provisions in the database can be retrieved.

For numerical clauses, we created multiple test cases, stored them in a temporary SQL table, and used Python scripts to call the SQL database for compliance checking. The checking process first matched the check objects in the cases with the clauses, performing compliance checks only if the preconditions of the clauses (e.g., building completion year, location, object type) were satisfied (Figure 10).



Figure 10: Partial Test Cases (Numerical)

For logical clauses, we similarly stored cases in a temporary SQL table, then used Python scripts to call the deployed Drools remote server for compliance checking. We also inserted test cases for clauses into Business Central to verify whether our drl code rules were invoked (Figure 11).



Figure 11: Checking if drl Files are Invoked in Business Central

For topological clauses, we stored cases in a temporary SQL table and used Python scripts to call the deployed Neo4j for compliance checking.

4.4 Application Value of the Knowledge Graph

The constructed knowledge graph, designed to support BIM based analysis, contains over 300 nodes and 500 relationships. Graph analysis reveals strong couplings between design entities such as “Bedroom” and performance parameters including “Daylight Factor” and “Ventilation Opening Area.” In a typical BIM aided design scenario, if a designer modifies a window size, the system can utilize the graph’s path-search algorithms to not only evaluate compliance with daylighting standards but also proactively surface related clauses involving “façade aesthetics” or “cost indicators.” This demonstrates the potential of the graph to enable multidimensional, rule aware correlation analysis within the BIM workflow.

5 Scalability and Limitations

The proposed framework demonstrates strong generalizability across building codes from different regions and domains. The DSL Schema decouples logical structure from natural language, enabling systematic migration to other regulatory documents such as the International Building Code (IBC), European construction standards, and domain specific fire safety regulations without reconstructing the core architecture. Since the four category classification (numerical, logical, topological, fuzzy) covers the vast majority of regulatory provisions encountered in typical building codes, core prompt templates require no modification when adapting to new codes. Adaptation follows a calibration workflow: prepare 20 – 30 annotated clauses as few shot examples, feed them into the LLM to learn new terminology and patterns, validate output quality, then process the full code with manual verification only for ambiguous cases. This calibration takes 2-3 hours per new code, demonstrating the framework’s automation efficiency for large scale code digitization.

Compared to existing LLM, rule based, or ontology approaches which suffer from hallucinations, manual effort, or limited scalability our hybrid framework integrates LLM semantic understanding with DSL precision. This enables verifiable rule extraction and provides a modular architecture for rapid reconfiguration across diverse building codes.

Nevertheless, the current system has limitations. It relies on manual verification to ensure semantic

accuracy, preventing full end to end automation. This manual step, while necessary to eliminate LLM hallucinations and guarantee correct DSL generation, represents a bottleneck in achieving a fully automated pipeline for rule formalization.

6 Conclusion

This paper proposes a complete technical pathway for transforming building codes into a structured digital knowledge base within a BIM oriented framework. Using GB 50096-2011 as a case study, the resulting hybrid database and knowledge graph form a reasoning capable rule system that can serve as the core engine for BIM based automated compliance checking. The system supports automated verification of numerical, logical, and topological clauses and provides auxiliary decision support via a vague clause knowledge base.

Future work will focus on enhancing BIM integration through multidimensional IFC parsing for richer semantic and geometric data, automated version comparison, DSL schema refinement, and automating manual verification in the DSL pipeline via three strategies: a retrieval augmented generation mechanism with few shot examples, a rule based post-processor for recurring errors, and a self consistency validation loop using bidirectional translation. These enhancements will reduce manual effort while maintaining accuracy, strengthening end to end automation and deployability, alongside a scalable compliance-checking application on Dify.

7 Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grant number 52408015), National Key Research and Development Program of China (Grant number 2024YFC3808901).

References

[1] R. Amor and J. Dimyadi. The promise of automated compliance checking. *Developments in the Built Environment*, 5:100039, 2021.

[2] Q. Ren, D. Zhang, M. Li, S. Chen, D. Tian, H. Li, and L. Liu. Automatic quality compliance checking in concrete dam construction: Integrating rule syntax parsing and semantic distance. *Advanced Engineering Informatics*, 60:102409, 2024.

[3] S. Macit İlal and H. M. Günaydın. Computer representation of building codes for automated compliance checking. *Automation in Construction*, 82:43–58, 2017.

[4] A. Yurchyshyna and A. Zarli. An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction. *Automation in Construction*, 18(8):1084–1098, 2009.

[5] P. Pauwels, S. Zhang, and Y.-C. Lee. Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, 73:145–165, 2017.

[6] J. Peng and X. Liu. Automated code compliance checking research based on BIM and knowledge graph. *Scientific Reports*, 13(1), 2023.

[7] L. Jia, M. Chen, C. Chen, and Y. Jin. A semantic web and IFC-based framework for automated BIM compliance checking. *Buildings*, 15(15):2633, 2025.

[8] S. Li, J. Wang, and Z. Xu. Automated compliance checking for BIM models based on Chinese-NLP and knowledge graph: an integrative conceptual framework. *Engineering, Construction and Architectural Management*, 32(6):3832–3856, 2024.

[9] Y. Wang, Y. Liu, H. Cai, J. Wang, and X. Zhou. An automated fire code compliance checking jointly using building information models and natural language processing. *Fire*, 6(9):358, 2023.

[10] F. Ciccozzi, L. Addazi, S. A. Asadollah, B. Lisper, A. N. Masud, and S. Mubeen. A comprehensive exploration of languages for parallel computing. *ACM Computing Surveys*, 55(2):1–39, 2022.

[11] E. D. Sozzo, D. Conficconi, A. Zeni, M. Salaris, D. Sciuto, and M. D. Santambrogio. Pushing the level of abstraction of digital system design: A survey on how to program FPGAs. *ACM Computing Surveys*, 55(5):1–48, 2022.

[12] S. Wu, W.-Z. Lu, C. Qin, B. Wen, D. Wu, and Y. Xiang. A goal-oriented planning approach for two-dimensional cutting components in architectural design: Coupling BIM and Parametric design (PD). *Journal of Building Engineering*, 76:107156, 2023.

[13] D. Jun and W. Zi Jia. A novel architectural design paradigm: Collaborative construction with design thinking semantic networks and large language models. *Preprints*, 2025051238, 2025.

[14] J. Ko, J. Ajibefun, and W. Yan. Experiments on generative AI-powered parametric modeling and BIM for architectural design. *arXiv preprint arXiv:2308.00227*, 2023.

[15] S. Leder and A. Menges. Architectural design in collective robotic construction. *Automation in Construction*, 156:105082, 2023.

[16] Y. Liu, M. Habibnezhad, and H. Jebelli. Brainwave-driven human-robot collaboration in construction. *Automation in Construction*, 124:103556, 2021.

In-Context Trajectory Learning for Construction Scheduling Using Large Language Models

Shanika Manamperi[1], Wei Peng[2] and Guomin Zhang[3]

[1,2,3]School of Engineering, RMIT University, Australia
s4061684@student.rmit.edu.au, wei.peng3@rmit.edu.au and kevin.zhang@rmit.edu.au

Abstract

Adoption of Artificial Intelligence (AI) is growing across the construction industry, but its use remains uneven. While AI is being increasingly adopted for information management workflows, its application to construction scheduling remains limited. It is a critical gap because construction scheduling is inherently dynamic and must respond to disruptions such as workforce constraints, procurement delays, coordination issues, and changing site conditions. Although traditional scheduling approaches have improved over time, they remain limited in supporting adaptive decision-making under uncertainty. To address this gap, this study proposes a trajectory-based learning approach for construction scheduling. A technical simulation environment was developed by modelling construction actions as trajectories. The proposed framework is demonstrated through a case study of ductwork installation in a six-story building. A dataset of 50 trajectories with 577 steps of thought-action-observation was constructed to capture diverse execution scenarios. An In-Context Learning (ICL) approach leveraging large language models (LLMs) was adopted to learn from training trajectories and predict subsequent construction actions. Zero-shot and few-shot prompting strategies were evaluated comparatively. The initial next-action prediction task was further extended with an additional experiment on predicting short sequences of future actions to better reflect real-world construction operations. Results show that Llama 3B achieved 76% accuracy in next-action prediction under the 5-shot setting, while GPT-5.4 demonstrated stronger reasoning capability in the extended experiment. This study highlights the potential of trajectory-based learning as an on-site decision-support mechanism and provides a preliminary foundation for future intelligent construction scheduling agents.

Keywords –

Construction; In-context learning; Large language models; Scheduling; Trajectory learning

1 Introduction

Despite technological advancements and applications in the industry, accurate construction scheduling remains a critical challenge. This is basically due to the inherent complexity and dynamic nature of construction work. In addition, other influencing factors such as labour, material and equipment availability, weather, change orders and economic fluctuations also contribute to unrealistic scheduling. Traditional scheduling approaches range from the Critical Path Method (CPM) to construction scheduling software like MS Project and Primavera which remain deterministic with poor reasoning capability [1]. Research has also focused on applying Artificial Intelligence (AI) and Machine Learning (ML) to construction scheduling in terms of predicting completion percentages and evaluating project performance [2]. Moving forward, current research has applied Large Language Models (LLMs) to construction scheduling in terms of capturing delay causes and automatic schedule generation [3]. Additionally, integration of digital technologies like Digital Twins (DTs) with LLMs can be seen in enhancing robots for dynamic task allocation in construction [4]. Some studies have developed LLM-based frameworks for construction with the integration of robotics by leveraging Industry Foundation Classes (IFC) [5]. These frameworks have been successful in guiding robots to perform tasks like brick-laying and 3D printing [5]. Despite these advances, existing approaches remain limited in their ability to handle construction uncertainties in an adaptive manner. Most AI and LLM-based scheduling approaches depend on predefined methods offering deterministic schedule generation instead of learning through sequential behaviours. This highlights a research gap in developing scheduling approaches that can reason under uncertainties and better address current industry needs. To address this gap, this study introduces an agentic approach for construction scheduling that learns sequencing behaviour from trajectories under uncertainty.

LLMs are transformative technologies with applications not only in construction but also in other sectors like finance and manufacturing [6]. In-context

learning (ICL) is a mechanism through which LLMs perform a task by learning from instructions and examples given directly in the prompt. ICL eliminates the need for complex fine-tuning which frequently acts as a bottleneck in existing ML approaches [7]. Hence, an ICL method was applied in this study leveraging LLMs, using training trajectory data as the context for the user prompt. This study demonstrates HVAC (Heating, Ventilation, and Air Conditioning) installation as a case study where both positive (successful) and negative (unsuccessful) trajectories were provided as in-context examples. Dynamic construction interactions which include task dependencies, rework and coordination between site personnel were incorporated to infer decision patterns. Ultimately, this study examines whether ICL enables LLMs to reason over construction scheduling trajectories.

2 Methodology

2.1 Trajectory data

This study models the interaction between the agent and the environment as trajectories which provides the structural representation used for learning. Trajectories capture the complete or partial execution of a construction task over time through a sequence of decisions and outcomes. In this study schedule related operations are represented as thought-action-observation (t-a-o) trajectories, where thought denotes what needs to be done, action denotes the managerial or operational decision taken, and observation denotes the resulting outcome (Figure 1) [8].

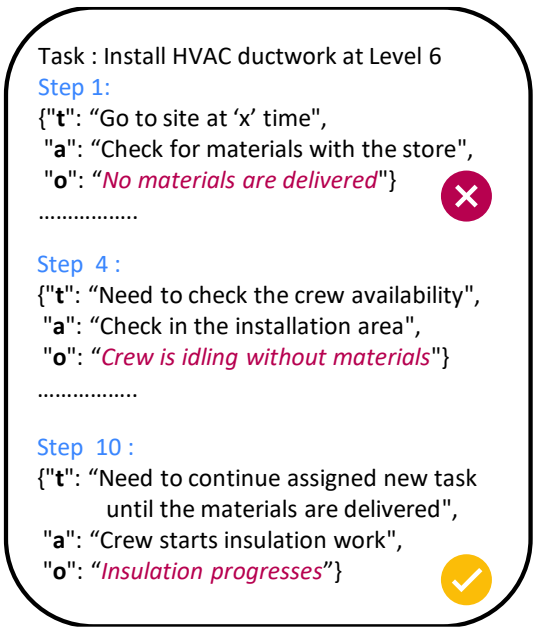


Figure 1. Trajectory representation

Dynamic scheduling factors such as crew availability, safety hazards, material delivery delays, and environmental conditions are embedded explicitly at the step level within the observation component. This enables the trajectories to capture real-world construction uncertainties and intermediate decision consequences. These trajectories are encoded into the task environment in a structured order, where the state space ($\mathcal{S}$) describes the context at a given step and the action space ($\mathcal{A}$).

This study used an open-source time-motion dataset which includes a long chain of events resulting from video footage [9]. It captures the activities of MEP (Mechanical, Electrical, and Plumbing) workers in a six-storey multifamily building in Finland which showed elevated levels of wasted efforts. A total of 50 trajectories were constructed for the HVAC installation task, including 12 positive trajectories and 38 negative trajectories. Positive trajectories contain correct execution steps throughout the full sequence. Negative trajectories include either trajectories with an unsuccessful final outcome, or trajectories with a successful final outcome but incorrect intermediate steps. Negative trajectories were included since the process of human learning not only involves observing successful executions but also includes experiencing failures [10].

2.2 ICL approach

ICL was adopted to evaluate LLMs for construction action prediction in schedule decision support. Instruction-finetuned open-source LLMs, namely Llama-3.2-3B-Instruct and Qwen 2.5-3B-Instruct were used in this study. These models were selected because they represent relatively lightweight and practically deployable LLMs suitable for early-stage evaluation of on-site decision-support scenarios. These models were accessed through a Hugging Face API token [11]. Training trajectories were given as contextual examples within the user prompt enabling models to learn scheduling patterns. Data were split based on the 80/20 rule. The held-out test trajectories served as the ground truth with the original t-a-o construction sequences.

During ICL, LLMs analysed unseen test trajectories and predicted the next action under uncertainty. However, this decision is solely based on the given trajectories of different scenarios. Zero-shot and few-shot prompting approaches were incorporated for comparative evaluation. In the zero-shot setting, the model receives no in-context trajectory examples, whereas in the few-shot setting, it learns from a small number of example trajectories provided in the prompt. Factors like the diversity and quality of examples shown in the in-context demonstrations may improve the quality of model output [7]. To assess the capability of ICL in distinguishing trajectory patterns, two conditions were studied: positive only and mixed positive-negative trajectory contexts.

Prompt was designed as using positive trajectory and (positive + negative) trajectory contexts to predict the next best action. Regarding example selection, few-shot examples were re-sampled for each run, and each model was evaluated across all test trajectories. The prompt was specified to generate only positive outcomes. For each condition, zero-shot, 1-shot, 5-shot and 10-shot prompting configurations were examined. Final reward was computed based on the Outcome Reward Model (ORM). ORM uses binary labels with "1" representing successful task completion and "0" for uncompleted tasks [12]. This study assigns a reward of "1" if the ground-truth label is "positive", whereas it assigns a value of "0" if the ground-truth label is "negative". Evaluation was conducted at the trajectory level by aggregating step-level rewards across each predicted trajectory and computing the average reward against the ground-truth trajectory. To evaluate whether the system can differentiate positive and negative trajectories, accuracy and a confusion matrix were computed, including true positives (TP) and true negatives (TN). Model outputs were evaluated over multiple runs, and accuracy was recorded. Subsequently, accuracy for each shot configuration was calculated by taking the average accuracy across all runs.

In real-world construction processes, scheduling decisions often involve not only the next action but also a short sequence of subsequent actions. To better reflect this practical need, the ICL setup for next-action prediction was extended with an additional experiment that evaluates the prediction of future action sequences with prior trajectory actions as input context. In addition

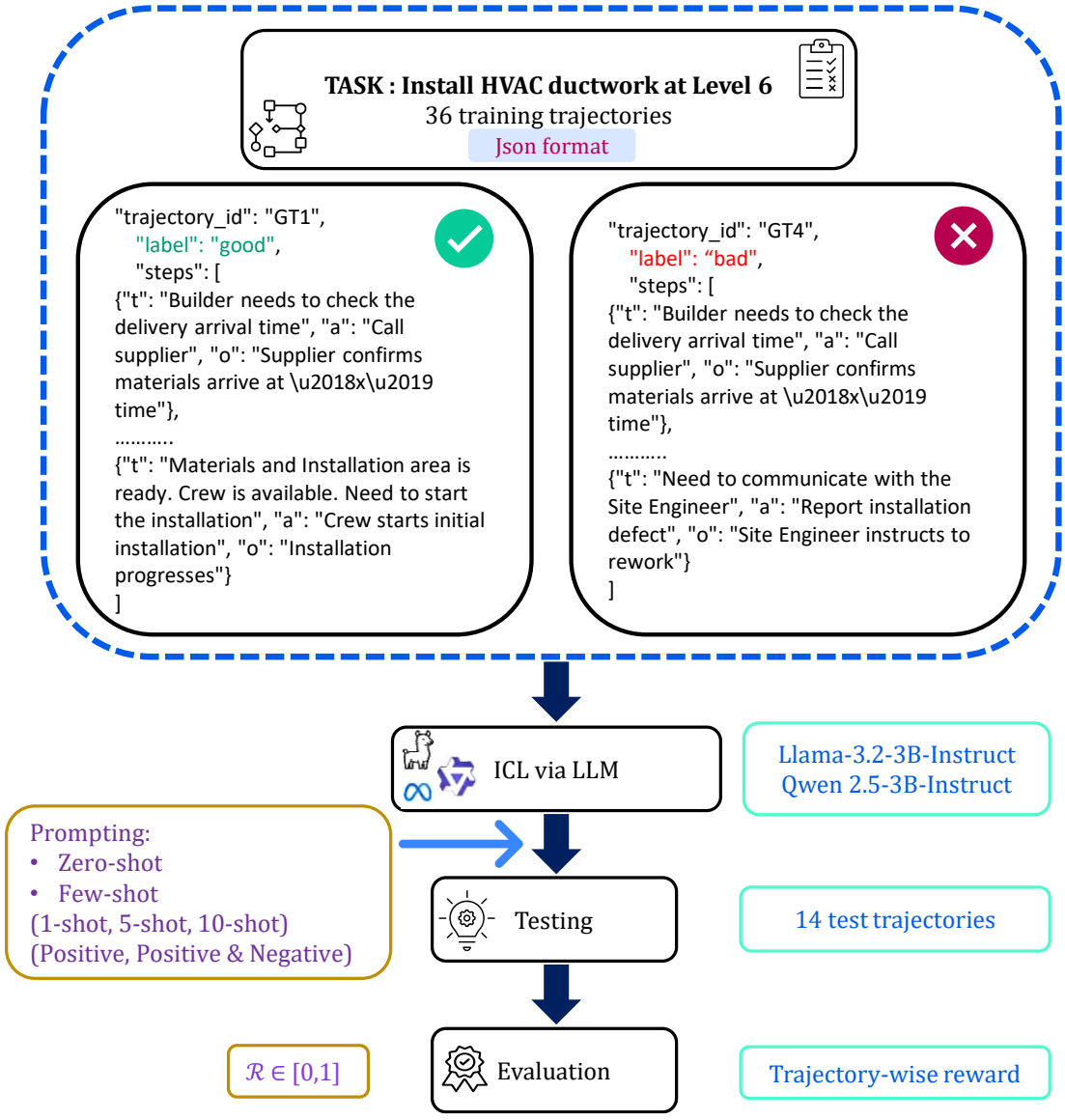


Figure 2. Methodology for preliminary ICL approach

to Llama and Qwen, this experiment included Mistral-7B-Instruct-v0.2, GPT-4o-mini, and GPT-5.4 to compare performance across both open-source and API-based models. Predicted trajectories were assessed manually.

3 Results and discussion

When the same unseen trajectory was tested with LLMs under positive and negative conditions the following outputs were obtained as indicated in Table 1.

Table 1. Accuracy levels of positive-negative trajectory context

Prompt	Qwen 2.5-3B-Instruct	Llama 3.2-3B-Instruct
Zero-shot	30%	40%
1-shot	76%	66%
5-shot	60%	76%
10-shot	58%	70%

Table 1 shows that neither model performed well under the zero-shot setting. Qwen achieved its best performance under the 1-shot configuration, with an accuracy of 76%, compared with 66% for Llama. In contrast, Llama performed better at the 5-shot and 10-shot settings, suggesting that it benefited more from additional in-context examples. However, both models showed lower performance at 10-shot than at 5-shot, indicating that adding more examples did not necessarily improve prediction accuracy. This suggests that excessive in-context examples may introduce less relevant or distracting context, reducing the models' ability to use the prompt effectively. Overall, Qwen performed best in the 1-shot setting, whereas Llama achieved its strongest performance in the 5-shot setting.

One possible reason for the variation in few-shot performance is output truncation. As the number of few-shot examples increases, the prompt becomes more complex and consumes more tokens during inference. Since each trajectory contains around 10 steps on average, larger few-shot settings increase both prompt length and response complexity. To manage this, generation was limited to 500 output tokens. As a result, the study mainly evaluated action prediction accuracy rather than the quality of the generated explanation. Example explanations are presented only for illustration, and their content may vary depending on the number and similarity of the in-context examples. An example from the 5-shot Llama setting is shown in Figure 3.

In addition to comparing different LLM architectures, an ablation study was conducted to examine the effect of model scale within the same model family by comparing Qwen 2.5-7B-Instruct and Qwen 2.5-3B-Instruct. As shown in Figure 4, the 7B model achieved higher accuracy than the 3B model in most settings, except for the 1-shot configuration. Zero-shot performance was nearly twice as high for the 7B model, suggesting stronger instruction-following capability. The 7B model also performed better under the 5-shot and 10-shot settings, indicating improved use of multiple in-context examples. These results suggest that increasing model scale may improve robustness to noisy context and strengthen the use of trajectory examples.

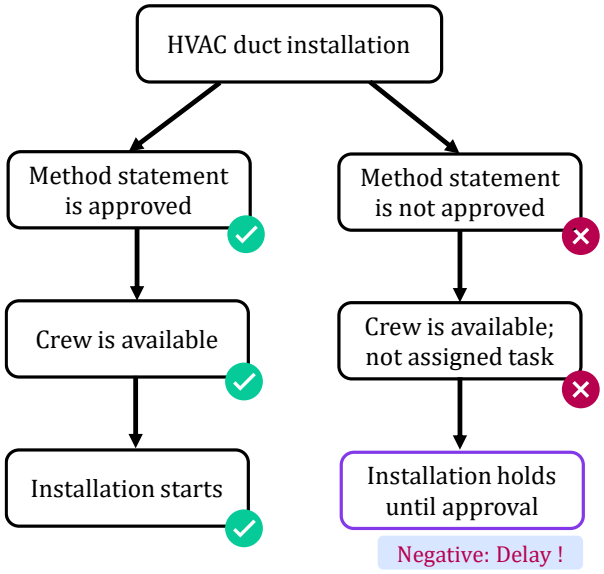


Figure 3. Next action prediction (5 shot Llama setting)

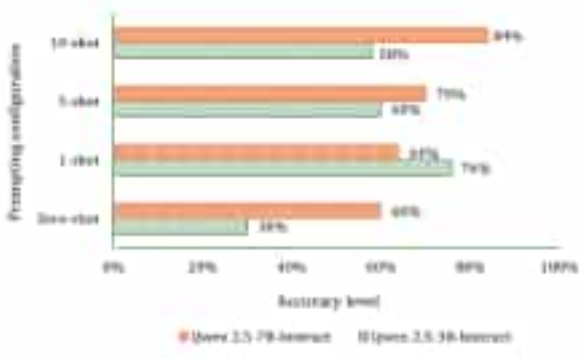


Figure 4. Accuracy levels of the ablation study

Figure 5 presents the cosine similarity values obtained in the extended experiment on predicting future action sequences across different LLMs and prompting configurations. Qwen and Llama showed patterns generally consistent with the next-action prediction results from the previous experiment, while Mistral performed at a similar level with slight variation across prompt settings. GPT-based models showed stronger performance than the open-source models, suggesting that trajectory-prediction performance may improve with model scale. Figure 6 shows an example future trajectory prediction generated by GPT-5.4.

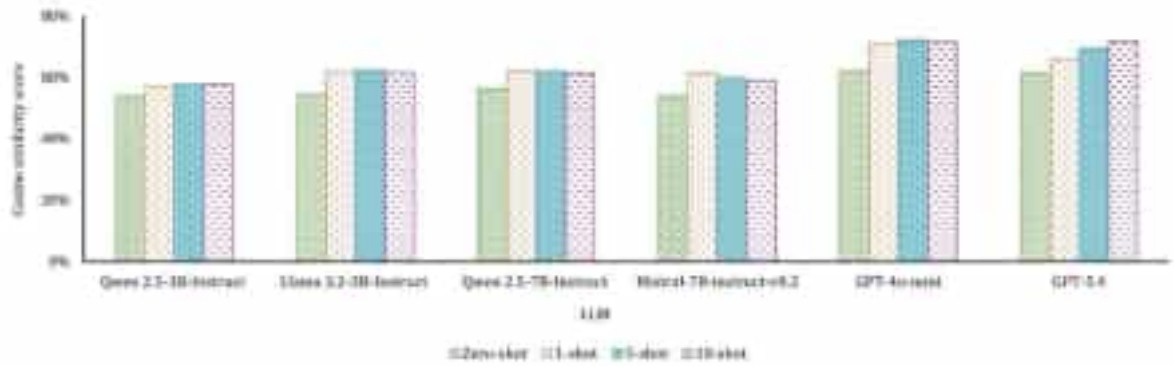


Figure 5. Cosine similarity scoring of different LLMs

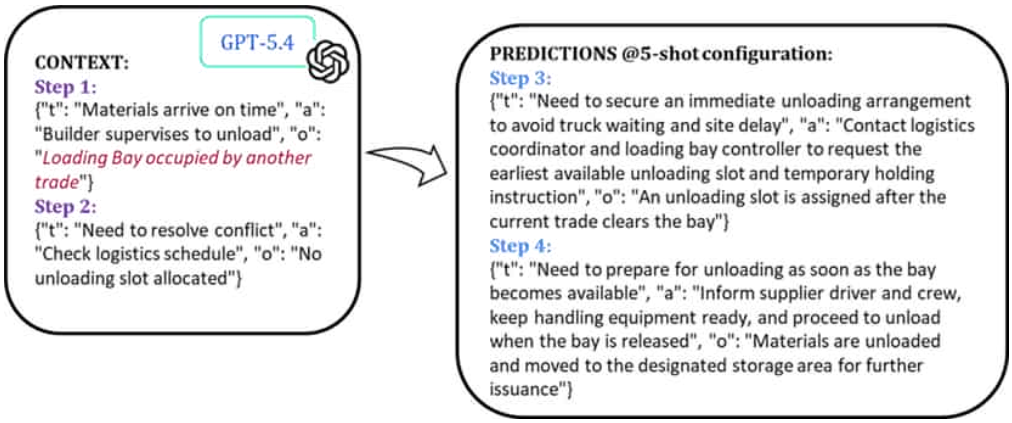


Figure 6. Prediction of trajectory using GPT-5.4

This study presents a trajectory-based ICL approach for construction scheduling. Rather than replacing conventional scheduling tools, the proposed method is designed as a decision-support mechanism that helps site teams interpret evolving situations and identify feasible follow-up actions. This work contributes to contrastive learning from positive and negative trajectory patterns, including intermediate-step failures that are often overlooked in conventional deterministic schedules.

One of the key limitations is the limited number of trajectories which limits the diversity of decision patterns available for unseen scenarios. Hence, it is recommended to train the model with a considerably bigger dataset, covering a wide range of cascading scenarios under uncertainty.

As an initial step towards developing a scheduling agent, this study provides a basis for future development. The next step is to advance this work towards an LLM-based scheduling agent by incorporating stepwise rewards and a mixed post-training method that includes supervised instruction fine-tuning and reinforcement learning. It would also be valuable to include larger scale LLM baseline models in future evaluations.

4 Conclusion

Timely project completion remains a persistent challenge in the construction industry adhering to its inherent complex nature and uncertainty. The proposed approach is designed to support adaptive scheduling decisions under disruptions by predicting future action sequences from prior trajectory context.

This study leads to the following key findings. First, the performance of ICL is influenced by both the number and quality of the in-context examples. Despite the small dataset, ICL appears to offer a practical alternative in settings where construction scheduling data are scarce. Additionally, Llama showed stronger performance in next-action prediction at the preliminary stage, while commercial LLMs such as GPT-5.4 performed better in the extended future-trajectory prediction experiment. This suggests that model architecture and scale may influence trajectory-learning performance.

This paper presents an exploratory study towards such a system, in which the agent learns construction sequences from both positive and negative trajectories and is evaluated on its ability to predict subsequent actions and short future trajectories. In practical terms, the contribution of the approach lies in its potential to support schedule adjustment under disruptions by recommending feasible next actions that reduce delay.

References

[1]. C. I. Lagos, R. F. Herrera, A. F. Mac Cawley, and L. F. Alarcón, "Predicting construction schedule performance with last planner system and machine learning," *Automation in Construction*, vol. 167, p. 105716, 2024/11/01, doi: https://doi.org/10.1016/j.autcon.2024.105716.

[2]. M.-Y. Cheng, Y.-H. Chang, and D. Korir, "Novel Approach to Estimating Schedule to Completion in Construction Projects Using Sequence and Non sequence Learning," *Journal of Construction Engineering and Management*, vol. 145, no. 11, p. 04019072, 2019, doi:https://doi.org/10.1061/(ASCE)CO.1943-7862.0001697.

[3]. A. Pal, D. Murguia, and C. Middleton, Automatic Inference of Construction Delays through Analysis of Weekly Progress Reports Using LLMs. 2024.

[4]. M. Deng, B. Fu, L. Li, and X. Wang, Integrating LLMs and Digital Twins for Adaptive Multi-Robot Task Allocation in Construction. 2025. doi: https://doi.org/10.48550/arXiv.2506.18178

[5]. S. Du, Y. Gao, W. Tong, and Y. Weng, "Towards Agent: LLM-based Framework for Robotic Construction by Leveraging IFC," in 4*2nd International Symposium on Automation and Robotics in Construction (ISARC 2025)*, pp. 358-368, Canada, 2025, doi: https://doi.org/10.22260/ISARC2025/0048.

[6]. M. Raiaan et al., "A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges," 2023. [Online]. Available: 10.36227/techrxiv.24171183. Preprints

[7]. Y. Emami, H. Zhou, S. Nabavirazavi, and L. Almeida, "LLM-Enabled In-Context Learning for Data Collection Scheduling in UAV-assisted Sensor Networks," 2025. [Online]. Available: https://doi.org/10.1109/JIOT.2025.3615410. Preprints

[8]. W. Xiong et al., Watch Every Step! LLM Agent Learning via Iterative Step-level Process Refinement. 2024, pp. 1556-1572. doi: https://arxiv.org/abs/2406.11176

[9]. C. Görsch, O. Seppänen, A. Peltokorpi, and R. Lavikka, "Unlocking Productivity: Revealing Waste and Hidden Disturbances Impacting MEP Workers," *Journal of Construction Engineering and Management*, vol. 150, no. 9, p. 04024108, 2024, https://doi.org/10.1061/JCEMD4.COENG-14204.

[10]. Y. Song, D. Yin, X. Yue, J. Huang, S. Li, and B. Y. Lin, "Trial and Error: Exploration-Based Trajectory Optimization for LLM Agents," in Annual Meeting of the Association for Computational Linguistics, 2024. doi: https://arxiv.org/abs/2403.02502

[11]. H. Face. "Hugging Face Hub Documentation." On-line. https://huggingface.co/docs, Accessed: 25/12/2025

[12]. P. Wang et al., "Math-Shepherd: Verify and Reinforce LLMs Step-by-step without Human Annotations," pp. 9426-9439, Bangkok, Thailand, August 2024: Association for Computational Linguistics, in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), doi: https://doi.org/10.18653/v1/2024.acl-long.510

A domain ontology for representing core modular building components aligned with Linked Building Data and Design for X principles

Salik Ahmed Khan[1], Yang Zou[2], Minh Kieu[1]

[1]Department of Civil and Environmental Engineering, University of Auckland, New Zealand
[2]School of Architecture, Building and Civil Engineering, Loughborough University, United Kingdom
skha864@aucklanduni.ac.nz, y.zou2@lboro.ac.uk, minh.kieu@auckland.ac.nz

Abstract –

Design for manufacturing and assembly (DfMA), extended through more generic Design for X (DfX) is an important component for realising the benefits of modular and industrialised construction. This requires balancing heterogeneous objectives derived from diverse, often fragmented data sets. In the absence of a unified data model, designers often resort to bespoke solutions on a per-case basis, missing the opportunity for optimally mass-customised solutions. Semantic web technologies have proven the capability to unify diverse data sources through the formalisation and re-use of knowledge across lifecycles and projects. But its application within modular construction has remained limited. This paper proposes the Modular Building Components Ontology (MBCO) to formally represent components' hierarchy in line with Linked Building Data (LBD) and DfX principles. The Ontology 101 methodological framework is adopted. It reuses and aligns high-level vocabularies like Building Topology Ontology (BOT) to enhance interoperability. It provides a modular core structure with optional extensions to further knowledge (like environmental indicators, constructability, mass customisation, etc.) without bloating the schema. Reasoning-based checks and SHACL shapes ensure consistency and data quality. We illustrate feasibility through a case study that populates a residential modular unit and show how SPARQL queries and lightweight rules can extract objective metrics like component commonality and production-assembly constraints directly from the graph. The results show that such an ontology can facilitate automated simultaneous multi-objective optimisation using diverse data sources, thereby operationalising DfX in modular construction.

Keywords –

Modular Buildings; Ontology; Linked Building Data; Design for X

1 Introduction

Modular and industrialised building construction has promised a huge potential with the recent advancements in the architecture, engineering, and construction (AEC) industry. The benefits realisation, however, has remained limited due to several challenges, design being one of the foremost. This is because the industrialised construction process largely relies upon modular design properties to achieve benefits through mass customisation. The disjointed process of offsite production and onsite assembly makes the consideration of downstream requirements within design more important, giving rise to the Design for Manufacturing and Assembly (DfMA) philosophy. More recently, this has been replaced by a notion of Design for X (DfX) [1], where the considerations are expanded to sustainability, modular suitability, etc. Multi-objective design optimisation, akin to the manufacturing domain, has thus gained momentum to address this gap [2]. The problem with that, however, is the fragmented nature of information linked to multiple objectives. For example, components' environmental data, manufacturing process, module transportation requirements, and crane assembly constraints all have different standards for representing information, and real-time inclusion to support the designer is often difficult, if not absent at all [3,4]. Semantic web technologies provide a mechanism to unify such information into a consistent model through the Linked Building Data (LBD) initiative for the AEC domain [5]. Existing applications within modular construction have primarily focussed on standardising the DfX-based design review process. However, the automation of optimal design generation promises wider benefits in view of the recent AI-based methods. These methods heavily rely on the representation of design, with graph-based methods showing promising results [6]. However, no formalisation of a graph-based modular building component structure could be found which would

streamline these efforts and exploit the benefits of standardisation. A need, thus, arises to formalise a core representation of modular buildings within the LBD context, based on which the requirements for automated design generation may be uniformly defined.

This paper tries to address this gap through the development of a core ontology to represent modular building component structure, aligned with existing efforts and allowing for subsequent extensions to support a unified design optimisation workflow. Ontology development 101 methodological framework provides a comprehensive, step-by-step guide for this process, and has been utilised for the current work. The Web Ontology Language (OWL) is used for the specifications, and it works on the Resource Description Framework (RDF) data model. The proposed ontology is meant to be positioned as middle-to-upper level, formalising domain-specific concepts, while aligning with existing high-level definitions like Building Topology Ontology (BOT) [7] for building context and spatial containment, and Ontology for Property Management (OPM) [8] for component properties. Core classes like Volumetric Module, Planar Module, etc. extend BOT concepts, while allowing for further domain extensions to suit multiple DfX criteria. Similarly, a set of minimal attributes are associated with the classes, which may be easily extended to capture details like geometry, materials, manufacturer details, etc. A modular architecture is employed to define a core structure, which is extendable through Linked Data hooks to external knowledge bases, without bloating the schema. Manufacturability and Mass Customisation based DfX criteria definitions are further constructed to demonstrate usability and automated reasoning across domains. Shapes Constraints Language (SHACL) has been used to define rules for consistency and manufacturability to ensure data quality and real-world applicability. The feasibility is illustrated through populating a residential modular building in line with the ontology and computing objective values of multiple criteria like component commonality, module dimensions, and weight limits from the RDF graph.

2 Background

The current work focusses on two primary theoretical foundations to develop the case for this research – (1) Design within modular buildings and the various considerations required for achieving the promised benefits of modular construction; (2) The application of semantic web and linked data technologies within modular construction to implement the proposed design principles.

2.1 Design in Modular Construction

Based on the requirements for design in modular construction, two distinct considerations can be pointed out – DfMA and Mass Customization.

DfMA involves incorporating downstream manufacturing and assembly requirements during design and originated in manufacturing. In AEC, its integration with BIM has been shown to improve process flow, although regulatory, cultural, and knowledge-related barriers remain [9]. These have also identified DfMA's potential for autonomous design and construction manufacturing, while demonstrating that existing work places greater emphasis on design review than on the a priori inclusion of design requirements, consistent with Lee et al. [1]. A broader Design for X (DfX) approach has further been proposed, where "X" may represent criteria such as modularity, sustainability, commonality, variety, production cost, transportability, etc. In this form, DfX aligns more closely with mass customisation as compared to DfMA.

Mass customisation seeks to combine the efficiencies of mass production with a desired level of variety. It is commonly enabled through product platform architectures and the customer-oriented design decoupling point model. Jensen [10] identified four modular construction platform strategies based on the decoupling point: engineered-to-order, modify-to-order, configured-to-order, and select-variant-to-order, in increasing order of standardisation. This trade-off between standardisation and customisation is central to DfX. The present work addresses its formalisation through design representation.

2.2 Linked Data for Modular Buildings

Semantic web and linked data technologies have shown promise in addressing information fragmentation in AEC, leading to standard knowledge dictionaries for representing spatial topology, product structure, geometry, and element properties [7,8,11]. For modular buildings, however, although product-based structuring has been recognised at a high level, domain-specific formalisation remains limited. Early efforts include the Product Semantic Representation Language (PSRL) and semantic web-based component catalogue representations highlighting the need for standardised components [12]. Subsequent studies proposed semantic support for DfMA-based design, manufacturing integration, and cost estimation [13]. Lee et al. [1] similarly aimed to standardise DfX review for modular buildings, but focused on process concepts without adequately representing the design artefact itself.

Other work has linked DfMA with mass customisation through client-engagement modelling [14], or proposed ontologies for constructability and mass customisation [15,16]. However, these approaches are either too component-centric for inter-domain DfX use or limited to lower-level product library descriptions

Table 1: Comparison of MBCO with existing efforts for formalised graph-based modular building representation. MLE=Modular, Lightweight, and Extensible Ontology Structure

References	Ontology Proposed	LBD Alignment	MLE Structure	DfX Criteria Inclusion	Abstraction Level
Gan [7]	No	No	No	Multi	-
IfcOwl [18]	Yes	Yes	No	No	All
DfX [1,14,15,20]	Yes	No	No	Single	Middle
Cao et al. [20]	Yes	Yes	No	Single	Lower
Rasmussen et al. [18]	Yes	Yes	Yes	No	Lower
MBCO	Yes	Yes	Yes	Multi	Middle

despite alignment with ifcOWL [17,18]. Further ontology-based manufacturability studies [19,20] define core component concepts, but do not adequately capture module–element compositional relationships or align well with existing linked building data ontologies. A lean, comprehensive, and standardised cross-domain design representation is therefore still lacking.

2.3 Research Gaps

The following research gaps can be summarised based on existing literature:

1. There is no standard ontology for representing modular buildings' component hierarchy concepts. Particularly, an alignment with upper-level ontological frameworks like BOT, BPO, etc. is missing.
2. There is a lack of a middle-level domain specific ontology definition for modular buildings. The existing efforts on modular building ontologies target the lower-level definitions, often based on individual cases or specific component instances, aligning more with the definition of a knowledge graph rather than an ontology. This also limits the multi-domain reasoning capability for DfX as each study defines all definitions based on its own singular criterion (e.g., manufacturability [19], constructability [15], cost assessment [13], etc.).
3. Most work on ontology-based DfX for modular buildings is focussed on the design review process [1], and not on design generation process, which limits the automation of design.

To address these gaps, this study develops the Modular Building Component Ontology (MBCO) as a federated extension to BOT in the domain of modular construction, aligning with the principles of domain extensible shared ontology [21], and Linked Data deployment scheme. This is meant to compliment the work of LBD Community Group[1] of the World Wide Web Consortium (W3C) in the direction of modular buildings. Table 1 compares MBCO with existing efforts aimed at formalising a graph-based representation for modular buildings across five dimensions. Here, LBD alignment assesses the reuse of existing ontologies. The modular, lightweight, and extensible (MLE) criterion assesses the ontology structure against linked open data principle of defining ontologies as light-weight modules which are easy to maintain and extend as required.

The positioning of the current study relative to the wider ontology development efforts within the AEC domain is illustrated in Figure 1. In this context, MBCO is meant to be a bridge connecting upper-level AEC domain ontologies with lower-level ontologies built for specific objectives.

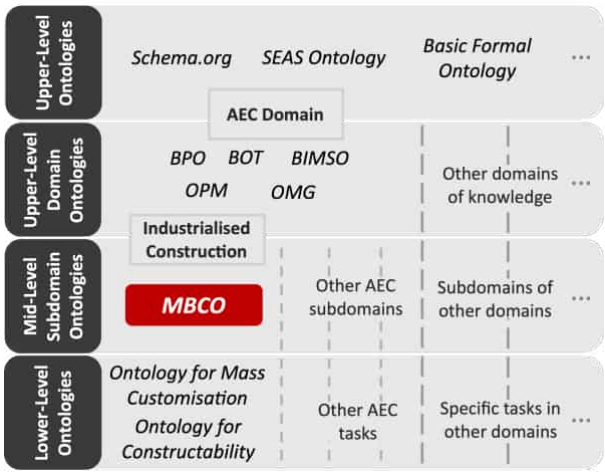


Figure 1. Positioning of MBCO among the wider ontology development efforts.

3 Methodology

The development of the proposed ontology follows a six-step process. This is guided by the Ontology Development 101 framework [22]. This method was selected because of its widespread use within

[1] https://www.w3.org/community/lbd/

construction applications and the provision of a tool-supported, step-by-step procedure incorporating best practices for ontology building [4,22]. This includes: (1) the definition of domain, scope, and competency questions, (2) reuse and alignment with existing ontologies, (3) enumerating key terms, (4) defining classes, (5) class properties & their types, and (6) creating instances and evaluation of ontology. Protégé is an open-source tool recommended by Ontology Development 101, and is used here, with Web Ontology Language (OWL) based reasoning to verify logical consistency and inferred classifications. The resulting ontology is organised in a modular manner – a lean core schema for component hierarchy and connectivity, with optional extensions to external datasets (like environmental indicators, etc.) to avoid over-complicating the core vocabulary while enabling multi-domain reasoning workflows.

4 Modular Building Component Ontology

4.1 Domain, Scope & Competency Questions

MBCO is developed to provide a core formal, reusable representation of modular building component hierarchy that supports design reasoning and automation, and which can be extended with specialised DfX ontologies. Here we use the term component in a loose definition to refer to any part of a modular building.

The scope of the ontology is restricted to defining (i) a core modular component decomposition structure, and (ii) a minimal set of extendable attributes for multi-criteria DfX incorporation. To further clarify the scope, the following competency questions (CQs) are developed:

CQ1. What are the types and numbers of all modules that constitute a given modular building?
CQ2. What smaller modules and/or elements are contained inside a larger module?
CQ3. What modules are adjacent to or intersecting with a given module?
CQ4. How to assign metadata to modules, elements, or their connections?
CQ5. How to assign attributes to constituents of a modular building to serve a given DfX criterion?
CQ6. How can attributes of modules/elements be reused across multiple DfX criteria?

Further, separate modular ontologies can be integrated for each DfX criterion that needs to be considered. Here, we show this by the examples of manufacturability and mass customization used in combination with the core MBCO. For example, competency questions for assessing mass customisation criteria using MBCO may be considered as follows:

CQ7. Which components are reused across multiple modules and with which properties?
CQ8. What is the aggregated manufacturing cost based on a given modular composition of a building?

4.2 Reuse of Existing Ontologies

Reuse of and alignment with existing ontologies and knowledge definitions is one of the key principles of linked data implementation [22]. To avoid duplicate effort, existing libraries of reusable ontologies like Linked Open Vocabularies[2], DAML[3], and Ontobee[4] were searched for terms related to modular buildings, industrialised construction, prefabrication, and design. Further, Scopus [5] database was searched for relevant works of a non-ontological nature, or which might not be published as open use ontologies. Although some recent efforts have tried to implement ontology-based design representations for DfX [15,19], but no effort was found providing a core, reusable modular building structure, in-line with W3C linked data deployment scheme which is reusable across multiple DfX criteria.

To ensure interoperability with the LBD ecosystem, MBCO reuses and aligns established vocabularies rather than redefining generic building concepts related to topology, products, or geometry. BOT is adopted to represent building topology, including zones, elements, and their containment and adjacency relationships. This caters to partially address CQ2, CQ3, and CQ4. Further, BPO is aligned to represent part-whole assembly relationships addressing CQ2. For the core concepts of MBCO, the product decomposition taxonomies from [6,23] and ontology concepts from [15,19,20] are reused to varying extents to address CQ1 through CQ6, as explained below. To attach metadata, we utilise concepts from ontology for property management (OPM) and ontology for manging geometry (OMG) to address CQ4.

4.3 Enumerating Important Terms

Important terms related to modular buildings' component hierarchy have been identified through knowledge sources including textbooks , articles, and open-source data dictionaries describing modular design [23–25]. Continuing from the scope definition, the important terms for MBCO may be viewed from two interlinked perspectives – (i) terms representing the core modular product decomposition structure in view of CQ1 through CQ4, like modular building, volumetric module, planar module, framing element, connection, interface, material, geometry, typology, sub-components, etc., and (ii) terms allowing for multiple DfX criteria inclusion in

[2] https://lov.linkeddata.es/dataset/lov
[3] https://www.daml.org/ontologies/
[4] https://ontobee.org/
[5] https://www.scopus.com/

view of CQ5 and CQ6, like DfX criterion, DfX performance indicator, etc.

4.4 Defining Classes and Class Hierarchy

The classes and their hierarchy are defined based on the enumerated concepts and the reused ontologies. A combination development process [22] is employed, which starts with the description of the most salient concepts like *VolumetricModule* and *PlanarModule*. These are then generalised (bottom-up) into a *Module* class and are specialised (top-down) into *ModuleType* class. *ModuleType* is introduced here as a generic concept to distinguish modules at any level of hierarchy. For example, *PlanarModule* may be specialised into Exterior wall, Interior wall, Slab, Framing, etc., or *VolumetricModule* may be specialised into Bathroom pod, Kitchen pod, or Balcony module, etc. But these concepts are left to be decided by the user, in this case, the designer or manufacturer. This is in line with addressing CQ1 and CQ6 and maintaining MBCO as a middle level domain ontology. *VolumetricModule* is then aligned as a subclass of *bot:Zone*, while *PlanarModule* is specified under *bot:Element*. The BOT classes allow the topological expressivity to address CQ2 and CQ3. Additionally, *mbco:ModularBuilding*, a subclass of *bot:Building* is deemed necessary to allow for building level attributes specific to modular buildings. Base attributes of a *Module* like material, geometry, etc. are supported directly through *opm:Property* class, which may further be linked to *omg:Geometry* class [11], addressing CQ4. Similarly, the types of connections between modules, elements, and modules and elements play an important role in defining the component breakdown structure in modular buildings, often influencing design choices related to mass customisation, cost, and constructability. *Connection* class has been defined here to specify these requirements as a subclass of *bot:Interface*. Further, *ConnectionType* class is defined in relation to *Connection* class to fully address the needs for CQ4.

Similar process is repeated for defining the DfX related classes addressing CQ5 and CQ6, where *DfXCriterion* represents a given module's DfX criterion under assessment, like manufacturability, constructability, etc. *DfXIndicator* class further encapsulates the performance indicators for individual DfX criteria like dimensional constraints for manufacturability, lifting constraints for constructability, typology constraints for mass customisation, etc. of a given module or element. These classes need to be utilised in combination with the task-specific ontologies extending the MBCO for realising comprehensive assessments.

4.5 Defining Properties and their Types

Properties describe the internal structure of the classes and are essential to realising the competency questions [22]. Object properties link classes to other classes, while datatype properties link classes to specific values as necessary. To operationalise the class model in MBCO, we define a small set of object properties deemed essential to answer the competency questions, as summarised in Table 2:

1. Modular decomposition across building levels is anchored in BOT. *mbco:hasModule* links a *bot:Storey* to any *mbco:Module*, while *mbco:hasVolumetricModule* specialises this for *mbco:VolumetricModule* and is declared a subproperty of both *mbco:hasModule* and *bot:containsZone*; similarly, *mbco:containsPlanarModule* is a subproperty of *mbco:hasModule* and *bot:hasElement*, enabling query-compatible retrieval while preserving BOT semantics.
2. *mbco:hasProperty* connects a *Module* to *opm:Property* definition for allowing attributes like

Table 2: Summary of object properties in MBCO (all entities defined in *mbco:* domain except specified otherwise)

Competency Addressed	Object Properties	Sub-Property of	Domain	Range
CQ3, CQ4	*connectionOf*	*bot:interfaceOf*	*Connection*	*Module, bot:Element*
CQ1, CQ2	*containsPlanarModule*	*hasModule, bot:hasElement*	*bot:Zone*	*PlanarModule*
CQ4	*hasConnectionType*	-	*Connection*	*ConnectionType*
CQ5, CQ6	*hasDfXCriterion*	-	*ModularBuilding, Module*	*DfXCriterion*
CQ5, CQ6	*hasDfXIndicator*	-	*DfXCriterion*	*DfXIndicator*
CQ1	*hasModule*	-	*bot:Storey*	*Module*
CQ1	*hasModuleType*	-	*Module*	*ModuleType*
CQ4	*hasProperty*	*seas:hasProperty*	*Module*	*opm:Property*
CQ1, CQ2	*hasVolumetricModule*	*hasModule, bot:containsZone*	*bot:Storey*	*VolumetricModule*

material, geometry, etc. Where the extended hierarchy of properties is not required, *props* predicates can be directly used from IFC models as done by the IFC-to-LBD convertor [26].

3. Inter-module connectivity is captured via *mbco:connectionOf* to represent realised interfaces between modules.
4. Explicit type assignment relations are provided using *mbco:hasModuleType* and *mbco:hasConnectionType*. These properties support integration with controlled vocabularies (e.g., manufacturer catalogues, platform families, standard connection taxonomies). In practice, *ModuleType* and *ConnectionType* can evolve into enumerated classes or be aligned to higher level concept schemes (like *skos:Concept*), depending on whether strict logical typing or flexible terminological tagging is desired.
5. DfX assessment is modelled using *mbco:hasDfXCriterion* and *mbco:hasDfXIndicator*, creating a structure for computable multi-criteria evaluation as envisioned by [1].

No datatype properties are introduced yet to keep the core ontology lightweight and easily extensible. Quantitative or logical DfX assessment values (e.g., costs, embodied carbon, commonality scores, constraint checks, etc.) can be added later either through task-specific extension modules or external vocabularies. A total of 9 classes and 9 object properties comprise the final core structure of MBCO, as illustrated in Figure 2.

5 Case Study – Creating Instances and Evaluation

To illustrate how MBCO can be used in practice, we model a minimal modular building design (Figure 3) consisting of two volumetric modules and eight planar modules: six wall panels (three solid walls and three opening walls) and two slabs. The goal is to show (i) how a designer can quickly create a consistent instance graph using the MBCO relationships, and (ii) how the same instance data can be evaluated using lightweight, executable checks and computations. The capabilities of CQ1 through CQ6 are utilised for realistic design applications. In this case study, *SolidWall*, *OpeningWall*, *Slab*, *BathroomPod*, and *KitchenPod* are represented as module types (instances of *mbco:ModuleType*). Module properties were attached using IFC-to-LBD-style [26] *props:* predicates so that wall and slab dimensions could be queried directly in a manner consistent with IFC-derived linked data. Additionally, a total of nine connections were instantiated: five of type-1 between volumetric modules and walls, and four of type-2 between walls. The graph thus becomes capable of reasoning over interactions between the modules along with the decomposition structure. The resulting instance graph includes 28 named nodes and 120 RDF triples (i.e. relationships), a part of which is illustrated in Figure 4.

Ontology evaluation ensures logical consistency and testing for fitness for the purpose. In this study, automated consistency checking is done using Protégé HermiT reasoner due to its speed and ease of implementation. It showed no error, confirming the consistency of the ontology. Further, basic task-based evaluation checks are conducted as follows: (i) a lower-level commonality [27] score is computed using the dimensions, types, and connections of planar modules retrieved using SPARQL queries on the populated RDF graph, (ii) manufacturability checks are done on all wall panels to ensure they do not exceed a production line maximum allowable length (set to 8m), and (iii) constructability checks are performed using gross weight of volumetric modules against a crane lifting capacity (set to 4500kg), using SHACL shapes for reasoning.

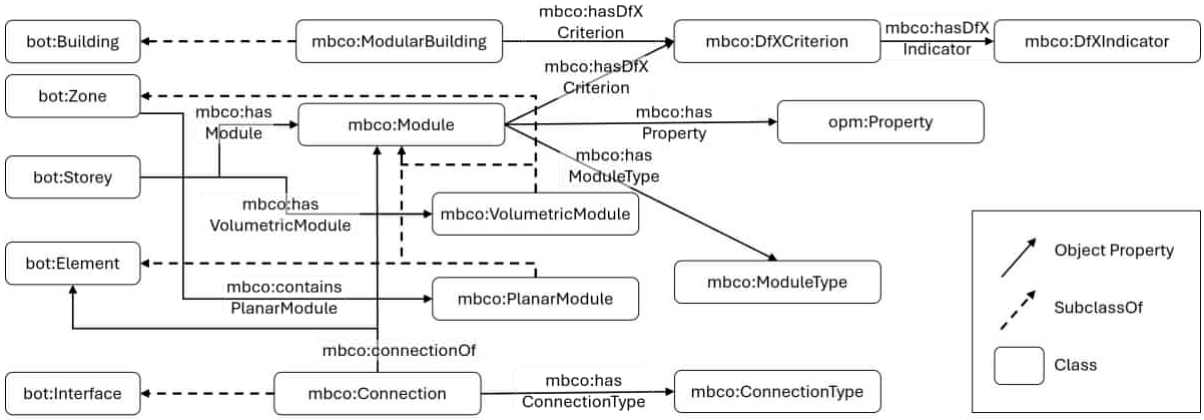


Figure 2: The Modular Building Component Ontology (MBCO)

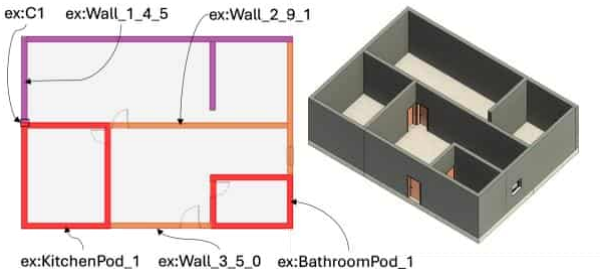


Figure 3. Simple Modular Building Example Case

The computed commonality value of each element and the SHACL reasoning results are populated back into the instance graph using *DfXCriterion* and *DfXIndicator* instances. The SHACL reasoning checks both missing values and values exceeding capacity. Two walls (13.3m and 9.1m) are flagged as being too long against the line limit of 8m. The kitchen pod is flagged as being too heavy against the crane lifting capacity. This demonstrates how MBCO can support the workflow from instance creation through to executable manufacturability and constructability checking and graph-native commonality evaluation.

6 Conclusion

This paper presented the Modular Building Component Ontology (MBCO) as a core, lightweight, interoperable semantic foundation for representing modular buildings in early-stage design. By positioning MBCO as a middle "core" layer aligned to established building ontologies (notably BOT), the ontology captures essential modular concepts like volumetric and planar modules, their containment relationships, and realised interfaces while remaining extensible to downstream task ontologies and external datasets. The case study demonstrated how a simple modular plan can be instantiated efficiently and then evaluated using executable mechanisms: a SPARQL-based commonality metric to quantify reuse across planar modules, and SHACL rules to validate manufacturability and constructability constraints such as production-line length and crane lifting weight limits for various modules. Together, these examples show how MBCO supports a practical workflow from structured representation to computable assessment without embedding project-specific constraints into the core schema. MBCO therefore provides a reusable bridge between modular building semantics and DfX-style decision support, enabling more consistent data integration, automated checking, and possibility for graph-based optimisation. Future work will expand the type catalogues and connection semantics, introduce optional lower-level extension ontologies for performance attributes, and validate the approach on richer industrial case studies integrating BIM/IFC pipelines and multi-objective design optimisation. The data from this research is available at https://github.com/khansalik/MBCO.

References

[1] S.-K. Lee, M.-S. Jeong, T.-Y. Cho, B.-Y. Chung, and J.-H. Yu. Ontology-Based Framework for Context-Aware DfX Criteria Inference Using BIM. *Journal of Management in Engineering*, 41:04025039, 2025.

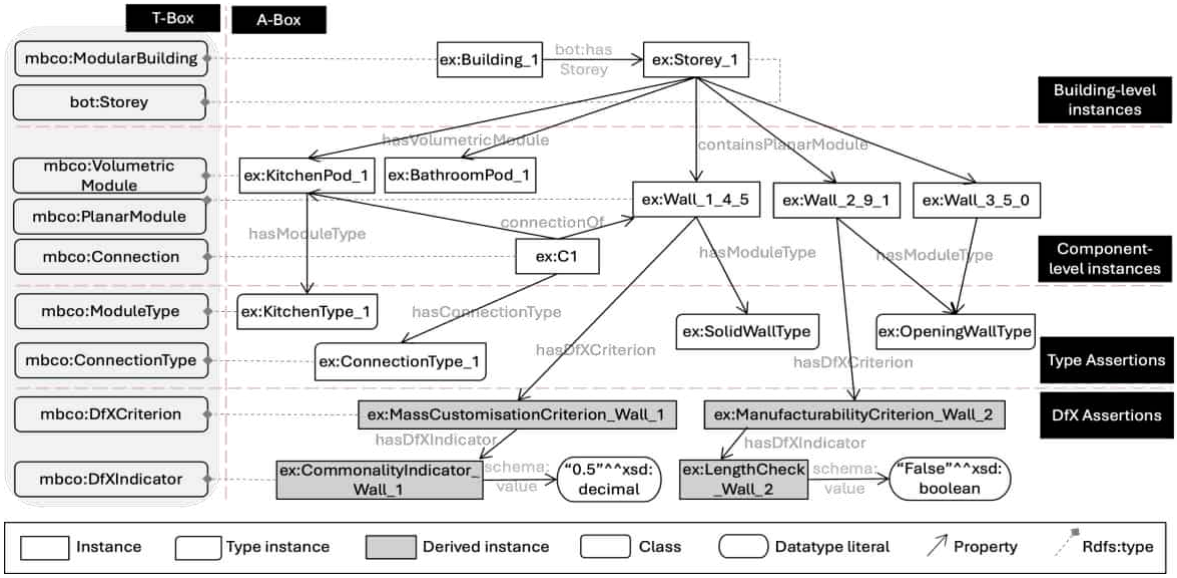


Figure 4: Instance subgraph for the case study populated in accordance with MBCO

[2] P. Ghannad and Y.-C. Lee. Optimizing Modularization of Residential Housing Designs for Rapid Postdisaster Mass Production of Housing. *Journal of Construction Engineering and Management*, 149:04023046, 2023.

[3] T. Salama. Configuration of Hybrid Modular Construction for Residential Buildings. *IJIMT*, pages 106–112, 2017.

[4] A.K. Agrawal, Y. Zou, L. Chen, H. Jin, and M.A. Abdelmegid. An Ontology for Representing Crawler Crane Operational Space Requirement on Semantic Web. In *32nd Annual Conference of the International Group for Lean Construction (IGLC 32)*, pages 648–659, Auckland, New Zealand, 2024.

[5] P. Pauwels, S. Zhang, and Y.-C. Lee. Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, 73:145–165, 2017.

[6] V.J.L. Gan. BIM-based graph data model for automatic generative design of modular buildings. *Automation in Construction*, 134:104062, 2022.

[7] M.H. Rasmussen, M. Lefrançois, G.F. Schneider, and P. Pauwels. BOT: The building topology ontology of the W3C linked building data group. *Semantic Web*, 12:143–161, 2021.

[8] M.H. Rasmussen, M. Lefrancois, M. Bonduel, C.A. Hviid, and J. Karlshøj. OPM: An ontology for describing properties that evolve over time. In *Proceedings of the 6th Linked Data in Architecture and Construction Workshop*, 2018.

[9] C. Widanage and K.P. Kim. Integrating Design for Manufacture and Assembly (DfMA) with BIM for infrastructure. *Automation in Construction*, 167:105705, 2024.

[10] P. Jensen. *Configuration of Platform Architectures in Construction*. Doctoral thesis, Luleå tekniska universitet, 2014.

[11] A. Wagner, M. Bonduel, P. Pauwels, and R. Uwe. Relating geometry descriptions to its derivatives on the web. In *Proceedings of the conference*, pages 304–313, 2019.

[12] J. Beetz and B. de Vries. Building product catalogues on the Semantic Web. In *Managing IT in Construction/Managing Construction for Tomorrow*, 2009.

[13] S.-K. Lee, K.-R. Kim, and J.-H. Yu. BIM and ontology-based approach for building cost estimation. *Automation in Construction*, 41:96–105, 2014.

[14] S. Bakhshi, M.R. Chenaghlou, F. Pour Rahimian, D.J. Edwards, and N. Dawood. Integrated BIM and DfMA parametric and algorithmic design based collaboration for supporting client engagement within offsite construction. *Automation in Construction*, 133:104015, 2022.

[15] J. Luo, P. Zadeh, and S. Staub-French. Ontology-Based Design Features for Representing Constructability in Architectural Design: Toward BIM in Off-Site Construction. *Journal of Construction Engineering and Management*, 150:05024010, 2024.

[16] J. Cao and D. Hall. Ontology-based Product Configuration for Modular Buildings. In *International Symposium on Automation and Robotics in Construction (ISARC) Proceedings*, pages 171–176, 2020.

[17] M.H. Rasmussen, C.A. Hviid, J. Karlshøj, and M. Bonduel. Managing space requirements of new buildings using linked building data technologies. In J. Karlshøj and R. Scherer, editors, *eWork and eBusiness in Architecture, Engineering and Construction*, pages 399–406. CRC Press, 2018.

[18] P. Pauwels and W. Terkaj. EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*, 63:100–133, 2016.

[19] J. Cao, E. Vakaj, R.K. Soman, and D.M. Hall. Ontology-based manufacturability analysis automation for industrialized construction. *Automation in Construction*, 139:104277, 2022.

[20] B. Qi and A. Costin. BIM and Ontology-Based DfMA Framework for Prefabricated Component. *Buildings*, 13:394, 2023.

[21] M. Niknam and S. Karshenas. A shared ontology approach to semantic representation of BIM data. *Automation in Construction*, 80:22–36, 2017.

[22] N.F. Noy and D.L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology*. 2001.

[23] I.J. Ramaji and A.M. Memari. Product Architecture Model for Multistory Modular Buildings. *Journal of Construction Engineering and Management*, 142:04016047, 2016.

[24] M. Lawson, R. Ogden, and C. Goodier. *Design in Modular Construction*. CRC Press, 2014.

[25] buildingSMART International. Title of the web content: buildingSMART Data Dictionary. On-line: https://www.buildingsmart.org/users/services/buildingsmart-data-dictionary/, Accessed: 31/01/2026.

[26] M. Bonduel, J. Oraskari, P. Pauwels, M. Vergauwen, and R. Klein. The IFC to Linked Building Data Converter - Current Status. In *Proceedings of the 6th Linked Data in Architecture and Construction Workshop (LDAC 2018)*, 2018.

[27] S.A. Khan, Y. Zou, M. Liarokapis, and M. Kieu. On Component Commonality in Prefabricated Building Design. In *Proc. Int. Workshop Intell. Comput. Eng., EG-ICE, European Group for Intelligent Computing in Engineering (EG-ICE)*, pages 590–599, 2024.

Construction of BIM-FM Data Linkage Methods Based on Real Estate Ownership Forms and Implementation in Management Systems

Ryoyu Tanaka[1] Kosei Ishida[2] Toshimasa Itaya[3] Kenji Ishihara[4]

[1] Waseda University, Graduate School of Creative Science and Engineering, Department of Architecture
[2]Waseda University, Department of Architecture, Associate Professor, Doctor of Engineering
[3]Waseda University, Graduate School of Creative Science and Engineering, Visiting Professor, Doctor of Engineering
[4] Xymax Real Estate Research Institute Senior Researcher, Doctor of Engineering
ryoyu.tanaka@moegi.waseda.jp, k_ishida@waseda.jp

Abstract

In the field of Facility Management (FM), the increasing complexity of operations and the demand for Environmental, Social, and Governance (ESG) compliance require strategic, lifecycle-oriented data management. However, current practices often suffer from information fragmentation and a lack of integration between spatial and FM data. This study addresses these issues by developing a cloud-based Building Information Modeling (BIM)-FM integration system that offers decision support to 2 specific real estate ownership forms: Corporate Real Estate (CRE) and Real Estate Investment Trust (REIT). By defining distinct management requirements for each type, we constructed a system using Autodesk Platform Services (APS) that dynamically links 3D BIM models with external management databases without overburdening the BIM files. The proposed system was verified through a prototype applied to office buildings in Tokyo. The results indicate that the system improves operational efficiency in CRE by visualizing cost sources and enhances governance in J-REITs by providing investors and asset managers with transparent, real-time access to granular property data. This approach establishes a practical model for utilizing BIM in the maintenance phase to maximize real estate value.

Keywords –

BIM; Cloud; CRE; Digital Twin; Decision Support System; FM; REIT

1 Introduction

1.1 Research Background

In recent years, facility maintenance and construction management operations have become increasingly complex due to the extension of facility lifespans and multifunctionality. Additionally, there is a growing demand for strategic management that considers the entire building lifecycle, including responses to Environmental, Social, and Governance (ESG) management and carbon neutrality. This has heightened the importance of management methods that capture various tasks cross-sectionally and handle information continuously, going beyond preservation and operation.

On the other hand, information handled in these operations often remains recorded on paper or managed individually by departments, causing information fragmentation, update delays, and personalization, which hinders rapid decision-making. Building Information Modeling (BIM) is an effective means to solve these issues.

BIM functions as a database integrating spatial and attribute information, enabling unified management throughout the lifecycle.

Nevertheless, most existing research and practices regarding BIM focus on information sharing during the design and construction phases, and concrete operational models for the facility management phase have not been established. In particular, sufficient consideration has not been given to the fact that the quality of decision-making and necessary data granularity differ depending on the purpose of real estate ownership.

Therefore, this study classifies real estate ownership forms into CRE and REIT and constructs a BIM-FM

integration system based on their respective management requirements. The purpose is to contribute to the maximization of real estate value in a stock-based society.

1.2 Previous Research

Previous research [1] attempted to construct an environment where facility managers without specialized skills could intuitively utilize BIM data. Specifically, a prototype of the system was developed using Autodesk Platform Services (APS). Figure 1 illustrates the platform developed in the previous study.



Figure 1. System developed in the previous study

In that study, targeting existing office buildings in Tokyo, a data linkage method was established to associate different forms of FM data—such as power consumption, seat usage, and repair records—with room IDs and seat IDs on the BIM model. This system enabled the visualization of energy consumption and spatial occupancy rates for each floor as a heatmap on a web browser without requiring specialized BIM software. This demonstrated the utility of the system as an integrated visualization system for building information by reducing the technical barriers to BIM utilization.

1.3 Research Objectives

The prototype constructed in previous research focused on the technical verification of 3D data visualization. However, it didn't sufficiently examine the lack of sustainability in update operations due to reference data relying on local environments, nor did it assess its suitability for decision-making processes in actual real estate management. In particular, sufficient consideration was not given to the fact that the requirements and judgment indicators sought by managers differ fundamentally depending on the form of real estate ownership.

Therefore, this study inherits the technical foundation of the previous research while securing operational sustainability by fully migrating the data management infrastructure to the cloud and newly introduces management requirements based on real estate type. The objective is to develop a more practical decision-making support model by implementing specific business logic into the system—such as operational efficiency in CRE and profitability evaluation in REIT—going beyond information visualization.

1.4 Research Methodology

This study is divided into the following three-stage process. First is the definition of management requirements based on real estate types. Target properties are selected from CRE and REIT, and the data granularity and technical requirements necessary for decision-making in each category are clarified.

Second is the construction of a BIM-linked system. Based on the cloud visualization technology developed in previous studies, a database structure and ID linkage logic satisfying the defined requirements are implemented.

Third is verification and evaluation. The developed system is applied to medium-sized office buildings in Tokyo, representing both CRE and REIT. The utility of the system for decision support and operational efficiency is verified in scenarios from both CRE and investor perspectives. Table 1 shows the chapter structure and content of this paper.

Table 1. Chapter structure and content of this paper

1. Introduction	1.1 Background 1.2 Previous Research 1.3 Objectives 1.4 Methodology	
2. Requirements Definition	2.1. Real Estate Classifications in Japan	
	CRE 2.2.1. Workflow Analysis 2.3.1. Requirements Definition	REIT 2.2.2. Work Analysis 2.3.2. Requirements Definition
3. System Development	3.1 System Architecture 3.2 Data Design 3.3 Visualization Design	
4. Verification	CRE 4.1. Workflow Verification	REIT 4.2. Workflow Verification
5. Conclusion	5.1. Summary 5.2 Future Outlook	

2 Requirements Definition

2.1 Classification od Real Estate Types and their Domestic Asset Scale

According to the Cabinet Office's Annual Report on National Accounts (FY2023 estimates) [2], the total value of "non-financial assets" (real assets including land, buildings, and structures) in Japan's national wealth has reached approximately 3,065 trillion yen. Based on the ownership entity, these assets are broadly classified into three categories: "Households," "Public Real Estate (PRE)," and "CRE".

Their ownership forms are broadly classified into two categories: PRE and CRE. In this study, grasping the asset scale of each category is important for estimating the social ripple effects of system implementation.

1. PRE
   This refers to assets owned by the national or local governments, such as government buildings, public housing, and educational facilities. The estimated asset scale is approximately 769 trillion yen. However, as the management principle is public welfare, this is excluded from the scope of this study.
2. CRE (broad defined)
   This refers to real estate owned by private enterprises. According to the Cabinet Office data, the total asset scale reaches approximately 786 trillion yen. In this study, we further classify this broad CRE into two types based on the purpose of ownership:

   - Operational Assets (narrow defined)
     Real estate owned for business activities (headquarters, factories, logistics, etc.). This accounts for the majority of the stock (approx. 762 trillion yen), where cost optimization is the primary goal.

   - Investment Assets (REIT)
     Real estate operated for rental income or capital gains. While the asset scale of J-REITs is approximately 24 trillion yen, this sector requires high transparency and accountability to investors.

In this study, we specifically target CRE, which has a large asset scale and high ripple effects on economic activities, and REIT, which is a growth area, to construct a BIM linkage model optimized for their respective management requirements.

2.2 Issues in Existing Real Estate Management Operations

2.2.1 Scope and Issues of Real Estate Management in CRE

CRE management operations consist of business processes that circulate within the enterprise. Issues and reports extracted from the site are aggregated to the property administration and FM departments and are processed as optimization measures for management resources, such as cost management and layout planning. The business content is broadly divided into the following three categories:

1 Management and Administrative Tasks:
Contract management, quality control, office receipts and disbursements, and complaint handling.

2 Operational Tasks:
Energy management, parking management, FFE (Furniture, Fixtures, and Equipment) management, environmental hygiene and cleaning, and general information and security.

3 Maintenance Tasks:
Equipment maintenance, exterior maintenance, building maintenance, construction planning and supervision.

Based on the organization of these tasks, there are two primary issues in CRE real estate management.

The first issue is the fragmentation of information and increased management costs. As mentioned above, tasks such as contract management, energy management, and equipment maintenance differ in nature, and in many cases, the departments in charge or the outsourced vendors are also compartmentalized. Consequently, data such as repair history, energy usage, and contract conditions are scattered, creating a situation where unified management is difficult. As a result, this becomes a factor in delaying judgments regarding cost reduction and quality maintenance from an overall optimization perspective.

The second issue is the personalization of maintenance operations. At building and equipment maintenance sites, there is a strong tendency for decisions to be made based on the empirical rules of skilled personnel. Since know-how regarding anomaly detection and preventive maintenance is not formalized as data, there is an inherent risk that operational quality will decline due to staff turnover or retirement. Furthermore, the shortage of Building Management (BM) personnel due to the declining birthrate and aging population may also contribute to the deepening of this issue.

2.2.2 Scope and Issues of Real Estate Management in REIT

J-REIT operations have a multi-layered structure consisting of the Site (PM), Operation (AM), and Investors. The business content is broadly divided into the following three categories:

1 Management and Financial Tasks
Lease contract management, tenant contract management, deposit management, budget and balance management, billing and payment management, and fund planning.

2 Operational Tasks
Energy management, parking management, FFE management, and environmental hygiene and cleaning management.

3 Maintenance Tasks
Equipment maintenance, exterior maintenance, building maintenance, construction planning and

supervision.

Regarding the issues in REIT management, information fragmentation, delayed decision-making, and the personalization of tasks—identified in the previous section on CRE—are considered common challenges. However, a specific issue unique to REIT is the limitation of reporting operations to the AM. In conventional paper-based monthly reports, vast amounts of operational data generated on-site were aggregated on a monthly basis and submitted in static media such as PDFs. However, the lack of direct access to raw data and means of verification prevents AMs and investors from objectively evaluating the validity of Capital Expenditure (CAPEX) and operating costs. Under increasingly sophisticated disclosure requirements, this hinders the ability to logically and immediately prove the effectiveness of investments or their contribution to improving asset value.

2.3 System Requirements Definition

2.3.1 Requirements Definition for CRE

Based on the previous section, we defined the specific system requirements for CRE.

1 Operational Workflow Transformation

In the conventional operational flow, different departments such as General Affairs, Property Administration, and Accounting managed data in their respective ledgers. For a Facility Manager to grasp the current status, a lot of man-hours were required to collect Excel files and drawings from these vertically segmented departments and to manually cross-reference them. After the introduction of this system, data from different sources is automatically integrated via spatial information; therefore, manual data verification becomes unnecessary, enabling operational efficiency and a transition to a higher standard of management.

2 Data Integration Requirements

- Power Consumption
  Time-series data of electricity usage recorded every 15 minutes on each floor from BEMS or smart meters.
- Construction and Repair Data
  Past repair history, costs, construction details, and long-term repair plans.
- Seat Occupancy Rate
  Utilization status per seat extracted from existing office seat reservation systems.

3 Key Functions:

- Cost Heatmap
  Visualizes the density of power and repair costs by color-coding them on the space to identify areas of energy waste.
- Occupancy Monitoring
  Identifies dead spaces with low usage frequency based on seat occupancy data to provide decision-making material for layout changes.

2.3.2 Requirements Definition for J-REIT

Based on the previous section, we defined the specific system requirements for REIT.

1 Operational Workflow Transformation

In conventional PM operations, basic data reported from the sites such as inspection records, repair history, and energy performance—were scattered without being linked to physical spatial information. Consequently, when PMs reported asset status to Asset Managers (AMs), it was difficult to quantitatively demonstrate the causal relationship between profitability and physical condition. Therefore, explaining equipment deterioration risks and life-extension effects—the basis for CAPEX proposals—required significant man-hours to manually re-aggregate dispersed data and supplement the logical structure. After the introduction of this system, the flow shifts to one where data in diverse formats from BM is automatically integrated on the cloud via common spatial information. This allows PMs to immediately visualize the factors causing fluctuations in Net Operating Income (NOI) in correlation with repair history and equipment status. Consequently, it makes an environment for fulfilling accountability at a higher level, enabling PMs to present risk avoidance effects and return on investment of CAPEX to AMs logically and with high resolution.

2 Data Integration Requirements

- Power Consumption
  Monthly meter reading data used for tenant billing.
- Construction/Repair Data
  Detailed construction order history including CAPEX/OPEX classification, and long-term repair plans.
- Contract Information
  Rent roll (rent, contract period, tenant information, security deposits, etc.).

3 Key Functions

- Construction Planning and Budget Management
  Visualizes construction results that caused budget variances and future repair plans in 3D space to support judgments on the validity of construction implementation.

- Contract Information Visualization

Visualizes vacancy information in 3D space to visually grasp vacant sections or sections approaching contract expiration, thereby supporting leasing activities.

In this study, we adopted a unified system architecture that satisfies these distinct requirements using a single platform foundation. The specific design of this common platform is detailed in the next chapter.

3 System Construction Methodology

3.1 Overall System Design and Architecture

To address the detailed technical design and operational workflow, a comprehensive schematic of the system architecture, database schema, and data update workflows is provided in Figure 2. As illustrated, the architecture consists of four layers. The Database Layer (Google Cloud BigQuery) stores the spatial master and operational records in separate tables, managed with key identifiers such as Spatial Object ID. The Backend API Layer (Node.js/Express) exposes endpoints (e.g., GET /api/construction/projects, GET /api/rooms/all) to process requests from the frontend layer. The BIM model is updated only during major renovations, while the Property Manager uploads financial and operational data (e.g., rent, energy, repair logs) on a regular basis, such as monthly or annually.

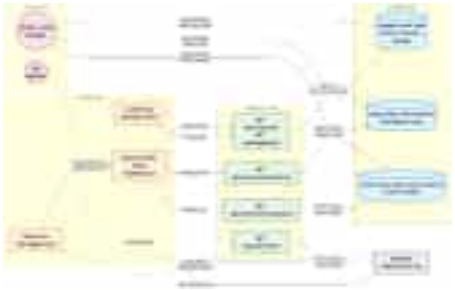


Figure 2. Comprehensive system architecture, database schema, and update workflows

3.2 Database Composition

3.2.1 Master Data Design

In designing the master data, the policy was adopted to limit the information held within the BIM model to the minimum IDs necessary for linking with the external database. Previous research has attempted methods of inputting management attributes and some non-geometric information into properties within the BIM model. However, updating frequently fluctuating information on BIM software each time requires specialized licenses and operational skills, which becomes a major factor causing update delays and discrepancies with actual conditions during the operation phase. To avoid this issue and minimize the future editing burden on the BIM model, this system uses BIM schedules to extract only the Spatial Object ID. These are defined as the primary keys for the spatial master in the system-side database. Fluid management information such as rent unit price, contract period, and tenant name is entirely managed on the cloud database side and dynamically linked via the Spatial Object ID. This establishes an environment where attribute information updates that do not involve shape changes can be executed immediately and inexpensively from a web browser without going through the BIM model.

3.2.2 Spatial Object ID Scheme

To accurately link spatial elements within the BIM model with management information in an external database, a unique identifier is essential. Previous BIM implementation case studies [3] have cited inconsistent tagging as a factor that hinders asset data integration, demonstrating the effectiveness of an ID scheme that hierarchically combines physical location and system classification.

Based on these findings, this system adopts a spatial ID naming rule consisting of the following five levels, hereafter referred to as "room_id" in this study, designed to optimize spatial identification and filtering efficiency in database queries. Table 2 summarizes the detailed definitions and examples of this scheme.

Table 2 Spatial Object ID Naming Levels

Object ID Naming Levels	Description
Level 1: Floor Code	Floor level indication where the space is located. (e.g., "2F" for Second Floor)
Level 2: Large Zone Name	AreaName: Functional classification of the major zone. (e.g., "OfficeZone", "Staircase")
Level 3: Large Zone Number	Two-digit number identifying specific major zones. (e.g., "01", "02")

Formatted Table

Formatted: Font: 8 pt

Formatted: Font: 9 pt

Formatted: Font: 8 pt

Formatted: Font: 9 pt

Formatted: Font: 8 pt

Formatted: Font: 9 pt

Level 4: Small Zone Name	Descriptive name of the specific functional unit. (e.g., "MeetingRoom", "Workspace")
Level 5: Small Zone Number	Two-digit number identifying specific zones. (e.g., "01", "02")

The codes for each level are jointed using underscores (_) as delimiters to construct the final format as follows:

[Floor]_[LargeZoneName]_[LargeZoneNumber]_[SmallZoneName]_[SmallZoneNumber]

In this study, this room_id is assigned based on structural spans to serve as a unchanging spatial unit. In the database, tenant contracts and other management data are managed as a collection of these associated room_id units. To ensure the continuity of historical data during floor plan changes caused by contract renewals or interior reconfigurations, the system responds by updating the assignment of these persistent room_id units to the corresponding management records. This room_id is stored as a Revit instance parameter, functioning as the primary key for external database integration.

3.2.3 Data Formatting

Raw data imported from accounting systems or PM reports are lists of numbers that do not possess spatial coordinates. In this system, a data formatting pipeline using SQL join processing was constructed to assign the ~~IfcGuid~~ DBID to this data via the spatial master defined in the previous section. The DBID is an internal integer identifier generated by APS, which is strictly required to select, color, or manipulate specific 3D objects within the web viewer. Specifically, queries were designed to execute joins against the spatial master table extracted from BIM, using floor information and section names contained in the raw data as foreign keys. Through this processing, for example, for a construction record such as "2024-Month X, AC Repair Cost, Office 1", the ~~IfcGuid~~ DBID corresponding to Office 1 is automatically mapped. In the event of a mismatch between the room_id and the external data due to input errors in the raw data, the DBID returns a null value. Since it is only necessary to verify these data matching errors during the initial stage of defining the spatial model, these missing data are not displayed on the dashboard. However, these null values can be checked in the frontend console.

3.3 Data Visualization Algorithm

The role of the visualization algorithm is to convert values in the database into visual attributes perceptible to the user. Internally, the system utilizes the aforementioned DBID to match the retrieved database records with specific 3D objects in the APS Viewer, rendering them based on the following logic. For coloring spatial objects on the BIM model, two methods are used depending on the nature of the data: Discrete Value Mapping and Continuous Value Mapping.

1 Discrete Value Mapping
  Applicable Data:
  Contract status, construction status (planned/completed), etc.
  Logic:
Determines whether the data string perfectly matches a specific condition.

2 Continuous Value Mapping
  Applicable Data:
  Power consumption, rent unit price, etc.
  Logic:
  Normalizes numerical values to a range of 0.0 to 1.0 based on maximum and minimum values and maps them to a color hue circle.
The calculation is performed using Equation (1).

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

x_{norm}: The normalized value (ranging from 0.0 to 1.0).
x_i: The individual data value of the target object.
x_{min}: The minimum value within the dataset.
x_{max}: The maximum value within the dataset.

The calculated normalized value *x(norm)* is substituted into the color ramp function *f(x)* to determine the RGB value.

4 Verification

4.1 Case Study for CRE

4.1.1 Overview of Target Facility

To verify the utility of the proposed system, a prototype system was constructed and applied to two representative office buildings in Tokyo, each corresponding to a different ownership form.

In this study, we verified the system's effectiveness in solving the issues of information fragmentation and difficulty in status grasping faced by CRE. The target facility is a medium-sized office building located in Minato-ku, Tokyo, selected to simulate a typical corporate asset. Details of the building are below.

Formatted: Font: 8 pt
Formatted: Font: 9 pt
Formatted: Font: 8 pt
Formatted: Font: 9 pt

- Scale: B1/10F
- Total Floor Area: Approx. 7,800 m²
- Structure: Steel (partly CFT)

4.1.2 Verification of Effectiveness Scenario

The system's effectiveness was verified through a scenario integrating energy management and maintenance tracking. First, Figure 3 shows that selecting a specific date on the dashboard simultaneously displays electricity usage and seat occupancy rates for each floor. This visualization allows facility managers to immediately identify inefficient zones where energy is consumed despite low occupancy. In the context of CRE, this directly supports space optimization decisions, such as identifying target areas for layout coordination or adjusting HVAC operational hours, thereby fulfilling the primary goal of operational cost minimization.



Figure 3. Visualization of energy consumption data

Second, regarding repair history, floors were color-coded by repair frequency in the 3D viewer. As illustrated in Figure 4, it was confirmed that selecting a specific floor displayed a list of past construction records, and clicking a record immediately highlighted the corresponding section while showing detailed specifications. This seamless transition from macro-level status grasping to micro-level detail verification demonstrates the system's capability to resolve information fragmentation and significantly reduces the time required for cause identification.



Figure 4. Visualization of repair history and details

4.2 Case Study for REIT

4.2.1 Overview of Target Facility

In the second case study, we verified the system's effectiveness in solving the issues of untransparent information and accountability faced by REIT management. The target facility is an existing office building located in Shinjuku-ku, Tokyo, selected to simulate an investment property under asset management. Details of the building are as below.

- Scale: 9F
- Total Floor Area: Approx. 6,000 m²
- Structure: SRC

4.2.2 Verification of Effectiveness Scenario

The system's effectiveness was verified through a scenario where PM and AM utilize a unified information environment. First, operational data such as electricity usage and construction history trends were visualized on the management dashboard. Figure 5 shows the management dashboard for construction history and KPIs. This interface enables intuitive tracking of the building's physical condition and budget status over time. For REIT management, this bridges the gap between physical deterioration and financial impact, enabling AMs and investors to make data-driven decisions regarding the validity of CAPEX proposals and to prioritize renovations that directly contribute to maintaining asset value and maximizing NOI.



Figure 5. Visualization of construction history data

Furthermore, regarding leasing management, contract

status was mapped directly onto the 3D model. Figure 6 shows the visualization of contract data and specific section details. In this view, it was confirmed that users could visually identify vacancy risks and, by clicking specific sections, immediately access detailed contract terms.



Figure 6. Visualization of contract data and specific section details

4.3 Summary of Verification Results

To evaluate the accuracy of the ID linkage method, the mapping results from both case studies are summarized in Table 3. ~~Finally, the verification results for both case studies are summarized in Table 4.~~

A total of 665 operational records for CRE and 259 for REIT were mapped to the spatial objects extracted from their respective BIM models. As a result, the system achieved a 100% success rate across all categories.

Table 3. The mapping results from both case studies

		Valid Records	Successfully Linked	Success Rate (%)
CRE	**Energy & Contract**	189	189	100%
	Construction History	429	429	100%
	Construction Plan	47	47	100%
REIT	**Energy & Contract**	147	147	100%
	Construction History	52	52	100%
	Construction Plan	60	60	100%

Furthermore, the practical verification results for both case studies are summarized in Table 4, which outlines the key issues and main outcomes for each scenario.

Table 4 Summary of verification results

Case	Issue	Main Outcome
Case 1: CRE	Data Fragmentation	Demonstration of streamlined workflows for identifying cost sources and analysing space usage.~~Reduced time for identifying cost sources and analyzing space usage.~~
Case 2: REIT	Information Asymmetry	Provision of a unified platform to support explanation capability and transparent decision-making.~~Improved explanation capability and transparent decision-making.~~

5 Conclusion

5.1 Summary

In this study, we developed a prototype system that integrates BIM and financial data on a web browser to address the issues of information fragmentation and reporting opacity in real estate management. Through a case study applied to office buildings in Tokyo, we technologically verified the system's effectiveness in two specific aspects: accelerating the identification of cost sources for CRE and ensuring data transparency and accountability for REITs. The results demonstrated that dynamically linking spatial information with management data via the cloud is a valid technical approach for supporting decision-making and resolving information asymmetry. However, this research is currently limited to functional verification within specific scenarios. Establishing measurable baselines to quantitatively evaluate the actual reduction in operational time and the degree of improvement in decision-making remains a subject for future research.

5.2 Future Outlook

5.2.1 Expansion to Portfolio Management

The current system targets the visualization of a single property, but operational practices require unified management of the entire portfolio consisting of dozens to hundreds of properties. As a future outlook, we aim to extend this system to support multi-asset management and implement functions to cross-sectionally compare and analyze the operation status and repair risks of multiple properties. This enables decision-making support from a management perspective, such as maximizing NOI for the entire portfolio.

Formatted: Font: 9 pt, (Asian) Chinese (Simplified, Mainland China)

Formatted: Font: 8 pt, (Asian) Chinese (Simplified, Mainland China)

Formatted: ISARC Table Caption

Formatted: Font: 9 pt, (Asian) Chinese (Simplified, Mainland China)

Formatted Table

Formatted: Font: 9 pt, (Asian) Chinese (Simplified, Mainland China)

Formatted: Font: 8 pt, (Asian) Chinese (Simplified, Mainland China)

5.2.2 Application to Investor Relations (IR) and Improvement of Market Evaluation

We are also considering utilizing this system for information disclosure to external investors. To achieve this, it is necessary to select the information to be disclosed and design data access restrictions. Since the primary objective of investors is to manage asset value and cash flow, the data is provided at a different level of granularity while sharing the same underlying data source as the PM. Although the contracts, energy usage, construction history, and construction plans visualized in this study remain the subjects of visualization, to protect tenant privacy and confidential information, the system will visualize large-scale renovations exceeding a certain threshold and display general floor-by-floor energy costs, occupancy rates, and rent-based revenue trends using heatmaps. Additionally, access restrictions are enforced at the frontend presentation layer. When accessing the application, the JavaScript side checks the user's login role. If the role is an investor, the frontend converts the view and restricts the rendering of detailed data tables.

References

[1] R. Tanaka, "Development of a Facility Maintenance and Management Data Analysis Environment Using BIM," Annual Research Meeting of the Architectural Institute of Japan Kanto Chapter, 2024.

[2] Cabinet Office, Government of Japan. "Annual Report on National Accounts for 2023," 2024.

[3] Eastman, C. et al. BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers. Wiley, 2018.

[4] buildingSMART. Industry Foundation Classes (IFC) Specifications.

[5] Autodesk Platform Services Official Docs, Model Derivative & Viewer API.

Retrieval-Augmented Generation and Multi-Agent Arbitration for Construction Code Compliance

Hao Yin[1], Xichen Chen[2], Yang Zou[3], and Liupengfei Wu[4]

[1]Department of Computing, Faculty of Computer and Mathematical Sciences,
The Hong Kong Polytechnic University, Hong Kong SAR, China
[2]School of Future Environments,
Auckland University of Technology, Auckland, New Zealand
[3]School of Architecture, Building and Civil Engineering,
Loughborough University, Loughborough, UK
[4]Division of Industrial Data Science, School of Data Science,
Lingnan University, Hong Kong SAR, China

hi-bruce.yin@connect.polyu.hk, xichen.chen@aut.ac.nz, y.zou2@lboro.ac.uk, mikewu@ln.edu.hk

Abstract -

Construction code retrieval is a recurring bottleneck in AEC compliance workflows, where decisions require clause-addressable evidence and cross-discipline reference completion under incomplete project attributes. We present a clause-centric retrieval-augmented generation (RAG) pipeline that treats retrieval as auditable evidence discovery and enforces an evidence-bounded selection contract: outputs may cite only clause identifiers from the candidate set C_k with evidence spans anchored to retrieved text for mechanical validation. A trigger-based multi-agent orchestrator invokes an arbiter only under expert disagreement or validation flags. Experiments on a labelled query set (dev200/test100/all300) over a 600-code corpus (141681 clause units) show that, at the original top-10 operating point on dev200, dense retrieval reaches Hit@1=0.550 versus 0.210 for BM25. A retrieval-budget sensitivity analysis over $k \in \{1, 3, 5, 10\}$ further shows that dense gold-in-C_k rises from 0.550/0.500 at $k = 1$ to 0.935/1.000 at $k = 10$ on dev200/test100, while post-hoc feasibility audits of archived selector and arbitration outputs indicate non-linear downstream gains as candidate pools deepen. Under identical evidence inputs, evidence-bounded selection favours cost-efficient models (Qwen-plus Hit@1=0.430; USD 0.896 per 1K queries) over higher-cost alternatives (GPT-4o Hit@1=0.345; USD 5.763 per 1K queries). On held-out tests, arbitration is triggered in 37% of cases and revises the majority decision in 32.4% of triggered cases. The resulting clause packages link design decisions to regulatory clauses with auditable evidence, improving accountability and reducing manual clause hunting. Prototype code and experimental data are released at `https://github.com/bruce0210/RagConstructionAssistant`.

Keywords -

Construction Codes; Clause-Level Retrieval; RAG; Evidence-Bounded Generation; Multi-Agent Arbitration

1 Introduction

Construction code clause retrieval is a routine step in design review and compliance checking [1]. Reviewers typically need clause-addressable evidence, an explicit statement of scope and exceptions, and a compact bundle of cross-referenced clauses that can span multiple disciplines. When project attributes that govern applicability are missing, reliable practice favours an auditable set of candidate clauses with stated assumptions, rather than a single confident answer.

Large language models (LLM) and RAG offer a natural interface for code look-up, but many AEC assistants remain answer-centric: retrieved passages are used as supporting material, and citations are not enforced as mechanically verifiable links to a bounded evidence pool [2, 3, 4]. Moreover, missing scope attributes and cross-discipline coupling can yield multiple plausible clauses; deployment therefore needs governance that surfaces disagreement and evidence gaps for triage instead of hiding them in a single response.

To address these constraints, a clause-centric RAG pipeline is introduced that treats retrieval as evidence discovery and keeps downstream reasoning evidence-bounded. Codes are unitised into clause evidence objects with stable identifiers and provenance; retrieval returns a fixed candidate set C_k; discipline experts produce structured clause packages with applicability labels and evidence spans; and an arbiter is invoked only under explicit disagreement or evidence-quality violations. Figure 1 overviews the end-to-end pipeline.

The pipeline is evaluated on a labelled six-discipline query set with clause-identifier ground truth. Tables 3–5 summarise fixed-point retrieval, selection, and arbitration results at the original $k = 10$ operating point, while Section 5.5 and Figure 4 report retrieval-budget sensitivity and post-hoc feasibility audits for archived downstream outputs.

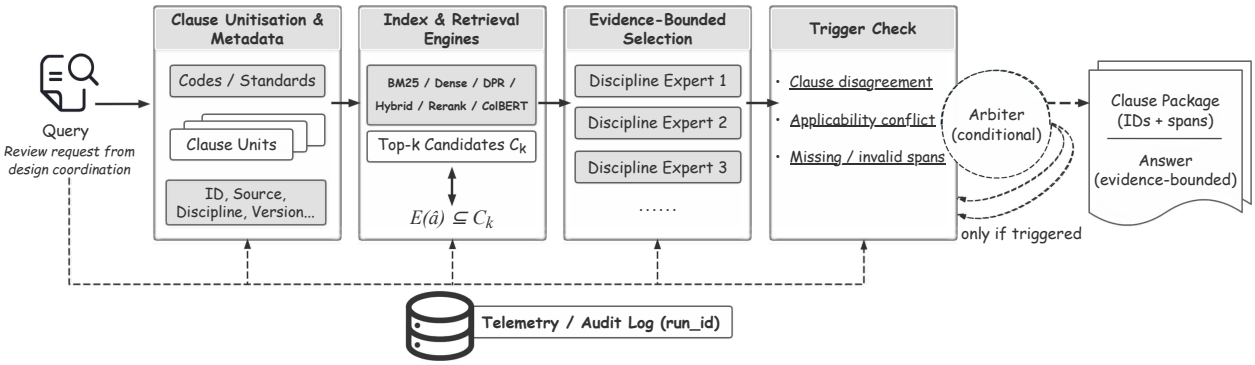


Figure 1. Auditable clause-centric RAG pipeline

The contributions are threefold. First, a clause-centric retrieval unit is defined for construction code retrieval and compliance checking, where each candidate is a citeable clause with stable identifiers and provenance metadata for plan-checking, design coordination, and cross-discipline review. Second, an evidence-bounded clause-selection contract (Formula 1) is enforced through structured validation: returned clause IDs must lie in C_k, and evidence spans must map to retrieved text. Third, a trigger-based multi-agent arbitration mechanism addresses cross-discipline ambiguity by invoking an arbiter only when experts disagree on clause/applicability or evidence is missing; trigger rates and outcomes are reported on dev200/test100/all300.

2 Literature Review

Construction code consultation sits at the intersection of regulatory knowledge management and decision support. Unlike general information seeking, the expected output is reviewable evidence traceable to specific clauses and defensible under audit [5]. This section therefore reviews clause-oriented compliance support, RAG for AEC knowledge services, and reliability and governance for LLM-assisted decision support.

2.1 Clause-Level Evidence Traceability

Digital compliance support in AEC has long recognised that regulatory text is structured by scopes, exceptions, and cross-references that affect applicability [6]. Traditional rule-checking approaches often rely on formalised representations aligned with building models [7], which can be effective when attributes are complete and scope is well specified, but usually require substantial authoring effort and remain brittle to code updates and local interpretations [8].

For practical construction retrieval, engineers need to locate the governing clause, confirm scope and exceptions, and assemble a minimal evidence bundle for review. The main bottleneck is therefore evidence traceability at clause granularity. Document-level retrieval or paragraph-level quoting can obscure the exact normative unit supporting a decision and complicate both evaluation and review workflows [9]. This motivates clause-addressable evidence units with stable identifiers and provenance, making the evidence object explicit, mechanically checkable, and easier to cite and track in review.

2.2 RAG for AEC Knowledge Services

RAG-style systems have been explored for AEC knowledge search over project documents, specifications, standards, and BIM-linked information [2, 3, 10]. They improve topical relevance and reduce reliance on parametric memory, but construction codes expose two limitations of generic RAG patterns: the review-relevant unit is usually a clause or subclause rather than an arbitrary chunk, and retrieval must account for applicability under scope conditions and cross-discipline coupling. Large passages can mask these constraints, while small chunks without stable identifiers struggle to provide review-ready citations.

2.3 LLM Reliability and Governance

Reliability in LLM-assisted decision support is often addressed with citations, structured outputs, and verifier-style checks, but compliance consultation requires a stronger guarantee than merely "having citations": cited evidence must remain restricted to, and traceable within, a bounded retrieved pool. This motivates a mechanically checkable evidence boundary (Formula 1) and span-level evidence tied to concrete text. Multi-agent designs are most useful here as governance rather than blanket ensembling: disagreement and evidence-quality violations should surface as operational risk signals that trigger conditional scrutiny, consistent with engineering review workflows and judge-style arbitration frameworks [11].

2.4 Summary of Gaps

Existing AEC compliance support and RAG systems rarely provide review-ready, clause-addressable evidence packages. Many rely on document or chunk retrieval and allow models to cite beyond a fixed retrieved pool, which weakens verifiability and auditability [4]. Ambiguity from missing scope attributes and cross-discipline coupling is also often absorbed into a single response, obscuring failure modes and governance needs under cost and latency constraints.

These gaps motivate the clause-centric evidence design, evidence-bounded selection contract, and trigger-based arbitration protocol introduced in Section 3.

3 Methodology

This section specifies a clause-centric RAG workflow for construction-code clause retrieval in compliance review, treating retrieval as auditable evidence discovery and constraining clause selection with an evidence-bounded contract. Figure 1 summarises the end-to-end pipeline.

3.1 Overview and Evidence-Bounded Formulation

Given a natural-language query q, retrieval returns a bounded candidate set of code clauses $C_k = \{c_1, \ldots, c_k\}$ from a clause-indexed corpus, and a selector produces the final clause set supporting the response. An evidence-bounded contract is enforced: the generated answer $\hat{a}$ may cite only clauses inside C_k. Let $\mathcal{E}(\hat{a})$ be the set of clause identifiers cited as evidence in $\hat{a}$, then

$$\mathcal{E}(\hat{a}) \subseteq C_k. \tag{1}$$

This constraint makes citations verifiable and bounds uncontrolled generation when project attributes are underspecified. Applicability in construction codes is shaped by scope conditions, exceptions, and cross-references; accordingly, clause unitisation provides citeable units with stable identifiers and provenance metadata, expert outputs include applicability labels and evidence spans anchored in C_k, and arbitration is triggered only when clause/applicability disagreement or invalid evidence spans indicates elevated ambiguity.

3.2 Clause Unitisation and Corpus Schema

Instead of treating each standard as an unstructured document, the corpus is represented as clause units [12]. Each clause unit stores (a) its canonical text, (b) a stable internal identifier (clause_id) used for indexing and evaluation, and (c) structured metadata such as clause number (clause_no), source document identifier, and optional media references (e.g., figures extracted from the source). This representation supports clause-addressable citations and allows downstream components to operate on comparable evidence granularity across different codes and disciplines. Clause boundaries are produced by a lightweight rule-based structural parser that detects clause-numbering patterns and common chapter/article/appendix markers, then normalises numbering and merges continuation lines to form a single clause unit. Unitisation can introduce noise in cases of inconsistent numbering, scanned layouts, or table/figure regions; therefore, raw source text is retained as provenance and spot checks are used to validate clause_id–text alignment.

3.3 Candidate Retrieval for Auditable Evidence

The retrieval stage maps q to a bounded candidate set C_k by ranking clause units using one or more retrieval engines. Each candidate $c_i \in C_k$ is packaged as an evidence item with its clause_id, clause_no, source, and a minimal text span sufficient for inspection. This evidence packaging is the only admissible evidence pool for downstream selection under the contract in Formula 1.

Construction-code queries commonly combine domain terminology with scope constraints, exceptions, and cross-discipline coupling; therefore, retrieval is framed as evidence discovery rather than a single best-match search.

3.4 Evidence-Bounded Clause Selection

Given q and C_k, each discipline expert performs clause selection and returns a structured clause package rather than free-form text. The output is defined as $o_j(q) = (S_j, A_j, \Sigma_j)$, where S_j is a short ranked list of selected clause identifiers, A_j is an applicability label for the query as stated, and Σ_j are evidence spans anchored in the retrieved candidate texts. Requiring Σ_j operationalises the evidence-bounded contract in Formula 1 and supports mechanical checks (e.g., whether cited spans fall inside C_k and whether a claimed selection is backed by explicit text), providing governance signals for downstream arbitration and error analysis.

3.5 Trigger-Based Arbitration and Governance

Arbitration serves as a governance step for ambiguity-prone cases rather than a default refinement stage, and it is invoked if the parameter-free trigger $\tau(q) = 1$ in Formula 3 holds: experts disagree on clause IDs or applicability, or schema/evidence checks fail ($\nu_j(q) = 0$). For example, if experts return $c_a \in C_k$ with applicability "yes" and $c_b \in C_k$ with "no", the arbiter stays within C_k, rejects any package whose clause IDs or spans fail validation, and then retains the package whose anchored span more directly supports the queried scope condition or exception. Thus, a scope-aligned clause may replace the majority top clause while keeping the final package

evidence-bounded; see Table 6 for an arbiter-changed example. Given $\{o_j(q)\}$, the arbiter outputs a final clause package by prioritising evidence-supported clauses, resolving cross-discipline references, and discarding selections not anchored to C_k, keeping the decision evidence-bounded and auditable.

$$m(q) = \text{mode}\big(\{s_j(q)\}_{j=1}^{M}\big), \tag{2}$$

$$\tau(q) = \mathrm{I}\Big[(\exists\, j,l:\ s_j(q) \neq s_l(q)) \vee (\exists\, j,l:\ a_j(q) \neq a_l(q)) \vee (\exists\, j:\ \nu_j(q) = 0)\Big], \tag{3}$$

$$\chi(q) = \mathrm{I}[s^*(q) \neq m(q)], \tag{4}$$

where $\mathrm{I}[\cdot]$ is the indicator function and $s^*(q)$ is the arbiter-selected clause when $\tau(q) = 1$. Formulas 2–4 define the majority top clause, trigger indicator, and arbiter-change flag.

3.6 Evaluation Protocol and Metrics

Evaluation adopts dev200 for calibration and test100 for final reporting. For each query q with a gold set of relevant clause identifiers $G(q)$, retrieval quality is measured by Hit@k on identifiers, together with MRR and nDCG@k. For the multi-expert stage, the majority selector is compared with the trigger-based arbiter output; Hit@1 is reported overall and on the triggered/changed subsets, and the arbitration accuracy delta is defined as $\Delta\text{Hit@1} = \text{Hit@1}_{arb} - \text{Hit@1}_{maj}$ (overall and triggered) under the same evidence pool. Governance rates over a query set Q are defined in Formula 5.

$$\begin{aligned} \text{TrigRate} &= \frac{1}{|Q|}\sum_{q\in Q}\tau(q), \\ \text{ChangeRate} &= \frac{1}{|Q|}\sum_{q\in Q}\tau(q)\chi(q), \\ \text{Change} \mid \text{Trig} &= \frac{\sum_{q\in Q}\tau(q)\chi(q)}{\sum_{q\in Q}\tau(q)}. \end{aligned} \tag{5}$$

4 Prototype Implementation

This section describes a working prototype that operationalises the methodology in Section 3. The implementation emphasises auditability, reproducibility, and modular replacement of retrieval and model components. Figure 2 shows the system architecture and data flow.

4.1 System Architecture and Data Flow

A query q enters the pipeline as a retrieval request. The retrieval layer ranks clause units and returns a bounded candidate set C_k as a packaged evidence bundle (clause identifiers, source references, and minimal inspectable text spans). The orchestration layer applies the evidence-bounded selector under Formula 1; if trigger conditions indicate clause/applicability disagreement or evidence violations, arbitration is invoked with q, C_k, and expert outputs. The system returns a clause-addressable package and a concise explanation bounded to cited evidence, while emitting structured telemetry keyed by a run identifier for replay and error analysis, as shown in Figure 2.

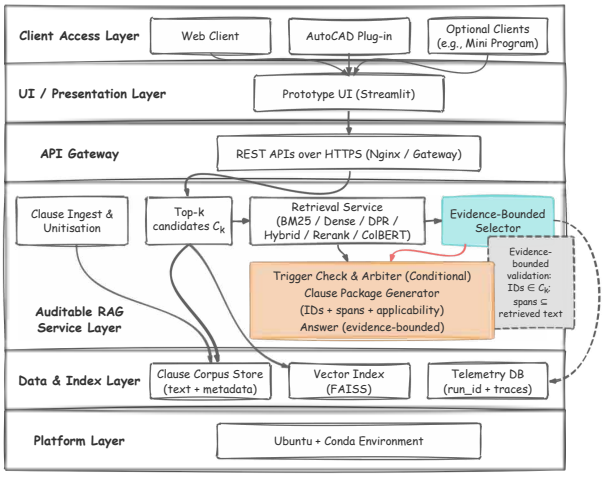


Figure 2. System architecture of the auditable clause-centric RAG prototype. Evidence-bounded validation checks that cited clause IDs belong to C_k and evidence spans are anchored to the retrieved clause text; violations raise governance triggers.

4.2 Web Interface and Telemetry Logging

The web UI prioritises evidence inspection: retrieved candidates are rendered with clause identifiers and provenance, and the final output is a clause package with explicit clause IDs and evidence spans for rapid cross-checking. All stages are recorded in a relational telemetry store for replay and audit, capturing the query, retrieval configuration, candidate-set hashes, model settings, and the final clause package. Table 1 summarises the core telemetry entities captured for reproducible experiments.

4.3 Arbitration Service and Prompt Orchestration

The orchestration layer separates multi-agent expert selection from arbitration under strict input-output schemas. A lightweight validator enforces the evidence-bounded contract by checking that cited clause identifiers belong to C_k and that evidence spans align with the returned candidate texts; arbitration is invoked only when the trigger rules

Table 1. Telemetry artifacts captured for auditable runs

Artifact	Stored in	Audit role
Run record	`trace_logs` (PostgreSQL)	Groups a run and links query, configs, and outputs via `run_id`.
Candidate pool	`*_details.jsonl`	Reproduces retrieval candidates for offline metrics and error analysis.
Expert outputs	`*.agents.jsonl`	Inspects discipline votes, applicability labels, and evidence spans.
Arbitration trace	`*.adjudication.jsonl`	Explains trigger reasons and produces the final clause package.

in Section 3.5 are satisfied. Prompts use a shared schema containing the query, minimal context, bounded candidates C_k (clause id/provenance/spans), and a fixed output contract (selected IDs, applicability, anchored spans). When triggered, the arbiter receives q, C_k, and the expert clause packages, and returns a final clause package with clause identifiers, aligned evidence spans, and an applicability judgement.

4.4 CAD Integration Entry Point

To integrate construction-code clause retrieval into drawing production and coordination, the prototype provides a lightweight CAD entry point for embedded consultation inside AutoCAD. A practitioner can invoke the custom command `"RAGCA"` during drafting to query relevant code clauses without leaving the drawing environment. The server-side pipeline returns the same evidence bundle and clause package as the web interface, keeping results auditable under the evidence-bounded contract and supporting compliance checking during drawing-based coordination. Figure 3 illustrates the within-CAD consultation flow.

5 Experiments and Results

This section reports fixed-point retrieval, selection, and arbitration results at the original $k = 10$ operating point, and then adds a retrieval-budget sensitivity analysis to examine how stricter candidate budgets affect reachability and downstream decision feasibility.

5.1 Datasets and Experimental Setup

Evaluation uses a labelled six-discipline query set (architecture, structure, heating, ventilation, and air conditioning (HVAC), plumbing, electrical, fire), with dev200 for calibration, test100 for held-out assessment, and all300 for their union. Retrieval operates over a clause-indexed corpus of 600 construction codes (141681 clause units); within each comparison, all methods share the same candidate budget k to ensure comparable evidence inputs.

Table 2. Dataset statistics and split configuration. Primary discipline is used for each query. "Cross" denotes integrated multi-discipline scenarios. Single-discipline queries = $\#Q - \#Cross$.

Split	#Q	#Cross	Arch.	Str.	HVAC	WSD	Elec.	Fire
Dev200	200	36 (18.0%)	31	30	35	29	39	36
Test100	100	24 (24.0%)	19	20	15	21	11	14
All300	300	60 (20.0%)	50	50	50	50	50	50

Abbreviations: Arch.=Architecture; Str.=Structure; WSD=Water supply and drainage; Elec.=Electrical; Fire=Fire protection.

Table 3. Clause retrieval results under a fixed top-10 candidate budget (dev200, $n = 200$)

Method	Hit@1	Hit@5	Hit@10	nDCG@10
Dense	**0.550**	**0.850**	**0.935**	**0.691**
Hybrid	0.385	0.725	0.820	0.504
Rerank	0.315	0.600	0.760	0.414
DPR	0.305	0.585	0.685	0.370
ColBERT	0.290	0.535	0.660	0.351
BM25	0.210	0.460	0.565	0.267

Main selector and arbitration comparisons use top-k candidates with $k = 10$ as the original operating point. This supports controlled comparison but does not by itself characterise retrieval-depth sensitivity. Increasing k would generally improve upper-bound recall, but also increases evidence inspection cost and can dilute top-rank precision. Section 5.5 therefore analyses $k \in \{1, 3, 5, 10\}$ using exact dense gold-in-C_k rates on dev200/test100 and post-hoc audits of archived $k = 10$ selector/arbitration outputs under stricter prefix pools C_k. Table 2 summarises split statistics and discipline coverage.

5.2 Retrieval Baseline Comparison

Retrieval baselines rank clause units to form the bounded evidence pool for downstream selection. Stronger retrieval raises the upper bound of evidence-bounded selection, although final correctness still depends on accurate clause selection.

Table 3 reports retrieval effectiveness on dev200 under a fixed top-10 candidate budget. Dense retrieval achieves Hit@1=0.550 and nDCG@10=0.691, versus 0.210 and 0.267 for BM25. Hybrid improves over BM25 on Hit@1 (0.385) and nDCG@10 (0.504) but remains below Dense under the same budget, suggesting that score fusion can still elevate lexically similar yet scope-mismatched clauses when attributes are underspecified.

5.3 Evidence-Bounded Selector Comparison

Evidence-bounded clause selection is evaluated with the same candidate budget ($k = 10$) to test whether a selector can identify the correct clause from a bounded evidence pool; dev200 is used because this stage is sensitive to prompt calibration and output-schema valida-

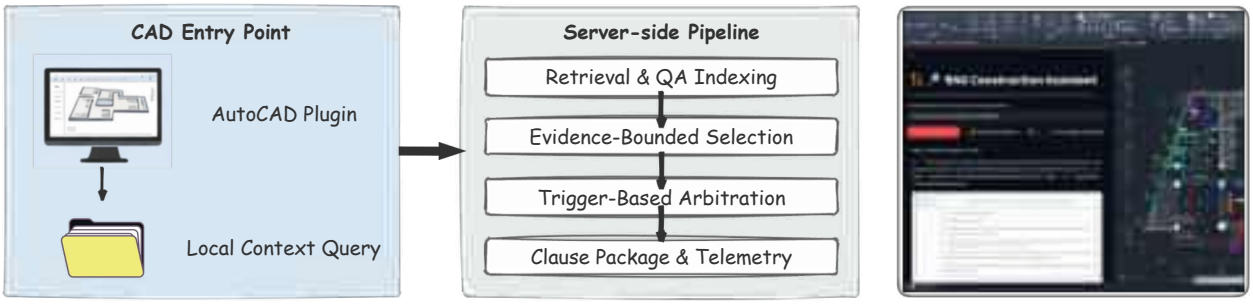


Figure 3. Within-CAD consultation flow reusing the server-side evidence pipeline

Table 4. Evidence-bounded selector results under a fixed top-10 candidate budget (dev200, $n = 200$). Cost is reported as USD per 1K queries.

Selector	Hit@1	MRR	Latency (s/query)	Cost
Qwen-plus	**0.430**	**0.523**	1.418	0.896
Qwen-max	0.395	0.501	2.498	2.944
GPT-4o	0.345	0.470	4.063	5.763
GPT-4o-mini	0.335	0.471	2.145	0.353

tion. Schema checks reject 1.0% of Qwen-plus outputs (2/200) due to out-of-candidate IDs, while other selectors have 0% invalid outputs; rejected cases are counted as misses. Table 4 compares API-based selectors under identical evidence inputs: Qwen-plus achieves Hit@1=0.430 (MRR=0.523) with 1.42 s/query and USD 0.896 per 1K queries, whereas GPT-4o achieves Hit@1=0.345 (MRR=0.470) with 4.06 s/query and USD 5.763 per 1K queries. Section 5.5 complements this fixed-point comparison with a post-hoc budget-feasibility audit of the archived Qwen-plus top-1 outputs under stricter prefix candidate pools.

5.4 Multi-Agent Arbitration Outcomes

Conditional arbitration is assessed on test100 to quantify when multi-agent expert disagreement indicates elevated ambiguity and how often arbitration revises majority selections under the fixed $k = 10$ evidence pool. Governance rates follow Formula 2 and Formula 5: arbitration triggers for 37% of queries and changes the majority top clause for 12% overall (32.4% among triggered cases). On test100, the main benefit concentrates on triggered queries: Hit@1 improves from 0.243 to 0.297 when arbitration is activated, while the overall gain is smaller (0.280 to 0.300) because non-triggered cases are already handled by the majority selector. Clause-ID disagreement accounts for all triggered cases (37/37 cases); applicability disagreement appears in 16 triggered cases as an additional signal, and no span-invalid triggers are observed in this benchmark (Table 5). Section 5.5 further audits how the archived final arbiter decisions remain feasible under stricter retrieval budgets.

Table 5. Arbitration statistics on test100 ($n = 100$). Trigger categories may overlap. Change is measured as whether the final arbiter-selected clause differs from the majority-voted top clause.

Split	Trig. rate	Change rate	Change|Trig.
Test100	37/100 (37.0%)	12/100 (12.0%)	12/37 (32.4%)

Split	Clause disagr.	Appl. disagr.	Span inv.
Test100	37	16	0

Notes: Trig.=arbiter invoked (computed from experts' top-1 clause outputs under the fixed $k = 10$ evidence pool); Change|Trig.=changes among triggered cases; Clause disagr.=experts select different clause IDs; Appl. disagr.=applicability conflict; Span inv.=missing/invalid evidence spans. Categories may overlap.

5.5 Retrieval-Budget Sensitivity and Post-hoc Feasibility Audit

To examine retrieval-depth sensitivity, Dense retrieval is analysed at $k \in \{1, 3, 5, 10\}$ using archived traces on dev200 and test100. Figure 4(a) shows that gold-in-C_k rises monotonically: on dev200, 0.550, 0.765, 0.850, and 0.935; on test100, 0.500, 0.820, 0.890, and 1.000 for $k = 1, 3, 5, 10$, respectively. The largest gain occurs between $k = 1$ and $k = 3$, with smaller gains thereafter, indicating diminishing marginal returns from deeper candidate pools.

Because the archived Qwen-plus selector and multi-agent arbitration runs in Tables 4 and 5 were executed at $k = 10$, downstream sensitivity is assessed through post-hoc feasibility audits under stricter prefix pools C_k. Figure 4(b) shows the results for the archived Qwen-plus run on dev200. Selected top-1 feasibility rises from 0.430 ($k = 1$) to 0.575 ($k = 3$), 0.660 ($k = 5$), and 0.740 ($k = 10$); all 86 correct top-1 decisions already correspond to retrieval rank 1. Figure 4(c) shows the results for the archived arbitration run on test100. The final arbiter-decision feasibility rises from 0.220 to 0.400, 0.620, and 0.880, while feasibility among the 30 correct final arbiter decisions rises from 43.3% to 60.0%, 83.3%, and 100.0%, respectively. Retrieval depth therefore increases reachability, but downstream gains are not linear in k: many correct selector and arbiter decisions are concentrated in highly

ranked candidates, whereas a smaller subset depends on deeper evidence pools.

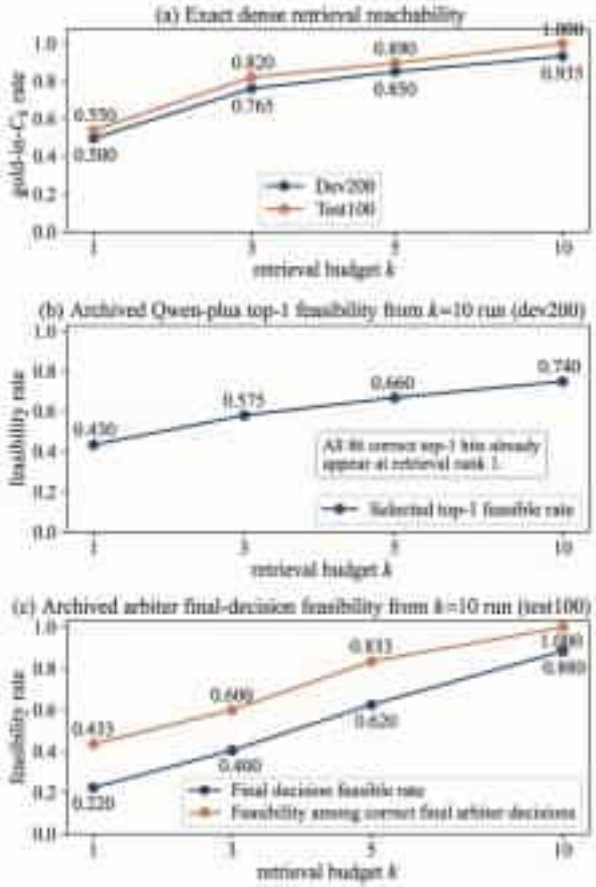


Figure 4. Retrieval-budget sensitivity and post-hoc feasibility audit over $k \in \{1, 3, 5, 10\}$. (a) Exact dense gold-in-C_k rate on dev200 and test100. (b) Post-hoc feasibility of archived Qwen-plus top-1 outputs from the $k = 10$ dev200 run under stricter prefix candidate pools. (c) Post-hoc feasibility of archived final arbiter decisions from the $k = 10$ test100 run under stricter prefix candidate pools.

5.6 Qualitative Case Studies

Two case studies illustrate auditable clause packaging and triggered arbitration under underspecified or cross-discipline conditions. Each reports the query, retrieved top-k clause IDs, the final clause package, and the arbitration trace when triggered; Table 6 provides the compact reporting template. Case B shows that arbitration corrects the majority choice under underspecified hazard-zone attributes by selecting an evidence-aligned clause within the retrieved pool, rather than a topically relevant but scope-mismatched alternative.

Table 6. Compact case study reporting for two representative queries on test100. Top-10 clause IDs are retrieved under a fixed candidate budget; the final clause package records the selected clause and an evidence span (translated excerpt).

Case A (no arbitration)	
Split / qid	Test100 / qid=12
Query	What is the minimum required clear width for a residential building corridor (passage)?
Retrieved top-10 clause IDs	`25429,3683,125838,76891,60520,60519,12735,11011,76499,37937`
Final clause package (ID + evidence span)	25429: "The clear width of a corridor passage shall not be less than 1.20 m."
Applicability (final)	yes
Trigger flags	–
Arbiter change (vs. majority top clause)	No (not triggered)
Case B (arbiter changed)	
Split / qid	Test100 / qid=230
Query	For Zone 20 in a hazardous (explosive) atmosphere, how should electrical equipment be selected per code requirements?
Retrieved top-10 clause IDs	`29097,71780,88188,122388,39587,71781,69799,101799,35034,39588`
Final clause package (ID + evidence span)	71781: "The explosion-proof class/group of equipment shall not be lower than that of the explosive gas mixture present in the environment."
Applicability (final)	yes
Trigger flags	clause disagreement; applicability disagreement
Arbiter change (vs. majority top clause)	Yes (71780→71781)

6 Discussion

Section 5 suggests that reliability in construction-code clause retrieval improves when evidence is bounded and disagreement is treated as a governance signal, rather than resolved through unconstrained generation. Dense retrieval raises the evidential ceiling by placing correct clauses earlier in C_k, while clause-level units keep verification tied to citeable clauses with stable identifiers. The retrieval-budget sensitivity analysis further qualifies the fixed-point results: reachability rises sharply from $k = 1$ to $k = 5$ and then slows, so $k = 10$ should be read as an operating point rather than a universal optimum. Post-hoc audits likewise show that downstream gains are not linear in k: many correct selector and arbiter decisions are already supported by highly ranked candidates, whereas a smaller subset depends on deeper evidence pools. The evidence-bounded selection contract keeps grounding mechanically checkable by requiring cited clause identifiers and evidence spans to remain traceable to retrieved text. For deployment, this supports a low-cost default selector with conditional arbitration for elevated-ambiguity cases. Residual failure sources remain missing scope attributes, long cross-references, and cross-edition drift. Reference-aware retrieval and lightweight attribute elicitation remain practical directions for improving robustness on tightly coupled multi-discipline queries.

7 Conclusion

This paper presented a clause-centric RAG pipeline for building construction code retrieval, where retrieval is treated as auditable evidence discovery and downstream models operate under an evidence-bounded clause-selection contract. By representing standards as clause-level evidence units with stable identifiers and provenance metadata, the system produces clause-addressable evidence bundles for rapid review and mechanical validation.

Experiments across multiple disciplines show that improving top-ranked evidence raises the ceiling for evidence-bounded selection, but that ceiling grows non-linearly with retrieval budget. Exact dense-retrieval sensitivity over $k \in \{1, 3, 5, 10\}$ shows strong reachability gains from $k = 1$ to $k = 5$, with smaller gains thereafter; post-hoc audits of archived $k = 10$ selector and arbitration outputs further show that many correct downstream decisions are already concentrated in highly ranked candidates, while a smaller subset depends on deeper evidence pools. Conditional multi-agent arbitration further serves as a governance layer by concentrating additional reasoning effort on high-risk cases identified through disagreement and evidence-quality triggers.

Together, these design choices support practical deployment under latency and cost constraints while keeping outputs traceable to explicit clause identifiers. Future work will focus on structured elicitation of missing project attributes and systematic handling of cross-standard cross-references for tightly coupled multi-discipline queries.

References

[1] Nicholas Nisbet, Zijing Zhang, and Selçuk Çidik. Real-time assessment of regulatory compliance of construction sites. In *EC3 Conference 2024*, volume 5, pages 0–0. European Council on Computing in Construction, 2024.

[2] Xiaorui Xue, Jiansong Zhang, and Yunfeng Chen. Question-answering framework for building codes using fine-tuned and distilled pre-trained transformer models. *Automation in Construction*, 168:105730, 2024.

[3] Jack Wei Lun Shi, Wawan Solihin, and Justin KW Yeoh. Fine-tuning a large language model for automated code compliance of building regulations. *Advanced Engineering Informatics*, 68:103676, 2025.

[4] Hao Yu, Aoran Gan, Kai Zhang, Shiwei Tong, Qi Liu, and Zhaofeng Liu. Evaluation of retrieval-augmented generation: A survey. In *CCF Conference on Big Data*, pages 102–120. Springer, 2024.

[5] Alsaffar Alhadi, Beach Dr Tom, and Rezgui Yacine. Enhancing asset management: Integrating digital twins for continuous permitting and compliance-a systematic literature review. *Journal of Building Engineering*, 99:111515, 2025.

[6] Thomas Beach, Jonathan Yeung, Nicholas Nisbet, and Yacine Rezgui. Digital approaches to construction compliance checking: Validating the suitability of an ecosystem approach to compliance checking. *Advanced Engineering Informatics*, 59:102288, 2024.

[7] Philipp Hagedorn, Judith Fauth, Sven Zentgraf, Sebastian Seiß, Markus König, and Ioannis Brilakis. Ontobpr: An ontology-based framework for performing building permit reviews using standardized information containers. *Advanced Engineering Informatics*, 66:103369, 2025.

[8] Stefan Fuchs, Michael Witbrock, Johannes Dimyadi, and Robert Amor. Using large language models for the interpretation of building regulations. *arXiv preprint arXiv:2407.21060*, 2024.

[9] Thomas Beach, Jonathan Yeung, Ali Ghoroghi, and Yacine Rezgui. Moving automated compliance checking to the operational phase of the building life-cycle: analysis and feasibility study in the uk. *International Journal of Construction Management*, 25(7):840–849, 2025.

[10] Odin Iversen and Lizhen Huang. Leveraging large language models for bim-based automated compliance checking. *Automation in Construction*, 182: 106707, 2026.

[11] Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024. URL `https://aclanthology.org/2024.emnlp-main.992/`.

[12] Hansi Hettiarachchi, Amna Dridi, Mohamed Medhat Gaber, Pouyan Parsafard, Nicoleta Bocaneala, Katja Breitenfelder, Gonçal Costa, Maria Hedblom, Mihaela Juganaru-Mathieu, Thamer Mecharnia, et al. Code-accord: A corpus of building regulatory data for rule generation towards automatic compliance checking. *Scientific data*, 12(1):170, 2025.

An automated framework to quantify exposure of informal urban settlements using UAV and Scan-to-BIM

Mirian Merma[1,2], Christopher Nuñez[1], Carlos Davila[2], Miguel Estrada[2], Marck Regalado[1], Raul Gonzalez[1], Kenny Requelme[1] and Angel Quesquen[2]

[1]Facultad de Ingenieria Civil, Universidad Nacional de Ingenieria, Lima 15333, Peru
[2]GeoGiRD Research Group, Facultad de Ingenieria Civil, Universidad Nacional de Ingenieria, Lima 15333, Peru

mirian.merma.o@uni.pe, christopher.nunez.v@uni.pe, carlos.davila.d@uni.pe, estrada@uni.edu.pe, marck.regalado.e@uni.pe, raul.gonzalez.o@uni.pe, kenny.requelme.c@uni.pe, aquesquenr@uni.pe

Abstract -

Rapid informal urban expansion in high-slope terrains presents a critical challenge for disaster risk management, particularly where foundational digital documentation is absent. This study proposes an automated framework for quantifying urban growth and geodynamic hazard exposure by integrating UAV photogrammetry with a Scan-to-BIM workflow. By processing high-resolution point clouds through noise reduction and normal vector calculation to segment horizontal roof planes, building footprints were extracted and converted into Industry Foundation Classes (IFC) models at a volumetric massing level using IfcOpenShell. Longitudinal analysis revealed a substantial rise in dwelling density across the study periods, with the total built-up area expanding significantly. Critically, spatial intersection with terrain models demonstrated that the most aggressive expansion occurred on the steepest slope categories, highlighting increased population exposure to rockslides according to national risk standards. While the study identifies limitations in detecting irregular geometries and unconventional materials, the proposed framework establishes a reproducible, interoperable BIM–GIS bridge that overcomes the temporal and spatial limitations of traditional monitoring in undocumented urban territories.

Keywords -

Urban Digitalization; UAV Photogrammetry; Scan-to-BIM; Informal Settlements

1 Introduction

The process of urbanization at the global level has experienced rapid growth over the last century. In 1990, only 15% of the urban population lived in urban areas [1]. However, this trend increased significantly during the 20th century. For example, in 2018, more than 58% of the world's population lived in urban areas. Current projections indicate that by 2050, more than approximately 75% of the world's population will live in urban areas [1].

In Peru, given that the formal housing supply has failed to meet demand and due to a lack of urban planning, informal settlements have become the main form of urban growth as massive migration from rural to urban areas [2].

Informal settlements refer to the organization of people in search of housing who occupy vacant lots and take collective action to resolve their urban and social organization problems themselves [3]. In cities such as Lima, the capital of Peru, these settlements have expanded into hillside areas, which are typically by unfavorable geographical conditions and, as a result, are exposed to natural hazards[4].

Currently, multitemporal analysis of urban growth based on remote sensing techniques, including 3D change detection [5], time series analysis of land use [6], and advanced approaches such as multitemporal mergers and segmentation using fuzzy logic [7], has become a high-level tool for monitoring informal urban growth at the regional and metropolitan scale using optical satellite and radar images. However, its application in dense informal contexts has some limitations.

One of the main limitations of using optical satellite images is cloud cover, which affects the acceptance of solar radiation reflected by the surface, reducing the quality of the spectral information available and limiting continuous multitemporal analysis. For example, in Lima, persistent cloud cover for several months of the year makes it difficult to capture clear and consistent images over time [2].

In addition, the construction characteristics of informal settlements may introduce additional challenges. The predominant materials used in housing roofs, such as concrete and dry wood, have spectral signatures similar to those of bare soil, which causes spectral confusion in classification processes and makes it difficult to accurately delimit building footprints. This spectral similarity, combined with high building density and the absence of separation between dwellings on a single block, makes it difficult to delimit building footprints. As a result, urban footprint maps created using spectral signatures have unsatisfactory levels of accuracy [8]. In contexts such as in Lima, characterized by desert conditions and exposed soil surfaces, the spectral similarity between housing roofs and soil is intensified [2], increasing the uncertainty of urban footprint extraction.

Although satellite radar images overcome some limitations associated with cloud cover and dependence on solar radiation, gaps remain in relation to spatial and temporal

scale and geometric distortions present in informal urban monitoring [4]. In particular, there is no consensus on the optimal spatial and temporal resolution needed to adequately capture rapid informal urbanization processes [9]. This limitation is evident in studies such as [4], which use Sentinel-1 SAR images to monitor informal growth in areas such as Morro Solar and Lomo de Corvina, where the analysis is based on a set of 14 images acquired between December 2020 and May 2021. This temporal resolution is insufficient to capture rapid changes in real time, while the spatial resolution of 10 m restricts the analysis to estimating occupied areas, without allowing for direct quantification of the number of dwellings.

On the other hand, in informal contexts, the availability of BIM models is practically non-existent, and creating them manually is tedious, costly, and time-consuming [10]. However, generating BIM models from point clouds in these environments can be highly beneficial for gathering information on current conditions, designing interventions, communicating proposals, and raising awareness among residents about environmental issues [11]. Studies such as [12] have shown that the integration of BIM models, visualized on Geographic Information System (GIS)-based platforms, improves the efficiency of urban monitoring and enables applications such as risk assessment in buildings and roads, as well as the development of early warning systems.

In this context, there is a clear need for alternative approaches that allow high-resolution spatial information to be captured with greater temporal flexibility, capable of characterizing the configuration of informal settlements and generating BIM models, particularly those located on slopes exposed to natural hazards.

To address this gap, this study proposes an automated framework for digitizing informal urban environments, integrating data acquired using Unmanned Aerial Vehicle (UAV) with Scan-to-BIM techniques. This approach allows for the automatic extraction of building footprints from point clouds, the generation of BIM models, and the multitemporal quantification of informal housing across different analysis periods. The methodology was applied in hillside areas exposed to rockfalls and landslides, enabling the assessment of informal urban growth and its relationship with the associated hazards, and providing detailed spatial information that can support urban planning and disaster risk management processes. Furthermore, the integration of BIM and GIS provides conceptual support for the representation and analysis of multiscale urban information, aligned with the principles of City Information Modeling and smart cities [13].

2 Related Work

Recent studies have explored the use of data acquired through UAVs and photogrammetric techniques for monitoring informal buildings in urban environments. In [14, 15], the use of oblique photogrammetry with drones was proposed to generate three-dimensional models of the urban environment, allowing the creation of Digital Surface Models (DSM), orthophotographic maps, and 3D models. By analyzing differences between DSMs, they were able to automatically extract and quantitatively analyze informal constructions. Similarly, in [16], a methodology was developed for detecting urban changes by combining differences between dense DSMs and true orthophoto maps (TDOM) generated from aerial images. This approach made it possible to identify changes in both floor plan and height, overcoming the limitations of conventional methods through the use of advanced techniques such as Semi-Global Matching (SGM) and Multi-View Stereo (MVS), and demonstrating high spatial accuracy at the regional scale.

On the other hand, various studies have used synthetic aperture radar (SAR) images and multitemporal approaches for urban growth analysis. In [17], a methodology was presented for the automatic detection of urban changes based on the integration of SAR time series and multispectral data, evaluating the performance of different optical and radar sensors. In the Peruvian context, [2] used SAR images acquired by the Sentinel-1 constellation since 2017 to identify concentrations of urban growth in the Lima metropolitan area, complementing the analysis with high-resolution optical images for the visual identification of informal settlements. Complementarily, in [3], time series of Sentinel-1 SAR images were applied to map the extent of recent occupations in areas with unfavorable soil conditions in the face of seismic events, demonstrating the potential of radar for urban monitoring in vulnerable contexts.

Although these approaches have proven effective for monitoring and detecting informal urban growth, most studies focus on two-dimensional analyses or identifying general patterns of urban expansion. They also have limitations in terms of the precise delimitation and individual quantification of informal dwellings, as well as the automated generation of three-dimensional models that can be integrated into BIM environments, particularly in hillside areas exposed to the risk of rockfalls and landslides.

3 Methodology

Figure 1 illustrates the proposed methodology, which consists of three main components. First, UAV technology is used to efficiently acquire high-resolution data across different analysis periods. These images are processed us-

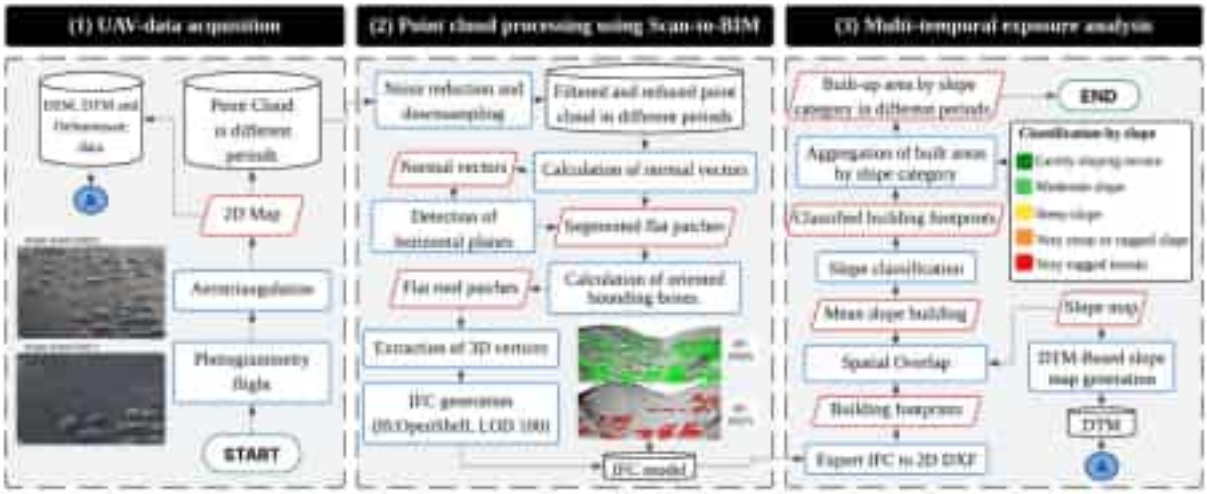


Figure 1. Proposed framework for quantifying exposure in informal urban settlements using UAVs and Scan-to-BIM.

ing photogrammetric techniques to generate point clouds, orthomosaics, and Digital Terrain Models (DTM), which form the geometric and topographic basis of the study.

Second, the generated point clouds are processed using an automated Scan-to-BIM workflow. This process includes noise reduction, normal vector calculation, and the detection of horizontal planes associated with the roofs of the houses. From these geometries, the footprints of the buildings are extracted and BIM models are generated in IFC format using IfcOpenShell, allowing for the digital representation of the houses in the absence of previous documentation.

Finally, the multitemporal analysis is carried out. The IFC models are exported to DXF format to obtain the footprints of the dwellings in a two-dimensional environment, which are intersected with slope maps derived from the DTM. Each footprint is assigned an average slope value, which is then classified into the five ranges of slope established by the National Center for Disaster Risk Estimation, Prevention and Reduction (CENEPRED) of Peru. The aggregation of areas by slope level for each period allows for a temporal comparison of urban occupation based on its exposure. This classification characterizes the terrain's susceptibility to geodynamic hazards, such as rockslides [4] . Finally, the aggregation of residential areas by slope level for each analysis period allows for a temporal comparison of urban occupation, assessing its exposure to hazards associated with highly susceptible terrain.

3.1 Study area

The district of Comas (Figure 2b), located in the province of Lima (Figure 2a), department of Lima, Peru, covers a total area of 48.75 km2 and has 14 zones and 32 urbanizations, with a population density of 9,989.3 inhabitants per square kilometer [18] (see Fig. 2). This study focuses on the Año Nuevo settlement, Zone A (Figure 2c), and covers an area of approximately 0.7 km² [4].

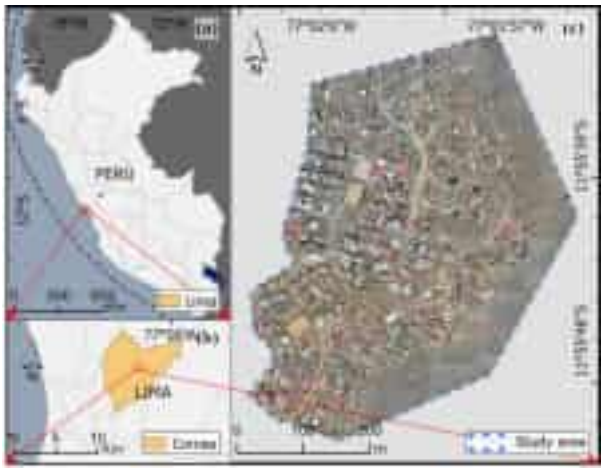


Figure 2. Location map of the study area.

3.2 UAV data acquisition

Data acquisition for the generation of point clouds and 2D maps was performed using photogrammetry for the periods of 2017 and 2025. The flight parameters are pre-

sented in Table 1, which summarizes the specifications of the UAV flights implemented.

Table 1. Technical specifications of UAV flights.

Data	2017 Flight	2025 Flight (%)
UAV model	DJI Phantom 3	DJI Mavic 3 Enterprise (M3E)
Sensor	Integrated RGB camera, FC300X model	High-resolution integrated RGB camera
Camera resolution	12 MP (4000 × 3000 px)	20 MP (5280 × 3956 px)
GSD	7.7 cm/pixel	4.5 cm/pixel

3.3 Scan-to-BIM workflow

The photogrammetric data was processed using DJI Terra 5.1.1, from which the main photogrammetric products were generated, including the DSM, DTM, orthomosaic, and the dense point cloud. Based on the generated point cloud, the Scan-to-BIM methodology was used to produce as-built BIM models of existing buildings, enabling digital documentation and geometric representation of the built environment in areas where no previous BIM information existed [19, 20].

3.3.1 Point cloud preprocessing

The raw datasets for 2017 and 2025 were subjected to noise reduction and spatial downsampling at a 0.05 m resolution. This resolution was selected as a compromise between computational efficiency and the preservation of geometric features relevant for roof segmentation, as finer resolutions did not significantly improve segmentation results while substantially increasing processing cost. This preprocessing balanced computational efficiency with geometric fidelity, significantly reducing the point density while preserving the structural features necessary for morphological analysis (see Tab. 2).

Table 2. Point cloud preprocessing parameters for the 2017 and 2025 datasets.

Dataset	Raw Points	Processed Points	Reduction (%)
2017	30,313,508	11,413,767	62.3%
2025	41,380,614	11,170,087	73.0%

3.3.2 Automated roof detection and segmentation

Roof surfaces were identified by detecting horizontal planes within the processed point clouds. Using the Open3D library, surface normals were estimated via KD-tree–based nearest-neighbor searches. Planar patches were segmented using the Oriented Bounding Box (OBB) method, supported by covariance analysis and local density criteria.

This simplification was guided by the study area, where roofs are predominantly flat or near-horizontal, making them the most reliable feature for building detection. While effective in this context, it may limit applicability in areas with more complex roof geometries. This process identified 1,659 horizontal planes for the 2017 dataset and 1,908 for 2025, forming the geometric basis for the building models.

3.3.3 Semantic alignment and interoperability strategy

The segmented point cloud data were transformed into a structured digital environment by extracting the 3D vertices of detected roof planes to define individual building contours. These geometries were converted into parametric BIM entities at LOD 100, representing the general volumetric massing of the dwellings. The reconstruction process was automated using IfcOpenShell, enabling a reproducible workflow for the generation of longitudinal models for 2017 and 2025.

The workflow was designed to comply with the FAIR principles for geospatial data [21] and to support BIM–GIS integration. The use of IFC as a native data structure enables direct integration with IfcOpenShell for geometry generation, attribute handling, and standardized data exchange, ensuring reproducibility and compatibility with established BIM processing tools [22, 23]. Although each dwelling corresponds conceptually to a single building, the geometry was instantiated as IfcSpace entities as a pragmatic modeling simplification within the implemented workflow. We acknowledge that this choice does not comply with IFC spatial hierarchy conventions, and that each entity should be assigned to an IfcBuilding according to IFC 4.3.

To reduce reprojection errors and information loss during IFC–CityGML transformations, global georeferencing was embedded directly within the IFC output [23]. By adopting LOD 100, the methodology supports efficient processing of high-density informal settlements while enabling standardized and interoperable BIM model generation despite irregular urban morphologies [24, 25].

This approach demonstrates that the absence of prior documentation or the irregular morphology of the built environment does not preclude the creation of standardized, interoperable BIM models [26].

3.3.4 Multitemporal Analysis

Multitemporal analysis was performed to evaluate the evolution of urban occupation in hillside areas between 2017 and 2025, based on the relationship between built-up areas and terrain slope. The main input used was the footprints of the dwellings extracted from the IFC models generated using the Scan-to-BIM methodology, which made it possible to represent urban occupation in a homogeneous environment for both periods.

The footprints of the buildings were integrated with slope maps obtained from the digital terrain model (DTM), allowing each built unit to be assigned a representative average slope value using zonal statistical methods. These values were then classified into the five slope levels established by CENEPRED. Finally, the built-up areas were aggregated by slope level for each period of analysis, enabling a temporal comparison of urban growth based on the topographical conditions of the terrain.

3.3.5 Quantitative Accuracy Evaluation

Numerical validation was performed by comparing the extracted features with a manual ground truth using pixel-by-pixel confusion matrices. The 2025 results show solid performance (precision: 0.919; F1 score: 0.703), thanks to the high resolution and precision of the DJI Mavic 3 drone camera, which minimizes noise. In 2017, although precision was 0.900, the F1 score dropped to 0.384. This lower precision is attributed to a more dispersed and noisy point cloud, resulting from lower-resolution sensors and a higher flight altitude, which, combined with unconventional surface materials, increased false negatives. This evaluation quantifies the reliability and limitations of the proposed automation system in the face of variations in the density and accuracy of the source data.

Table 3. Performance metrics for automated building footprint extraction.

Year	Raw Accuracy	Precision Points	Recall	F1 Score (%)
2017	0.9	0.432	0.345	38.4
2025	0.919	0.578	0.896	70.3

4 Results and discussions

The Scan-to-BIM workflow successfully transformed spatial point cloud data into georeferenced, interoperable BIM models in IFC format. By extracting building footprints and altitudes from detected horizontal planes (roofs), volumetric representations of the informal settlement were generated for the years 2017 and 2025.

4.1 Longitudinal modeling analysis

The automated reconstruction process identified a clear increase in the number of reconstructed dwellings over the eight-year study period.

- 2017 Dataset: Around 1000 dwellings were reconstructed as parametric IfcSpace entities, representing the spatial distribution and volumetric configuration of the settlement at the baseline year (Fig. 3).
- 2025 Dataset: Using identical reconstruction parameters, almost 1600 dwellings were generated. This corresponds to an increase of 682 dwellings, or around 60%, relative to 2017, providing a quantitative indicator of built-up expansion and informal growth within the study area (Fig. 4).



Figure 3. Reconstructed dwelling distribution for the 2017 dataset.

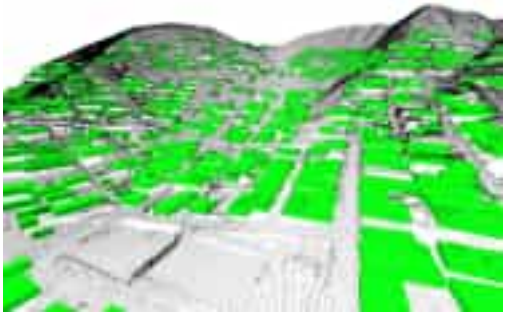


Figure 4. Reconstructed dwelling distribution for the 2025 dataset.

4.2 Multitemporal analysis

The multitemporal analysis made it possible to generate thematic maps of urban occupation for the years 2017 and 2025, as shown in Figures 5 and 6, which represent the footprints of buildings within the study area, each footprint associated with an average slope value and classified into five levels according to the ranges established by

CENEPRED, which allowed for the spatial visualization of the area where the dwellings are located in relation to the topographical conditions of the terrain.

In addition, summary tables were prepared that quantify the total area occupied by dwellings for each slope level and for each analysis period, as shown in table 4. These results made it possible to compare urban occupation between the two years, revealing differences in the spatial distribution of built-up areas according to slope level.

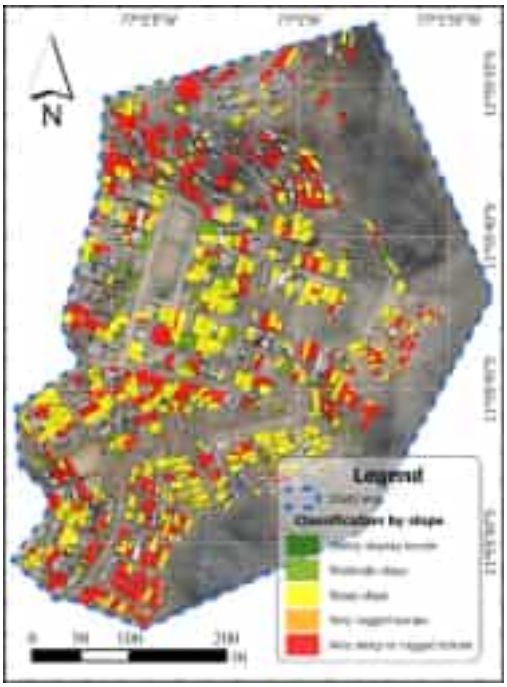


Figure 5. Map of building footprints in 2017 classified by terrain slope levels.

Table 4. Distribution of built-up area by terrain slope level, 2017 and 2025.

Classification by slope	**Slope**	**Area 2017** (m^2)	**Area 2025** (m^2)
	< 5º	173,75	416,25
	5º - 15º	4.005,75	6.287,25
	15º - 25º	21.737,25	21.813,75
	25º - 45º	76,25	924,75
	> 45º	18.766,75	33.142,75
Total area (m^2)		44.759,75	62.584,75

5 Conclusions

This study proposes an automated framework for quantifying the exposure of informal urban settlements over

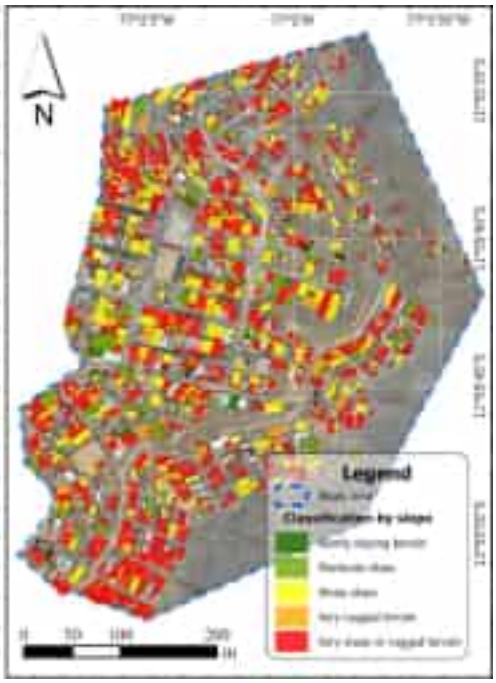


Figure 6. Map of building footprints in 2025 classified by terrain slope levels.

different time periods using UAVs and Scan-to-BIM. The results suggest that the proposed approach allows for: (1) automated digitization of building footprints, (2) generation of BIM models in LOD 100 and their export in IFC format, and (3) multitemporal analysis of urban growth and exposure to hazards. In this sense, the methodology reduces some of the limitations associated with the spatial and temporal resolution of traditional satellite image-based approaches, by providing high-resolution spatial data with greater temporal flexibility. It also minimizes the need for manual and repetitive processing of point cloud data and geospatial analysis.

Despite these advantages, the study also highlights limitations related to the accurate detection of informal dwellings, especially in cases where roof geometries are not flat or are irregular. The presence of unconventional roofing materials, such as corrugated iron and wood, as well as the lack of minimum separation between dwellings, affected the correct delimitation of building footprints. In addition, high noise levels in the point cloud led to the misclassification of surfaces that do not correspond to dwellings, which influenced the accuracy of urban occupation estimates. In this context, it is suggested that future research incorporate automated tools for the detection of non-flat surfaces, as well as methods for automatic classification and noise reduction in the point cloud.

References

[1] Naledzani Mudau and Paidamwoyo Mhangara. Mapping and assessment of housing informality using object-based image analysis: A review. *Urban Science*, 7(3), 2023. ISSN 2413-8851. doi:10.3390/urbansci7030098. URL https://www.mdpi.com/2413-8851/7/3/98.

[2] Luis Moya, Marta Vilela, Javier Jaimes, Briggite Espinoza, Jose Pajuelo, Nicola Tarque, Sandra Santa-Cruz, Pablo Vega-Centeno, and Fumio Yamazaki. Vulnerabilities and exposure of recent informal urban areas in lima, peru. *Progress in Disaster Science*, 23:100345, 2024. ISSN 2590-0617. doi:https://doi.org/10.1016/j.pdisas.2024.100345. URL https://www.sciencedirect.com/science/article/pii/S2590061724000358.

[3] L. Moya, F. Garcia, C. Gonzales, M. Diaz, C. Zavala, M. Estrada, F. Yamazaki, S. Koshimura, E. Mas, and B. Adriano. Brief communication: Radar images for monitoring informal urban settlements in vulnerable zones in lima, peru. *Natural Hazards and Earth System Sciences*, 22(1):65–70, 2022. doi:10.5194/nhess-22-65-2022. URL https://nhess.copernicus.org/articles/22/65/2022/.

[4] Frank Angel Mallma Acuña. Herramientas geomáticas para la evaluación de zonas urbanas amenazadas por deslizamiento de material en laderas, 2019. URL https://hdl.handle.net/20.500.14138/2005.

[5] Wen Xiao, Hui Cao, Miao Tang, Zhenchao Zhang, and Nengcheng Chen. 3d urban object change detection from aerial and terrestrial point clouds: A review. *International Journal of Applied Earth Observation and Geoinformation*, 118:103258, 2023. ISSN 1569-8432. doi:https://doi.org/10.1016/j.jag.2023.103258. URL https://www.sciencedirect.com/science/article/pii/S1569843223000808.

[6] Peng Dou and Zhen Han. Quantifying land use/land cover change and urban expansion in dongguan, china, from 1987 to 2020. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:201–209, 2022. doi:10.1109/JSTARS.2021.3133703.

[7] Ahmed Alejandro Cardona-Mesa, Rubén Darío Vásquez-Salazar, Juan Camilo Parra, César Olmos-Severiche, Carlos M. Travieso-González, and Luis Gómez. A methodology for the multitemporal analysis of land cover changes and urban expansion using synthetic aperture radar (sar) imagery: A case study of the aburrá valley in colombia. *Remote Sensing*, 17(3), 2025. ISSN 2072-4292. doi:10.3390/rs17030554. URL https://www.mdpi.com/2072-4292/17/3/554.

[8] Gonzalo Rodolfo Peña Zamalloa, Yumin Tan, and Lingfeng He. Modelling human settlement growth and fringe patterns in the andes through remote sensing and deep learning. *Cities*, 170:106660, 2026. ISSN 0264-2751. doi:https://doi.org/10.1016/j.cities.2025.106660. URL https://www.sciencedirect.com/science/article/pii/S0264275125009631.

[9] Gundegmaa Chinzorig, Md. Omar Sarif, and Ayyoob Sharifi. An analysis of the effectiveness of urban master plans in controlling and regulating urban growth in ulaanbaatar, mongolia. *Cities*, 170:106713, 2026. ISSN 0264-2751. doi:https://doi.org/10.1016/j.cities.2025.106713. URL https://www.sciencedirect.com/science/article/pii/S0264275125010169.

[10] T. Zhou, L. Lv, J. Liu, and J. Wan. Application of uav oblique photography in real scene 3d modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021:413–418, 2021. doi:10.5194/isprs-archives-XLIII-B2-2021-413-2021. URL https://isprs-archives.copernicus.org/articles/XLIII-B2-2021/413/2021/.

[11] Victoria Rautenbach, Yvette Bevis, Serena Coetzee, and Carin Combrinck. Evaluating procedural modelling for 3d models of informal settlements in urban design activities. *South African Journal of Science*, 111(11/12):10, Nov. 2015. doi:10.17159/sajs.2015/20150100. URL https://sajs.co.za/article/view/3804.

[12] C.-H. Yang, C. Stemmler, N. Wolf, M. Koppe, T. Rudolph, and A. Müterthies. Fusion of bim and sar for innovative monitoring of urban movement – towards 4d digital twin. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-G-2025:1627–1632, 2025. doi:10.5194/isprs-archives-XLVIII-G-2025-1627-2025. URL https://isprs-archives.copernicus.org/articles/XLVIII-G-2025/1627/2025/.

[13] Jianyong Shi, Zeyu Pan, Liu Jiang, and Xiaohui Zhai. An ontology-based methodology to establish city information model of digital twin city by

merging bim, gis and iot. *Advanced Engineering Informatics*, 57:102114, 2023. ISSN 1474-0346. doi:https://doi.org/10.1016/j.aei.2023.102114. URL https://www.sciencedirect.com/science/article/pii/S1474034623002422.

[14] T. Zhou, L. Lv, J. Liu, and J. Wan. Application of uav oblique photography in real scene 3d modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021:413–418, 2021. doi:10.5194/isprs-archives-XLIII-B2-2021-413-2021. URL https://isprs-archives.copernicus.org/articles/XLIII-B2-2021/413/2021/.

[15] Carlos Davila, Angel Quesquen, Fernando Garcia, Brigitte Puchoc, Oscar Solis, Julian Palacios, Jorge Morales, and Miguel Estrada. Assessment of building vulnerability to tsunami in ancon bay, peru, using high-resolution unmanned aerial vehicle imagery and numerical simulation. *Drones*, 9(6), 2025. ISSN 2504-446X. doi:10.3390/drones9060402. URL https://www.mdpi.com/2504-446X/9/6/402.

[16] David Marzi, Erith Munoz Rios, Antonietta Sorriso, Fabio Dell'Acqua, and Paolo Gamba. Joint multitemporal sar and optical mapping of urban changes. *International Journal of Microwave and Wireless Technologies*, 16(3):359–370, 2024. doi:10.1017/S175907872400031X.

[17] X. Zhu, G. Pang, P. Chen, Y. Tao, Y. Zhang, and X. Zuo. Research on urban construction land change detection method based on dense dsm and tdom of aerial images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W10:205–210, 2020. doi:10.5194/isprs-archives-XLII-3-W10-205-2020. URL https://isprs-archives.copernicus.org/articles/XLII-3-W10/205/2020/.

[18] Municipalidad de Comas. Geografía, s. f. URL https://www.municomas.gob.pe/distrito/geografia. Consultado el 16 de enero de 2026.

[19] Sebastian Ochmann, Richard Vock, and Reinhard Klein. Automatic reconstruction of fully volumetric 3d building models from oriented point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 151:251–262, 2019. ISSN 0924-2716. doi:https://doi.org/10.1016/j.isprsjprs.2019.03.017. URL https://www.sciencedirect.com/science/article/pii/S0924271619300863.

[20] Chao Wang, Yong K. Cho, and Changwan Kim. Automatic bim component extraction from point clouds of existing buildings for sustainability applications. *Automation in Construction*, 56:1–13, 2015. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2015.04.001. URL https://www.sciencedirect.com/science/article/pii/S0926580515000734.

[21] Linda van den Brink. *Geospatial Data on the Web*. Dissertation (tu delft), Delft University of Technology, 2018.

[22] Ying Liang and Yi Tan. A systematic literature review of ifc-to-citygml conversion. In Dezhi Li, Patrick X. W. Zou, Jingfeng Yuan, Qian Wang, and Yi Peng, editors, *Proceedings of the 28th International Symposium on Advancement of Construction Management and Real Estate*, pages 1265–1278, Singapore, 2024. Springer Nature Singapore. ISBN 978-981-97-1949-5.

[23] Mohammed Sani and Alias Rahman. Gis and bim integration at data level: A review. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W9:299–306, 10 2018. doi:10.5194/isprs-archives-XLII-4-W9-299-2018.

[24] F. Bianconi, M. Filippucci, F. Cornacchini, and A. Migliosi. The impact of google's apis on landscape virtual representation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W9-2024:91–98, 2024. doi:10.5194/isprs-archives-XLVIII-4-W9-2024-91-2024. URL https://isprs-archives.copernicus.org/articles/XLVIII-4-W9-2024/91/2024/.

[25] Edoardo De Santis and Pieter Pauwels. Actor-based framework for interoperable bim data management. In *Proceedings of the 2025 European Conference on Computing in Construction*, volume 6 of *Computing in Construction*, Porto, Portugal, July 2025. European Council on Computing in Construction. ISBN 978-9-083451-31-2. doi:10.35490/EC3.2025.314. URL https://ec-3.org/publications/conference/paper/?id=EC32025_314.

[26] Nuno Abreu, Andry Pinto, Aníbal Matos, and Miguel Pires. Procedural point cloud modelling in scan-to-bim and scan-vs-bim applications: A review. *ISPRS International Journal of Geo-Information*, 12(7), 2023. ISSN 2220-9964. doi:10.3390/ijgi12070260. URL https://www.mdpi.com/2220-9964/12/7/260.

Smart Bridge Sensing for Automated Vehicle Weight Estimation in Complex Traffic Scenarios

Dawid Piotrowski[1], Marcin Jasiński[1], Artur Nowoświat[1], Piotr Łaziński[1], Qian Chen[2], Tony Yang[2]

[1]Faculty of Civil Engineering, Silesian University of Technology, Poland
[2]Faculty of Applied Science, The University of British Columbia, Canada
dawid.piotrowski@polsl.pl, marcin.jasinski@polsl.pl, artur.nowoswiat@polsl.pl,
piotr.lazinski@polsl.pl, qian.chen@ubc.ca, yang@civil.ubc.ca

Abstract

Structural Health Monitoring (SHM) systems installed on bridges are increasingly being used to analyze traffic parameters. This includes estimating the gross vehicle weight and may be referred to as Weigh-in-Motion (WIM) methods. However, current research relies predominantly on numerical analyzes, whereas available field experiments are typically performed under quasi-controlled conditions. Such studies often assume that a single vehicle is moving along the predefined path at a known speed. In contrast, real-world traffic is inherently non-deterministic, often making it impossible to precisely define all initial conditions. This article presents the SHM system installed on a network arch bridge in Wolin, Poland. Strains induced by passing vehicles are measured using fiber-optic sensors based on Fiber Bragg Grating (FBG) technology. A neural network (NN) algorithm, whose architecture was optimized using Particle Swarm Optimization (PSO), is then used to estimate vehicle weight, with particular emphasis placed on robustness to traffic-related uncertainties. The system directly incorporates the structural specificity of the bridge and involves multi-gate weight estimations by strain values recorded by a minimal number of sensors. To evaluate the performance of the proposed approach, various tests were performed using real-world traffic experiments, involving multiple vehicles and varying speed. The system achieved a mean absolute percentage error of 4.6%, also for several vehicles on the bridge, complex traffic scenarios, and dynamic effects included.

Keywords –

Structural Health Monitoring; Bridge Weigh-in-Motion; Machine Learning

1 Introduction

Structural Health Monitoring (SHM) is currently one of the fastest growing fields, combining engineering, science, and information technology. These systems integrate various measurement techniques and are widely applied to different types of structures. In bridge engineering, SHM has been increasingly used to estimate road traffic load in the form of Weigh-In-Motion (WIM) systems. It should address the issues associated with the non-deterministic nature, uncertain positioning, and variety of traffic scenarios. The available studies in the literature confirm that WIM is still evolving and remains an important area of industrial research [1].

However, existing WIM applications are subject to several limitations. A significant part of studies are based solely on numerical analyzes [2], while available field experiments are often conducted under quasi-controlled or laboratory conditions [3], [4]. Consequently, this study extends the WIM method, demonstrating the possibility of utilizing the structural specificity of the bridge and allowing the number of sensors required to be reduced. Furthermore, by using three independent measurement gates, the proposed method enables the determination of vehicle weight under unconstrained traffic conditions. The presented approach is resistant to random impact of multiple vehicles and dynamic effects while maintaining a high level of accuracy.

2 Literature review

In many cases, experiments assume the exact position and speed of the vehicle [5], [6]. In reality, traffic flow is inherently random, and therefore the results of such studies may not be fully representative of real-world traffic conditions. For example, Brzozowski et al. [7] investigated the influence of various external factors on the accuracy of weight estimates. Their results show that the vehicle traffic path is one of the most significant contributors to measurement error. For lateral load displacements of up to 50 cm, the total weight prediction

error was approximately 10%, while for displacements exceeding 170 cm, the error increased dramatically to almost 30%.

Furthermore, many current WIM solutions require the definition of specific preconditions [8],[9], distributed measurements, or a large number of sensors, especially the influence line-based applications [10]-[13]. Although a higher level of accuracy can be achieved, the requirement to define input parameters such as the number, spacing of axles, or vehicle speed is often difficult to satisfy in real-world traffic. Moghadam [14] presented an example of the Nothing-on-Road (NOR) WIM application, which assumes the time delay between the strain peaks caused by the traversing axles. The study was conducted using numerical data under controlled conditions. Similarly, Wang et al. [15] proposed a method to estimate vehicle parameters using the vehicle-bridge interaction. In this approach, additional input data was defined in the form of roughness of the road surface, which was estimated using a particle filter. Moreover, the analysis was limited to two-axle vehicles and scenarios involving only a single vehicle. Under these assumptions, the maximum weight estimation error was 11.9%.

Consequently, WIM methods still lack experimental validation in which vehicle position, speed, and traffic composition are not known a priori. The method shown in this paper estimates vehicle weight in unconstrained traffic using three independent measurement gates and a limited number of sensors, while remaining accuracy in random multi-vehicle scenarios.

3 Methodology

3.1 Bridge case study description

This study introduces an SHM system installed on a network arch bridge in Wolin, Poland, in which one of the objectives is the detection of the gross weight of vehicles. The bridge analyzed is a steel structure with a theoretical span of 165.0 m. Its total width is 16.9 m and the height of the arch is 24.0 m (Figure 1).

3.2 System description

The system consists of a network of sensors in which data is acquired and processed using Machine Learning (ML) algorithms. The specificity of the bridge influenced the selection of the sensors, as the most significant feature of the structure is the local nature of the influence area of the crossbeams. Consequently, the increase in the strain in the crossbeams is specifically caused by vehicles located in their immediate vicinity. This ensures that the readings are not affected by vehicles farther away. To improve the robustness of the system and reduce the probability of two vehicles moving side by side in two lanes at the same, three consecutive measurement gates, each consisting of three sensors, are designed, located at 1/4, 1/2, and 3/4 of the span. This configuration allows for estimation of vehicle weight by three independent measurements recorded in different places along the bridge. The network is supplemented with two additional sensors located in the middle of the tie-beams. (Figure 2).



Figure 1. Analyzed arch bridge in Wolin, Poland

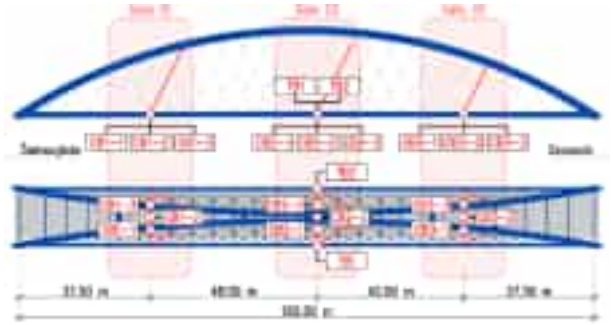


Figure 2. Distribution of the sensors on the bridge

The sensors were spot-welded directly to the surface of the element after the paint layer had been ground off. The sensors on the crossbeams are placed on the top surface of the bottom flange, whereas those on the tie-beams are installed underside of the section. The sensors were subsequently protected against corrosion using a non-hardening, high-adhesion masking sealant. Finally, they were covered with a 0.05 mm thick aluminum foil coated with a 3 mm layer of kneadable putty (Figure 3). The system enables synchronous acquisition of data from all channels allowing for time-correlation analysis.

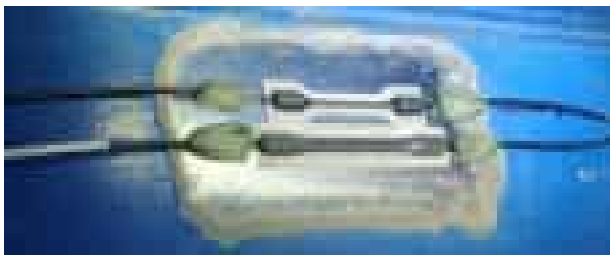


Figure 3. Example configuration of the sensor

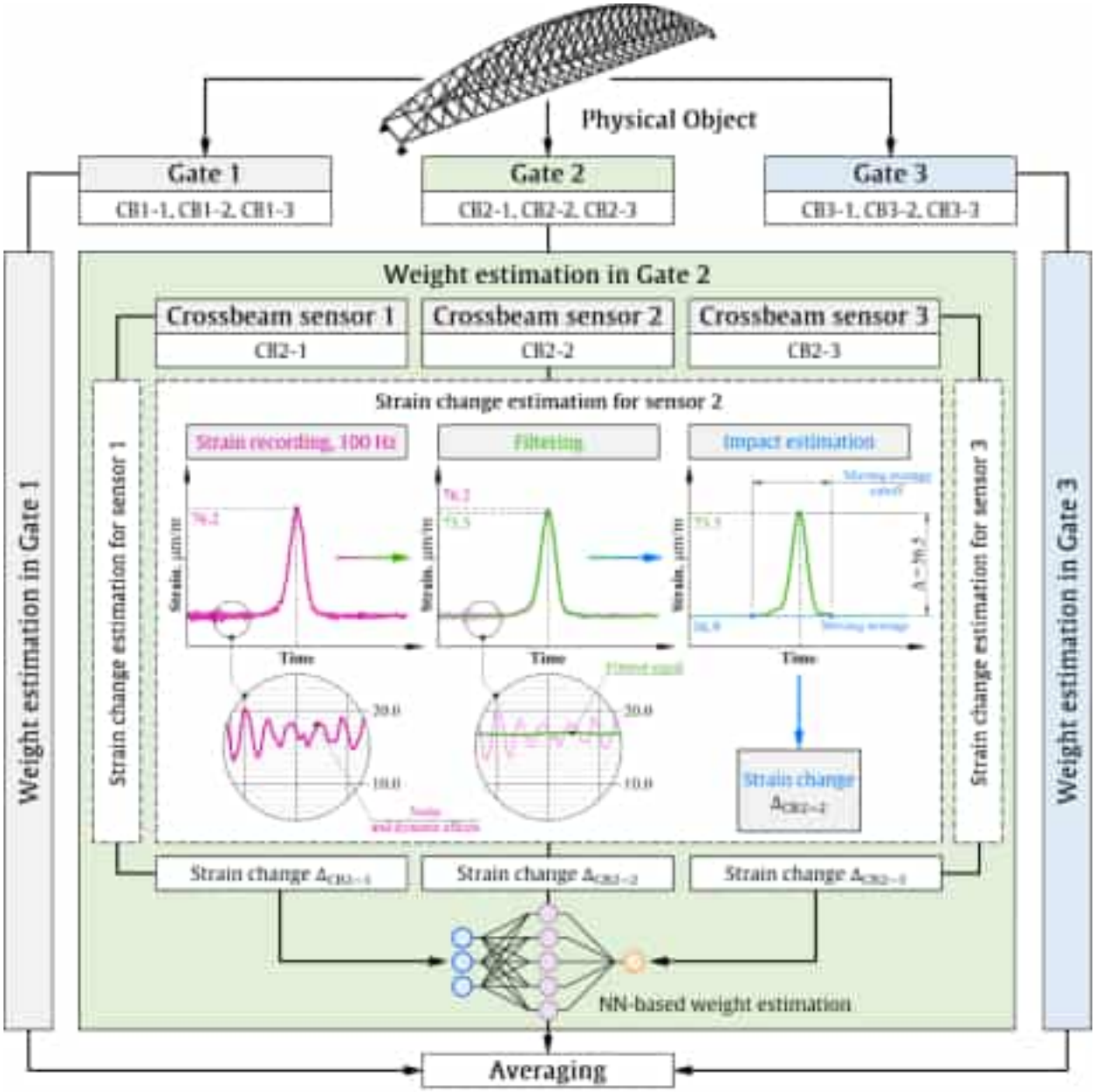


Figure 4. Overview of the proposed weight estimation methodology

The overview of the weight estimation methodology is shown in Figure 4. Each of the nine sensors records the strain with a frequency of 100 Hz. The time-series data are processed by an Infinite Impulse Response (IIR) Butterworth filter. It applies the low-pass design and processes the signal forward and backward to achieve zero-phase distortion. Using the processed signal, the moving average is calculated for each of the sensors independently, excluding outliers that usually indicate the beginning of the load impact. The moving average serves as a base against which the maximum strain value is compared to estimate the strain change due to vehicular load. The strain increases recorded by the sensors are grouped by timestamp and gate and, as such, are passed to the neural networks for weight estimation.

The proposed design incorporates three independent neural networks, each assigned to the specific gate and consisting of three input units. The networks are trained using data obtained through Finite Element (FE) simulation. An important part of the current procedure is an updated bridge FE model, calibrated on the basis of the static load test results. It involved minimization of the difference between the measured and simulated response, including strains and deflections due to a known and precisely arranged load.

The procedure of the FE model calibration involved the results of static load tests with nine trucks positioned in various scenarios, including three loading schemes above each of the measurement gates, individually, and two asymmetrical cases with trucks aligned to the left and right edges of the road. A Genetic Algorithm (GA) was applied, with the objective of minimizing the difference between the strains measured by the SHM system and those predicted by the FE model. The GA consisted of 20 generations of 150 candidates and reached convergence in the last generations. The input parameters of the GA were the elastic moduli of all structural elements with the assumed limit values of 160 – 240 GPa for steel and 5 – 45 GPa for concrete elements. Although the elastic moduli boundaries include values that exceed the standard values for structural steel and concrete, it was assumed that they should reflect different factors that affect the overall stiffness. The effects in question may include, but are not limited to, assembly clearances, inaccuracies in geometrical, material, and structural properties, and stiffness of the non-structural elements like pavement, caps, curbs, barriers or railings.

3.3 Optimization of the NN architecture

Choosing an appropriate neural network architecture is crucial to achieving high accuracy in estimating the gross weight of vehicles. This process involves defining the number of layers, the number of neurons in each layer, the activation function, and the learning rate, and can be formulated as a classical optimization task.

One of the most widely used optimization algorithms is the Particle Swarm Optimization (PSO), which is inspired by the behavior of a herd of animals cooperating to find an optimal solution. In PSO, a population of particles iteratively moves through the solution space, adjusting their position based on their individual best experience and the best position in the entire swarm.

Adopting appropriate values for the coefficients and inertia weight is a crucial element of the PSO algorithm and directly influences the accuracy of the results obtained. In order to balance global exploration and local exploitation, the inertia weight ω was dynamically updated at each iteration of the algorithm (Equation (1)).

$$\omega^{t+1} = \omega_{\max} - \left(\frac{\omega_{\max} - \omega_{\min}}{t_{\max}}\right) \cdot t \qquad (1)$$

where, t_{max} denotes the maximum number of iterations, and t the current iteration. The inertia weight value was assumed in the range between of $\omega_{\min} = 0.4$ and $\omega_{\max} = 0.9$.

To determine the appropriate NN architecture for weight prediction, each PSO particle was represented in a 5-dimensional solution space. This consisted of: the number of layers ranging from 1 to 5, the number of neurons in the first hidden layer, the reduction factor for number of neurons applied to subsequent layers, the learning rate, and the activation function used, encoded in the numerical form: 0: ReLU, 1: leaky ReLU, 2: sigmoid, 3: hyperbolic tangent, 3: ELU, 4: SELU, 5: GELU, and 6: Swish.

The architectural search space was constrained to improve stability and reduce computational cost. The number of neurons in the first hidden layer was selected from a discrete set of powers of two from 32 to 256. The second parameter determining the number of neurons in subsequent layers was the reduction factor α. Instead of directly optimizing the reduction factor, an unconstrained auxiliary variable z was used in the PSO and then mapped to the admissible range via the sigmoid transformation (Equation (2)).

$$\alpha = 0.4 + 0.45 * \text{sigmoid}(z) \qquad (2)$$

Applying such a transformation results in a reduction factor ranging from 0.40 to 0.85 and ensures smooth exploration of the entire solution space. Additionally, layers with a small number of neurons destabilize the gradient propagation and increase the variance of the results, therefore a soft architectural regularization was used. For each particle in which a given hidden layer had fewer than 16 neurons, the Root Mean Square Error (RMSE) value was increased by a relative penalty l proportional to both the current prediction error and the deviation from the minimum layer width (Equation (3)).

$$l = 0.2 \cdot \text{RMSE} \cdot \frac{16 - n}{16} \qquad (3)$$

where n is the number of neurons in the hidden layer.

The learning rate was optimized in the logarithmic domain, with the search space defined between $\log_{10}(0.0005)$ and $\log_{10}(0.1)$ which reflects the multiplicative nature of the learning rate and allows the algorithm to explore different orders of magnitude in a balanced manner. The final NN architectures and their RMSE obtained on the FE-sourced test set are presented in Table 1.

Table 1. NN architecture after PSO algorithm

Gate	**Number of layers**	Reducing factor	1st layer	2nd layer	3rd layer	4th layer	5th layer	Learning rate	Activation function	RMSE (kg)
1	4	0.8483	256	217	184	156	–	0.0008	GELU	97.0
2	3	0.5720	256	146	84	–	–	0.0119	GELU	115.3
3	4	0.6896	128	88	61	42	–	0.0061	Swish	110.8

4 Results

The accuracy of the system has been verified using vehicles of known weights and different configurations of axles, passing the bridge in various scenarios. The basic scenario, representing the target traffic conditions expected during the operational phase and detailed in Section 3.1, consists of a single truck passing the bridge in the left or right lane. Exemplary passages of multiple trucks and involving varying speed are presented in Sections 3.2 and 3.3, respectively.

4.1 Single-truck runs

The single-truck tests involved nine types of vehicles with gross weights in the range between 27.8 t and 41.5 t and a wheel arrangement including 3 to 6 axles (Table 2). Figure 5 shows the strain change in consecutive gates during a run of a 4-axle truck of 32.7 t total weight. The vehicle moves in the direction of the target traffic, first passing through Gate 3, then Gate 2, and finally Gate 1. The most significant strain increase is observed in the middle sensors, CB1-2, CB2-2, and CB3-2, in Gates 1, 2, and 3, respectively. The magnitude of strain increase in the outer sensors is consistent in all three gates and indicates the pass of the truck on the lane by side of the sensors labeled CB1-3, CB2-3, and CB3-3. The gross vehicle weight is independently estimated by each gate, resulting with values 34.4 t in Gate 1, 33.8 t in Gate 2, 33.4 t in Gate 3, and the average of 33.9 t (error: 3.5%).

The results of the rest of the single-truck runs, 18 in total, and the mean performance metrics of the system are given in Table 2. The mean absolute percentage error of 4.6% and respective mean absolute error of 1.5 t is achieved in the given scenarios, while maintaining the mean percentage error of –2.6% and the mean error of 0.7 t when considering the sign of predictions against the true values. Two tests out of the performed ones, namely the 6-axle truck of 41.4 t in the right lane and the 3-axle truck of 28.7 t in the left lane, are clearly outstanding. This kind of effect can be accounted for additional uncertainties related to transversal positioning of vehicles, axle configuration, and the overall impact of the axles spacing compared to the range of the influence area.

Table 2. Weight estimation and performance metrics in single-truck tests: ACC – Accuracy, MPE – Mean Percentage Error, MAPE – Mean Absolute Percentage Error, ME – Mean Error, MAE – Mean Absolute Error

		Weight estimation, w				**Metrics**				
Vehicle	**Lane**	Gate 1 (t)	Gate 2 (t)	Gate 3 (t)	**Average**, w_{mean} (t)	ACC (–)	MPE (%)	MAPE (%)	ME (t)	MAE (t)
6-axle	Right	37.1	37.4	37.2	37.2	0.899	10.13	10.13	–4.2	4.2
[**41.4 t**]	Left	39.9	41.3	41.5	40.9	0.986	1.353	1.353	–0.6	0.6
5-axle	Right	39.8	39.9	39.2	39.6	0.976	2.379	2.379	–1.0	1.0
[**40.6 t**]	Left	43.2	39.6	42.2	41.7	1.026	–2.625	2.625	1.1	1.1
5-axle	Right	39.4	40.1	39.8	39.8	0.958	4.224	4.224	–1.8	1.8
[**41.5 t**]	Left	42.7	41.7	42.9	42.4	1.022	–2.213	2.213	0.9	0.9
4-axle	Right	34.4	34.2	33.7	34.1	1.031	–3.097	3.097	1.0	1.0
[**33.0 t**]	Left	36.2	31.9	32.8	33.6	1.017	–1.702	1.702	0.6	0.6
4-axle	Right	33.2	33.1	32.4	32.9	1.023	–2.335	2.335	0.8	0.8
[**32.2 t**]	Left	34.0	32.1	32.7	32.9	1.024	–2.393	2.393	0.8	0.8
4-axle	Right	34.4	33.8	33.4	33.9	1.035	–3.516	3.516	1.2	1.2
[**32.7 t**]	Left	34.3	34.1	32.5	33.6	1.028	–2.805	2.805	0.9	0.9
3-axle	Right	31.9	30.9	30.8	31.2	1.088	–8.777	8.777	2.5	2.5
[**28.7 t**]	Left	33.4	34.8	31.7	33.3	1.161	–16.14	16.14	4.6	4.6
3-axle	Right	30.2	29.5	30.1	29.9	1.078	–7.813	7.813	2.2	2.2
[**27.8 t**]	Left	27.7	28.4	31.4	29.2	1.052	–5.191	5.191	1.4	1.4
3-axle	Right	30.3	30.7	30.7	30.6	1.062	–6.245	6.245	1.8	1.8
[**28.8 t**]	Left	28.8	29.1	28.4	28.8	0.999	0.061	0.061	0.0	0.0
					Mean:	1.026	–2.595	4.611	0.7	1.5
					Standard deviation:	0.056	5.567	3.951	1.8	1.2

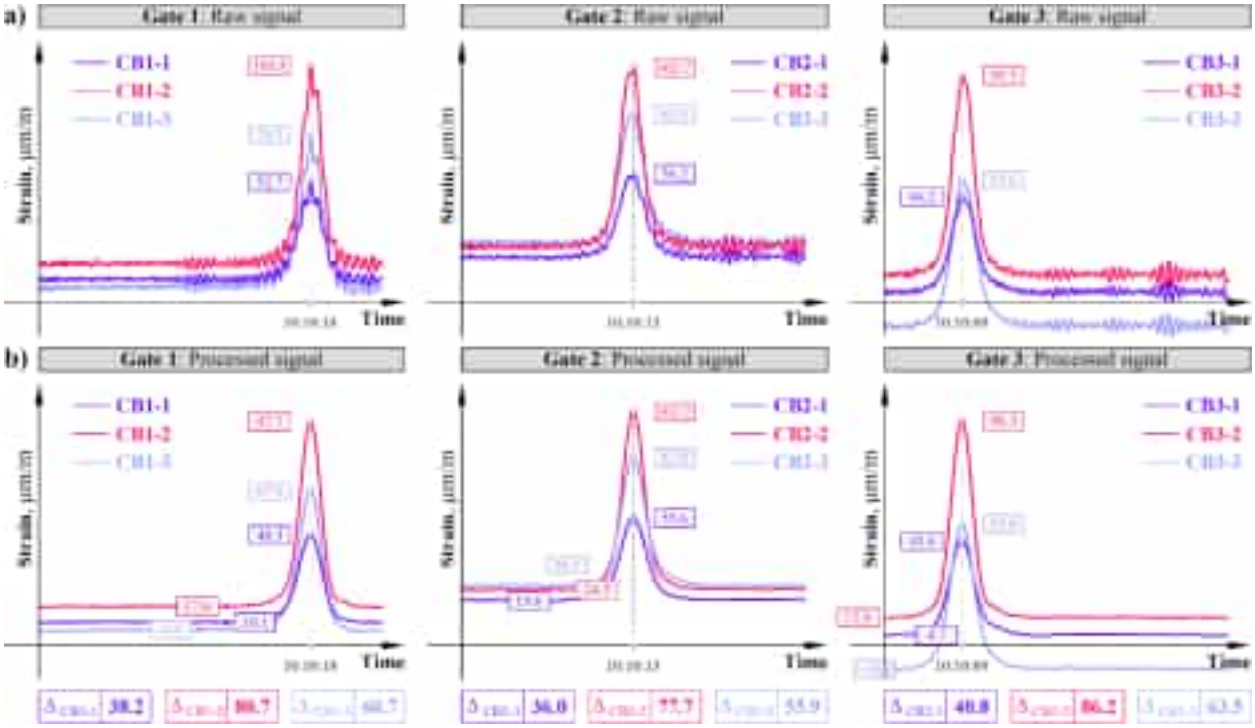


Figure 5. Exemplary strain change in a single-vehicle test in three consecutive measurement gates: a) raw signal, b) filtered signal, moving average, and the respective strain change

4.2 Multiple-truck runs

Although still relatively rare, the system should be resistant to vehicles arranged in groups and crossing the bridge in a convoy. To measure accuracy in this kind of scenarios, several tests were performed, among which the passage of 4-axle trucks of gross weights 33.0 t, 32.2 t and 32.7 t in the right lane was reported in the paper. The overview of the raw and filtered signals is shown in Figure 6. The performance of the system maintains the desired accuracy given by the MPE of –2.2%, MAPE of 2.2%, and both ME and MAE of 0.7 t, despite the clear overlapping of the second and third impacts. Thanks to the reduced range of the influence surface for crossbeam strains, relatively close vehicles can be separated and estimated with no additional measures, just due to the structural specificity of network arch bridges. The weight estimations reported for the three trucks in the test convoy are given in Table 3.

4.3 Runs of varying speed

Considering that specific applications are prone to the varying speed of vehicles, including the influence line-based ones, the given approach is evaluated against different velocities of trucks passing the bridge. An integral part of field tests is the vehicle moving at a speed of 30 km/h through an artificial obstacle, in this case a transversal step on the carriageway. In the given context, the step allows inducing additional dynamic excitations

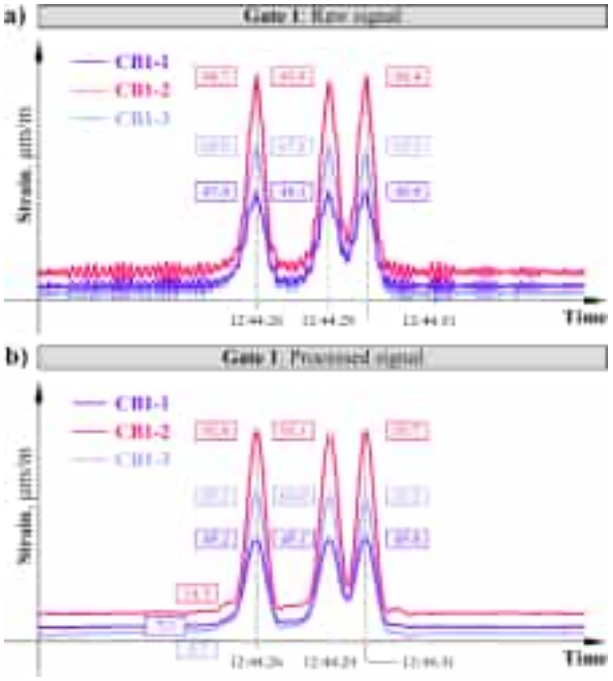


Figure 6. Exemplary strain change in a multiple-vehicle test in the Gate 1: a) raw signal, b) filtered signal

and signal disturbance and evaluate the system against unpredictable side-effects. The tests performed, among which the passage of the 4-axle truck in the left lane is

reported in the paper, indicate the consistent performance in varying speed conditions. In Figure 7, compared to Figure 5, the strain increase in the sensors CB1-1 and CB1-2 in Gate 1, CB-2-1 and CB2-2 in Gate 2, and CB3-1 and CB3-2 in Gate 3 are alike, indicating the placement of the vehicle being away from the CB1-3, CB2-3, and CB3-3 sensors in the consecutive gates. One should note the distinctive signal artifacts in the 30 km/h passage with the artificial step mounted right above the gate. Despite the perturbations, the filter is able to reduce additional effects, and produce clear enough signal for further processing by the neural network.

Table 4 and Figure 7 report gross vehicle weights for varying passage velocities, indicating consistent results given by MPE of –0.8%, MAPE of 2.9%, ME of 0.3 t, and MAE of 0.9 t. Based on the results, the presented approach allows for achieving good estimation accuracy regardless of the accompanying dynamic effects and variations in vehicle type.

Table 3. Weight estimation and performance metrics in multiple-truck tests, 4-axle vehicles in the right lane

		Weight estimation, w				**Metrics**				
Vehicle	**Lane**	Gate 1 (t)	Gate 2 (t)	Gate 3 (t)	**Average**, w_{mean} (t)	ACC (–)	MPE (%)	MAPE (%)	ME (t)	MAE (t)
[**33.0 t**]	Right	34.7	33.9	33.8	34.1	1.033	–3.309	3.309	1.1	1.1
[**32.2 t**]	Right	32.9	31.1	34.3	32.8	1.019	–1.886	1.886	0.6	0.6
[**32.7 t**]	Right	32.3	32.0	35.3	33.2	1.015	–1.467	1.467	0.5	0.5
					Mean:	1.022	–2.221	2.221	0.7	0.7

Table 4. Weight estimation and performance metrics for varying velocities, gross vehicle weight of 32.7 t

		Weight estimation, w				**Metrics**				
Vehicle	**Lane**	Gate 1 (t)	Gate 2 (t)	Gate 3 (t)	**Average**, w_{mean} (t)	ACC (–)	MPE (%)	MAPE (%)	ME (t)	MAE (t)
30 km/h	Left	34.2	34.1	33.0	33.8	1.032	–3.199	3.199	1.0	1.0
50 km/h	Left	32.4	31.5	31.7	31.9	0.974	2.608	2.608	–0.9	0.9
80 km/h	Left	34.3	34.1	32.5	33.6	1.028	–2.791	2.791	0.9	0.9
					Mean:	1.008	–0.788	2.865	0.3	0.9

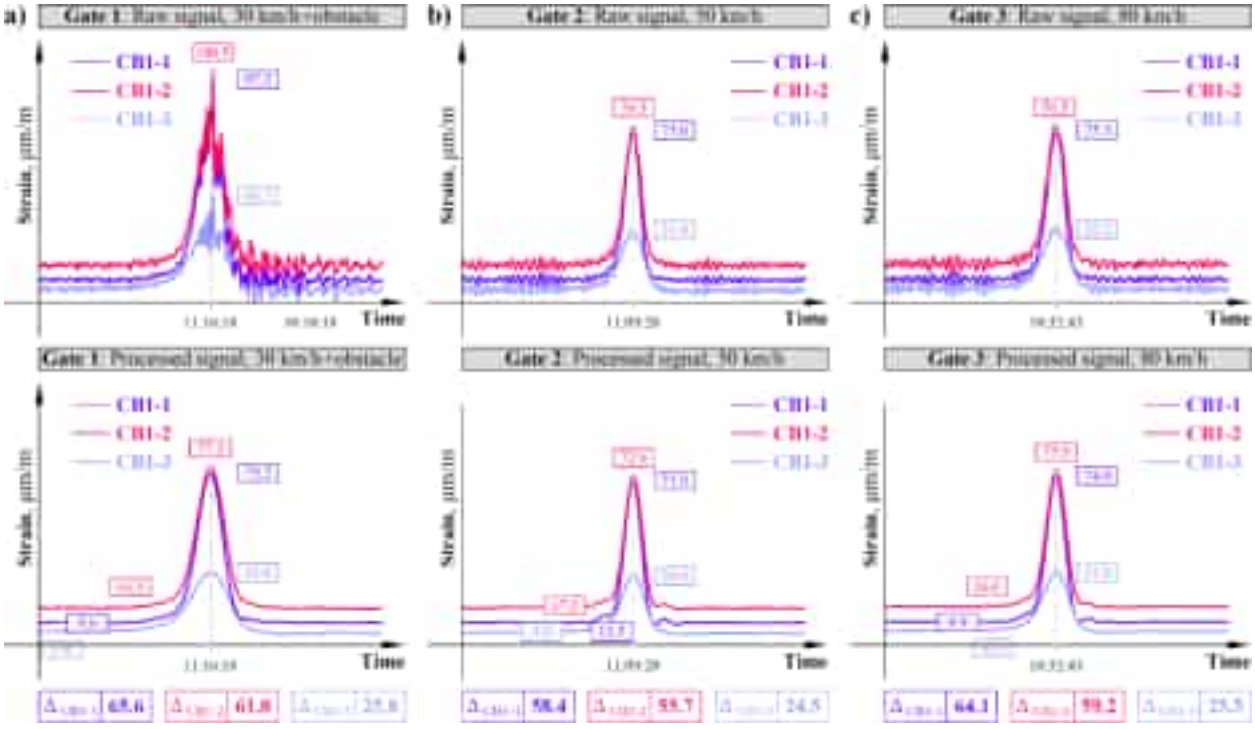


Figure 7. Exemplary strain change in a single-vehicle test for different passage velocities : a) 30 km/h with the dynamic impact due to artificial obstacle, b) 50 km/h, c) 80 km/h

5 Limitations

The proposed WIM approach constitutes a promising solution for long-term bridge monitoring under realistic traffic conditions, however several aspects may benefit from further investigation. The current validation was conducted on a single bridge structure, and future work will focus on extending the methodology to other bridge types. In addition, the neural networks were trained using data from a calibrated finite element model, which proved effective but could be further enhanced with larger datasets obtained directly from field measurements.

Moreover, although a range of vehicle configurations was considered, expanding the dataset may further improve the robustness of the model. These aspects do not affect the overall applicability of the method but rather indicate its potential for further development.

6 Conclusions

This research integrates the bridge SHM system using FBG sensors and ML to estimate the gross vehicle weight in real-word traffic. Unlike many existing WIM studies, the proposed approach was experimentally validated under realistic traffic conditions, without requiring prior knowledge of vehicle path or speed as input to the estimation algorithm.

The system achieved an average MAPE of 4.6% for single-vehicle run and 2.2% for multiple-vehicle runs, confirming its robustness to varying traffic scenarios. Moreover, the influence of vehicle speed and dynamic effects was shown to be limited and did not significantly degrade the prediction accuracy.

A key feature of the proposed methodology is the use of three independent measurement gates distributed along the span. The structural features of the network arch bridge, primarily the short-range influence area of the crossbeam strain, proved to beneficial in separating closely spaced vehicles, constituting an important component of the framework.

References

[1] D. Paul, K. Roy. Application of bridge weigh-in-motion system in bridge health monitoring: a state-of-the-art review. *Structural Health Monitoring*, 22(6):4194-4232 (2023).

[2] X. Xiao, D. Pi, Q. Zhu. A bridge weigh-in-motion algorithm for fast-passing railway freight vehicles considering bridge-vehicle interaction. *Mechanical Systems and Signal Processing*, 181 (2022) 109493

[3] S. Alamandala, R.L.N. Sai Prasad, Rathish Kumar Pancharathi et al. Study on bridge weigh in motion (BWIM) system for measuring the vehicle parameters based on strain measurement using FBG sensors. *Optical Fiber Technology*, 61 (2021) 102440.

[4] W. He, T. Ling, E. O'Brien et al. Virtual Axle Method for Bridge Weigh-in-Motion Systems Requiring No Axle Detector. *Journal of Bridge Engineering*, 24(9) 2019.

[5] C. Tan, B. Zhang, H. Zhao et al. An Extended Bridge Weigh-in-Motion System without Vehicular Axles and Speed Detectors Using Nonnegative LASSO Regularization. *Journal of Bridge Engineering*, 28(5) 2023.

[6] N. Zolghadri, M. W. Halling, N. Johnson et al. Field Verification of Simplified Bridge Weigh-in-Motion Techniques. *Journal of Bridge Engineering*, 21(10) 2016.

[7] K. Brzozowski, A. Maczyński, A. Ryguła, T. Konior. A weigh-in-motion system with automatic data reliability estimation. *Measurement*, 221 (2023) 113494.

[8] X. Jian, Y. Xia, S. Sun et al. Integrating bridge influence surface and computer vision for bridge weigh-in-motion in complicated traffic scenarios. *Structural Control and Health Monitoring*, 2022, 29

[9] L. Zhang, S. Wang, L. Deng et al. A general solution to bridge weigh-in-motion equations based on maximum likelihood estimation. *Engineering Structures*, 343 (2025), 121193

[10] A. J. Ferguson, D. Hester, F. Huseynov et al. On the normalisation and mapping of influence lines. *Mechanical Systems and Signal Processing*, 237, 112883 (2025)

[11] W. He, J. Liu, S. Song et al. A non-contact vehicle weighing approach based on bridge weigh-in-motion framework and computer vision techniques. *Measurement*, 225, 113994 (2024)

[12] H. Wu, H. Zhao, J. Liu et al. A filtering-based bridge weigh-in-motion system on a continuous multi-girder bridge considering the influence lines of different lanes. *Frontiers of Structural and Civil Engineering*, 14:1232-1246 (2020)

[13] I. Yoshida, S. Mustafa, K. Maruyama et al. Bridge weigh-in-motion considering dynamic response in observation noise with application to multiple driving conditions. *Structural Safety*, 103, 102350 (2023)

[14] A. Moghadam. A novel speed-adaptive framework for structural load evaluation using notihing-on-road bridge weigh-in-motion. *Structures*, 82 (2025) 110714.

[15] H. Wang, T. Nagayama, B. Zhao, D. Su. Identification of moving vehicle parameters using bridge responses and estimated bridge pavement roughness. *Engineering Structures*, 153:57-70, 2017.

Identification and Prediction of Heat-Related Illnesses in Construction

Mehdi Torbat Esfahani[1], Jonathan Gayechuway[2], and Ibukun Awolusi, Ph.D.[3*]

[1]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
[2]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
[3*]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
mehdi.torbatesfahani@my.utsa.edu, jonathan.gayechuway@my.utsa.edu, ibukun.awolusi@utsa.edu

Abstract –

Heat-related illnesses (HRIs) pose a persistent threat to construction workers' safety. Effective risk assessment enables early detection of hazardous conditions and timely prevention of illness, yet incident logs and narrative reports are seldom leveraged for computational risk analysis. This research develops a data-driven framework to (i) identify HRIs from routine safety records and (ii) predict heat-related risk based on historical incident and environmental conditions to support planning and site controls. The approach combines text mining and supervised classification to detect heat-related cases from narratives, supported by structured incident details, including task, project context, and HRI indicators. Incident time and location are combined with environmental heat-stress indices to analyze environment–task–risk relationships. A predictive model is developed to estimate near-term risk levels for various construction activities. Results reveal clear seasonal and regional patterns, with higher risks in outdoor and heavy-civil operations during prolonged heat waves. The classification model performed well, with an F1 score of about 0.72–0.81 across trade groups, and probabilistic heat-related risk prediction offers actionable lead time for planning adjustments. The risk scores provide a foundation for potential integration into construction scheduling systems to support work–rest planning, hydration strategies, shade provisioning, and task resequencing during high-heat periods. This work provides a scalable computational framework that transforms routine safety records into predictive insights, enabling proactive, climate-aware risk mitigation in construction.

Keywords –

Construction; Heat-related Illnesses; Heat Stress; Prediction; Risk Assessment

1 Introduction

The construction industry remains one of the most labor-intensive and risk-prone sectors worldwide, where workers are frequently exposed to physically demanding tasks, hazardous environments, and adverse climatic conditions [1-3]. Among environmental stressors, occupational heat exposure has emerged as a critical and escalating threat due to rising global temperatures, intensifying heat waves, and prolonged outdoor work durations [1-6]. Elevated heat exposure has been associated with reduced physical performance, impaired cognitive function, increased fatigue, and heightened risk of occupational injuries, particularly in labor-intensive construction activities such as heavy civil, infrastructure, and building operations [7-9]. As climate change continues to amplify thermal stress across geographic regions, heat-related health risks in construction are expected to intensify, reinforcing the need for improved prevention and mitigation strategies [10-12].

Effective identification and prediction of heat-related illnesses (HRIs) are central to protecting worker health and supporting proactive safety management in construction environments [13, 14]. Timely detection of hazardous heat conditions enables the implementation of preventive interventions, such as optimizing the work–rest cycle, planning hydration, rescheduling tasks, and deploying cooling or shading resources [15-17]. Predictive heat-risk modeling further strengthens operational decision-making by providing early warnings that help reduce injury rates, minimize productivity losses, and lower medical and compensation costs [8, 18, 19]. As regulatory agencies and industry stakeholders increasingly recognize heat exposure as a critical occupational hazard, data-driven identification and forecasting mechanisms are becoming integral to modern construction safety frameworks [20, 21].

Previous studies have proposed a variety of approaches for assessing heat exposure and managing

thermal risk in occupational settings, including physiological monitoring, wearable sensing technologies, heat-stress indices, and real-time environmental advisory systems ([4, 22]. Research has examined the influence of meteorological variables, workload intensity, acclimatization, and organizational controls on worker heat strain and injury outcomes [7, 8, 23]. More recently, machine learning and artificial intelligence techniques have been applied to predict heat strain, occupational accidents, and safety risks, demonstrating promise for automated hazard detection and decision-support applications [12, 18]. In parallel, organizational and policy-based interventions have been explored to enhance heat awareness, regulatory compliance, and protective behaviors among construction workers [24].

Many existing models rely on physiological sensors, wearable devices, or experimental data collected under controlled conditions, which constrains scalability and limits practical deployment across large construction workforces [11, 16, 22]. Despite these advances, several critical knowledge gaps remain. First, existing research largely relies on structured datasets or sensor-based measurements, while unstructured narrative safety records remain underutilized for identifying heat-related risks. Second, few studies integrate historical construction injury data with environmental heat-stress indicators to establish explicit environment–task–risk relationships. Third, prior work has focused primarily on general heat exposure or physiological responses, with limited attention to task-specific risk stratification and severity-oriented modeling using real-world incident data [25, 26]. Incident narratives frequently contain rich contextual information about environmental exposure, work tasks, symptoms, and contributing risk factors, yet these unstructured data sources remain underutilized in computational safety analytics. In addition, few studies have integrated historical construction injury records with environmental heat-stress metrics to develop probabilistic heat-related risk predictive frameworks tailored specifically to construction operations [14, 17]. These gaps indicate a need for scalable, data-driven methodologies that transform existing safety documentation into actionable heat-risk intelligence.

This study develops a computational framework to identify HRIs to predict probabilistic heat-related risk by integrating historical injury records with environmental heat-stress indicators. The study seeks to answer three research questions: How effectively can HRIs events be detected using text-based analysis of occupational incident narratives? Which environmental and task-related factors are most strongly associated with heat-related risk in construction activities? Can historical safety records be leveraged to forecast near-term heat risk to support proactive planning and site-level intervention?

2 Literature Review

Table 1 summarizes key prior studies on occupational heat stress, predictive safety modeling, and data-driven construction risk assessment, highlighting their data sources, methodological approaches, and limitations.

Table 1. Prior Studies on Heat Stress in Construction

Study	Focus / Data Source	Key Findings & Limitations
[7]	Experimental heat exposure data	Identified physiological heat stress thresholds; limited real-world field scalability
[4]	Climate and occupational health data	Linked heat indicators to worker health outcomes; not construction-specific
[8]	Construction field injury data	Associated heat exposure with injury risk; limited predictive capability
[22]	Wearable physiological sensor data	Enabled real-time heat strain monitoring; costly and difficult to scale
[27]	Environmental heat index datasets	Proposed heat-stress assessment metrics; no injury outcome integration
[19, 28]	Construction injury records	Applied machine learning to predict safety risk; relied on structured data only
[14]	Historical construction safety data	Modeled task-based injury risk; did not incorporate heat exposure
[18]	Multi-source safety datasets	Improved automated hazard detection using AI; limited climate integration
[25]	Incident narrative text data	Extracted safety risk factors using natural language processing (NLP); not heat-specific
[26]	Construction safety narratives	Demonstrated text-based risk prediction; HRIs not addressed

2.1 Occupational Heat Stress in Construction

Heat stress has been widely recognized as a major occupational health hazard, particularly in labor-intensive industries such as construction, where prolonged outdoor exposure and physically demanding tasks elevate physiological strain [4, 7]. Prior research has linked high ambient temperatures and humidity to increased fatigue, dehydration, reduced cognitive performance, and heightened risk of injury among construction workers [8, 9]. These effects are often exacerbated by inadequate rest schedules, insufficient hydration, and limited access to cooling resources [23].

Several studies have examined the broader implications of climate change on occupational heat exposure, highlighting the growing frequency and intensity of heat events and their disproportionate impact on outdoor workers [6, 11].

2.2 Heat Risk Monitoring and Measurement Approaches

A range of methods has been proposed to quantify and monitor heat exposure in occupational environments. Traditional approaches rely on environmental heat-stress indices, such as temperature–humidity metrics, Wet Bulb Globe Temperature proxies, and meteorological forecasting models [16]. Other studies have focused on physiological monitoring, employing wearable sensors to track heart rate, body temperature, sweat rate, and metabolic workload in real time [22].

While sensing technologies offer high-resolution physiological data, their deployment at scale is often constrained by cost, maintenance requirements, privacy concerns, and worker compliance challenges [11, 22].

2.3 Predictive Modeling and Machine Learning for Safety Risk

Machine learning and artificial intelligence techniques have increasingly been applied to occupational safety research, enabling predictive modeling of accident risk, hazard exposure, and injury likelihood [12, 18]. In construction contexts, predictive models have been developed to estimate injury risk based on task attributes, environmental factors, worker demographics, and historical incident data [14].

Several studies have demonstrated the potential of supervised learning algorithms, including decision trees, support vector machines, and ensemble models, for forecasting safety outcomes and supporting proactive intervention [19]. However, many existing models rely on structured datasets and do not fully leverage unstructured narrative data or historical safety documentation, limiting their ability to capture nuanced causal and contextual risk factors [25, 29].

2.4 Use of Safety Records and Data for Risk Analysis

Occupational safety incident reports contain rich qualitative information describing task conditions, environmental context, worker behavior, and contributing factors. Recent research has explored the use of NLP to extract risk indicators, classify injury mechanisms, and automate hazard identification from narrative safety records [25, 26]. Despite growing interest in narrative-based safety analytics, few studies have applied text-mining techniques to identify HRIs or to link narrative injury data to environmental heat metrics [12, 29].

Despite these advances, existing studies have not fully integrated narrative-based safety data with environmental heat metrics to develop scalable, task-specific, and probabilistic heat-risk assessment frameworks in construction.

3 Materials and Methods

This study employed a retrospective, data-driven analytical framework to identify HRIs cases in construction and to predict probabilistic heat-related risk using historical occupational safety records and environmental heat indicators. The methodology integrates natural language processing, supervised machine learning, and predictive modeling to extract heat-related risk signals from narrative-based incident reports and structured injury attributes. Figure 1 illustrates the overall methodological workflow from data acquisition to predictive modeling.



Figure 1. Methodological Workflow for HRIs Identification and Risk Prediction

3.1 Data Sources and Study Sample

This study adopted a retrospective, data-driven research design to identify HRI cases in construction and

to develop a predictive framework for probabilistic heat-related risk. Occupational injury and incident records were obtained from the OSHA Severe Injury Reports database, covering construction-related incidents across multiple industry subsectors, including commercial, industrial, residential, highway, utility, and pipeline construction. Each record contains structured attributes such as event date, geographic location, industry classification, injury type, and severity indicators, along with an unstructured narrative field describing the circumstances, contributing factors, and outcomes of each incident. Only construction-related cases with complete narrative descriptions were retained for analysis. Table 2 summarizes the primary data sources and analytical variables used in the study.

Table 2. Data Sources and Key Variables Used in the Study

Data Source	Variable Category	Key Variables
OSHA Severe Injury Reports	Incident & injury data	Event date, location, NAICS, injury type
OSHA Final Narrative	Unstructured text	Heat exposure cues, symptoms, task context, risk factors
NOAA / Meteostat weather	Environmental heat	Temperature, humidity, heat index, WBGT proxy
Derived temporal features	Time & seasonality	Month, season, heat-wave indicator
Machine learning features	Modeling inputs	TF–IDF text features, structured injury fields, heat metrics

3.2 Narrative Processing and Heat-Related Case Identification

Incident narratives in the Final Narrative field served as the primary source for identifying heat-exposure conditions, symptoms, and contributing risk factors. Text data were pre-processed using standard natural language processing procedures, including normalization, tokenization, lemmatization, and removal of non-informative terms. Heat-related linguistic cues and symptom descriptors were compiled to support the semantic interpretation of narratives.

A supervised classification approach was used to distinguish heat-related from non-heat-related incidents. A subset of records was manually labelled based on narrative evidence of heat exposure, thermal stress symptoms, or heat-induced medical outcomes. The labeling process was conducted by two independent reviewers with domain knowledge in construction safety. Incidents were classified as heat-related if the narrative contained explicit references to elevated ambient temperature or heat exposure, or described symptoms consistent with heat-related illness, such as dizziness, heat exhaustion, or loss of consciousness. Inter-annotator agreement was evaluated using Cohen's kappa, and any discrepancies were resolved through discussion to achieve consensus labeling. Textual features were extracted using term-frequency–inverse-document-frequency representations and combined with relevant structured injury attributes.

Model hyperparameters were selected via a 5-fold cross-validation grid search. Key parameters, including regularization strength, class weighting, and TF–IDF feature settings (e.g., n-gram range and minimum term frequency), were systematically varied to identify the configuration that maximized F1-score. This approach ensured model performance while mitigating overfitting.

3.3 Environmental Heat Exposure Integration

Incident records were linked to historical meteorological conditions using reported event dates and geographic coordinates. Weather data were retrieved from gridded historical climate datasets to obtain hourly and daily measurements of ambient temperature, relative humidity, and related atmospheric variables at the time and location of each OSHA incident. When precise hourly timestamps were unavailable, daily averages and maximum heat values were used to approximate environmental exposure conditions.

Composite heat-stress indicators, including the heat index and a Wet Bulb Globe Temperature (WBGT) proxy, were computed to characterize the thermal load at the time of each event. These indices were derived from established thermophysiological formulas that combine temperature and humidity to estimate perceived heat stress. Additional temporal variables, such as season, month, and heat-wave occurrence, were generated to capture climatic patterns and prolonged heat exposure.

3.4 Predictive Modeling of Heat Risk

To estimate probabilistic heat-related risk in construction operations, predictive models were developed to assess the likelihood of heat-related incidents based on environmental exposures, industry subsector, temporal factors, and narrative-derived risk indicators. Supervised learning algorithms, including logistic regression and ensemble tree-based models, were trained to generate probabilistic heat-risk scores. Feature-importance analysis identified key predictors of heat risk, thereby enabling interpretation of the environmental and operational drivers of heat-related incidents.

4 Results

4.1 Dataset Summary and Incident Characteristics

The final OSHA construction dataset included 216 reportable incidents recorded between 2015 and 2023, spanning eight construction subsectors. Highway construction accounted for the largest share of records (30.6%), followed by commercial building construction (23.1%) and water and sewer line construction (13.4%). Incident narratives averaged 52.3 ± 18.6 words, providing textual detail for narrative-based analysis. Temporal distribution indicated that 61.8% of incidents occurred between May and September, consistent with peak seasonal heat exposure. Table 3 summarizes the distribution of incidents across construction subsectors.

Table 3. Distribution of OSHA Construction Incidents by Subsector

Construction Subsector	NAICS Code	Incident Count	Percentage (%)
Highway / Street / Bridge	237310	66	30.6
Commercial Building	236220	50	23.1
Water / Sewer Line	237110	29	13.4
Power / Communication Line	237130	25	11.6
Industrial Building	236210	20	9.3
Oil & Gas Pipeline	237120	12	5.6
Other Heavy / Civil	237990	9	4.2
Single-Family Residential	236115	5	2.3
Total	—	*216*	*100.0*

4.2 Narrative-Based Analysis of Incident Causes

Analysis of the Final Narrative field revealed distinct causal patterns between heat-related and non-heat-related incidents. Among heat-related cases (n = 47), the most frequently reported contributing factors were heat exposure or high ambient temperature (78.7%), physical overexertion (63.8%), dehydration or inadequate fluid intake (46.8%), and prolonged sun exposure (42.6%). Commonly documented physiological symptoms included dizziness or fainting (51.1%), heat exhaustion (38.3%), and loss of consciousness (23.4%), indicating acute heat strain as a primary mechanism of injury.

Task-level narrative evidence showed that outdoor, labor-intensive operations, including asphalt paving, concrete placement, trenching, roofing, and utility line installation, occurred in 69.1% of heat-related incidents. Furthermore, limited access to shade or cooling resources was noted in 34.0% of heat-related narratives, and insufficient rest breaks were cited in 28.9%, suggesting gaps in heat-mitigation practices.

In contrast, non-heat-related incidents (n = 169) were predominantly associated with falls (31.4%), struck-by events (26.0%), equipment-related injuries (22.5%), and material-handling incidents (18.9%), indicating fundamentally different causal mechanisms. These findings show that heat-related incidents have distinct exposure conditions, symptom profiles, and operational risk factors, underscoring the value of narrative-based analytics for differentiating injury etiology in construction safety data.

4.3 Heat-Related Illness Identification Performance

The supervised NLP classifier identified 47 incidents (21.8%) as HRIs within the OSHA construction dataset. Model performance, evaluated using five-fold cross-validation, demonstrated stable classification capability, yielding a mean accuracy of 0.84 (95% CI: 0.81–0.87), precision of 0.79 (95% CI: 0.75–0.83), recall of 0.76 (95% CI: 0.72–0.80), and an F1-score of 0.81 (95% CI: 0.78–0.84). The classifier achieved an ROC–AUC of 0.87, indicating high discriminative power between heat-related and non-heat-related narratives.

Subsector-specific performance analysis revealed modest variation across trade groups, with F1-scores ranging from 0.72 in residential construction to 0.81 in heavy civil construction, suggesting consistent generalizability across operational contexts. Collectively, these results demonstrate that narrative-based machine learning can reliably distinguish heat-related occupational incidents from other injury events. Table 4 summarizes the classification performance metrics.

Table 4. Performance of the HRIs Classification Model

Metric	Value
Accuracy	0.84
Precision	0.79
Recall	0.76
F1-score	0.81
ROC–AUC	0.87
Heat-related cases identified	47
Share of dataset (%)	21.8

4.4 Environmental Heat Exposure at Time of Incidents

Meteorological observations obtained from NOAA/Meteostat were successfully linked to 94.9% of OSHA construction incidents recorded between 2015 and 2023, enabling a quantitative assessment of ambient

thermal exposure at the time of injury. Comparative analysis revealed that heat-related incidents occurred under significantly greater heat stress than non-heat incidents. Wet Bulb Globe Temperature (WBGT) proxy was 28.9°C for heat-related incidents versus 24.7°C for non-heat incidents (mean difference = 4.2°C; Cohen's d = 1.21), reflecting substantially elevated physiological heat strain. As shown in Figure 2, 44.7% of heat-related cases occurred in July and August, corresponding to periods of peak regional thermal load.

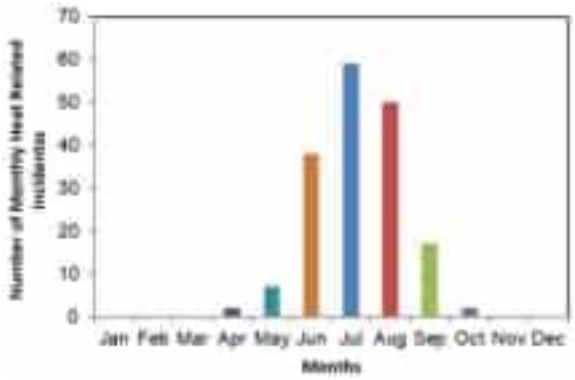


Figure 2. Monthly Distribution of Heat-Related OSHA Incidents

4.5 Predictive Heat-Risk Model Performance

The risks of HRIs were modelled as a binary outcome at the incident level, as presented in Equation 1, where indicates whether the OSHA incident $i$was classified as heat related. The probability of HRIs occurrence was estimated using a regression model in which explicitly defined risk factors ,derived from OSHA Narrative text and structured incident attributes ,served as predictors:

$$P(HRI_i = 1) = 1 / (1 + exp(-(\beta_0 + \beta_1\, DEHYD_i + \beta_2\, OVEREX_i + \beta_3\, OUTDOOR_TASK_i + \beta_4\, MONTH_i + \beta_5\, SECTOR_i))) \quad (1)$$

Where $P(HRI_i = 1)$denotes the predicted probability that the incident $i$is heat-related, and $\beta_0$through β_8, are the estimated model coefficients. The binary predictors $HEATEXP_i$, $DEHYD_i$, $OVEREX_i$, SUN_i, $SHADE_i$, and $REST_i$ indicate the presence (coded as 1) or absence (coded as 0) of narrative-identified risk factors extracted from the OSHA Narrative field. These indicators represent explicit mentions of heat exposure or elevated ambient temperature, dehydration or insufficient fluid intake, physical overexertion, direct sun exposure, limited access to shade or cooling resources, and inadequate rest breaks, respectively. Temporal variation is captured by $MONTH_i$, encoded as categorical month indicators with January serving as the reference category, while industry context is represented by $SECTOR_i$, encoded as construction subsector indicators with a defined reference subsector.

Table 5 reports the fitted coefficients, odds ratios, standard errors, and 95% confidence intervals. Physical overexertion emerged as the strongest predictor of severe heat illness, with an estimated coefficient of $\beta = 2.95$corresponding to an odds ratio of 19.11 (95% CI: 7.66–47.65; $p < 0.001$). This indicates that incidents involving exertion-related language were associated with a nearly 20-fold increase in the odds of severe heat outcomes.

Table 5. Logistic Regression Coefficients for Predicting Severe Heat-Related

Predictor	β	Interpretation
Intercept	−2.104	Baseline odds of severe heat illness when all predictors = 0
Dehydration mentioned (DEHYD)	−3.111	Dehydration mentions are associated with lower odds of severe outcomes, likely reflecting reporting patterns in non-collapse cases
Overexertion mentioned (OVEREX)	2.950	Overexertion increases the odds of severe heat illness by ~19×
Outdoor high-intensity task (OUTDOOR _TASK)	0.543	Outdoor heavy work increases odds by ~72% (not statistically significant)
Month (numeric)	−0.094	Each month later in the year reduces the severity odds by ~9%
Construction subsector (SECTOR)	Category -specific	Risk varies across subsectors relative to the reference sector

The model output is a continuous probability value between 0 and 1, interpreted as an incident-specific heat risk score. To compute $P("HRI"_i = 1)$ for any OSHA record, a reader may substitute observed predictor values into the linear predictor term, sum the corresponding coefficient contributions, and apply the logistic transformation. Figure 3 also shows the relative effect sizes of narrative-derived risk factors on the odds of severe heat-HRIs.

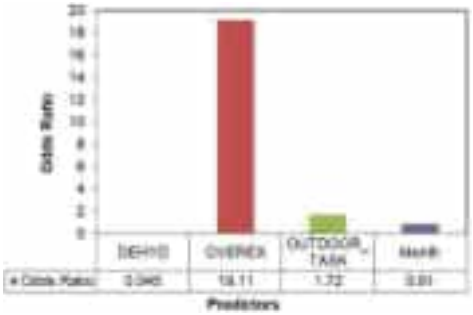


Figure 3. Odds Ratio Effect Sizes of Narrative-derived Risk Factors Predicting Severe HRI

4.6 Risk Stratification by Construction Activity

Risk stratification analysis revealed substantial heterogeneity in predicted HRI probabilities across construction activities, reflecting variation in physiological workload, exposure duration, and task intensity. Predicted risk scores from the fitted logistic model were aggregated by activity type to quantify differential vulnerability across operational contexts. Activities involving high physical exertion and prolonged outdoor exposure had the highest predicted heat-related risk. Asphalt paving, roofing, trenching, concrete placement, and utility-line installation ranked among the highest-risk task categories, with mean predicted severe-heat probabilities exceeding 0.65 and upper-quartile probabilities surpassing 0.80.

In contrast, activities with lower metabolic demand, such as interior finishing, equipment inspection, and light electrical work, showed lower risk, with mean probabilities below 0.35. The mean difference in predicted risk between high-intensity outdoor tasks and lower-intensity activities exceeded 0.30. Subsector-level stratification further indicated elevated risk concentrations in heavy civil construction, followed by utility and pipeline installation, consistent with their greater reliance on continuous outdoor labor under thermally stressful conditions. Conversely, commercial building and interior construction exhibited comparatively lower predicted heat-related risk profiles.

5 Conclusion

This study developed a data-driven model to identify, quantify, and stratify heat-related injury risk in construction, leveraging OSHA incident narratives, structured injury records, and predictive modeling. By extracting mechanistic risk factors from injury narratives and modeling their association with HRIs' severity, the analysis advances beyond descriptive or seasonal assessments toward a task-specific and causally informed understanding of heat-risk dynamics.

The results indicate that physical overexertion is the primary driver of HRI, with elevated odds of acute outcomes among incidents involving high metabolic workload. Risk stratification further demonstrates that outdoor, labor-intensive activities, including paving, roofing, and utility installation, exhibit a high predicted severity. The resulting predictive model provides a probability-based risk score that supports heat-mitigation planning, work–rest scheduling optimization, and the prioritization of high-risk construction tasks.

Several limitations warrant consideration. First, the analysis relies on OSHA narrative records, which may contain reporting inconsistencies or underrepresentation of certain risk factors due to variability in documentation practices. Second, the dataset's relatively small sample size, particularly within certain construction subsectors, may limit the statistical reliability and generalizability of subgroup-level comparisons; therefore, stratified risk estimates should be interpreted with caution. Third, the absence of direct physiological measurements, such as core temperature and heart rate, limits the ability to fully capture individual heat-strain responses. Third, meteorological exposure estimates depend on the resolution of historical weather data and may not reflect microclimatic conditions at specific job sites.

Future research should extend this framework by integrating high-resolution environmental monitoring, wearable physiological sensors, and longitudinal exposure histories to enable more precise, real-time heat-risk estimation. Expanding the dataset across regions, industries, and climate regimes would further improve model robustness and applicability.

References

[1]. Abdul Muthalif, M.Z., D. Shojaei, and K. Khoshelham, *Interactive Mixed Reality Methods for Visualization of Underground Utilities.* PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 2024. 92(6): p. 741-760.

[2]. Chan, A.P.C., et al. *A research framework for assessing the effects of heat stress on construction workers*. 2011. Research Publishing Services.

[3]. Rowlinson, S. and A.Y. Jia, *Application of the predicted heat strain model in development of localized, threshold-based heat stress management guidelines for the construction industry.* Annals of Occupational Hygiene, 2014. 58(3): p. 326-339.

[4]. Kakamu, T., et al., *Heat-related illness risk and associated personal and environmental factors of construction workers during work in summer.* Scientific Reports, 2021. 11(1): p. 6.

[5]. Messeri, A., et al., *Heat Stress Perception among Native and Migrant Workers in Italian Industries: Case Studies from the Construction and*

Agricultural Sectors. International Journal of Environmental Research and Public Health, 2019. 16(7): p. 13.
[6]. Mirhosseini, S.A., R. Chang, and H. Omrany. *Heat-exposure risks in the construction industry: An analysis of fatal and non-fatal incidents in the United States between 2000 and 2023*. in *Proceedings of the International Conference of Architectural Science Association*. 2024.
[7]. Farshad, A., et al., *Heat stress level among construction workers.* Iranian Journal of Public Health, 2014. 43(4): p. 492-498.
[8]. Idris, M.A., et al., *Examining Management and Employees' Perceptions of Occupational Heat Exposure and the Effectiveness of a Heat Stress Prevention Intervention on Safety and Well-Being among Natural Gas Construction Workers: A Qualitative Field-Based Study.* International Journal of Environmental Research and Public Health, 2024. 21(9): p. 1255.
[9]. Kim, Y. and Y. Ham, *Spatio-temporal heat risk analysis in construction: Digital twin-enabled monitoring*. Automation in Construction, 2024. 168.
[10]. Kawakami, R., et al., *Cooler break areas: Reducing heat stress among construction workers in Japan.* Building and Environment, 2024. 262: p. 111821.
[11]. Shin, Y., et al., *Impact of Heat Stress on Individual Cognitive States: Utilizing EEG Metrics in Immersive VR-Based Construction Safety Training.* Journal of Management in Engineering, 2024. 40(6).
[12]. Xu, H., et al., *Construction of a composite cooling network for the mitigation of urban heat risk in Fuzhou.* Frontiers in Environmental Science, 2024.
[13]. Jia, A.Y., et al., *Institutions and institutional logics in construction safety management: the case of climatic heat stress.* Construction Management and Economics, 2017. 35(6): p. 338-367.
[14]. Li, L., et al. *Investigation on heat stress of construction workers in summer in Chongqing, China*. 2019. EDP Sciences.
[15]. Chan, A.P.C., et al., *Developing a heat stress model for construction workers.* Journal of Facilities Management, 2012. 10(1): p. 59-74.
[16]. Kakamu, T., et al., *Heart rate increase from rest as an early sign of heat-related illness risk in construction workers.* International Journal of Industrial Ergonomics, 2022. 89: p. 5.
[17]. Ouyang, Y., G. Huang, and S. He, *Impacts of adverse environmental factors on construction workers' attention allocation during hazard identification: a study of noise and heat exposure.* Engineering, Construction and Architectural Management, 2024.
[18]. Fuertes, A. and M. Khatabakhshrad. *Health and Safety Risk Management in a Changing Climate: Contrasting Supervisors and Workers Perceptions and Approaches to Heat Coping Strategies on Construction Sites*. 2024. Association of Researchers in Construction Management.
[19]. Yi, W. and A.P.C. Chan, *Which Environmental Indicator Is Better Able to Predict the Effects of Heat Stress on Construction Workers?* Journal of Management in Engineering, 2015. 31(4): p. 9.
[20]. Qiao, Y., et al., *Wearable sensor for continuous sweat biomarker monitoring.* Chemosensors, 2022. 10(7): p. 273.
[21]. Shakerian, S., et al., *Assessing occupational risk of heat stress at construction: A worker-centric wearable sensor-based approach.* Safety Science, 2021. 142.
[22]. Calkins, M.M., et al., *A case-crossover study of heat exposure and injury risk among outdoor construction workers in Washington State.* Scandinavian Journal of Work Environment & Health, 2019. 45(6): p. 588-599.
[23]. Dutta, P., et al., *Perceived heat stress and health effects on construction workers.* Indian Journal of Occupational and Environmental Medicine, 2015. 19(3): p. 151-158.
[24]. Jia, Y.A., S. Rowlinson, and M. Ciccarelli, *Climatic and psychosocial risks of heat illness incidents on construction site.* Applied Ergonomics, 2016. 53: p. 25-35.
[25]. Ojha, A., et al. *Quantifying the Implications of Humidity and Temperature on Heat Stress Exposure of Construction Workers: A Worker-Centric Physiological Sensing Approach*. 2024.
[26]. Torbat Esfahani, M., I. Awolusi, and Y. Hatipkarasulu, *Heat Stress Prevention in Construction: A Systematic Review and Meta-Analysis of Risk Factors and Control Strategies.* International Journal of Environmental Research and Public Health, 2024. 21(12).
[27]. Ma, K., et al., *Construction of a mouse model of spermatogonial stem cell transplant recipient by high-temperature heat stress.* Zhonghua nan ke xue = National journal of andrology, 2019. 25(9): p. 771-779.
[28]. Yi, W., Y. Zhao, and A.P.C. Chan. *A Tailored Artificial Intelligence Model for Predicting Heat Strain of Construction Workers*. in *IOP Conference Series: Earth and Environmental Science*. 2022. Institute of Physics.
[29]. Qi, C., et al., *Study on Hierarchical Prevention Method of Heat Hazard in Geothermal Tunnel Construction.* Chinese Journal of Underground Space and Engineering, 2024. 20(5): p. 1717-1750.

Information-Theoretic Uncertainty Quantification for Vision-Based Construction Ergonomic Assessment

Roy Lan[1] and Ibukun Awolusi[2]

[1]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
[2]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
roy.lan@my.utsa.edu, ibukun.awolusi@utsa.edu

Abstract –

Work-related musculoskeletal disorders (WMSDs) remain a critical cost driver in construction safety. While computer vision enables scalable ergonomic monitoring via automated REBA assessment, current deterministic systems suffer from hazardous overconfidence, providing definitive risk scores even when occlusion or lighting compromises data integrity. This study presents an information-theoretic framework that decomposes prediction uncertainty into aleatoric (sensor noise) and epistemic (model ignorance) components. Aleatoric uncertainty is quantified via surprisal-based confidence mapping from pose keypoints, while epistemic uncertainty is approximated through stochastic perturbation of the scoring component, consistent with standard Monte Carlo dropout in learned-model settings. Validated on construction worker frames across steel, concrete, and roofing tasks, the framework achieves 96% prediction interval coverage (target: 95%) with 0.76 REBA point mean absolute error. Task-specific analysis reveals 3.0× uncertainty variation, with roofing exhibiting 1.76 total uncertainty versus 0.59 for steel. An Operational Design Domain strategy map derived from uncertainty profiles provides quantitative deployment guidelines, distinguishing scenarios suitable for autonomous monitoring from those requiring human oversight.

Keywords –

Artificial Intelligence; Computer Vision; Construction Safety; Ergonomics; Trustworthy; Uncertainty Quantification

1 Introduction

Work-related musculoskeletal disorders remain a major burden in construction, driving substantial days-away-from-work cases and associated costs. [1]. Traditional observational ergonomic assessment methods, such as the Rapid Entire Body Assessment (REBA), provide validated frameworks for postural risk evaluation [2]. However, manual assessment is labor-intensive, subjective, and infeasible for continuous real-time monitoring across large construction sites [3].

Recent advancements in computer vision (CV) offer a scalable alternative. Markerless pose estimation frameworks, such as MediaPipe and OpenPose, enable automated skeletal keypoint detection from standard RGB video [2, 4, 5]. This capability has driven a surge in automated ergonomic assessment tools that map geometric joint angles directly to risk scores [2, 3]. While these systems demonstrate promising accuracy in controlled settings, their deployment in real-world construction environments reveals a critical limitation: they lack mechanisms to quantify prediction reliability [6, 7].

Current ergonomic monitoring systems are deterministic, they produce point estimates (e.g., "REBA Score: 4") regardless of input data quality [1, 6]. Consider a roofing worker partially occluded by scaffolding. A standard system may detect only 60% of body keypoints yet interpolate the missing joints to predict "Low Risk." This prediction receives identical treatment as a clearly visible steel worker with 95% keypoint visibility, despite vastly different measurement confidence. The inability to distinguish between "genuinely low risk" and "low risk due to missing data" undermines trust and prevents deployment in liability-sensitive construction environments [3, 6].

A potential mitigation approach would require a shift from deterministic scoring to probabilistic risk assessment [8, 9]. Bayesian deep learning distinguishes between two fundamental uncertainty sources. Aleatoric uncertainty represents irreducible measurement noise from environmental factors such as occlusion, motion blur, and lighting degradation. Epistemic uncertainty captures reducible model ignorance when encountering postures outside the training distribution [10]. Decomposing these components enables targeted

improvement strategies: high aleatoric uncertainty indicates sensor limitations requiring better camera placement, while high epistemic uncertainty reveals This work addresses that trust gap by shifting from deterministic scoring to probabilistic ergonomic risk assessment. We leverage the standard distinction in Bayesian deep learning between two sources of uncertainty [10, 11]. Aleatoric uncertainty reflects irreducible variability driven by measurement and environment (occlusion, blur, distance), while epistemic uncertainty reflects reducible model uncertainty when encountering postures or conditions underrepresented in training[11]. Disentangling these components is operationally useful because they imply different mitigation actions. High aleatoric uncertainty motivates sensing or viewpoint improvements, while high epistemic uncertainty motivates targeted data collection or model adaptation.

Accordingly, this study presents an uncertainty-aware framework that augments pose-based REBA assessment with calibrated reliability signals suitable for construction deployment. The key contributions are:

1. Dual-uncertainty estimation for pose-to-REBA pipelines, combining an information-theoretic mapping from pose keypoint confidence to aleatoric uncertainty using surprisal, and a stochastic perturbation strategy to approximate epistemic uncertainty in the current scoring implementation, with direct Monte Carlo dropout interpretation in learned scoring models.
2. Empirical construction uncertainty profiles, validated on 2,041 expert-labeled frames spanning three task archetypes (steel erection, concrete work, and roofing), establishing task-specific uncertainty signatures and their decomposition.
3. A quantitative Operational Design Domain strategy map, which translates uncertainty coordinates into deployment guidance by stratifying tasks into "Safe for Automation," "Human-in-the-Loop," and "Do Not Deploy" regions based on validated uncertainty thresholds.

2 Methodology

2.1 System Architecture

Figure 1 illustrates the uncertainty-aware REBA assessment pipeline. The system processes construction video frames through four stages: (1) MediaPipe pose estimation extracts 33 keypoint locations with per-joint confidence scores ci ∈ [0,1]; (2) A rule-based REBA calculator computes point predictions $\mu REBA$ following ISO 11226 guidelines [2]; (3) Two parallel uncertainty quantification streams estimate aleatoric (σ_a) and epistemic (σ_e) uncertainty independently; (4) A fusion module combines uncertainty components via Pythagorean sum (Equation 4), followed by isotonic calibration to ensure 95% empirical coverage.

The architecture is detector-agnostic. Any pose estimation backbone providing confidence scores can substitute MediaPipe (e.g., OpenPose, HRNet, ViTPose), and any REBA scoring algorithm (rule-based or learned) can replace our calculator. The uncertainty quantification layers operate independently of these components. This modularity enables retrofitting existing ergonomic systems without architectural modifications.

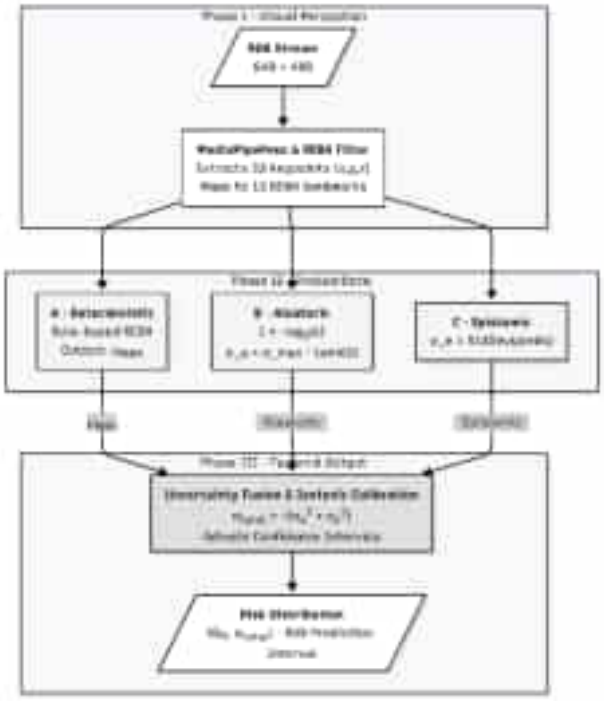


Figure 1. Uncertainty-aware REBA assessment pipeline

2.2 Aleatoric Uncertainty Quantification

Aleatoric uncertainty captures irreducible measurement noise from occlusion, motion blur, and lighting degradation. We ground this quantification in information theory. For $keypoint\ i$ with confidence c_i, the surprisal (self-information) quantifies the information deficit:

$$I(c_i) = -log_2(c_i) \quad (1)$$

High surprisal indicates low confidence, reflecting ambiguous measurements. This formulation has intuitive properties: (1) Perfect confidence (c_i = 1) yields zero surprisal (I = 0 bits), indicating no information deficit; (2) Low confidence ($c_i = 0.5$) produces I = 1 bit of surprisal, signaling binary uncertainty; (3) The logarithmic scale naturally compresses the [0,1] confidence range into an unbounded information metric.

We aggregate surprisal across N REBA-relevant keypoints (shoulders, elbows, hips, knees, ankles) and map to standard deviation via bounded nonlinearity:

$$\sigma_{\mathrm{a}} = \sigma_{\max} \cdot tanh(1/N \, \Sigma_{\mathrm{i=1}} \, I(c_{\mathrm{i}})) \quad (2)$$

Where σ_{max} is a scaling parameter controlling maximum uncertainty (validated empirically in Section 3.2), and $tanh$ provides bounded output [0, σ_{max}] preventing unbounded growth. The $tanh$ nonlinearity reflects physical reality: uncertainty saturates when data becomes completely unreliable rather than growing indefinitely. Unlike arbitrary confidence scaling [7], this approach quantifies uncertainty in units of REBA points scaled to the categorical structure of the assessment (15-point scale with integer scores).

The keypoint selection strategy warrants discussion. REBA assessment requires joint angles at the neck, trunk, legs, and arms. We filter MediaPipe's 33 keypoints to the 13 REBA-relevant landmarks: left/right shoulders, elbows, wrists, hips, knees, ankles, plus nose (head proxy). Facial landmarks and hand keypoints are excluded as REBA scoring does not depend on finger positions or facial features. This filtering reduces noise from irrelevant keypoints while preserving all biomechanically essential measurements.

2.3 Epistemic Uncertainty Quantification

Epistemic uncertainty captures model ignorance for postures outside the training distribution. We employ Monte Carlo (MC) dropout [12], performing T stochastic forward passes through the REBA scoring model with dropout activated at inference:

$$\sigma_{\mathrm{e}} = \sqrt{(1/T \, \Sigma_{\mathrm{t=1}}{}^{\mathrm{T}} \, (\hat{\mathrm{y}}_{\mathrm{t}} - \bar{\mathrm{y}})^2)} \quad (3)$$

Where $\hat{y}t$ is the REBA prediction from pass t, and $\bar{y}$ is the mean prediction across T passes. We use T = 30 based on pilot studies showing variance convergence beyond this threshold. The dropout rate is set to p = 0.2, applied to intermediate feature representations in the REBA calculator. Intuitively, MC dropout asks the scoring component the same question multiple times under small internal perturbations. If the predicted REBA scores remain consistent across passes, the model is confident. If they vary substantially, the input is likely outside the model's familiar operating region, resulting in higher epistemic uncertainty.

For novel postures (e.g., extreme roofing angles unseen in training), prediction variance increases. Critically, epistemic uncertainty can be reduced by collecting additional training data, unlike aleatoric uncertainty, which is bounded by sensor capabilities. This distinction enables strategic improvement: high-aleatoric tasks require better sensors or viewpoints, while high-epistemic tasks require more training examples.

It is important to note that for rule-based REBA calculators utilized in this study, dropout is applied following angle computation but prior to the application of categorical thresholds. This simulates uncertainty in the rule boundaries themselves. For learned models (e.g., neural network REBA predictors), dropout is applied directly to hidden layers following standard practice [12].

2.4 Uncertainty Fusion and Calibration

Assuming independent error sources, total uncertainty combines via Pythagorean sum:

$$\sigma_{\mathrm{total}} = \sqrt{(\sigma_{\mathrm{a}}{}^2 + \sigma_{\mathrm{e}}{}^2)} \quad (4)$$

We adopt independence as a practical approximation for uncertainty fusion. Aleatoric uncertainty primarily reflects measurement degradation from sensing conditions such as occlusion and blur, whereas epistemic uncertainty reflects limitations in data support and scoring-model knowledge. Although some coupling may arise under severe occlusion, the Pythagorean fusion provides a simple and interpretable combination rule, and post-hoc calibration with the observed empirical coverage indicates that this approximation is adequate for the present dataset.

The system outputs a Gaussian distribution $N(\mu_{REBA}, \sigma_{total})$ from which 95% prediction intervals are derived as $[\mu - 1.96\sigma, \mu + 1.96\sigma]$. However, raw uncertainty estimates often exhibit systematic bias (over- or under-confidence). We apply post-hoc isotonic regression [14] to calibrate intervals to 95% empirical coverage. Isotonic calibration monotonically adjusts σtotal values to achieve target coverage without compromising sharpness (interval width). This differs from temperature scaling, which applies a global multiplicative factor; isotonic regression learns a non-parametric mapping $\sigma_{cal} = f(\sigma_{total})$ preserving relative uncertainty rankings.

The calibration uses a dedicated 400-frame holdout set reserved exclusively for isotonic fitting. The remaining 1,641 frames constitute the test set on which all metrics in Section 3 are evaluated.

2.5 Experimental Setup

2.5.1 Dataset Description

To evaluate the proposed framework, we collected a dataset of 2,041 video frames from active construction sites. Frames were sampled at 2 Hz during 30-minute recording sessions to reduce temporal dependence while maintaining sufficient task coverage. Ground-truth REBA labels were assigned by two expert raters with sufficient training and domain knowledge of the REBA protocol, and inter-rater agreement is reported in Table 1.

Of the 2,041 frames, 400 were reserved for isotonic calibration holdout, while the remaining 1,641 frames constitute the test set.

Table 1. Dataset composition and computer vision challenge characteristics

Task Type	Frames	GT Source	Dominant CV Challenge
Steel Erection	681	Expert REBA (κ=0.89)	Baseline (clear visibility)
Concrete Work	680	Expert REBA (κ=0.87)	Motion blur, dust occlusion
Roofing	680	Expert REBA (κ=0.85)	Extreme angles, partial occlusion
Total	2,041	κ=0.87 overall	Multi-condition validation

2.5.2 Task Selection Rationale

To ensure the generalizability of our findings, we categorize the construction trades not merely by job title, but by their computer vision risk archetype. This taxonomy allows us to evaluate the framework's robustness against specific failure modes:

1. The Baseline Archetype (Steel Erection): Characterized by upright, static postures with clear lines of sight. This serves as the control group, representing ideal deployment conditions where deterministic models typically perform well.
2. The Dynamic Archetype (Concrete Work): Characterized by rapid, repetitive movements and environmental particulates (dust). This archetype stress-tests the model's ability to handle Aleatoric Uncertainty arising from motion blur and Epistemic Uncertainty arising from complex biomechanical loads.
3. The Geometric Archetype (Roofing): Characterized by severe self-occlusion (crouching, kneeling) and non-standard camera angles (overhead or steep inclination). This represents the "Edge Case," stress-testing the information-theoretic surprisal mechanism (Eq. 2) under conditions of high data loss.

2.5.3 Evaluation Metrics

Three metrics were used to assess uncertainty: Mean Absolute Error (MAE) evaluated point prediction accuracy in REBA points (refer to Equation (5)), Prediction Interval Coverage Probability (PICP) measured the proportion of ground truth scores within 95% intervals, with a target of $\geqslant$95% (see Equation (6)), and Expected Calibration Error (ECE) assessed calibration quality across confidence bins with lower values indicating better alignment between predicted confidence and empirical accuracy.

$$MAE = 1/N \sum_{i=1}^{N} |\hat{y}_i - y_i| \quad (5)$$

$$PICP = 1/N \sum_{i=1}^{N} \mathbb{1}[y_i \in [\hat{y}_i - 1.96\sigma_i, \hat{y}_i + 1.96\sigma_i]] \quad (6)$$

where N = 2,041 test frames, $\hat{y}_i$ is the predicted REBA score, y_i is the ground truth, and $\mathbb{1}[\cdot]$ is the indicator function. These metrics capture the accuracy-uncertainty trade-off: perfect point predictions with zero uncertainty ($\sigma = 0$) achieve low MAE but fail PICP (coverage = 0%), while excessively conservative estimates ($\sigma \to \infty$) achieve PICP $\geq$ 95% but produce useless intervals spanning the entire 15-point scale.

We also report interval sharpness (mean width) as a secondary metric. Narrow intervals are preferable when coverage is maintained, as they provide more actionable guidance to practitioners.

3 Results

3.1 Overall Performance

Table 2 presents aggregate performance metrics across the 2,041-frame test set. The framework achieves 96.0% prediction interval coverage, exceeding the nominal 95% target by 1.0 percentage point. This marginal over-coverage indicates well-calibrated uncertainty estimates with slight conservatism. Mean absolute error (MAE) measures 0.76 REBA points, demonstrating that point predictions remain accurate while providing probabilistic bounds.

Table 2. Overall performance metrics

Metric	Value
Prediction Interval Coverage (PICP)	96.0%
Mean Absolute Error (MAE)	0.76 REBA pts
Expected Calibration Error (ECE)	0.075
Median Total Uncertainty (σ_{total})	1.09 REBA pts
Mean Interval Width	4.34 REBA pts

Median total uncertainty of 1.09 REBA points translates to mean prediction intervals spanning 4.34 points (calculated as $2 \times 1.96\sigma$). On the 15-point REBA scale, this represents 28.9% coverage, indicating intervals are neither trivially wide (which would sacrifice utility) nor dangerously narrow (which would miss ground truth values). The 96% empirical coverage validates that our intervals are properly sized: tightening them to 4.0 points would reduce coverage below the 95% safety threshold, while widening to 5.0 points would provide marginal coverage gains (97.2%) at the cost of reduced actionability.

The expected calibration error of 0.075 indicates well-calibrated predictions. This metric measures the absolute difference between predicted confidence and

empirical accuracy across binned predictions. Values below 0.10 are generally considered acceptable for safety-critical applications [8]. Our ECE demonstrates that reported uncertainties reflect true prediction reliability rather than arbitrary inflation factors.

3.2 Hyperparameter Selection and Validation

The σ_{max} parameter in Equation 2 controls the maximum aleatoric uncertainty magnitude. We evaluated candidate values from 1.5 to 5.0 REBA points using a 400-frame holdout calibration set (20% of data, excluded from all other analyses). Figure 2 illustrates the coverage-precision trade-off.

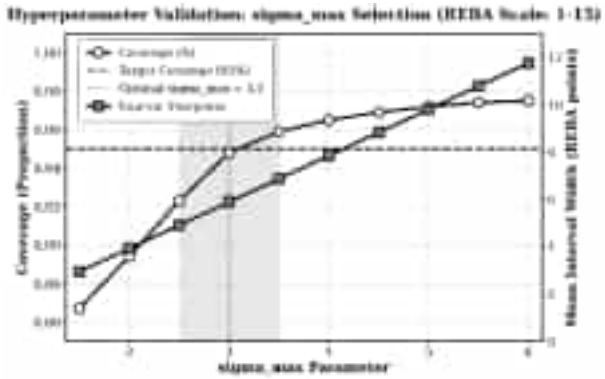


Figure 2. Hyperparameter validation for σ_{max} showing the trade-off between prediction interval coverage (safety) and interval width (precision). The optimal value σ_{max} = 3.0 balances 95% coverage with minimal interval width.

At σ_{max} = 3.0, the system achieves 95.0% coverage on the holdout set with 7.0 REBA point average pre-calibration width. Post-calibration via isotonic regression reduces this to 4.34 points on the test set while maintaining 96% coverage. Values below 2.5 produce unsafe under-coverage (<92%), as insufficient uncertainty bounds fail to capture prediction variability. Values above 4.0 create excessively conservative intervals (>9 points raw width) with diminishing coverage returns: increasing from 4.0 to 5.0 yields only 1.8% additional coverage while inflating intervals by 40%.

This knee-point response at σ_{max}= 3.0 demonstrates that the parameter is empirically grounded rather than arbitrary. The Pareto frontier analysis reveals that 3.0 provides the optimal balance: sufficient coverage for safety assurance without excessive conservatism that would render predictions non-actionable. Sensitivity analysis shows the selection is robust to modest perturbations (σ_{max} = 2.8 or 3.2 produces <2% coverage variation).

3.3 Task-Specific Uncertainty Decomposition

Figure 3 reports the decomposition of predictive uncertainty by construction task archetype. The results show strong task heterogeneity, with total uncertainty increasing from steel (σ_{total} = 0.59 REBA points) to concrete (σ_{total} = 1.09) and roofing (σ_{total} = 1.76). Roofing exhibits approximately 3.0× higher total uncertainty than steel (1.76 versus 0.59). Pairwise two-sample t-tests with Bonferroni correction for three comparisons (α = 0.0167) confirm that all task differences are statistically significant ($p < 0.001$ for each comparison).

Steel erection represents the most favorable deployment condition, exhibiting balanced contributions from aleatoric and epistemic uncertainty (σ_a = 0.42, σ_e = 0.36). This profile is consistent with high pose observability and stable viewing conditions, indicating that the system operates close to its baseline performance envelope in this archetype.

Concrete work shows a marked increase in uncertainty driven primarily by the aleatoric component (σ_a = 0.96, σ_e = 0.44). Relative to steel, aleatoric uncertainty increases by approximately 2.3× (0.96 versus 0.42), while epistemic uncertainty increases modestly. This pattern indicates that measurement degradation, rather than model insufficiency, is the dominant factor in concrete environments.

Roofing exhibits the highest uncertainty and is strongly aleatoric-dominated (σ_a= 1.64, σ_e= 0.58). The large σ_a relative to steel indicates that task geometry and visibility constraints are the primary drivers of unreliability, while the elevated σ_e suggests a smaller, secondary contribution from less frequently observed postural configurations. Overall, the decomposition isolates the dominant uncertainty source by task and provides a principled basis for deployment decisions and targeted mitigation strategies.

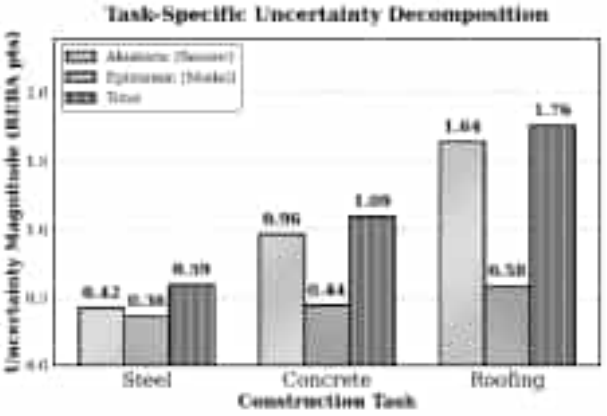


Figure 3. Task-specific uncertainty decomposition across construction activities. Roofing exhibits 3.0× higher

aleatoric uncertainty (1.64) compared to steel (0.42) due to geometric occlusion and extreme viewing angles.

3.4 Calibration Quality Assessment

Figure 4 evaluates uncertainty calibration using a reliability diagram that compares predicted confidence against empirical accuracy, where accuracy is defined as the fraction of samples whose prediction error falls within 1 REBA point. The calibration curve closely tracks the identity line, yielding an Expected Calibration Error of 0.075, which indicates that the calibrated confidence values align well with observed performance. For example, a predicted confidence of 0.60 corresponds to 0.583 empirical accuracy, and a predicted confidence of 0.80 corresponds to 0.787 empirical accuracy.

Calibration deviations remain modest across the confidence range and lie within the ±10% tolerance band. The largest bin-wise deviation is 4.2%, occurring in the mid-confidence regime, indicating no systematic overconfidence or underconfidence after calibration. Confidence bins are constructed by partitioning predictions into deciles of predicted confidence, followed by computing empirical accuracy within each bin. The inset histogram further shows that most predictions concentrate at higher confidence levels, supporting stable calibration behavior in the operating region, which is most frequently encountered during inference. Task-stratified calibration remains consistent across task conditions, with ECE values of 0.062 for steel, 0.078 for concrete, and 0.081 for roofing. All three values remain below 0.10, supporting that the calibration procedure generalizes across tasks without requiring task-specific tuning.

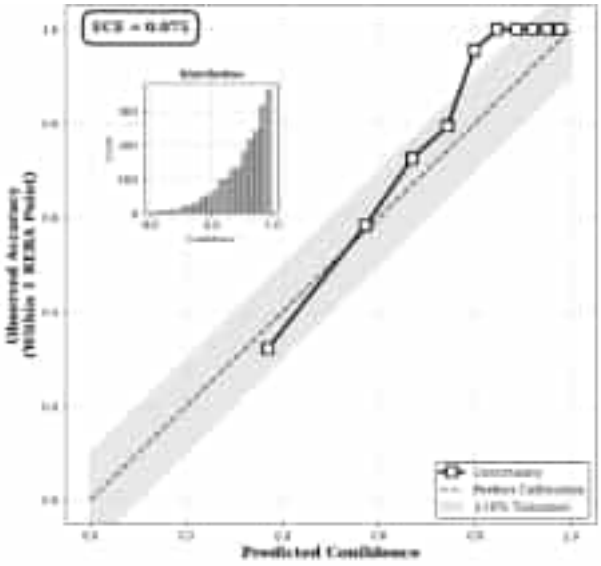


Figure 4. Uncertainty calibration reliability diagram. Expected Calibration Error (ECE) of 0.075 indicates well-calibrated predictions where reported confidence aligns with empirical accuracy.

4 Discussion

4.1 Operational Design Domain Strategy

The empirical uncertainty profiles enable evidence-based deployment stratification. We propose a three-zone Operational Design Domain framework parameterized by total uncertainty thresholds, illustrated in Figure 5. The boundaries correspond to constant total uncertainty contours in the $\sigma_a - \sigma_e$ plane, where each task archetype is represented by its centroid and a ±1 SD dispersion ellipse. The thresholds should be interpreted as dataset-grounded operating guidelines for the studied conditions rather than universal cutoffs for all construction environments. In particular, σ_{tot} = 1.0 marks the regime where predictions remain consistently reliable with limited dispersion, while σ_{tot} = 1.8 represents the high-uncertainty regime where error rates become operationally unacceptable for safety-critical decision making.

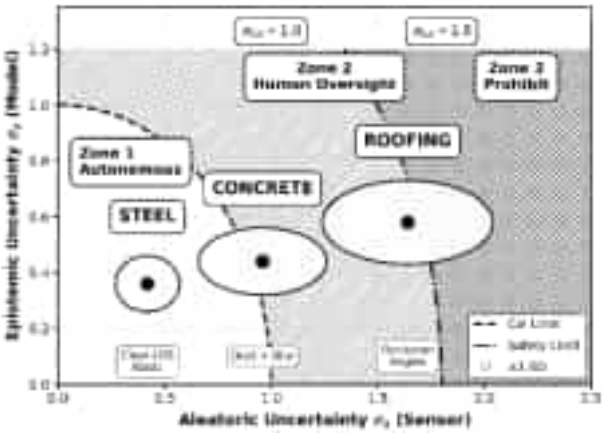


Figure 5. Operational Design Domain map in the $\sigma_a - \sigma_e$ uncertainty plane.

Zone 1: Autonomous Operation (σ_{tot} < 1.0). Zone 1 captures tasks with sufficiently low uncertainty to support autonomous monitoring without routine human oversight. Steel erection lies in this region with σ_{tot} = 0.59. This deployment classification is supported by three validation characteristics: prediction-interval coverage of 97.4%, mean absolute error of 0.52 REBA points, and a false negative rate of 2.1% for high-risk postures (REBA ≥ 8). Under this regimen, the system can support automated screening and notification workflows in which infrequent false positives are acceptable, while instances of missed high-risk events occur only rarely.

Zone 2: Supervised Operation ($1.0 \leq \sigma_{tot} < 1.8$). Zone 2 represents a human-in-the-loop regime where the system can triage frames efficiently, but expert confirmation is required for decisions with safety consequences. Concrete work occupies Zone 2 with σ_{tot} = 1.09, accompanied by reduced coverage of 95.8%, increased MAE of 0.81 REBA points, and a higher false negative rate of 4.8%. Roofing occupies the upper portion of Zone 2 with σ_{tot} = 1.76, indicating that vision-based estimates remain usable for screening in favorable viewing conditions, but operate close to the boundary where reliability degrades.

Zone 3: Prohibit Vision-Only Deployment ($\sigma_{tot} \geq 1.8$). Zone 3 defines operating conditions where vision-only assessment should be rejected due to unacceptable uncertainty. Although the roofing centroid lies just below this boundary, the dispersion and task geometry imply that a non-trivial subset of roofing conditions can cross into Zone 3, particularly under severe occlusion or extreme camera angles. In this regime, validation evidence indicates degradation consistent with unsafe classification behavior, including coverage falling below 95%, MAE rising to levels that can shift Medium Risk postures into Low-Risk categories, and a false negative rate that is operationally unacceptable.

For Zone 3 conditions, alternative sensing and workflows are required. Wearable IMU-based measurements can provide joint-angle estimates that are less sensitive to visual occlusion, while multi-view camera placement can reduce occlusion and potentially shift some scenarios back into Zone 2, subject to site-specific validation. Manual assessment remains necessary when uncertainty-based screening indicates that automated inference is unreliable.

4.2 Limitations and Future Directions

Generalization is bounded by the current evaluation setting. Validation relied on monocular video and three construction task archetypes, so the extent to which the present uncertainty thresholds transfer to additional trades, viewpoints, environmental conditions, and site layouts remains to be established. Future validation should therefore include broader task categories such as interior finishing, ladder-based work, congested indoor operations, and low-visibility or time-varying lighting conditions.

The framework was assessed with REBA, and extension to RULA, OWAS, or the NIOSH lifting equation will require revalidating the uncertainty mapping under different scoring structures. In addition, the current epistemic component is implemented as a stochastic approximation around a rule-based scoring stage, which is less directly grounded than standard MC dropout in learned models. Alternative approaches, such as ensemble-based or bootstrap-based uncertainty estimation, therefore, warrant comparison in future work. We also acknowledge that the present study does not benchmark against simpler uncertainty baselines such as confidence-thresholding or scalar post-hoc calibration methods. Such comparisons are valuable for isolating the incremental benefit of the proposed information-theoretic mapping and uncertainty decomposition and are an important direction for future work.

Active learning represents a promising extension, using high-uncertainty cases to prioritize expert labeling, although the task decomposition suggests sensing and viewpoint improvements may yield larger near-term gains where aleatoric uncertainty dominates.

5 Conclusion

This work demonstrates that uncertainty-aware REBA assessment can move vision-based ergonomics from a single deterministic score to a calibrated, decision-ready output that quantifies reliability for the evaluated construction task archetypes. Validation shows strong task dependence driven primarily by aleatoric effects under challenging geometry and visibility, and the proposed three-zone ODD framework operationalizes these findings into practical deployment rules for autonomous use, supervised review, or rejecting vision-only assessment.

References

[1] Z. Jiao, K. Huang, Q. Wang, G. Jia, Z. Zhong, and Y. Cai. Improved REBA: deep learning based rapid entire body risk assessment for prevention of musculoskeletal disorders. *Ergonomics*, 2024. doi:10.1080/00140139.2024.2306315.

[2] S. O. Jeong and J. Kook. CREBAS: Computer-based REBA evaluation system for wood manufacturers using MediaPipe. *Applied Sciences*, 13(2), 2023. doi:10.3390/app13020938.

[3] R. Lan and I. Awolusi. UAV-DL-based computer vision framework for construction worker pose estimation. *In Proceedings of the CIB World Building Congress*, 2025. https://docs.lib.purdue.edu/cib-conferences.

[4] R. Lan, I. Awolusi, and J. Cai. Computer vision for safety management in the steel industry. AI, 5(3):1192-1215, 2024. doi:10.3390/ai5030058.

[5] C. Zhou, R. Yu, and J. Kaner. Evaluating functional ability in older adults' object retrieval behavior from

kitchen furniture using OpenPose and REBA. *Scientific Reports*, 14(1), 2024. doi:10.1038/s41598-024-75470-6.

[6] D. Kee. Systematic comparison of OWAS, RULA, and REBA based on a literature review. *International Journal of Environmental Research and Public Health*, 19(1), 2022. doi:10.3390/ijerph19010595.

[7] B. Yalcin Kavus, P. Gulum Tas, and A. Taskin. A comparative neural networks and neuro-fuzzy based REBA methodology in ergonomic risk assessment: An application for service workers. *Engineering Applications of Artificial Intelligence*, 123, 2023. doi:10.1016/j.engappai.2023.106373.

[8] Y. Tao, H. Hu, F. Xu, and Z. Zhang. Ergonomic risk assessment of construction workers and projects based on fuzzy Bayesian network and D-S evidence theory. *Journal of Construction Engineering and Management*, 149(6), 2023. doi:10.1061/jcemd4.coeng-12821.

[9] A. Mumani, R. T. Stone, and A. M. Momani. An application of Monte-Carlo simulation to RULA and REBA. *Theoretical Issues in Ergonomics Science*, 22(6):673-688, 2021. doi:10.1080/1463922X.2021.1893406.

[10] A. Kendall and Y. Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Proceedings of the 31st Conference on Neural Information Processing Systems* (NeurIPS), Long Beach, CA, USA, 2017.

[11] E. Hüllermeier and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110(3):457-506, 2021. doi:10.1007/s10994-021-05946-3.

[12] Gal, Y. & Ghahramani, Z. (2016). "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning." Proc. 33rd ICML, pp. 1050–1059.

[13] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K.Q. (2017). "On Calibration of Modern Neural Networks." Proc. 34th ICML, pp. 1321–1330

[14] Zadrozny, B. & Elkan, C. (2002). "Transforming Classifier Scores into Accurate Multiclass Probability Estimates." Proc. 8th ACM KDD, pp. 694–699.

From Worker Narratives to Safety Intelligence: Extracting Construction Health Risks Using NLP and Knowledge Graphs

Karen DSouza[1], Yuting Chen[2], Gongfan Chen[2], Lingguang Song[2] and Don Chen[2]

[1]School of Science, Department of Computer Science, Purdue University, U.S.A.
[2]Albert School of Construction, Department of Engineering Technology and Construction Management, University of North Carolina at Charlotte, U.S.A.

kmdsouza@purdue.edu, ychen106@charlotte.edu, gchen12@charlotte.edu, lsong2@charlotte.edu, dchen9@charlotte.edu

Abstract -

Construction workers face multifaceted physical and psychological health risks due to challenging and often unpredictable working environments. This study introduces a computational framework that uses artificial intelligence and natural language processing to extract and structure occupational health insights from 7,441 Reddit posts and comments related to construction. Focusing on eight custom entity categories that include worker types, health symptoms, workplace stressors, institutional actions, and control measures, we applied rule-based entity recognition methods to identify recurring patterns and concerns discussed by workers. Frequency analysis highlighted prevalent health symptoms such as stress, dehydration, and exhaustion. A domain-specific knowledge graph for stressors, health symptoms and control measures revealed relationships between extracted entities, uncovering structural patterns, isolated risks, and frequently co-mentioned hazards. The framework demonstrates how informal, user-generated narratives can be transformed into structured knowledge to support occupational health surveillance. This work contributes scalable tools that can assist policymakers, unions, safety officers, and researchers in understanding evolving worker needs. By bridging Natural Language Processing (NLP) techniques with occupational safety, our method offers a new lens for data-driven intervention strategies rooted in real-world discourse.

Keywords -

Knowledge Graphs; Web Scraping, Natural Language Processing; Physical Health; Construction Workers; Risk Factors

1 Introduction

The construction industry is consistently ranked among the most hazardous occupational sectors worldwide, with workers frequently exposed to a combination of physical, environmental, and psychological stressors [1]. Despite extensive safety regulations and awareness campaigns, the dynamic and unpredictable nature of construction sites continues to pose serious health risks [2]. Traditional approaches to occupational health and safety rely heavily on formal incident reports, regulatory citations, and clinical diagnoses [3][4]. These reports often fail to capture the lived experiences of frontline workers. The sources may underreport certain types of safety issues such as physical health conditions, mental health struggles, and risk factors that lead to gaps in intervention and prevention strategies [5].

At the same time, the social digital platforms such as Reddit and X have introduced new avenues for workers to share uncensored, real-time narratives about their occupational experiences. These narratives include valuable information about risk exposure, institutional responses, and perceived gaps in workplace protections [6]. However, due to their informal and unstructured nature, such data sources are largely untapped in occupational safety research [7]. The unstructured and noisy nature of this data presents significant challenges for systematic analysis and interpretation.

Existing work in occupational health surveillance using natural language processing (NLP) has demonstrated the potential of social media to supplement official data [8]. Yet, few studies have tailored these techniques specifically to the construction context, and even fewer have attempted a structured, multi-entity analysis using domain-specific categories relevant to physical and mental health, job roles, and safety interventions. This study addresses these gaps by developing a computational framework for extracting and analyzing occupational risk narratives using NLP and named entity recognition (NER).

In recent years, NLP and knowledge representation techniques have shown promising potential in structuring large-scale unstructured text data and generative artificial intelligence [9]. Among these, knowledge graphs which as graph-based representations that encode semantic relationships among entities have emerged as powerful tools for extracting and analyzing contextual patterns from narrative data [10][11]. Coupled with transformer-based models for NER and relation extraction, knowledge graphs enable analyzable structures that reflect the complexity of

occupational discourse [12].

In this study, we propose a computational framework that leverages NLP and knowledge graph construction to analyze 7,441 Reddit posts and comments related to the construction sector. We first define a construction domain-specific schema encompassing eight key entities such as worker types, environmental stressors, health symptoms, and institutional actions. We then use these categories to extract and map relationships within the text. Our approach enables the discovery of emerging risk clusters, reveals central actors and symptoms within the discourse. The corresponding knowledge graphs highlight underreported hazards such as heat related health symptoms, and inadequate hydration practices.

2 Related Work

Occupational safety in construction has been studied extensively through traditional sources such as injury reports, surveys, and regulatory databases [13]. However, these methods often suffer from underreporting and lack of real-time insight into worker experiences. Recent research has started to explore alternative data sources, including social media platforms and online forums, as a means of capturing informal yet rich narratives about workplace conditions.

2.1 Social Media and Informal Occupational Data

The use of social media as a data source for occupational health surveillance is a growing field. Recent work done by researchers has (i) explored mining Reddit posts for mental health indicators, highlighting the platform's potential for early detection of stress and burnout. (ii) researchers applied NLP techniques to datasets from social media such as X data to monitor workplace hazards. These studies reinforce the value of informal, crowd-sourced narratives in complementing official safety reporting mechanisms.

For risk analysis, work has been done in the construction industry with researchers evaluating failure modes and effects analysis as well as highlighting the overall risks analysis [1][14][15][16].

2.2 Knowledge Graphs and Occupational Safety

Knowledge graphs have gained popularity as a structured approach for modeling relationships between concepts in unstructured text. In occupational safety contexts, these graphs have been used to map hazard domains, link injury types to causes, and integrate data from disparate sources for improved decision-making [17][18].

Domain adaptation of NER techniques has become crucial in contexts like construction, where entities may not align with those in general purpose domains. Hybrid approaches combining dictionary based methods and transformer models like BERT have proven effective in extracting occupationally relevant entities such as symptoms, worker roles, and safety measures. Beyond entity identification, relation extraction enables the construction of semantic graphs that reveal causal and correlational patterns.

3 Methodology

The data collection and processing pipeline framework is a crucial part of the project. The quality of the results depends heavily on the data processing techniques used.

3.1 Data Collection and Processing

To analyze occupational risk narratives from the construction sector, we extracted Reddit posts and comments using a keyword-based scraping strategy. Keywords were selected to reflect both domain-specific roles (e.g. construction, worker, airport) and health-related stressors (e.g. burnout, depression, heat, noise, fatigue, mental). These combinations were used to query Reddit via the PRAW API, resulting in over 130,000 subreddit matches. To increase domain relevance, filtering was applied to retain only posts from subreddits such as Construction, ConstructionManagers, HVAC, Plumbing, and civilengineering, yielding a dataset of 7,441 posts and comments.

All retrieved text was normalized to lowercase for consistency, and a manual filtering step was conducted to remove irrelevant matches, such as metaphorical uses of construction or posts unrelated to workplace safety. The dataset was filtered to retain only English-language text and exclude irrelevant or low-information content such as memes, spam, or non-descriptive titles. All text entries were cleaned to remove links, special characters, and excessive whitespace. Sentences were tokenized and split using standard NLP preprocessing techniques. This preparation was essential to ensure high-quality input for entity recognition and relation extraction in the subsequent steps.

3.2 Annotation

The next step was the data annotation process. A structured annotation protocol was applied to a corpus of 7,441 Reddit posts and comments related to construction work. The dataset consisted of worker types that included categories such as construction, postal, delivery, industrial and so on. However, the annotation focused on a key dimension: relevance to construction worker type data only. The process involved three annotators with domain knowledge to ensure robustness, minimize subjectivity, and support cross-validation.

The cross-validation annotator step was included as the last step in the annotation process. The cross-validation

notes included justifications for labeling decisions, highlighting ambiguity, context-specific interpretations, and suggestions for edge-case resolution. For instance, in one entry, physical strain was indicated by indirect language, prompting a note about interpretation boundaries.

The manual annotation serves as a ground truth foundation for subsequent entity extraction and relation modeling. It allows the computational models to be evaluated or trained with human verified categories and enables higher-quality graph construction by filtering out noise and irrelevance from informal social media narratives.

3.3 Named Entity Recognition (NER)

To extract structured knowledge from unstructured text data related to construction worker health and safety, we implemented a custom NER pipeline. Our approach relied on rule-based pattern matching, lemmatization, and curated dictionaries mapped to eight distinct entity categories relevant to occupational safety, health, and organizational behavior. We based our work on a subset of 525 narratives which were cross-validated targeting construction-specific content extracted from a larger corpus of 7,441 worker incident narratives.

Domain-specific dictionaries and spaCy's PhraseMatcher were used to supplement transformer-based models, allowing for more accurate recognition of low-frequency or colloquial phrases that may not appear in standard corpora. All entities were normalized to lowercase to reduce duplication and unify mentions. Next, we describe the methodology and criteria used for each of the eight entity categories. These categories are listed Table 1 along with representative examples.

Table 1. Entity Categories

Category	Examples
Worker Type Managers	Supervisor, project manager
Worker Type Frontline	Electrician, roofer
Health Symptoms	Fatigue, dizziness
Environmental Stressors	Heat, cold
Workplace Stressors	Long shift, understaffing
Individual Factors	Age, single
Institutional Actions	OSHA inspection, fine
Control Measures	Safety training
Additional Category	**Example**
Location	Dallas, Houston

3.3.1 Worker Types

The worker type category was divided into two subtypes: (i) frontline workers and (ii) managers. Frontline worker entities such as electrician, roofer, plumber, and laborer represent skilled tradespeople directly involved in physical construction tasks. They constituted **worker type frontline**. Managerial worker entities such as supervisor and project manager reflect administrative roles within construction settings. They were included in **worker type managers**. To improve precision and reduce ambiguity, these entity mentions were matched and normalized using occupation classification terms and examples referenced from the U.S. Bureau of Labor Statistics (BLS.gov) [19]. This alignment with standardized occupational definitions ensured that extracted roles mapped consistently to recognized construction labor segments.

3.3.2 Health Symptoms

Health symptoms were extracted using a dictionary-based lemmatized matcher for both physical and psychological health indicators. Common entities included fatigue, exhaustion, headache, dehydration, stress, and burnout. The matched symptom entities were later used in co-occurrence and knowledge graph-based analyses.

3.3.3 Environmental Stressors

Environmental stressors refer to physical conditions present on the job site. We built a controlled vocabulary around terms such as noise, heat, sunlight, chemical, dust, and air quality. Each term was identified using lemmatized rule-based matching, often in combination with contextual qualifiers. For example, sun exposure and UV radiation were collapsed under sunlight. These stressors were crucial in identifying the relationship between job site conditions and adverse health outcomes.

3.3.4 Workplace Stressors

This category focused on psychosocial and organizational stressors. Examples include harassment, trauma, overwork, pressure, and lack of support. Unlike environmental stressors, these were often mentioned in indirect language or idiomatic phrases (e.g. toxic work culture). We expanded the dictionary with variations and synonyms using pre-trained word embeddings to improve recall. Workplace stressors were shown to have strong co-occurrence with psychological symptoms such as anxiety and burnout.

3.3.5 Individual Factors

Personal vulnerabilities such as pre-existing conditions or life circumstances were grouped under this category. Detected terms included older worker, health issues, disability, and family responsibilities. These entities help provide a nuanced understanding of differential risk exposure.

3.3.6 Institutional Actions

This category captured formal institutional responses such as OSHA investigation, union involvement, citations, lawsuits, and fines. A regular expression and phrase-based matching system was used to detect regulatory terms and legal follow-ups. These entities were often connected to either severe outcomes or recurring risk scenarios and served as anchor nodes in our knowledge graph to highlight systemic issues.

3.3.7 Control Measures

We extracted procedural mitigation measures such as hydration, shade, rest breaks, safety training, and cooling measures. This category showed how actively safety protocols were implemented post-incident. Entity detection relied on a mix of procedural keywords and compliance terms sourced from OSHA and construction industry best practices.

Risk mitigation strategies were divided into 5 subcategories called elimination, substitution, engineering controls, administrative controls, personal protective equipment (PPE). Entities were tagged accordingly and used to evaluate the adequacy of mitigation strategies in each case. This tiered tagging allowed assessment of how often high-level (elimination, substitution) versus low-level PPE controls were employed.

3.3.8 Additional Categories

We identified one additional category (i) location. To preprocess location entities, geographic references within the text data were extracted using NER with spaCy's pretrained language model. The model detects entities tagged as GPE (Geo-Political Entities), which include names of cities, states, and countries. Each row of the dataset was processed to identify these locations. After extraction, the results were aggregated and filtered to compile a distinct list of frequently mentioned locations. This method allowed for both broad and granular recognition of geographic context capturing entries like Texas, Dallas and Travis County. The compiled list of location entities plays a key role in constructing a spatial layer within the knowledge graph, enabling location-aware analysis of occupational safety concerns across different regions.

Each recognized entity in the 8 categories and the 2 additional categories were stored as a list in a new column later converted into binary label vectors (1 if the entity was present in a document, 0 otherwise) for downstream modeling and correlation analysis.

3.4 Knowledge Graphs

The motivation to use knowledge graphs stems from the need to model complex, non-linear relationships in unstructured text. Reddit narratives often describe multifactorial risks, where a single worker may report a combination of symptoms caused by several stressors, within a particular environmental or organizational context. Traditional text classification or keyword-based techniques cannot represent these many-to-many, context-dependent interactions. In this context, knowledge graphs offer the following advantages in the construction domain: • Semantic structuring: Graphs preserve the relational meaning between entities, helping to reveal patterns such as co-occurrence of symptoms and institutional inaction. • Interpretability: Unlike black-box models, graphs allow human analysts and stakeholders to visually trace paths between causes and effects. • Graph analytics: We can apply network algorithms such as centrality, clustering, and shortest-path analysis to identify influential symptoms or recurring stressor combinations. • Scalability: As new data is collected, it can be continuously incorporated into the graph without retraining, enabling real-time risk tracking.

3.5 Rule-Based Symptom Extraction Performance Evaluation

To evaluate the performance of the rule-based entity extraction system, a subset of annotated data was used where relevance was confirmed by a primary annotator, resulting in 525 records. For the symptom category, binary ground truth labels were manually verified and compared against predictions generated by the rule-based matching approach. The presence of symptom-related entities was treated as a document-level classification task, where each record was labeled as either containing a symptom (1) or not (0). Precision, recall, and F1-score were computed to assess performance.

To compare rule-based extraction with a machine learning baseline, a BERT-based classifier was fine-tuned for binary health symptom extraction using the same annotated dataset. As shown in Table 3, the rule-based approach achieved perfect recall (1.00), indicating that all symptom-related instances were successfully identified. However, it exhibited slightly lower precision due to false positives. In contrast, the BERT model achieved a more balanced performance, with an F1-score of 0.91, by leveraging contextual understanding to reduce over-detection. While BERT improved precision-recall balance, the rule-based method remains advantageous for high-recall applications and interpretability. To ensure a fair comparison, both the rule-based approach and the BERT model were evaluated on the same held-out test set.

Table 2. Performance of Rule-Based Health Symptom Extraction

Class	Precision	Recall	F1-score
No Health Symptoms (0)	1.00	0.63	0.77
Health Symptoms (1)	0.89	1.00	0.94
Accuracy		0.91	
Macro Avg	0.95	0.81	0.86
Weighted Avg	0.92	0.91	0.91

Table 3. Comparison of Rule-Based and BERT Models for Health Symptoms

Model	Precision	Recall	F1-score
Rule-Based	0.89	1.00	0.94
BERT	0.87	0.96	0.91

4 Findings and Analysis

4.1 Frequency Analysis of Extracted Entities

After completing entity extraction, we analyzed the frequency of each recognized term within its respective category to identify dominant concerns and patterns in the construction domain.

For each entity column, the lists of matched keywords were flattened and aggregated to compute their frequency. We plotted the top 10 most common terms for each of the eight categories. This provided insight into which symptoms, stressors, and interventions were most frequently mentioned in the dataset. The findings from frequency analysis are described below. These frequency distributions served as the foundation for more advanced relational analysis through co-occurrence heatmaps and graph modeling. Figure 1 shows keywords like laborer, plumber, roofer and electrician among the top entities in frontline worker type. The environmental stressors in Figure 2 shows top entities like heat, cold, noise and sunlight. The top healthcare symptoms in Figure 3 shows stress, dehydration, exhaustion, and physical labor.

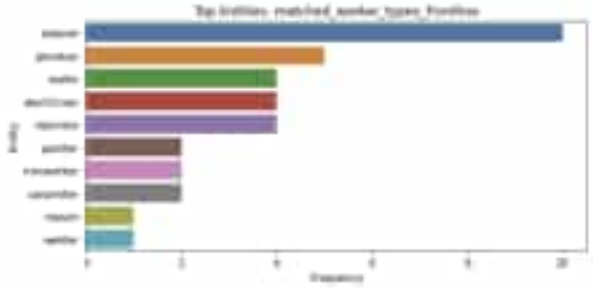


Figure 1. Frequency of Worker type frontline

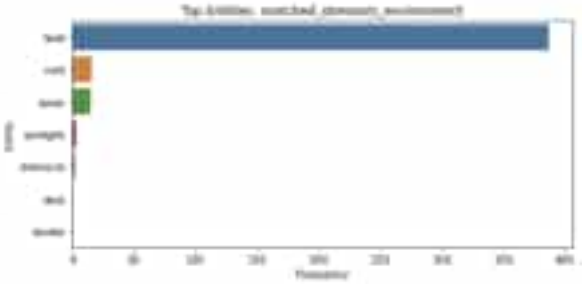


Figure 2. Frequency of Environment Stressors

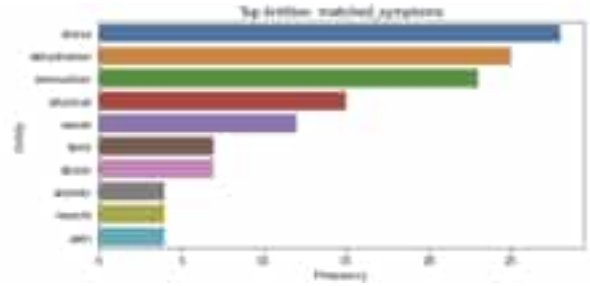


Figure 3. Frequency of Health Symptoms

4.2 Co-occurrence and Correlation Analysis

We next analyzed how frequently entities from different categories appeared together within the same document using a co-occurrence matrix. For example, we examined whether certain symptoms (e.g. burnout) were consistently mentioned alongside specific stressors (e.g. overwork or lack of support). The analysis computed pairwise co-occurrence counts between entity labels using binary-encoded columns. The resulting matrix was visualized as a heatmap, where color intensity indicates higher co-occurrence frequency.

- Heat showed the highest co-occurrence with work, reinforcing environmental causality.
- Heat correlated heavily with job, hourly.
- Heat was correlated highly with stress, injury, health.

This correlation structure informed the next step of constructing a knowledge graph to visualize entity relationships.

4.3 Insights from the Knowledge Graph

To better understand how extracted entities interact, we constructed an integrated knowledge graph linking stressors, health symptoms, and control measures. These three categories were selected to represent a conceptual pathway

from exposure to outcome to response. The knowledge graph in Figure 4 captures the co-occurrence of stressors, health symptoms and control measures in Reddit narratives, with node size indicating entity frequency and edge thickness reflecting co-mention strength. This visualization uncovers significant patterns. In this framework, stressors (e.g., heat, cold) represent environmental or organizational risk factors, health symptoms capture the resulting physical or psychological effects, and control measures represent mitigation strategies discussed in worker narratives.

By connecting these categories within a single graph, Figure 4 enables the identification of complete risk-to-response pathways rather than isolated associations. This provides a more holistic view of occupational health dynamics, illustrating how specific stressors lead to particular health outcomes and how those outcomes are addressed through practical interventions.

The integrated knowledge graph reveals several dominant stressor–symptom–control pathways.

- The most prominent pathway is heat → heat exhaustion → rest breaks, indicating that acute thermal stress is frequently associated with rest-based mitigation strategies. Rest breaks are a control measure.
- Similarly, heat → dehydration → hydration highlights the importance of water access as a primary response to heat exposure. Hydration is a control measure.
- Another recurring pathway, heat → physical exhaustion → rest breaks, suggests that prolonged exposure leads to fatigue, which is addressed through rest-based interventions.
- In addition to heat-dominated pathways, other environmental relationships are also observed. For example, the pathway sunlight → fainting → shade highlights how direct solar exposure can lead to acute physiological responses that are mitigated through access to shaded rest areas. This reinforces the importance of environmental controls alongside behavioral interventions in managing occupational heat stress.

These pathways demonstrate that worker-reported experiences follow consistent patterns linking environmental stressors to physiological responses and subsequent mitigation actions. The integrated knowledge graph reveals incidents of heat-related health risk severity reflected in worker narratives. Early indicators of thermal stress include symptoms such as sweating and fatigue, which are associated with initial mitigation measures such as cooling and rest. Symptoms such as dehydration emerge, often linked to hydration-based interventions. More severe conditions, including vomiting and physical exhaustion, indicate escalating physiological stress. In extreme cases, critical symptoms such as fainting are observed, typically associated with immediate protective measures such as access to shaded areas.

This progression highlights how occupational heat exposure manifests across multiple levels of severity and indicates the control measures that were reported in Reddit narratives.

4.4 Limitations

Reddit provides access to rich, real-time worker narratives, it has several limitations as a data source. The platform's user base may not be representative of the broader construction workforce, as it is likely skewed toward younger, more digitally engaged individuals. Additionally, the content is self-reported and may reflect subjective perceptions rather than verified incidents. As a result, the findings should be interpreted as indicative patterns of worker experiences rather than statistically representative conclusions. Despite these limitations, Reddit offers valuable insights into underreported issues such as fatigue, burnout, and psychosocial stress, complementing traditional safety reporting systems.

5 Conclusion

This work demonstrates how unstructured social media narratives can be transformed into structured occupational safety insights using a computational framework grounded in NER and knowledge graph modeling. By extracting eight domain-specific entity categories related to worker roles, symptoms, stressors, institutional actions, and mitigation strategies, we reveal risk patterns such as heat-related symptoms, fatigue, and psychological stress in construction settings. Frequency and co-occurrence analysis highlight strong relationships between environmental and workplace stressors and adverse health outcomes. The resulting knowledge graph exposes systemic gaps between health symptoms and workplace stressors. Together, these findings illustrate the value of integrating informal worker discourse into scalable occupational health surveillance and lay the groundwork for future transformer-based models to enhance precision, generalizability, and policy relevance.

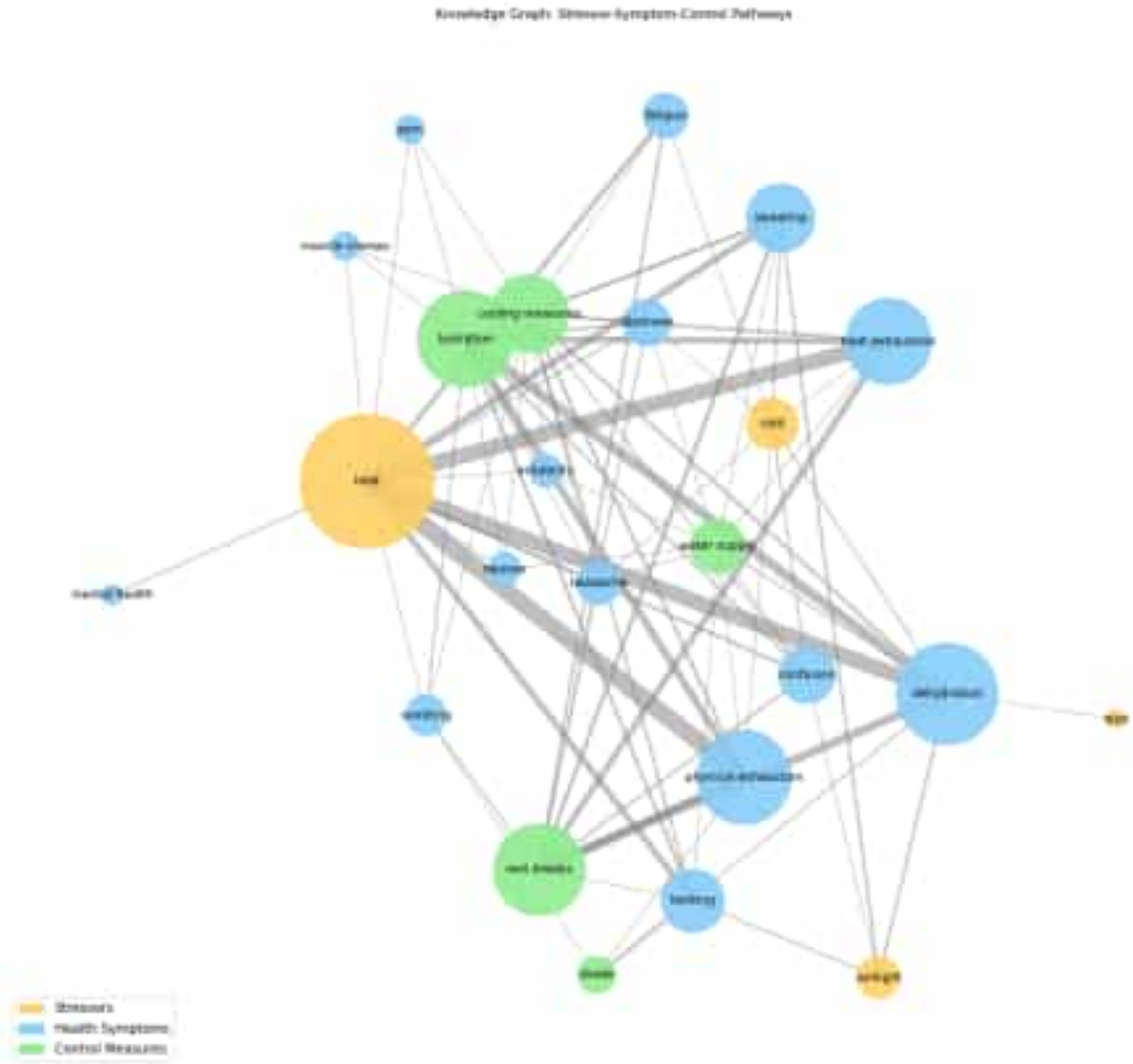


Figure 4. Knowledge Graph: Stressor-Symptoms-Control Pathways. Knowledge Graph illustrates integrated pathways between stressors (yellow), health symptoms (blue), and control measures (green). Nodes represent extracted entities, and edges indicate co-occurrence within worker narratives. The graph highlights dominant pathways such as heat → dehydration → hydration and heat → physical exhaustion → rest breaks, demonstrating how occupational risks are experienced and mitigated in practice.

References

[1] Devdatt P Purohit, NA Siddiqui, Abhishek Nandan, and Bikarama P Yadav. Hazard identification and risk assessment in construction industry. *International Journal of Applied Engineering Research*, 13(10): 7639–7667, 2018.

[2] KA Shamsuddin, Mohd Norzaimi Che Ani, AK Ismail, and MR Ibrahim. Investigation the safety, health and environment (she) protection in construction area. *International Research Journal of Engineering and Technology*, 2(6):624–636, 2015.

[3] Johan M Sanne. Incident reporting or storytelling? competing schemes in a safety-critical and hazardous work setting. *Safety Science*, 46(8):1205–1222, 2008.

[4] Heather M Zoller. Health on the line: Identity and disciplinary control in employee occupational health and safety discourse. *Journal of Applied Communication Research*, 31(2):118–139, 2003.

[5] Nadia A Al-Aubaidy, Carlos H Caldas, and Stephen P Mulva. Assessment of underreporting factors on construction safety incidents in us construction projects.

International journal of construction management, 22(1):103–120, 2022.

[6] Munmun De Choudhury and Scott Counts. Understanding affect in the workplace via social media. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 303–316, 2013.

[7] JY Teoh. How occupational therapy practitioners use virtual communities on the facebook social media platform for professional learning: A critical incident study. *Scandinavian Journal of Occupational Therapy*, 29(1):58–68, 2022.

[8] Michael Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 265–272, 2011.

[9] Alexis Figueroa-Baltazar and Anh D Chau. Leveraging generative ai to analyze construction industry trends through social media data. In *Computing in Civil Engineering 2025: Computational and Intelligent Technologies*, pages 36–45.

[10] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54 (4):1–37, 2021.

[11] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514, 2021.

[12] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L Opdahl, and Csaba Veres. Named entity extraction for knowledge graphs: A literature overview. *IEEE access*, 8:32862–32881, 2020.

[13] Zhipeng Zhou, Yang Miang Goh, and Qiming Li. Overview and analysis of safety management studies in the construction industry. *Safety science*, 72:337–350, 2015.

[14] Hao-Tien Liu and Yieh-lin Tsai. A fuzzy risk assessment approach for occupational hazards in the construction industry. *Safety science*, 50(4):1067–1078, 2012.

[15] Abel Pinto, Isabel L Nunes, and Rita A Ribeiro. Occupational risk assessment in construction industry–overview and reflection. *Safety science*, 49(5):616–624, 2011.

[16] Leslie Edwards and Leslie J Edwards. *Practical risk management in the construction industry*. Thomas Telford, 1995.

[17] Qi Yao, Rita Yi Man Li, Lingxi Song, and M James C Crabbe. Construction safety knowledge sharing on twitter: A social network analysis. *Safety science*, 143:105411, 2021.

[18] Wenjing Wu, Caifeng Wen, Qi Yuan, Qiulan Chen, and Yunzhong Cao. Construction and application of knowledge graph for construction accidents based on deep learning. *Engineering, construction and architectural management*, 32(2):1097–1121, 2025.

[19] U.S. Bureau of Labor Statistics. Standard occupational classification (soc) system, 2018. URL `https://www.bls.gov/soc/2018/home.htm`. Accessed: 2026-01-31.

Multi-Camera and BIM-Based Benchmarking of Building Evacuation Parameters

Diran Yu[1], Abdul-Majeed Mahamadu[1], Weiwei Chen[1,2], Yang Su[1], Fujia Lyu[1], Jieyu Chen[3], Jiaxu Huang[3], Nan Li[3]

[1] The Bartlett School of Sustainable Construction, University College London (UCL), UK
[2] Division D: Civil Engineering, Engineering Department, University of Cambridge, UK
[3] Department of Construction Management, Tsinghua University, China
diran.yu.22@ucl.ac.uk, a.mahamadu@ucl.ac.uk, weiwei.chen@ucl.ac.uk, yang.su@ucl.ac.uk, fujia.lyu.24@ucl.ac.uk, chenjiey22@mails.tsinghua.edu.cn, hjx21@mails.tsinghua.edu.cn, nanli@tsinghua.edu.cn

Abstract –

As VR evacuation training and performance-based egress design increasingly rely on simulation, their credibility depends on empirical, space-specific motion benchmarks. However, VR evacuation research often lacks empirical, area-specific motion parameters for calibration. We demonstrate benchmark values from a full-scale multi-floor drill, reconstructed from multi-camera video as 2D trajectories in lobbies and corridors and 3D trajectories on stairs. An automated pipeline with YOLOv11 detection, DeepSORT tracking, and BIM-based calibration provides kinematics and crowd metrics, including speed, acceleration, local density, flow q = ρv, and route choice. Corridors enabled the fastest travel with a mean of about 2.1 m/s. The lobby was slower at about 1.2 m/s with faster peripheral lanes. Stairs were slowest at about 0.52 m/s on landings and 0.39 m/s on flights. The fundamental diagram by local density ρ reveals three systems: free flow ρ<3 persons/m², collective flow 3⩽ ρ ⩽5, and congested flow ρ >5. As ρ approaches 3 – 4, accelerations decay while mean speed remains stable. At ρ ⩾5, speeds decline, and stop-and-go waves appear. Bottlenecks commonly formed at doors and stair portals, and exit usage was uneven. The Evacuation Parameter Benchmark Library derived here supports calibration of VR drills and agent-based models and informs performance-based egress design grounded in measured behaviour. The findings will be interpreted as building-specific empirical benchmarks derived from a moderate-scale drill, rather than universally transferable evacuation parameters.

Keywords –

Evacuation Simulation, Computer Vision, Pedestrian Dynamics, Trajectory Analysis, Building Safety, Data Benchmarking

1 Introduction and Background

1.1 Motivation and Scope

Immersive VR and agent-based evacuation simulation are increasingly used for training, design evaluation, and research because frequent full-scale multi-floor drills are costly, disruptive, and difficult to repeat under controlled conditions [1–3]. However, the credibility of VR-based findings depends strongly on calibration: the microscopic rules that govern walking, turning, merging, and vertical movement are sensitive to local geometry (width changes, corners, doors, landings), visibility, and interaction opportunities.

Reliable motion benchmarks that are space-specific are essential for calibrating walking, turning, and interaction in VR evacuation studies. Many implementations still apply a single speed cap or a uniform density–speed curve across heterogeneous spaces, which induces bias between lobbies, corridors, and stairwells [4–8]. Multi-camera tracking, person re-identification, and crowd-density estimation now enable reusable parameters to be extracted from real drills [4–8]. Nevertheless, consistent coordinates across cameras and floors, rigorous error control, and area-wise calibration are still underdeveloped, which limits transferability and credibility in VR and agent-based models [4,7]. The present study analyses multi-floor drill data, reconstructs 2D trajectories in lobbies and corridors and 3D trajectories on stairs, and provides multi-area benchmarks that can be used directly for model calibration.

1.2 Key Concepts and Terminology

A particle-to-collective view of pedestrian dynamics is adopted. Local density ρ in persons per square meter and mean speed v define the flow q through q=ρv.

Operating regimes are partitioned by ρ to align with engineering practice. Free flow corresponds to $\rho < 3$ with limited interactions. Collective flow corresponds to $3 \leqslant \rho \leqslant 5$ with coordination and speed regulation. Congested flow corresponds to $\rho > 5$, where impediments and stop-and-go may occur [4,6–8]. Stair geometry motivates a stairway slowdown factor, typically yielding lower speeds on flights than on landings. Lobbies and corridors often exhibit edge preference with faster peripheral lanes along walls or railings [5,6,8]. The density–speed–acceleration relations later support the identification of acceleration decay zones and speed drop thresholds and serve as parameter inputs for VR scenes.

1.3 Related Work and Research Gap

Video-driven crowd research has progressed along three questions: who is present, where people move, and how many are there. Multi-camera detection and re-identification enable continuous trajectories across views and provide quantifiable kinematics for building evacuation [4,5,7]. Machine-vision density estimation delivers local crowding without markers and supports capacity evaluation and risk assessment [8]. Detector families continue to evolve, and the latest YOLOv11 offers improved accuracy–efficiency trade-offs for large-scale automated video processing [9–11]. Despite these advances, space-resolved distributions of speed and acceleration for lobbies, corridors, and stairwells remain insufficiently documented. A unified spatial frame that merges multi-camera trajectories, BIM coordinates, and density–speed–acceleration relations with transparent quality controls is still uncommon [4,7].

1.4 Positioning and Contributions

This work serves as a multi-area evacuation benchmarking resource for VR calibration and performance-based egress design. Four contributions are delivered. First, an automated pipeline integrates multi-camera video with BIM alignment and outputs reusable 2D and 3D trajectories with quality records [4,5,7]. Second, speed and acceleration profiles are reported for lobbies, corridors, and stairs with landings separated from flights, and a stairway slowdown together with an edge-preference index is quantified. Third, a density-based fundamental diagram is constructed. Acceleration decays markedly as ρ approaches 3–4 while mean speed remains stable. When ρ reaches 5 or higher, speeds decline, and stop-and-go emerges, which indicates a congestion threshold [4–6,8]. Fourth, an Evacuation Parameter Benchmark Library is compiled that covers reaction delay, distributions of speed and acceleration, and cross-area references for density and flow to support parameter inputs and sensitivity analyses in VR and agent-based models.

2 Methodology

Figure 1 summarises the overall workflow, which consists of pre-processing, in-processing, and post-processing.

2.1 Pre-processing

Regarding governance and ethics, video arises from a multi-floor drill under a signed data-sharing memorandum and a low-risk ethics approval. Outputs are de-identified and reported as aggregates and spatial summaries [4–8].

In the scene and BIM frame, lobbies, corridors, and stairwells are covered. The BIM model supplies a common building frame and geometric semantics for camera calibration, area masks, and path topology annotation [4–6].

Thirdly, about camera and time alignment, intrinsics and extrinsics are compiled. When missing, extrinsic parameters are estimated from control points on reference imagery. Streams are synchronised and quality-checked for sharpness and duplication to ensure cross-view comparability.

Furthermore, in the consideration of clip filtering and reduction, motion-present segments are extracted first to reduce empty frames and total duration, which increases signal-to-noise and throughput downstream.

In addition, concerning the issues of spatial calibration and unification plan, planar areas adopt 2D homographs that map pixels into local plane coordinates. Stairwells adopt 3D calibration by sampling control points on landings and sloped flights and by fitting plane equations to support 3D projection and path recovery. Local coordinates are transformed into the building frame via BIM transforms. Robust estimation and reprojection-error thresholds enforce geometric consistency [12].

Lastly, about plenty of parameters, a configuration file fixes detector confidence, NMS, minimum track duration, caps for speed and acceleration, smoothing windows, and the cadence for density estimation. This enables consistent reruns across batches.

2.2 In-processing

To achieve pedestrian detection and multi-object tracking, a modern YOLO detector provides frame-wise person boxes. DeepSORT performs online tracking with appearance embeddings and a Kalman motion model. Dual thresholds on IoU and cosine distance limit identity drift and mismatches [9–11,13,14]. Dense periods use frame decimation and input-size adaptation to balance accuracy and efficiency.

To solve the issue caused by Cross-camera, within a unified timeline, tracklets are stitched across adjacent views using BIM topology constraints and spatiotemporal proximity. Re-association is permitted only within reachable topology and feasible walking windows to curb false merges [4–6].

To process and coordinate the unification need to set. Thus, tracks are first projected to local planes or stair slopes and then mapped to the building frame. RANSAC and reprojection-error thresholds guard geometric consistency and route suspect tracks to review [12].

Continue to trajectory cleaning and physical plausibility, speeds and accelerations are computed by central differences on equal-interval samples and denoised with a 0.5 s moving average [15]. Physical caps remove artefacts, with a speed cap near 4.0 m/s and an acceleration cap near 5.0 m/s. Short occlusions are bridged by multi-hypothesis interpolation with uncertainty recorded. Rules for backtracking, maximum temporal gap, boundary crossing, and abnormal heading change trigger human audits.

The final stage in in-processing, quality logs and versioning are essential. Each batch records detector settings, association rates, retained-track ratios, ID-switch rates, and geometric error distributions. Logs and metadata accompany data packages for comparison and traceability.

2.3 Post-processing analytics

2.3.1 Measures and visualisations

This subsection states the selection principles for measures and visualisations to achieve low bias, portability across scenes, and stable behaviour at high densities while preparing outputs for VR and agent-based calibration.

- Local density uses Voronoi personal space or a fixed-radius kernel. Voronoi yields lower scatter under heterogeneous layouts. Fixed-radius kernels simplify implementation and cross-scene comparison [16].
- Speed and acceleration are computed by central differences with Savitzky–Golay or moving-average smoothing to suppress pixel jitter and tracking noise while retaining temporal resolution [15].
- Uncertainty is expressed with quantile bands and bootstrap confidence intervals to handle skew and heavy tails beyond mean–SD summaries [16,17].
- Bivariate relations use hexbin or equal-frequency bins to limit overplotting. Kernel density estimation is optional when a continuous field is required [16,18].
- Structured indicators include a stair slowdown

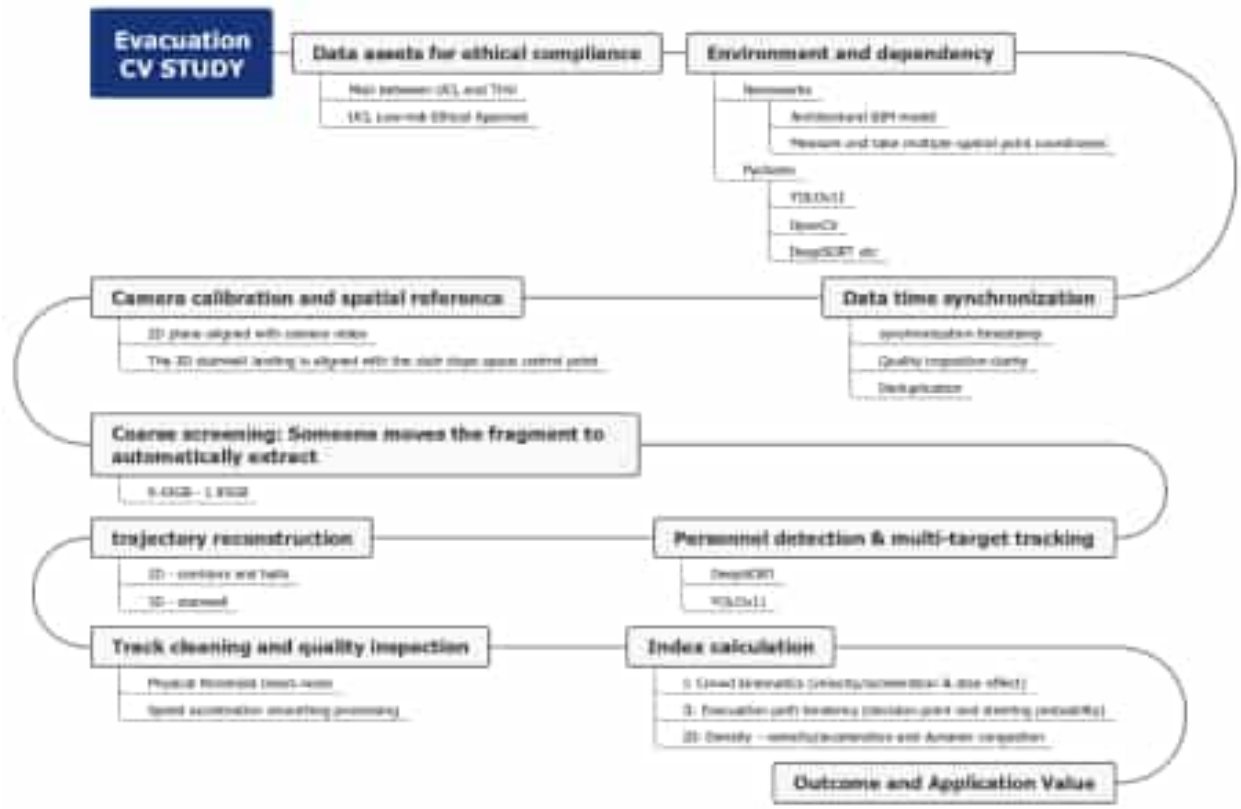


Figure 1 Workflow of Evacuation Computer Vision Tracking Study

factor, an edge-preference index, and a bottleneck score for calibration and sensitivity studies. No thresholds or regime cuts are preset in Methods. They are determined from the data in Results.

2.3.2 Unified formula block

A minimal set of formulas supports three result streams: crowd kinematics (area-wise speed and acceleration), evacuation path tendency (cell-based speed/acceleration heatmaps to reveal decisions and turns), and density–velocity/acceleration with dynamic congestion (1s, 1x1 m2 counts linked to kinematics). Thus, the core formulas are as follows:

$$v_i(t_k) = \frac{[x_i(t_{k+1}) - x_i(t_{k-1})]}{2\Delta t} \quad (1)$$

$$a_i(t_k) = \frac{[v_i(t_{k+1}) - v_i(t_{k-1})]}{2\Delta t} \quad (2)$$

$$\bar{v}_c = mean\{\, ||v_i|| : x_i \; in \; cell \; c \,\} \quad (3)$$

$$\bar{a}_c = mean\{\, ||a_i|| : x_i \; in \; cell \; c \,\} \quad (4)$$

$$\rho_c(t) = \frac{N_c(t;\, \Delta t = 1\; s)}{A_c} \quad (5)$$

$$q_c(t) = \rho_c(t) \cdot \bar{v}_c(t) \quad (6)$$

Use (1)–(4): Direct inputs for crowd kinematics by area and for cell-wise speed/acceleration heatmaps (decision/turning zones).

Use (5)–(6): Density axis at one-second cadence and local flow to build density–speed/acceleration relations and identify regimes in Results.

Symbols and parameters:

- x_i, v_i, a_i: position, speed, acceleration vectors
- A_c: cell area (e.g., 0.2 m × 0.2 m)
- $N_c(t;\, \Delta t = 1\; s)$: unique headcount in cell c within [t, t + 1 s)
- $\bar{v}_c$, $\bar{a}_c$: cell-wise mean speed and acceleration magnitude (quantiles may be used for robustness)
- $\rho_c(t)$: one-second density (persons/m²); $q_c(t)$: local flow

2.3.3 Parameterisation and reproducibility

This subsection summarises parameter settings and reproducibility rules so that outputs remain comparable and traceable across batches in time granularity, smoothing strength, and visualisation strategies.

- One-second density cadence balances frame-level jitter against pedestrian interaction timescales and aligns with common 0.5–1.0 s decision steps in VR or ABM [15,16].
- Smoothing window for kinematics is set near 0.5 s to reduce quantisation noise without lagging peaks materially [15].
- Density-axis cap is fixed at 8 persons per square meter to cover extreme congestion yet avoid unstable extrapolation during fitting [16–18].
- Distributional reporting uses quantile regression and bootstrap to generate trend bands and confidence intervals that remain comparable across floors and areas under skew and heavy tails [16,17].
- Visualisation robustness prefers hexbin or equal-frequency bins to reduce overplotting and sampling bias. KDE is optional for continuous fields, and legends report windows, bandwidths, sample sizes, and exclusion rules for interpretability and reruns [16,17].
- No preset regimes. Thresholds such as free or congested are not specified in Methods. Any segmentation is data-driven and reported in the results after sensitivity checks.

3 Results

This section reports multi-camera trajectory patterns and area-wise kinematics, distils density–movement relations and behavioural tendencies, and delivers reusable benchmarks for calibration.

3.1 Crowd kinematics: area benchmarks

Table 1 compiles representative parameters for calibration and cross-area comparison. Values result from multi-camera trajectory reconstruction with centralised differencing and unified smoothing; density cadence is one second to balance jitter suppression and interaction timescales.

Table 1 Key kinematic parameters by area

Area / Scenario	Mean speed v (m/s)	Mean |a| (m/s²)	Notes
Corridors (2F–4F)	2.072	0.079	Straight corridors permit brief running; 4F start-up means ≈ 1.2; 2F nearer the exit shows higher sustained speeds.
Lobby (ground floor)	1.233	0.071	Merging and route choice broaden the spread; door zones remain slow.
Stair landings	0.524	-0.088	Faster than flights with turning and re-entry at corners.
Stair flights	0.390	0.001	≈25% lower than landings; slowdown factor ≈ 0.75.

3.2 Evacuation path tendency: grid heatmaps

BIM-aligned 0.2 × 0.2 m tiles aggregate speed and

acceleration to expose spatial structure. The lobby exhibits a fast perimeter lane and a slow door basin, indicating boundary following and exit coupling. Corridors display a start-up burst to mid-segment dip to pre-door adjustment pattern. Stairwells show speed peaks at landings and acceleration pulses at corners and door thresholds, confirming the landing–flight rhythm. Lobby, stairs and corridor's path trajectory visualisation are shown in Figure 2 to Figure 6.

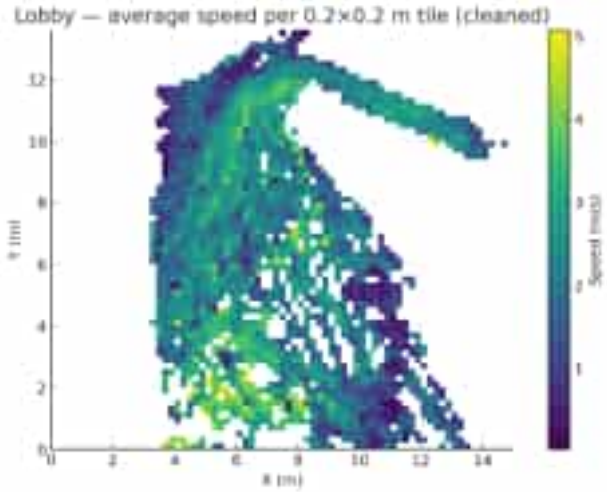


Figure 2 Lobby—average speed per 0.2 × 0.2 m tile (cleaned). Faster edge lane and a slow zone within ~1 m of the exit.

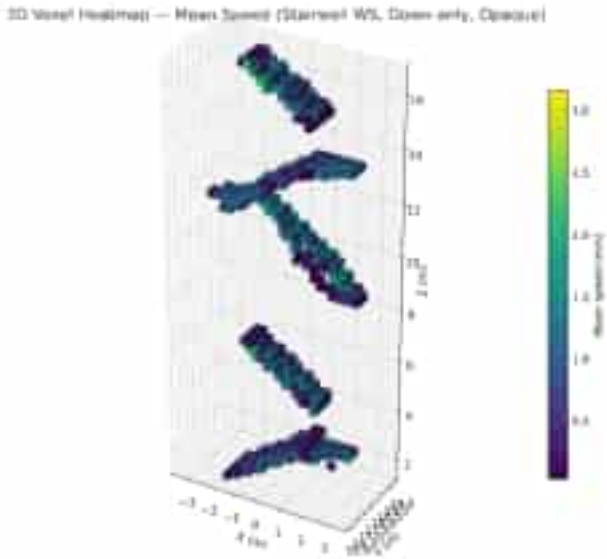


Figure 3 Stairwell WS—3D voxel heatmap of mean speed (down only). Periodic landing–flight texture with faster landings

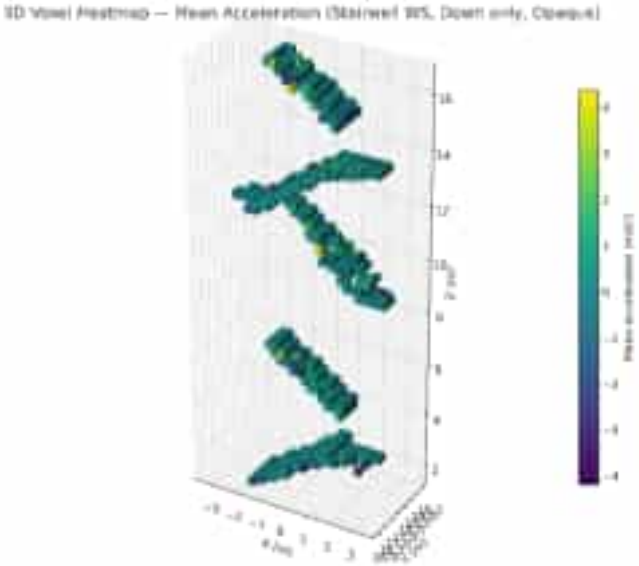


Figure 4 Stairwell WS—3D voxel heatmap of mean acceleration (down only). Pulses at landings and door thresholds indicate braking and restarting.

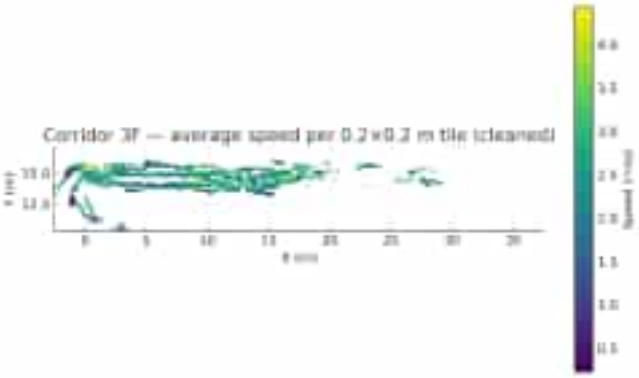


Figure 5 Corridor 3F—average speed per 0.2 × 0.2 m tile. Mid-segment steady; fluctuations near the stair door.

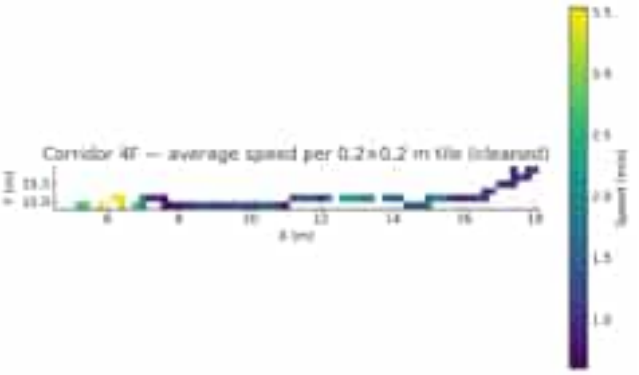


Figure 6 Corridor 4F—average speed per 0.2 × 0.2 m tile. Short start-up acceleration, slowdown near the wall, and accelerate again to the stair area.

3.3 Density–velocity/acceleration and dynamic congestion

Local density ρ is defined as the unique headcount per square meter within one second and is linked to tile-level speed and acceleration. The results are shown in Figure 7 to Figure 8 (Lobby) and Figure 9 to Figure 10 (Stairs).

Lobby. A strong negative density–speed relation is observed: the median speed is ~1–1.5 m/s at ρ≈1–2, falls to ~0.5–0.8 m/s at ρ≈4, and in micro-cells near the exit at ρ≈6–7, speeds are often < 0.5 m/s or momentarily zero. The acceleration–density relation shows maximal variability at moderate density ρ≈2–4, consistent with frequent avoidance and short stop–start movements.

Stairs. The speed–density decline is steeper than on flat areas. Even at ρ≈3–3.5 on landings, speeds converge to ~0.3–0.4 m/s, reflecting tread spacing and safety-distance constraints. Combined with acceleration trends, the evidence indicates that around ρ≈3–4 acceleration dampens markedly while speed changes little, consistent with controlled flow, and that for ρ≥5 speed drops sharply with intermittent movement—consistent with congested flow.

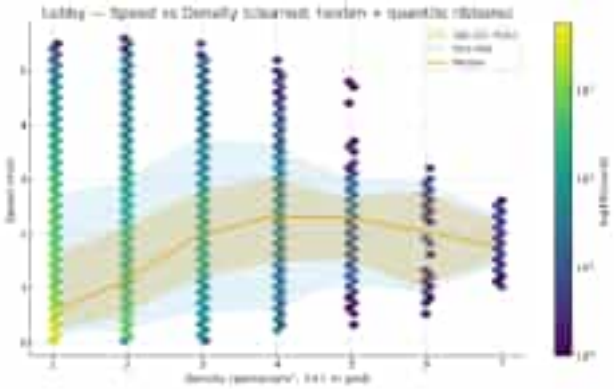


Figure 7 Lobby—Speed vs Density (hexbin with quantile ribbons).

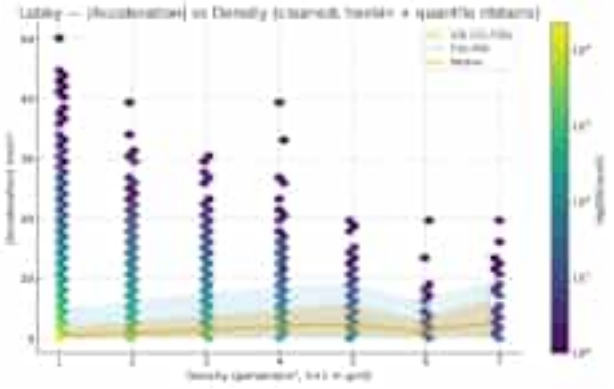


Figure 8 Lobby—|Acceleration| vs Density (hexbin with quantile ribbons).

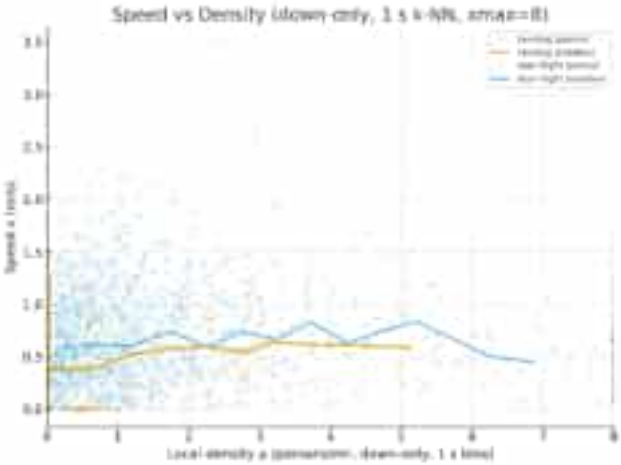


Figure 9 Stairs—Speed vs Density.

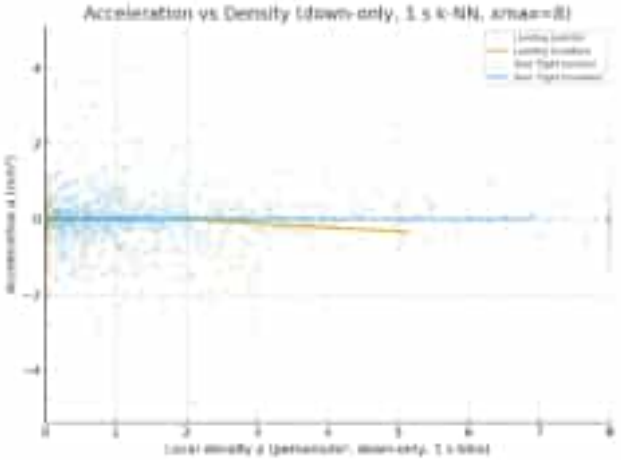


Figure 10 Stairs—Acceleration vs Density.

4 Discussion

The results show pronounced spatial dependence of movement. Corridors support high speeds. The lobby slows under merging and choice. Stairs exhibit a stable landing–flight contrast. Heatmaps (Figures 2–6) reveal fast perimeter lanes, turning nodes, and doorway coupling. Density vs. speed and density vs. acceleration relations (Figures 7 – 10) are consistent with three regimes: free flow ($\rho < 3$ persons/m²), controlled flow ($3 \leqslant \rho \leqslant 5$), and congested flow ($\rho > 5$). Table 1 provides reusable kinematic benchmarks for calibration and design.

4.1 Spatial movement patterns and congestion

Area-specific priors should replace uniform parameters. Table 1 gives representative means for corridors, lobby, and stairs that can anchor prior distributions and validation targets. The stair slowdown

factor (~0.75 from landing to flight) parameterises staircase submodels. Speed – density curves should be calibrated by area, with slope and intercept estimated from Figures 7 – 10. Quantile trends are preferable to mean-only fits to capture heavy tails and heterogeneity. Around $\rho \approx 3 - 4$, acceleration attenuates while speed changes little. This pattern identifies controlled flow and should be represented by lower stochastic acceleration and higher temporal persistence in velocity updates. Above $\rho \geqslant 5$, speed drops sharply and stop – go emerges, which calls for stronger interaction forces and tighter headway rules.

Path tendency maps (Figures 2–6) provide spatial priors for decision nodes and turn costs. Perimeter bands should carry a lower movement cost to reproduce edge preference, while door cells should carry a higher cost to reproduce slow zones. The landing–flight rhythm in the stair voxel maps supports a periodic turn–align–re-enter sequence with short acceleration pulses that can be encoded as state transitions. Local co-occupancy in corridors must be interpreted with care. Two-person tiles on 2F reflect brief overtakes rather than sustained crowding and should be modelled as transient events, not capacity increases. Route imbalance between twin stairs indicates familiarity and accessibility bias. These biases should be explicit in route choice utilities. Reaction delay should follow a right-skewed distribution with short upper-floor delays and a long lower-floor tail.

4.2 Design and operational implications

Throughput is governed by stair–corridor coupling. Speed troughs and acceleration pulses cluster at landings and door thresholds (Figures 3–4 and 5–6 near stair doors). Priority actions include clearing the first step beyond each door, increasing landing turning radius where feasible, and improving sight lines at corners. The lobby shows a fast perimeter lane and a slow door basin (Figure 2). Centre guidance markings and floor cues can draw flow off walls and reduce corner convergence. Doorway slow zones warrant checks of clear width, door swing, and obstacle placement. Twin-stair imbalance highlights that adding exits does not guarantee balanced use. Training and dynamic signage can redistribute demand in real time. High corridor speeds raise fall and collision risk during emergencies. Non-slip finishes, local slowdown cues, and staff positioning at corners and exits can mitigate these risks.

4.3 Limitations and future work

The drill involved about ninety participants and reflects moderate loading. Extreme congestion appears in micro-cells near the exit. Extrapolation to extremely high densities should be cautious. Occlusion and camera geometry can cause occasional track loss. Spatial quantisation and smoothing may damp noticeably short peaks in speed or acceleration. Geometry is site-specific, including stair width, tread rise–run, and doorway layout, so cross-building transfer requires recalibration. The benchmark represents the first step toward a VR evacuation parameter library. Expansion is needed across occupancies and layouts, broader building typologies, and outdoor domains such as forecourts and sidewalks. Higher-density drills should bound performance in congested regimes. Standardised schemas, quality control, and uncertainty reporting will support reuse. With progressive release of parameters and visual assets, calibrated immersive studies can feed validated insights back into performance-based egress design and operations.

5 Conclusion

An automated multi-camera pipeline aligned to BIM reconstructed 2D/3D evacuation trajectories across corridors, lobby, and stairs. Results deliver a reusable, multi-area benchmark for VR and agent-based calibration. Corridors supported higher speeds (overall mean $\approx$ 2.07 m/s). The lobby averaged $\approx$ 1.23 m/s under merging and choice. Stairs showed a stable geometric effect with means $\approx$ 0.52 m/s on landings and $\approx$ 0.39 m/s on flights, implying a slowdown factor $\approx$ 0.75. Density and speed were negatively related beyond $\approx$ 2 persons/m^2. Acceleration variability peaked at moderate density. Edge preference and doorway slowdowns dominated open areas. Twin stairs showed route imbalance. Reaction times were right-skewed with short upper-floor delays and a lower-floor long tail.

These outputs constitute the first step toward a VR evacuation parameter benchmark library. Area-specific speed and acceleration distributions provide priors and validation targets. The stair slowdown factor parameterises staircase submodels. Density–speed and density–acceleration curves set quantitative goals for interaction strength and stochastic noise. Edge preference and door penalties can be encoded in path-cost fields to reproduce observed streamlines. The intended outcome is a practical feedback loop where calibrated immersive studies inform real-world egress planning and operations.

Current limits include moderate loading, site-specific geometry, and potential occlusions. Next steps expand coverage across different occupancies and layouts, additional interior zones such as atria and large assembly spaces, and outdoor domains including forecourts and sidewalks. Higher-density drills will bound congested-regime performance. Standardised schemas, quality control, and uncertainty reporting will support reuse. Progressive public release of parameters and visual assets will grow the library into a shared reference for performance-based design and for immersive

technologies that return validated insights to practice. Accordingly, the contribution lies not in universal parameter claims, but in a reproducible workflow for deriving space-specific benchmarks from real drills.

Acknowledgement

This work was supported by the Horizon Europe Project, D-HYDROFLEX (Grant Number 101122357), the Horizon Europe Project, INHERIT (Grant Number 101123326), and the ISPF British Council funded EyeISENS project. This research was supported in part by the National Key R&D Program of China (2024YFE0114400) and supported by UCL BSSC Conference Funding.

Reference

[1] Cimellaro GP, Domaneschi M, de Iuliis M, Villa V, Caldera C, Cardoni A. Fire emergency evacuation in a school building through VR. vol. 1, 2019, p. 2046–55.

[2] Yu D, Mahamadu Abdul-Majeed, Chen, Weiwei. Integration of Immersive Technology and Digital Twins for Emergency Management in Buildings: A Systematic Literature Review 2024. https://doi.org/10.13140/RG.2.2.11050.84160.

[3] Yu D, Chen W, Meng T, Yu H, Mahamadu A-M. Applications of Immersive Technologies in Construction Computing: A Systematic Literature Review. Computing in Construction 2024;5.

[4] Ding Y, Chen X, Zhang Y, Huang X. Smart building evacuation by tracking multi-camera network and explainable Re-identification model. Engineering Applications of Artificial Intelligence 2025;148:110394. https://doi.org/10.1016/j.engappai.2025.110394.

[5] Zhang X, Chen X, Ding Y, Zhang Y, Wang Z, Shi J, et al. Smart real-time evaluation of tunnel fire risk and evacuation safety via computer vision. Saf Sci 2024;177. https://doi.org/10.1016/j.ssci.2024.106563.

[6] Li S, Tong L, Zhai C. Extraction and modelling application of evacuation movement characteristic parameters in real earthquake evacuation video based on deep learning. Int J Disaster Risk Reduct 2022;80. https://doi.org/10.1016/j.ijdrr.2022.103213.

[7] Ding Y, Chen X, Wang Z, Zhang Y, Huang X. Human behaviour detection dataset (HBDset) using computer vision for evacuation safety and emergency management. Journal of Safety Science and Resilience 2024;5:355–64. https://doi.org/10.1016/j.jnlssr.2024.04.002.

[8] Huang S, Ji J, Wang Y, Li W, Zheng Y. A machine vision-based method for crowd density estimation and evacuation simulation. Safety Science 2023;167:106285. https://doi.org/10.1016/j.ssci.2023.106285.

[9] Xue Z, Kong L, Wu H, Chen J. Fire and Smoke Detection Based on Improved YOLOV11. IEEE Access 2025;13:73022–40. https://doi.org/10.1109/ACCESS.2025.3564434.

[10] Jiang Y, Liu P, Han Y, Xiao B. YOLOv11-CHBG: A Lightweight Fire Detection Model. Fire 2025;8:338. https://doi.org/10.3390/fire8090338.

[11] Reddy VSR, Chintala KS, M Amsaprabhaa, S Vidhusha. A Vision-based Gait Analysis and Fall Detection Using YOLOv11 Model. 2025 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India: IEEE; 2025, p. 1–7. https://doi.org/10.1109/WiSPNET64060.2025.11004995.

[12] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 1981;24:381–95. https://doi.org/10.1145/358669.358692.

[13] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric. 2017 IEEE international conference on image processing (ICIP), IEEE; 2017, p. 3645–9.

[14] Kalman RE. A new approach to linear filtering and prediction problems 1960.

[15] Savitzky Abraham, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal Chem 1964;36:1627–39. https://doi.org/10.1021/ac60214a047.

[16] Silverman BW. Density estimation for statistics and data analysis. Routledge; 2018.

[17] Koenker R, Bassett G. Regression Quantiles. Econometrica 1978;46:33. https://doi.org/10.2307/1913643.

[18] Lewin-Koh N. Hexagon binning: an overview 2024.

Evaluating Carbon Payback Period in Parametric Facade Optimization Using Genetic Algorithms and Multi-Criteria Decision-Making

Dian Sapitri[1], Yun-Tsui Chang[2], Shang-Hsien Hsieh[1]
[1]Department of Civil Engineering, National Taiwan University, Taiwan
[2]School of Architecture, Feng Chia University, Taiwan
r13521621@ntu.edu.tw, ytchang@caece.net, shhsieh@ntu.edu.tw

Abstract

Facade shading systems can reduce cooling energy demand in office buildings located in tropical-humid climates, but the additional materials required may also increase embodied carbon emissions. Although many parametric facade optimization studies focus on operational energy and daylight performance, the temporal relationship between embodied carbon and operational carbon savings remains insufficiently explored. Carbon Payback Period (CPP) is therefore rarely integrated into parametric facade optimization and decision-making workflows. This study proposes an integrated framework combining Life Cycle Assessment (LCA), CPP, multi-objective optimization, and Multi-Criteria Decision-Making (MCDM) to support low-carbon facade design. A parametric facade model was developed for a medium office building in Taipei, Taiwan, with variations in window-to-wall ratio, glazing type, shading depth, and window configuration. Energy and daylight performance were simulated using Honeybee and Ladybug, while Wallacei was used to optimize Energy Use Intensity (EUI), Annual Sunlight Exposure (ASE), Whole-Life Carbon (WLC), Carbon Payback Period (CPP), and Spatial Daylight Autonomy (sDA). The resulting 20 Pareto-optimal alternatives were then evaluated using TOPSIS under multiple weighting scenarios. The results revealed clear trade-offs between operational efficiency and life-cycle carbon performance. Building Type 02 emerged as the most robust and balanced alternative, ranking first across the energy-priority, carbon-priority, and balanced scenarios and achieving a positive CPP of 8.95 years. These findings highlight the value of CPP as a complementary time-based indicator for verifying whether additional embodied carbon can be offset by future operational carbon savings in facade design decision-making.

Keywords –

Parametric facade design; Carbon payback period; Life cycle assessment; Genetic algorithms; Multi-criteria decision-making; Tropical-humid climate

1 Introduction

Global warming remains a major global challenge and has accelerated the use of life cycle assessment (LCA) to quantify emissions and support SDG 13 and long-term net-zero targets[1]. In the building sector, particularly in office buildings located in humid tropical climates, high cooling demand makes facade design a critical strategy for reducing operational energy use. In Taipei's hot-humid context, facade measures such as high-performance glazing and external shading can improve thermal and visual performance, but they may also increase embodied carbon.

Recent advances in parametric modeling, building performance simulation, and genetic algorithms have enabled the systematic exploration of facade alternatives based on indicators such as Energy Use Intensity (EUI), Spatial Daylight Autonomy (sDA), and Annual Sunlight Exposure (ASE)[2]. However, many studies still primarily focus on operational performance and give limited attention to embodied carbon throughout the building life cycle. Although LCA has increasingly been applied in building evaluation, aggregated carbon indicators often fail to reflect the temporal relationship between additional embodied carbon and future operational carbon savings. In this regard, the Carbon Payback Period (CPP), defined as the time required for operational carbon savings to offset additional embodied carbon, provides an important time-based perspective[3]. Yet CPP remains rarely integrated into parametric facade optimization workflows.

At the same time, multi-criteria decision-making (MCDM) methods such as TOPSIS are widely used to rank Pareto-optimal solutions under different decision priorities[4]. However, most facade-related MCDM studies still emphasize energy, cost, and daylight metrics, while time-based carbon indicators, such as CPP, receive limited attention. This gap is particularly relevant in humid tropical and subtropical climates such as Taipei,

where the interactions among shading, glazing, daylight availability, and life-cycle carbon performance remain insufficiently explored.

Therefore, this study proposes an integrated LCA-CPP-MCDM framework that combines parametric facade modeling, Wallacei-based multi-objective optimization, and TOPSIS evaluation for an office building in Taipei. CPP is incorporated in the optimization stage as a time-based carbon objective, but it is excluded from the final TOPSIS ranking because negative CPP values represent no-payback conditions rather than stable and directly comparable ranking criteria. Instead, CPP is used as a post-evaluation indicator to verify whether the highly ranked alternatives also achieve meaningful carbon payback performance. This approach provides a more comprehensive and time-sensitive basis for low-carbon facade design in humid tropical climates.

2 Literature Review

2.1 Parametric Design and Multi-Objective Optimization for Facade Performance

Facade design strongly influences building performance by affecting solar heat gain, daylight availability, visual comfort, and energy use. In tropical office buildings, key variables such as window-to-wall ratio (WWR), glazing type, and shading depth must be carefully balanced because improving one aspect of performance may reduce another. Facade design should therefore be treated as a multi-criteria problem rather than a single-objective task[5].

Parametric design helps manage this complexity by defining geometric and performance-related properties as adjustable variables. When integrated with building performance simulation, it supports the rapid generation and evaluation of design alternatives during the early design stage. Because facade objectives are often conflicting, multi-objective optimization (MOO) is widely used to identify Pareto-optimal solutions rather than a single best design. In this context, Wallacei, which implements NSGA-II in Grasshopper-based workflows, has become a relevant tool for facade optimization[6].

However, most parametric multi-objective optimization studies still focus mainly on operational indicators, particularly energy use and daylight performance, with limited attention to embodied carbon and broader life-cycle carbon trade-offs. This limitation is also evident in tropical office building studies, where optimization frameworks often prioritize operational efficiency and indoor environmental quality over long-term carbon implications. As a result, life-cycle carbon indicators, especially time-based measures such as Carbon Payback Period (CPP), remain insufficiently integrated into facade optimization. To address this gap, the present study incorporates LCA-based metrics, including Whole-Life Carbon (WLC) and CPP, into facade optimization and design decision-making.

2.2 Life Cycle Assessment in Parametric Design Workflows

Life Cycle Assessment (LCA) is increasingly important in building evaluation because it captures environmental impacts beyond operational energy use[7]. In facade design, this is particularly relevant because strategies that improve energy and daylight performance may also increase embodied carbon through additional materials, such as glazing, framing, and shading systems.

To support early-stage design, parametric Life Cycle Assessment (pa-LCA) links parametric models with environmental impact calculations, enabling rapid comparison of design alternatives[8]. However, its application remains limited by data uncertainty, interoperability issues, and the computational burden of repeated quantity take-off and carbon calculations. Accordingly, this study adopts a partial life-cycle scope based on EN 15978, covering A1–A3 for embodied carbon and B6 for operational energy use[9]. This boundary is appropriate for early-stage facade evaluation, where detailed life-cycle data are still unavailable. Within this framework, Whole-Life Carbon (WLC) is used to assess combined embodied and operational emissions, while Carbon Payback Period (CPP) captures the time dimension of carbon performance[10].

2.3 Carbon Payback Period (CPP)

Carbon Payback Period (CPP) is a time-based indicator that measures how long it takes for the additional embodied carbon caused by a design intervention to be offset by annual operational carbon savings[11]. In facade design, CPP is particularly relevant because variables such as window-to-wall ratio (WWR), glazing type, and shading depth affect both embodied and operational carbon, thereby creating a trade-off between upfront and long-term environmental impacts. CPP is therefore useful for evaluating whether a facade strategy can deliver meaningful carbon benefits within a relevant time frame[12].

A key strength of CPP is its ability to introduce a temporal perspective into carbon assessment. A facade alternative may reduce operational energy use and even achieve lower Whole-Life Carbon (WLC), yet still require a long time to offset its initial embodied carbon burden when annual carbon savings are limited[13]. In this way, CPP helps distinguish between solutions that provide near-term carbon benefits and those whose advantages only emerge over longer time horizons. This distinction is important because earlier carbon reductions

are generally more valuable for climate change mitigation than delayed ones[14].

Despite its relevance, CPP remains underused in parametric facade optimization studies, where most research continues to focus on energy use, operational carbon, or Whole-Life Carbon[15]. Therefore, this study adopts CPP as a complementary indicator to strengthen life-cycle carbon evaluation and support early-stage facade design decision-making. Positive CPP values indicate a valid carbon payback duration, meaning that the additional embodied carbon can be offset over time through annual operational carbon savings. Negative CPP values do not represent a valid payback duration; rather, they indicate that carbon payback is not physically meaningful under the given condition. This occurs when annual operational carbon savings are absent or insufficient to offset the initial embodied carbon increment.

2.4 Facade Optimization in Tropical-Humid Climates

Facade optimization in tropical-humid climates requires climate-specific evaluation because high solar radiation, temperatures, humidity, and daylight availability create performance conditions distinct from those in temperate regions[11]. In Taipei, facade design influences cooling demand, glare risk, and daylight distribution, making variables such as window-to-wall ratio (WWR), glazing type, and shading geometry critical to performance. These trade-offs are commonly assessed using metrics such as sDA, ASE, and DGP, which reflect daylight adequacy, excessive sunlight exposure, and visual comfort. Although previous studies in hot-humid climates have emphasized operational energy and comfort performance, life-cycle carbon and carbon payback remain less explored. Therefore, this study extends climate-responsive facade optimization by integrating energy, daylight, glare, and life-cycle carbon metrics for an office building in Taipei.

2.5 Research Gap

The reviewed literature demonstrates substantial progress in parametric facade optimization, life-cycle assessment integration, and multi-criteria decision-making. However, three critical gaps remain:

(1) First, CPP is rarely integrated into parametric optimization workflows and is typically applied only as a post-evaluation metric.

(2) Second, existing LCA-based facade studies rely heavily on aggregated carbon indicators, overlooking the temporal relationship between embodied and operational emissions.

(3) Third, tropical-humid climates remain underrepresented in integrated LCA–optimization studies, particularly those addressing facade shading systems.

To address these gaps, there is a clear need for an integrated framework that embeds CPP within parametric facade optimization and decision-making processes, enabling designers to evaluate not only energy and daylight performance but also the time-dependent carbon effectiveness of facade interventions. This study responds to this need by proposing a combined LCA–CPP–MCDM framework applied to a tropical-humid office building case study.

3 Method

This study adopts an integrated, multi-stage framework (see Figure 1) combining parametric facade modeling, life cycle assessment (LCA), carbon payback period (CPP) analysis, genetic algorithm optimization, and multi-criteria decision-making (MCDM). The overall workflow is structured into four sequential stages: (1) literature review and framework setup, (2) parametric modeling and simulation, (3) multi-criteria decision-making, and (4) validation and final decision support.

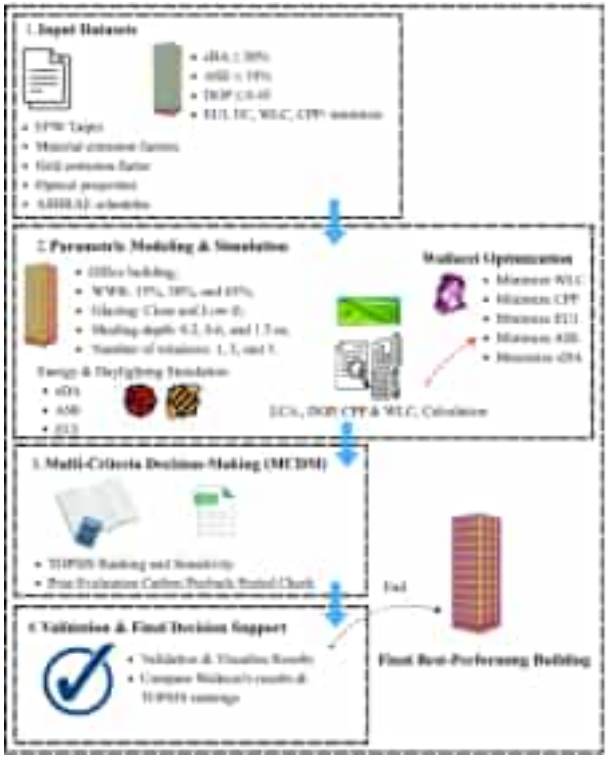


Figure 1. Research Workflow

3.1 Stage 1: Input Datasets

In the first stage, the study defines the input datasets, analytical boundaries, and performance targets for the integrated facade optimization framework. The case study is a medium office building in Taipei, where solar

radiation, cooling demand, and daylight availability strongly influence facade performance. The main inputs include the Taipei EPW weather file, material emission factors, grid emission factors, optical properties, and ASHRAE schedules, which together support energy, daylight, and carbon assessment. The facade is defined by four key variables: window-to-wall ratio (WWR), glazing type, shading depth, and number of windows. At this stage, the performance criteria are established as sDA ≥ 50%, ASE ≤ 10%, and DGP ≤ 0.45, while EUI, embodied carbon (EMB), WLC, and CPP are treated as indicators to be minimized.

3.2 Stage 2: Parametric Modeling and Simulation

In the second stage, a parametric facade model was developed in a visual programming environment by varying the window-to-wall ratio (WWR), shading depth, glazing type, and number of windows within predefined ranges, as summarized in Table 1.

Table 1. Building Geometry & Simulation Setup

No	Category	Value
1.	Building Type	Medium Office Building
2.	Location & Climate	Taipei, Taiwan; Tropical-Humid
3.	Geometry	15 × 15 m plan, 12 floors, floor height 4.0 m
4.	Spatial Model	Single zone per floor
5.	Orientation	North, East, South, and West
6.	Facade Variables	WWR: 15% / 30% / 65%; Glazing: Clear / Low-E; Shading depth: 0.2 / 0.6 / 1.5 m; Number of windows: 1 / 3 / 5
7.	Simulation Platform	Rhino + Grasshopper + Honeybee/Ladybug
8.	Weather Data	EPW (Taipei)
9.	Energy Simulation	Honeybee Energy (Energy Plus)
10.	Daylight Standards	EN 17037, LEED v4.1
11.	Energy Metric	EUI (kWh/m²·yr)
12.	Daylight Metrics	sDA, ASE, DGP
13.	Carbon Metrics	Embodied Carbon, CPP, WLC
14.	LCA Data Source	QTO and EF-based carbon coefficients
15.	Grid Emission Factor	Taiwan electricity grid
16.	Optimization Method	Wallacei Genetic Algorithm (GA)
17.	Optimization Objectives	Maximize sDA; minimize ASE, EUI, WLC, and CPP
18.	Decision-Making Method	TOPSIS (MCDM) with DGP as a supporting visual comfort criterion
19.	Facade Final Output	Optimal facade parameters

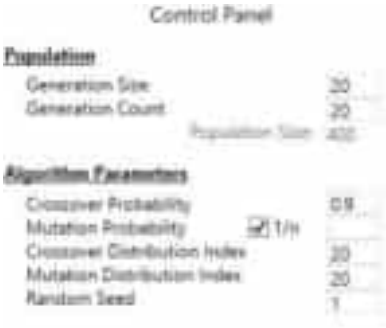


Figure 2. Wallacei optimization settings

Multi-objective optimization was then performed using the Wallacei genetic algorithm to minimize EUI, ASE, WLC, and CPP while maximizing sDA. The optimization settings are shown in Figure 2. During the evolutionary search, a simplified embodied carbon proxy was adopted to represent carbon sensitivity within the optimization loop, while Daylight Glare Probability (DGP) was evaluated separately at the post-optimization stage. For the selected Pareto-optimal alternatives, embodied carbon was recalculated using quantity take-off (QTO)-based geometry and emission-factor-based carbon coefficients. These recalculated embodied carbon values were then combined with annual operational carbon emissions to determine Whole-Life Carbon (WLC) and Carbon Payback Period (CPP), enabling an integrated assessment of energy, daylight, and carbon performance.

Accordingly, this study used two embodied carbon representations with distinct roles. QTO-based embodied carbon was used for the final calculation of WLC and CPP, whereas the simplified embodied carbon proxy in Wallacei served only as a surrogate indicator to test the influence of carbon sensitivity during the evolutionary search. Although the proxy may affect which solutions are explored and retained in the Pareto set, the final carbon assessment of the selected alternatives was based entirely on the more detailed QTO-derived embodied carbon values.

3.3 Stage 3: Multi-Criteria Decision-Making (MCDM)

In the third stage, the 20 Pareto-optimal facade alternatives generated by Wallacei were evaluated using TOPSIS to identify the most balanced final design. Four weighting scenarios were defined: energy-priority, carbon-priority, comfort-priority, and balanced.

TOPSIS ranked the alternatives using six criteria: sDA, ASE, DGP, EUI, WLC, and embodied carbon (EMB). Carbon Payback Period (CPP) was excluded from the final ranking because negative CPP values do not represent a valid payback duration but rather indicate a no-payback condition. This occurs when a design fails

to generate sufficient annual operational carbon savings to offset the additional initial embodied carbon, meaning that the carbon payback mechanism is not achieved in physical terms. Therefore, CPP was not used as a primary ranking criterion, but was instead treated as a post-evaluation indicator to verify whether the highly ranked alternatives also demonstrated meaningful carbon payback performance. A post-evaluation CPP check was also applied to all Pareto-optimal solutions. Together, these analyses provided a stronger basis for selecting the final facade design for validation and interpretation.

3.4 Stage 4: Validation and Final Decision Support

In the fourth stage, the optimization outcomes from Wallacei were compared with the TOPSIS ranking results to assess the consistency between search-based and decision-based evaluation. This stage also included visual validation and post-evaluation CPP verification to support the final selection of the best-performing facade alternative. Through this process, the final design was assessed not only on ranking performance but also on its practical balance among daylight, visual comfort, operational energy, and life-cycle carbon performance.

4 Optimization Results: Pareto Front Analysis

4.1 Multi-Objective Optimization Outcomes

The Wallacei optimization generated a range of facade design alternatives that balance energy, daylight, and life-cycle carbon objectives. Building Type 00 served as the baseline reference case, and 20 Pareto-optimal solutions were identified as non-dominated alternatives.

Figure 3 shows the distribution of these solutions in the objective space, and Figure 4 summarizes the facade parameters and performance values of the baseline and optimized cases. Together, these results define the feasible trade-off space among the main optimization objectives.

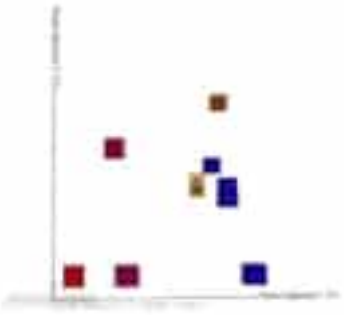


Figure 3. Wallacei Objective Space Results

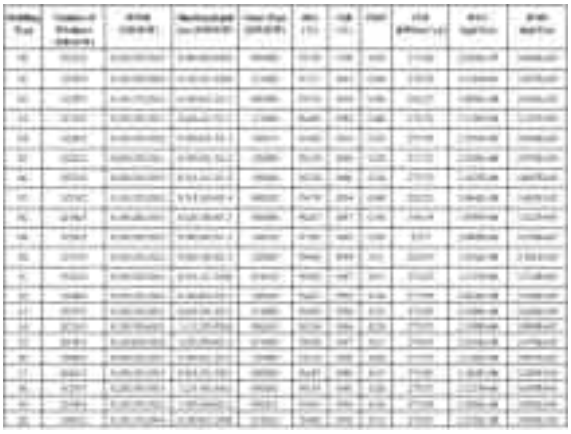


Figure 4. Pareto Dataset of Optimized Designs

4.2 Characteristics of Pareto-Optimal Facade Solutions

The 20 Pareto-optimal facade solutions show variations in WWR, shading depth, glazing type, and window configuration that affect energy, daylight, and carbon performance. However, daylight performance remains consistently high across the Pareto set, with all solutions satisfying the applied sDA, ASE, and DGP criteria. This indicates that daylight primarily serves as a compliance condition rather than as the main source of variation among the final alternatives.

The main differences among the Pareto solutions are observed in EUI, WLC, and embodied carbon. Lower EUI does not always correspond to lower WLC, showing that operational energy savings may be accompanied by higher facade-related carbon impacts. These results reflect the trade-off between operational efficiency and life-cycle carbon performance in the optimized facade set.

4.3 Pareto Trade-Offs between Energy, Daylighting, and Carbon Performance

The Pareto set shows that the main trade-off lies between EUI and WLC rather than between acceptable and unacceptable daylight performance. Because all solutions meet the applied daylight and glare criteria, the final comparison is mainly driven by differences in energy use and carbon impact.

CPP provides an additional layer of interpretation in this trade-off. Although it is included in the optimization stage, it is used only as a post-evaluation indicator in the final analysis, as negative CPP values indicate no-payback conditions and are not suitable for direct MCDM comparison. CPP, therefore, helps verify whether alternatives with favorable energy and carbon performance also achieve meaningful carbon payback.

4.4 Selection of Pareto Dataset for Decision Analysis

The 20 Pareto-optimal facade alternatives were retained as the input dataset for the subsequent TOPSIS analysis, while Building Type 00 was treated as a separate baseline reference case. For each alternative, facade parameters and performance indicators were compiled into a unified evaluation table. At this stage, final embodied carbon values were calculated using detailed QTO-based results from the optimized facade geometries. These values were then used to derive WLC and support the post-evaluation assessment of CPP. This data set provides the basis for transparent multi-criteria comparison of the optimized facade alternatives.

5 Multi-Criteria Decision-Making Results

5.1 TOPSIS Evaluation under Different Priority Scenarios

Following the Pareto front analysis, the 20 Pareto-optimal facade alternatives were evaluated using TOPSIS to support final decision-making. The decision matrix included six criteria: sDA, ASE, DGP, EUI, WLC, and embodied carbon (EMB). All criteria were normalized using vector normalization, and benefit and cost criteria were treated accordingly, as shown in Table 2. Four weighting scenarios were defined: energy-priority (EUI, ASE), carbon-priority (WLC, EMB), comfort-priority (sDA, DGP), and balanced. TOPSIS then produced a closeness coefficient for each alternative under each scenario.

Table 2. TOPSIS Normalization

Building Type	sDA	ASE	DGP	EUI	WLC	EMB
1	0.223	0.219	0.258	0.231	0.227	0.205
2	0.223	0.219	0.258	0.201	0.203	0.212
3	0.223	0.219	0.258	0.231	0.227	0.210
4	0.223	0.228	0.232	0.227	0.226	0.222
5	0.224	0.223	0.181	0.227	0.231	0.250
6	0.224	0.223	0.181	0.227	0.232	0.257
7	0.223	0.219	0.258	0.209	0.211	0.219
8	0.223	0.226	0.232	0.208	0.207	0.203
9	0.223	0.228	0.232	0.227	0.224	0.212
10	0.223	0.219	0.258	0.209	0.208	0.198
11	0.224	0.226	0.200	0.227	0.229	0.235
12	0.223	0.228	0.232	0.227	0.226	0.221
13	0.223	0.223	0.226	0.228	0.226	0.216
14	0.224	0.223	0.181	0.227	0.231	0.250
15	0.224	0.226	0.200	0.227	0.226	0.218
16	0.223	0.223	0.181	0.227	0.229	0.242
17	0.223	0.223	0.226	0.228	0.226	0.213
18	0.224	0.223	0.181	0.227	0.228	0.232
19	0.223	0.228	0.232	0.227	0.225	0.219
20	0.223	0.223	0.226	0.228	0.228	0.229

Table 3 presents the weighting scenarios used in the TOPSIS evaluation, including the energy-, carbon-, and comfort-priority settings.

Table 3. TOPSIS Weighting Scenarios

Scenario	sDA	ASE	DGP	EUI	WLC	EMB
Energy	0.05	0.15	0.05	0.55	0.10	0.10
Carbon	0.03	0.02	0.02	0.08	0.75	0.10
Comfort	0.20	0.10	0.40	0.10	0.10	0.10
Balanced	0.050	0.100	0.050	0.400	0.300	0.100

5.2 Ranking Results and Identification of the Optimal Facade Design

The TOPSIS results in Figures 5–8 indicate that the preferred facade alternative changes with the weighting scenario, confirming that the final ranking is highly dependent on decision priorities. Building Type 02 ranked first in the energy-priority, carbon-priority, and balanced scenarios, demonstrating robust performance across multiple evaluation contexts. In contrast, Building Type 18 ranked first only in the comfort-priority scenario, indicating stronger performance when daylight and visual comfort are emphasized. These findings suggest that no single facade alternative is universally optimal under all weighting scenarios. Overall, Building Type 02 can be regarded as the most robust option, whereas Building Type 18 is preferable when comfort is prioritized.



Figure 5. TOPSIS Results: Energy-Priority



Figure 6. TOPSIS Results: Carbon-Priority



Figure 7. TOPSIS Results: Comfort-Priority

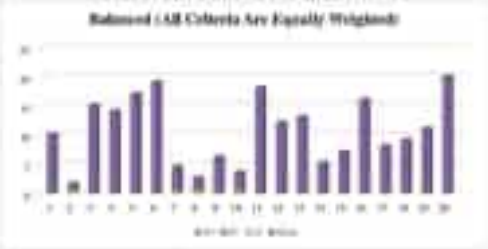


Figure 8. TOPSIS Results: Balanced

Table 4. Post-Evaluation Carbon Payback Period Check

BUILDING TYPE	CPP (YEARS)	CPP STATUS
1	-30.24	No payback
2	8.95	Achieved
3	-43.80	No payback
4	-953.39	No payback
5	-1670.47	No payback
6	-1893.55	No payback
7	17.92	Achieved
8	6.39	Achieved
9	-606.39	No payback
10	3.31	Achieved
11	-514.18	No payback
12	-914.38	No payback
13	-129.70	No payback
14	-1681.97	No payback
15	-299.12	No payback
16	-1444.11	No payback
17	-108.48	No payback
18	-1135.95	No payback
19	-847.28	No payback
20	-217.64	No payback

As shown in Table 4, positive CPP values indicate a valid carbon payback duration, meaning that the additional embodied carbon can be offset over time through annual operational carbon savings. By contrast, negative CPP values do not represent a valid payback duration; instead, they indicate a no-payback condition in which annual operational carbon savings are absent or insufficient to offset the initial embodied carbon increment.

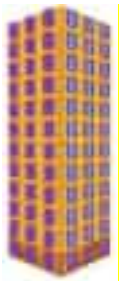


Figure 9. Final Selected Facade Design

Figure 9 presents the final selected facade design for Building Type 02. The figure visualizes the overall facade composition of the selected alternative, highlighting the arrangement of windows and shading features across the building form. This design was chosen as the final recommendation because it showed the most stable overall performance across the TOPSIS scenarios and also achieved a positive CPP value, indicating acceptable carbon payback performance over the building life cycle.

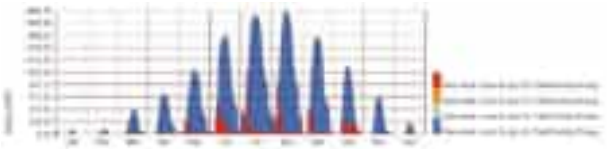


Figure 10. Monthly Cooling Energy Profile (HB-Energy)

Figure 10 presents the monthly cooling energy profile of the selected facade design. A pronounced seasonal variation can be observed, with cooling demand remaining minimal during the winter months, increasing steadily through spring, and peaking in July and August before declining again toward the end of the year. This distribution is consistent with Taipei's hot-humid climatic context, in which cooling loads are primarily driven by elevated summer air temperatures and increased solar exposure. The figure therefore highlights the strong seasonal dependence of cooling energy use in the evaluated office building.

6 Conclusion

This study developed an integrated LCA–CPP–MCDM framework combined with parametric modeling and Wallacei-based multi-objective optimization for office facade design in Taipei's hot-humid climate. The results revealed clear trade-offs among energy use, daylight availability, glare performance, and life-cycle carbon outcomes. Shading strategies were found to reduce cooling demand, but they could also increase embodied carbon, making the Carbon Payback Period (CPP) an important indicator for evaluating long-term carbon effectiveness. The optimization process generated 20 Pareto-optimal solutions, and the TOPSIS analysis identified Building Type 02 as the most robust and balanced alternative across the evaluated weighting scenarios. Building Type 02 also achieved a positive CPP of 8.95 years, indicating that its additional embodied carbon can be offset by annual operational carbon savings within a measurable period. This result was therefore classified as Achieved. However, CPP was not included as a direct TOPSIS criterion because negative CPP values are not physically interpretable as valid payback durations in a conventional ranking process. Therefore, CPP was treated as a post-evaluation indicator to verify whether the selected alternatives also achieved meaningful carbon payback performance.

This study was tested on a single office-building case in Taipei, which limits the direct generalizability of the findings. Nevertheless, the proposed workflow remains adaptable to other tropical and subtropical contexts through adjustment of weather data, grid emission factors, and building geometry inputs. Future work should extend the framework to multiple building types and climatic conditions to further examine the robustness of facade design trade-offs under different environmental and carbon scenarios.

Acknowledgment

The authors would like to thank the National Science and Technology Council (NSTC) for their support of this research and for providing the funding necessary for conference participation.

References

[1] "International Energy Agency, Tracking industrial energy efficiency and CO2 emissions, Energy Indicators (2007) http://www.iea.org/publications/ freepublications/publication/tracking emissions.pdf (accessed 20.02.13).".

[2] A. Dervishaj and K. Gudmundsson, "Parametric design workflow for solar, context-adaptive and reusable facades in changing urban environments," *J. Build. Perform. Simul.*, vol. 18, no. 2, pp. 161–190, 2025, doi: 10.1080/19401493.2024.2432916.

[3] C. Zhang, M. Hu, B. Laclau, T. Garnesson, X. Yang, and A. Tukker, "Energy-carbon-investment payback analysis of prefabricated envelope-cladding system for building energy renovation: Cases in Spain, the Netherlands, and Sweden," *Renew. Sustain. Energy Rev.*, vol. 145, no. July 2020, p. 111077, 2021, doi: 10.1016/j.rser.2021.111077.

[4] T. Chen and L. Tan, "Multiobjective optimization of external shading for west facing university dormitories in Kunming considering solar radiation and daylighting," *Sci. Rep.*, vol. 15, no. 1, pp. 1–18, 2025, doi: 10.1038/s41598-025-04465-8.

[5] M. Thalfeldt, E. Pikas, J. Kurnitski, and H. Voll, "Facade design principles for nearly zero energy buildings in a cold climate," *Energy Build.*, vol. 67, pp. 309–321, 2013, doi: 10.1016/j.enbuild.2013.08.027.

[6] A. T. Nguyen, S. Reiter, and P. Rigo, "A review on simulation-based optimization methods applied to building performance analysis," *Appl. Energy*, vol. 113, pp. 1043–1058, 2014, doi: 10.1016/j.apenergy.2013.08.061.

[7] M. Dasse, K. Slavkovic, A. Stephan, and E. Gobbo, "Towards a parametric early design approach for office buildings that integrates life cycle assessment and indoor environmental quality," *Build. Environ.*, vol. 289, no. November 2025, 2026, doi: 10.1016/j.buildenv.2025.113998.

[8] A. R. Ali, D. Derin, S. Blömeke, and C. Herrmann, "Parametric life cycle assessment model for absolute environmental sustainability assessment of lithium-ion batteries," *Sustain. Prod. Consum.*, vol. 57, no. February, pp. 80–94, 2025, doi: 10.1016/j.spc.2025.05.010.

[9] P. Briefing, "Policy Briefing," no. October, pp. 1–16, 2008.

[10] M. Zaker Esteghamati, H. Sharifnia, D. Ton, P. Asiatico, G. Reichard, and M. M. Flint, "Sustainable early design exploration of mid-rise office buildings with different subsystems using comparative life cycle assessment," *J. Build. Eng.*, vol. 48, 2022, doi: 10.1016/j.jobe.2022.104004.

[11] A. Stephan, R. H. Crawford, and K. de Myttenaere, "A comprehensive assessment of the life cycle energy demand of passive houses," *Appl. Energy*, vol. 112, pp. 23–34, 2013, doi: 10.1016/j.apenergy.2013.05.076.

[12] S. Abd Alla, V. Bianco, L. A. Tagliafico, and F. Scarpa, "Life-cycle approach to the estimation of energy efficiency measures in the buildings sector," *Appl. Energy*, vol. 264, no. March, p. 114745, 2020, doi: 10.1016/j.apenergy.2020.114745.

[13] M. A. Aydın and G. K. Oral, "Environmental impact assessment of building envelope insulation standards for decarbonization efforts in a developing country: a case study for Türkiye," *Energy Build.*, vol. 349, no. July, p. 116537, 2025, doi: 10.1016/j.enbuild.2025.116537.

[14] M. Norouzi, A. N. Haddad, L. Jiménez, S. Hoseinzadeh, and D. Boer, "Carbon footprint of low-energy buildings in the United Kingdom: Effects of mitigating technological pathways and decarbonization strategies," *Sci. Total Environ.*, vol. 882, no. January, 2023, doi: 10.1016/j.scitotenv.2023.163490.

[15] T. Srisamranrungruang and K. Hiyama, "Life carbon assessment of embodied and operational carbon with view assessment across facade systems in Japanese climate zones," *Results Eng.*, vol. 29, no. January, p. 109219, 2026, doi: 10.1016/j.rineng.2026.109219.

Bridging Building Information Modelling and Natural Language Processing: A Systematic Review on Current Applications and Limitations

Tessa Marie Oberhoff[1], Julian Cloos[1], Sven Mackenbach[1] and Katharina Klemt-Albert[1]

[1]Institute for Construction Management, Digital Engineering and Robotics in Construction
RWTH Aachen University, Germany
oberhoff@icom.rwth-achen.de; julian.cloos1@rwth-aachen.de; mackenbach@icom.rwth-aachen.de; klemt-albert@icom.rwth-aachen.de

Abstract –

Building Information Modelling (BIM) is widely recognised as a key driver of digital transformation in the architecture, engineering and construction industry. However, its full potential remains underutilised, partly due to the vast and complex nature of BIM-related data. At the same time, Natural Language Processing (NLP), a rapidly evolving field of artificial intelligence, offers advanced methods for interpreting and leveraging text-based information. The convergence of these two technologies holds considerable potential to facilitate human interaction with BIM data by making it more accessible. Despite this potential, research directly linking NLP and BIM remains limited. This paper attempts to bridge this gap by mapping and evaluating the current state of research on NLP applications within BIM workflows. To this end, a Systematic Literature Review is conducted in accordance with the methodological framework proposed by Fink, identifying 92 relevant publications. The metadata of these publications was analysed using VOSviewer, followed by a detailed content analysis. For each source, the BIM-related tasks optimised by NLP and NLP functionality applied to BIM data are identified, and a novel approach is proposed to classify the BIM-NLP integration into five categories: Extract, Query, Enrich, Modify, and Match. The analysis reveals that most existing literature explores NLP application in a peripheral or exploratory manner, with few instances of deep integration into BIM workflows. Nevertheless, the reviewed literature highlights the potential of NLP to improve data accessibility, automate information processing, and enhance human-machine interaction.

Keywords –

BIM–NLP integration, Large Language Models, Human–Machine Interaction, Data Accessibility

1 Introduction

The architecture, engineering, and construction (AEC) industry is a highly data-intensive domain. Throughout the entire building lifecycle, large volumes of heterogeneous information are generated and exchanged. Despite ongoing digitisation and the increasing adoption of Building Information Modelling (BIM), a substantial share of information exchange in the AEC sector continues to rely on textual information, from both unstructured documents and semi-structured but fragmented information systems. [1–3]

Natural Language Processing (NLP) provides methods for the automated analysis of such textual information [4, 5]. Recent advances in NLP, particularly the emergence of Large Language Models (LLMs), have significantly expanded the ability to capture semantic meaning, contextual relationships, and domain-specific knowledge from vast and unstructured text sources [6, 7]. In the context of the transition toward a data-driven and model-based construction industry, the combination of BIM and NLP has been identified as a technology with considerable potential [8–10]. BIM relies on structured information models that are commonly exchanged using the Industry Foundation Classes (IFC) standard, which is based on formal, machine-readable text [11]. Integrating NLP techniques with BIM, therefore, enables the linkage of textual information from other sources within one central information model of a built asset. This provides the potential for more data-driven processes, which can result in enhanced decision-making and greater automation across the entire lifecycle of a building.

To understand the potential of BIM–NLP integration current applications and limitations need to be analysed. But only a limited number of studies have explicitly reviewed the combined application of BIM and NLP. Di Giuda et al. [12] presented a literature review, identifying several potential application areas of NLP in AEC, but

with a strong focus on using NLP for requirements engineering at an early project phase to initialize BIM. Two subsequent studies by Locatelli et al. [2, 13] applied scientometric methods to analyse BIM–NLP integration. These analyses identified a rapidly increasing scientific interest in NLP and Semantic BIM as independent research domains and clustered the identified articles thematically. However, the clusters lack a unified clustering strategy, consisting of a mix of BIM-related tasks, NLP tasks, and review articles.

Overall, the existing review studies indicate growing academic interest in the integration of BIM and NLP but also reveal unresolved research gaps. In particular, there is a lack of a systematic, content-oriented literature review that critically examines NLP integration into BIM-centred workflows. In addition, due to rapid developments in Artificial Intelligence (AI) methods, NLP is a dynamically evolving field [5]. Consequently, the findings of existing studies are no longer up to date. Hence, the objective of this article is to conduct a Systematic Literature Review on the combined use of BIM and NLP in the AEC sector. The study aims to approach existing research from a content perspective, identify dominant application areas, and assess current limitations to outline directions for future research.

2 Theoretical Background

2.1 Natural Language Processing

NLP is a technique to analyse, understand, and work with human language with the help of computers [14]. NLP involves converting unstructured language data into a structured data format. This enables machines to identify and understand language and text to generate relevant responses [15]. Contrary to some definitions, NLP does not constitute a subfield of AI per se; however, it does utilise AI methods [4, 5], as shown in a schematic classification of different forms of NLP in Figure 1.

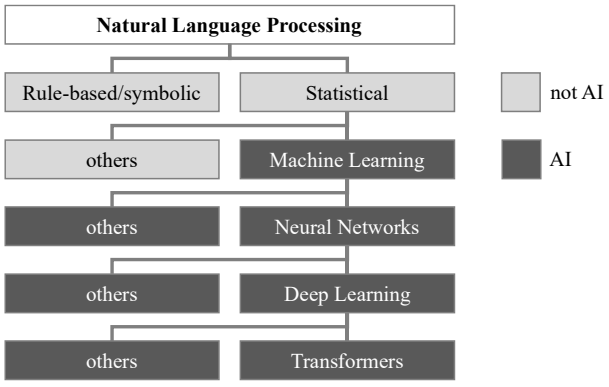


Figure 1. Schematic classification of NLP methods and models based on Canchila et al. [5]

According to Locatelli et al. [2] the first scientific article about NLP application in the field of AEC was published in 1989 by Logcher et al., in the form of an article bearing the title "Knowledge Processing for Construction Management Data Base"[16]. Since then, methods have evolved from rule-based/symbolic procedures, such as ontologies, to statistical procedures. In the field of statistical methods, it is possible to distinguish between classical statistical methods and machine learning methods, which are part of the field of AI. Within machine learning, one can differentiate between Neural Networks (NN) and other methods. Continuing along the chain, among NN methods, Deep Learning is the subfield leading to Transformer architectures, which have attracted considerable attention for several years [15]. Transformer models refer to a NN architecture introduced by Vaswani et al. [17]. It is a general-purpose architecture that can be applied to many tasks beyond language. Most contemporary LLMs (e.g. GPT-5, LLaMa, and BERT) are implemented using Transformer architectures [5]. LLMs, in turn, are models defined by their large scale in terms of parameters, training data, and computational resources, as well as their ability to perform a wide range of language-related tasks [18].

2.2 Building Information Modelling

BIM is a model-based methodology for the digital representation and management of physical and functional characteristics of built assets [19]. BIM enables the integration of geometric, semantic, and relational information into a coherent information model that can be used across different phases of the building lifecycle [19]. In this context, the 3D geometric model is commonly extended by additional information, so-called dimensions. Typically, temporal information to support construction sequencing and scheduling is referred to as 4D-BIM, and adding cost-related data to enable quantity-based cost estimation and financial control is referred to as 5D-BIM [20]. Beyond these dimensions, BIM models can be further enriched with information related to aspects such as sustainability, safety, or facility management [20]. The extension of BIM into multiple dimensions enables a more comprehensive and integrated management of construction projects by combining heterogeneous information within the digital model of the built asset. A central aspect of storing the relevant information is the underlying data structure of the BIM model. BIM models are typically exchanged using the IFC standard, which provides a formal, object-oriented schema for representing building components, their properties, and their relationships [11, 21]. IFC is based on machine-readable and hierarchical data structures that support interoperability between software systems [22]. These structured representations allow information to be

queried, validated, and processed algorithmically, forming the basis for NLP applications [11].

3 Methodology

To identify relevant publications, a Systematic Literature Review according to the methodology of Fink [23] is conducted. The first step in this methodology is defining research questions (RQs). In this article, three RQs are posed. The first RQ pertains to the identification of relevant literature and the initial characterisation of the same. The specific inquiry is as follows:

- RQ1: What scientific articles exist on the integration of BIM and NLP?

Following the next steps according to Fink [23], subsequently the databases for the search are defined. The databases Engineering Village (Compendex), IEEE Xplore, Scopus, Taylor & Francis, Wiley Online Library, and Web of Science are used, since they cover a large part of common databases for both general and engineering-specific topics. Additionally, the DBs allow suitable options for keyword entry, as well as filter and export functions. The search string for the Systematic Literature Review consists of one component on the topic of "BIM" and a second component on the topic of "NLP". Synonyms for the two main search words are additionally collected. In the NLP component, "LLMs" are included as the only specific method, as they have gained particular prominence in recent years. The synonyms are linked by the Boolean operator "OR" and the two components are linked by an "AND" operator.

This procedure identifies 430 articles, from which duplicates are first removed. According to Fink [23], the remaining articles are then subjected to a first and second filtering procedure (screening). Formal quality criteria are applied in the first filtering process. These quality criteria include:

- Article not accessible
- Article no longer available
- Article not in German or English

Material quality criteria are applied in the second filtering process. These include, above all, thematic relevance to the research question. For instance, reasons for exclusion include:

- Article does not relate to the construction industry
- Article does not relate to the use of NLP
- Article only mentions BIM in the outlook/introduction

The review process is shown in Figure 2, based on a PRISMA diagram [24]. It shows the number of articles included or excluded in each step, along with the reason for exclusion.

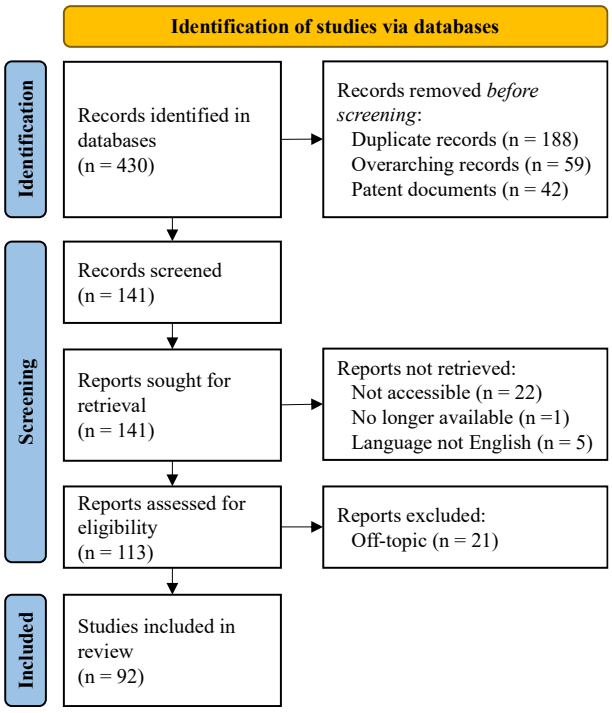


Figure 2. Conducted Systematic Literature Review process based on a PRISMA-diagram

As a result of the Systematic Literature Review, 92 articles remain for further analysis. To answer the first RQ, scientometric data is evaluated by VOSviewer [25]. This paper highlights two aspects: the year of publication and a classification of the publication type. The development of publication frequency indicates the interest in a research field and can show trends. Analysing the publication type, differentiating between "Original Research", "Reviews", and "Case Studies", can give a first impression of the content.

The bibliometric data is followed by an in-depth content analysis of the BIM–NLP applications described in the articles. Therefore, two more RQs are defined:

- RQ2: Which BIM workflow-related tasks are tackled by NLP integration in the identified articles?
- RQ3: Which NLP functionalities are used regarding BIM data in the identified articles?

To answer both RQs, clustering techniques are applied. For RQ2, the clusters are formed based on the BIM-related tasks within a building's lifecycle. At this point, the primary focus is on the identification of tasks within the BIM workflow that are impacted, while the specific functionality of NLP is not a primary concern. The task clusters are developed based on existing review research articles, and are then adapted to or supplemented by novel categories. For RQ3, the clusters are formed based on the NLP functionality applied to BIM data. These clusters are not related to specific NLP methods or

models, but rather characterise the actions performed by the NLP application on the BIM data. Therefore, new categories are developed.

4 Analysis

4.1 Year of Publication

Figure 3 shows how the identified publications are distributed over time. The diagram starts in 2014, as no articles published before this date were identified.

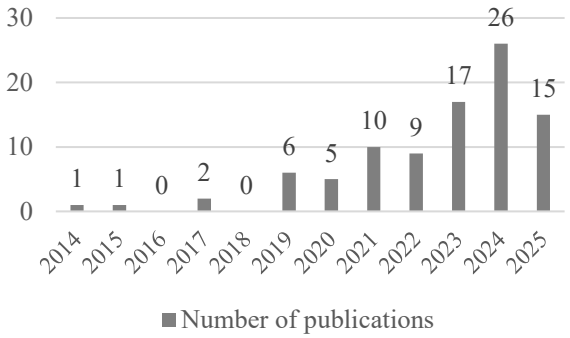


Figure 3. Development of the number of publications about BIM–NLP integration

It should be noted that the number of articles for 2025 only covers part of the year, as the research was conducted in May 2025. Based on the results from May, a higher number of publications can be expected for 2025 than for 2024. The development of publication numbers shows a rising research interest in the integration of BIM and NLP, confirming the findings of the reviews mentioned in Chapter 1. At the same time, the number of existing studies is still rather limited.

4.2 Type of Publication

The following section provides a concise overview of the categorisation of each publication into one or more types of publication. The results of the publication type analysis are shown in Table 1.

Table 1. Number and percentage of articles per publication type

Publication type	Number	% (N=92)
Original Research	78	84
Review	11	12
Case Study	3	3

The majority of the identified publications are classified as Original Research. These contributions investigate a wide range of applications related to the integration of BIM and NLP, typically focusing on specific methods, tools, or workflows. Given the diversity of application contexts and technical approaches, a detailed content-oriented analysis of these studies is provided in Chapters 4.3 and 4.4.

The second largest group consists of Reviews. Three of these publications [2, 12, 13], briefly introduced in Chapter 1, focuses explicitly on the combined application of BIM and NLP. One additional review addresses the use of NLP in the context of construction schedule management [9]. The remaining Reviews cover related but more specialized topics, including reviews of computational techniques supporting specific BIM-related tasks, such as quantity take-off, knowledge management, and compliance checking, as well as broader technological perspectives on digital twins, big data analytics, and AI in the AEC industry. Two reviews focus on specific NLP-related models or techniques, namely GPT models (short for: Generative Pre-trained Transformer models; a special kind of Transformer model) and Deep Learning approaches, while one addresses digital twin applications in tunnel construction.

Only a small number of publications are classified as pure Case Study. This limited representation indicates that, to date, relatively few BIM–NLP applications have been empirically evaluated in real-world settings. The field, therefore, currently remains largely exploratory, and there is an evident need for further validation of proposed approaches through practical implementations.

4.3 BIM Task Classification

This section provides a concise overview of the BIM workflow-related task clusters identified in the relevant publications. The results of the clustering are demonstrated in Table 2. The total number of clusters is nine. Seven publications, all Reviews, were excluded from this analysis as they deal with various BIM task clusters.

Table 2. Number and percentage of articles per BIM task cluster

BIM task cluster	Number	% (N=85)
Knowledge Management	23	27
Compliance Checking	16	19
BIM Authoring	12	14
Design and Planning	10	12
Facility Management	6	7
Risk and Safety Management	6	7
Cost Management	3	4
Robot Control	3	4
Schedule Management	3	4
Sustainability Assessment	3	4
various	7	-

Knowledge Management: This category comprises the capture, structuring, and reuse of project-related

knowledge, such as information from previous projects, guidelines, and experiential knowledge. Its purpose is to systematically provide knowledge to support and enhance other BIM-related tasks. The cluster is only assigned if the activities are not related to any of the following specific BIM task clusters. In previous review articles, the cluster was named differently, e.g. Information management.

Compliance Checking: This category refers to the verification of BIM models against regulatory requirements, standards, codes, and project-specific rules. It encompasses both formal rule-based checks and content-related compliance assessments. The cluster has the second highest number of assigned publications and was already mentioned in all previous reviews on BIM–NLP integration. This indicates that Compliance checking is one of the most well-established BIM workflow-related tasks, where NLP application is studied.

BIM Authoring: This category covers the creation, modification, and continuous maintenance of BIM models throughout the project lifecycle. It includes model structuring, attribute management, and information enrichment. In previous studies, only information enrichment of BIM models was mentioned. This shows that BIM modelling from scratch is new, but, as the high number of publications indicates, it is a topic of rising interest.

Design and Planning: This category focuses on supporting decisions during early project phases. It includes design alternative exploration, scenario analysis, and generative approaches that support decision-making. A similar task cluster, namely Design management, was identified in one previous study. The name of the cluster was adapted to focus on the design process and illustrate the difference from the cluster Compliance Checking.

Facility Management: This category describes the use of BIM information to support building operation after project completion. It includes activities related to maintenance, operation, and asset management, mostly related to decision support during those stages. The task cluster was only mentioned by the more recent of the previous studies and seems to be a rather slowly developing application of NLP.

Risk and Safety Management: This category addresses the identification, assessment, and prevention of risks and safety issues in a BIM context. Most publications in this cluster focus on the construction phase. Previous studies also mentioned risk and safety-related applications. Partly, publications related to this BIM task cluster were excluded during the Systematic Literature Review, due to the NLP application often focusing on safety reports as separate documents, as opposed to being incorporated into BIM workflows.

Cost Management: This category addresses activities related to cost estimation, cost planning, and cost control based on 5D BIM. As the number of publications shows, it is currently a rather small research field.

Robot Control: This category covers the application of robotic and automated systems in construction and building operation contexts. It includes automated execution activities and interactions between humans and robots in a BIM context.

Schedule Management: This category involves the planning, monitoring, and optimization of 4D BIM. It includes construction sequencing, progress tracking, and coordination of resources. Even though the number of assigned publications is relatively small, one of the Reviews is related to the BIM task cluster Schedule Management.

Sustainability Assessment: This category focuses on the evaluation of buildings across their lifecycle, especially in the context of Life Cycle Assessment (LCA). It intentionally does not include design recommendations to improve the sustainability of the building, nor does it consider energy optimization decision-making, as both activities are included in other BIM task clusters.

4.4 NLP Functionality Classification

In order to explore the functions performed by NLP in more detail, this section examines the functionality of NLP methods with BIM data, independent of the concrete BIM task supported. The NLP functionality clusters intentionally avoid using a hierarchical structure. To characterize the functionality, five distinct clusters were developed. The resulting clusters and the number and percentage of assigned publications are demonstrated in Table 3 and discussed following the table. All eleven publications classified as Review were excluded from the analysis, as they deal with various NLP functionality clusters. The remaining publications can be assigned to more than one NLP cluster, as some of the presented approaches incorporate more than one NLP application.

Table 3. Number and percentage of articles per NLP functionality cluster

NLP functionality cluster	Number	% (N=81)
Extract	32	40
Query	31	38
Match	20	25
Enrich	17	21
Modify	12	15
various	11	-

Extract: This category refers to approaches in which NLP is applied to derive structured or semi-structured information from BIM-related sources, such as regulatory texts, technical specifications, or reports. The extracted information is typically transformed into formats suitable for further processing and use in BIM-

based workflows. The BIM database itself is not analysed or changed by NLP methods in this cluster. Extract has the highest number of allocated publications. This may well be the case, given that NLP was originally developed to work with human language rather than machine-readable texts, such as IFC models.

Query: This category includes methods that allow users to access BIM-related information through natural language queries. NLP is employed to interpret user inputs and retrieve relevant data from BIM models, databases, or linked information systems, often via conversational or assistant-based interfaces, without altering the underlying model. Chatbots are identified to be a trending research topic in this cluster, leading to the second highest number of allocated publications.

Match: This category comprises methods that establish correspondences between BIM data and external information sources, such as regulations, schedules, databases, or real-world entities. NLP supports the alignment and comparison of heterogeneous representations, enabling activities such as data integration and cross-system mapping. Thus, the IFC file of the BIM model is unaffected.

Enrich: This category encompasses techniques that use NLP to augment BIM data with additional semantic information. Such approaches add or infer attributes, classifications, relationships, or ontological links, thereby improving the semantic completeness and interpretability of BIM models and associated datasets. BIM data is enhanced by adding information, but no previously existing data is changed or deleted.

Modify: This category covers approaches in which NLP is used to directly influence or perform changes to BIM data. It includes text- or speech-driven creation and modification of model elements, execution of modelling commands, and automated or semi-automated generation of BIM content. The limited number of publications assigned to this cluster could indicate the novelty of these NLP functionalities or the complexity, resulting from the profound impact on the BIM data structure in this cluster.

5 Discussion

The analysis reveals distinct patterns regarding both the BIM-related tasks addressed and the functionalities for which NLP is integrated into BIM-based workflows.

Among the BIM task clusters, Knowledge Management represents the largest share of publications. This finding reflects the strong reliance of BIM processes on textual information, such as requirements, reports, guidelines, and project documentation, which can be processed effectively using NLP techniques. The predominance of this category is consistent with earlier observations [12, 13]. Compliance Checking constitutes the second most frequent task cluster. The review shows that compliance-related studies often combine information extraction from regulatory texts with alignment or comparison against BIM data, indicating a close relationship between the Compliance Checking task cluster and the NLP functionality clusters Extract and Match. Tasks related to the clusters BIM Authoring and Design and Planning are also well represented. A shift toward more interactive and model-centred applications can be observed. Similar tendencies are discussed in one of the identified review articles [7]. In contrast, task clusters that typically depend on numerical/geometric data or require tight integration with domain-specific tools remain limited. This suggests that current methods are less applicable to such data-intensive tasks and are less applicable in complex BIM ecosystems.

Regarding NLP functionality clusters, the categories Extract and Query are most prevalent, followed by Match. The strong presence of Extract and Query is consistent with the dominance of the BIM task clusters Knowledge Management and Compliance Checking, both of which primarily involve handling textual information. The comparatively lower frequency of Enrich and Modify shows that semantic augmentation and model-altering applications remain limited. Extract and Query functionalities may dominate, due to requiring only limited interaction with BIM schemas, whereas the more advanced functionalities Modify and Enrich remain comparatively rare due to their need for strict compliance with structured data models and semantic consistency. This discrepancy can be interpreted as a representation mismatch problem: NLP systems operate on probabilistic, context-dependent representations, while BIM relies on deterministic, schema-constrained data structures. Bridging this mismatch is therefore a broader integration problem between statistical and symbolic systems.

To enable a transition from passive NLP interaction, such as querying, to active model modification, this study proposes several technical prerequisites. First, the development of schema-aware language models is essential, as current LLMs lack an explicit understanding of BIM (especially IFC) schemas. Future research should therefore focus on fine-tuning models on IFC-based datasets. Second, formal validation layers are required to ensure the correctness of model modifications. This includes the integration of rule-based validation systems for IFC compliance, as well as hybrid pipelines that combine NLP outputs with symbolic reasoning. Without such mechanisms, NLP-generated modifications risk producing invalid or inconsistent BIM models. [26]

The analysis also indicates significant barriers to real-world adoption, as reflected in the low number of case studies identified in the literature. One possible reason is data privacy and security concerns, as construction projects often involve sensitive contractual and design information [27]. Additionally, issues of trust play a

crucial role, as practitioners require transparent and verifiable outputs due to the high impact of modifications on BIM data, which the black-box nature of many AI-based NLP techniques precludes. Inaccuracies can propagate into design decisions, cost estimation, or construction execution, leading to considerable financial risks [28]. Instead of fully automated pipelines, human-machine-interaction, like Explainable AI and human-in-the-loop approaches are therefore seen as promising [29]. This trend can also be observed in the review, namely in the rising number of co-pilot systems.

Overall, the results indicate that current research mainly applies NLP at the interface or preprocessing level, while deeper integration into BIM workflows remains uncommon. Although recent publications demonstrate technically advanced prototypes, particularly in connection with Transformer architectures, their scope is often limited to controlled scenarios and lacks validation in complex, real-world projects.

6 Conclusion

This study presents a Systematic Literature Review on the integration of BIM and NLP, identifying 92 relevant publications and analysing them with respect to BIM-related tasks and NLP functionalities. The results confirm that BIM–NLP integration is an emerging but still fragmented research field. To characterise the nature of BIM–NLP integration, this study introduces a novel classification of NLP functionalities in connection with BIM data, consisting of the five clusters Extract, Query, Enrich, Modify, and Match. The analysis shows that most existing work focuses on information extraction and querying, whereas deeper integration forms, such as semantic enrichment and model modification, are less common. The discussion identifies barriers that hinder deeper integration of NLP into BIM workflows. A key finding is the mismatch between probabilistic NLP models and schema-based BIM representations, which limits the ability of current approaches to directly modify or enrich BIM data. Furthermore, data privacy concerns in handling sensitive project information, and a lack of trust due to limited reliability of NLP outputs obstruct real-world adoption.

Some limitations of this review should be acknowledged. The literature search was conducted up to May 2025, which restricts the coverage of very recent developments in a rapidly evolving field. In addition, the clustering of BIM-related tasks and NLP functionality involves interpretative decisions, particularly for studies addressing multiple objectives.

To bridge the gap between NLP and BIM, future research should focus on the development of schema-aware language models and NLP–rule-based validation frameworks including technical and human quality assurance. In addition, the evaluation of approaches in real-world scenarios is a high-priority step toward practical adoption. Ultimately, the integration of NLP into BIM has the potential to improve automation of complex information processes and enhance decision-making in a BIM context. However, realising this potential requires solutions that can operate reliably within existing BIM workflows.

References

[1] F. Opitz, R. Windisch, and R. J. Scherer. Integration of Document- and Model-based Building Information for Project Management Support. *Procedia Engineering*, 85:403–411, 2014. doi: 10.1016/j.proeng.2014.10.566.

[2] M. Locatelli, E. Seghezzi, L. Pellegrini, L. C. Tagliabue, and G. M. Di Giuda. Exploring natural language processing in construction and integration with building information modeling: A scientometric analysis. *Buildings*, 11(12):583, 2021. doi: 10.3390/buildings11120583.

[3] A. Shamshiri, K. R. Ryu, and J. Y. Park. Text mining and natural language processing in construction. *Automation in Construction*, 158:105200, 2024. doi: 10.1016/j.autcon.2023.105200.

[4] C. Feng, Y. Li, Z. Chen, and L. Guo. The Evolution and Breakthrough of Natural Language Processing: The Revolution from Rules to Deep Learning. In *Proceedings of the 5th International Conference on Computer Science and Management Technology*, Xiamen Guangdong, China, 2024. doi: 10.1145/3708036.3708089.

[5] S. Canchila, C. Meneses-Eraso, J. Casanoves-Boix, P. Cortés-Pellicer, and F. Castelló-Sirvent. Natural language processing: An overview of models, transformers and applied practices. *ComSIS*, 21(3): 1097–1145, 2024. doi: 10.2298/CSIS230217031C.

[6] X. Xiahou, G. Chen, Z. Li, X. Xu, and Q. Li. Knowledge Management in Construction Quality Management: Current State, Challenges, and Future Directions. *IEEE Transactions on Engineering Management*. 72:1069–1088, 2025. doi: 10.1109/TEM.2025.3550354.

[7] A. Saka, R. Taiwo, N. Saka, B. Salami, S. Ajayi, K. Akande, and H. Kazemi. GPT models in construction industry: Opportunities, limitations, and a use case validation. *Developments in the Built Environment*, 17:100300, 2024. doi: 10.1016/j.dibe.2023.100300.

[8] E. Ohene, G. Nani, M. F. Antwi-Afari, A. Darko, L. A. Addai. and E. Horvey. Big data analytics in the AEC industry: scientometric review and synthesis of research activities. *Engineering Construction &*

Architectural Management. 32(11):7299–7331, 2025. doi: 10.1108/ECAM-01-2024-0144.

[9] A. K. Singh, A. Pal, P. Kumar, J. J. Lin, and S.-H. Hsieh. Prospects of Integrating BIM and NLP for Automatic Construction Schedule Management. In *41st International Symposium on Automation and Robotics in Construction*, pages 238–245, Lille, France, 2024. doi: 10.22260/ISARC2023/0034.

[10] A. Corneli, L. Binni, F. Spegni, L. Messi, and B. Naticchia, NATURAL LANGUAGE PROCESSING FOR CONSTRUCTION SITES MANAGEMENT. In *Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference*, Heraklion, Crete, Greece, 2023. doi: 10.35490/EC3.2023.272.

[11] M. Selim, K. Nassar, and O. Hosny, Hg., Automating BIM (IFC) Data Analysis Using LangGraph. In *Proceedings of the 10th World Congress on Civil, Structural, and Environmental Engineering.* New Cairo, Egypt, 2025. doi: 10.11159/icsect25.182.

[12] G. M. Di Giuda, M. Locatelli, and E. Seghezzi. Natural language processing and bim in aeco sector: A state of the art. In *Proceedings of International Structural Engineering and Construction,* 7(2), 2020. doi: 10.14455/ISEC.2020.7(2).CON-22.

[13] M. Locatelli, E. Seghezzi, and G. M. Di Giuda. Exploring BIM and NLP applications: A scientometric approach. In *Proceedings of International Structural Engineering and Construction,* 8(1), 2021. doi: 10.14455/ISEC.2021.8(1).CON-12.

[14] G. G. Chowdhury. Natural language processing. *Annual Review Info Sci & Tec*, 37(1):51–89, 2003. doi: 10.1002/aris.1440370103.

[15] W. Khan, A. Daud, K. Khan, S. Muhammad, and R. Haq. Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends. *Natural Language Processing Journal*, 4:100026, 2023. doi: 10.1016/j.nlp.2023.100026.

[16] R. D. Logcher, M.-T. Wang and F. H.-S. Chen. Knowledge Processing for Construction Management Data Base. *Journal of Construction Engineering and Management*, 115(2):196–211, 1989. doi: 10.1061/(ASCE)0733-9364(1989)115:2(196).

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010, Long Beach, California, USA. doi: 10.48550/arXiv.1706.03762.

[18] Z. Wang, Z. Chu, T. V. Doan, S. Ni, M. Yang, and W. Zhang. History, Development, and Principles of Large Language Models – An Introductory Survey. *AI Ethics*, 5:1955–1971, 2025. https://doi.org/10.1007/s43681-024-00583-7.

[19] A. Borrmann, M. König, C. Koch und J. Beetz. Die BIM-Methode im Überblick. In *Building Information Modeling*, pages 1–31, Springer Fachmedien, Wiesbaden, 2021. doi: 10.1007/978-3-658-33361-4_1.

[20] A. Gärtner, T. Bender, M. Härtig, M. Krämer, M. May, and M. Schlundt. Building Information Modeling. In *CAFM-Handbuch*, pages 197–233, Springer Fachmedien, 2025. doi: 10.1007/978-3-658-49047-8_11.

[21] H. Dawood, J. Siddle, and N. Dawood. Integrating IFC and NLP for automating change request validations. *Journal of Information Technology in Construction*, 24:540–552, 2019. doi: 10.36680/J.ITCON.2019.030.

[22] buildingSMART International Limited, Industry Foundation Classes (IFC). Online: https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes, Accessed: 06/01/2026.

[23] A. Fink. *Conducting research literature reviews: From the internet to paper*. SAGE Publications, Los Angeles, 2020.

[24] M. J. Page et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ*, 372:n71, 2021, doi: 10.1136/bmj.n71.

[25] N. J. van Eck, and L. Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538, 2010. doi: 10.1007/s11192-009-0146-3.

[26] B. Liu and H. Chen. BIMCoder: A Comprehensive Large Language Model Fusion Framework for Natural Language-Based BIM Information Retrieval. *Applied Sciences*, 15(14):7647, 2025. doi: 10.3390/app15147647.

[27] T. Patel and V. Patel. Data privacy in construction industry by privacy-preserving data mining (PPDM) approach. *Asian J Civ Eng*, 21(3):505–515, 2020. doi: 10.1007/s42107-020-00225-3.

[28] R. Lopez and P. E. D. Love. Design Error Costs in Construction Projects. *Journal of Construction Eng. Manage*ment, 138(5):585–593, 2012. doi: 10.1061/(ASCE)CO.1943-7862.0000454.

[29] M.D.B. in Cugurra et al. The Cross-Fertilization Between the Human-in-the-Loop Approach and the Explainable AI Techniques Toward Trustworthiness. In *Artificial Intelligence, Data and Robotics*, E. Curry et al., pages 77–104, Cham: Springer Nature Switzerland, 2026. doi: 10.1007/978-3-032-10561-5_5.

Bridging Simulation and Practice: Surrogate-Based Production Scheduling for Stochastic Modular Construction

Nima Moghimi[1*], Sahar H. Shamaee[2], Qipei Mei[3], Vicente A. Gonzalez[4], Farook Hamzeh[5]

[1]Corresponding Author, Ph.D. Student, Department of Civil and Environmental Engineering, University of Alberta, Canada
[2]Graduate Research Assistant, Department of Civil and Environmental Engineering, University of Alberta, Canada
[3]Assistant Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
[4] Hal Kvisle Professor and Tier 1 Canada Research Chair in Digital Lean Construction, Infrastructure Human Tech (IHT) Lab, Strategic Projects Insight Centre in Engineering (SPICE), Department of Civil and Environmental Engineering, University of Alberta, Canada
[5]Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada
nmoghimi@ualberta.ca, hamedsha@ualberta.ca, qipei@ualberta.ca, vagonzal@ualberta.ca, hamzeh@ualberta.ca

Abstract

In space-constrained modular construction, stochastic work content renders deterministic scheduling ineffective, often leading to production deadlocks. While Simulation-Based Optimization (SBO) addresses these dynamics, it typically lacks the operational utility for routine daily planning due to its high computational intensity. To bridge this gap, this study adopts a Design-Science Research (DSR) approach to design and evaluate a Surrogate-Based Scheduling Framework. The methodology follows a structured cycle where a Hierarchical Stochastic Demand Model generates a wide spectrum of synthetic production scenarios, which are evaluated via high-fidelity simulation to determine optimal Inter-Project Time Buffers. These complex optimization results are subsequently instantiated into a lightweight planning rule via surrogate-based heuristic derivation. To validate the utility of this artifact, a case study was conducted on a wood-frame wall panel assembly line. The resulting "Peak Workload" heuristic (instantiated via Linear Regression) achieved an R2 of 0.657, demonstrating that a simple linear approximation sufficiently captures dominant variability trends despite the system's inherent non-linearity. Operationally, the heuristic reduced average Cycle Time by 37.7% compared to the baseline while maintaining equivalent throughput. These results confirm the feasibility of transforming complex buffer optimizations into transparent, data-driven policies that stabilize production flow without requiring real-time computation.

Keywords – Modular Construction; Simulation-Based Optimization; Production Scheduling; Surrogate Modelling; Data-Driven Heuristics; Work-in-Process (WIP)

1 Introduction

Modular construction offers a promising, factory-based alternative to the inefficiencies of traditional building practices; however, its full potential is often constrained by unique scheduling complexities [1]. The Engineer-to-Order (ETO) nature of modular projects means that each component has a distinct work content, introducing significant process time variability [2]. In space-constrained production facilities, this unpredictability creates a precarious planning environment. Schedules that fail to account for this stochasticity often cause production lines to become either deadlocked with excess inventory or idle from a lack of available work, undermining the very efficiency the factory setting is meant to provide [3].

To navigate this uncertainty, production planners rely on look-ahead scheduling to sequence upcoming projects. However, the foundational tools for this task, such as the Critical Path Method (CPM), are inherently deterministic and ill-suited for dynamic environments [4,5]. Planners often compensate for this rigidity by inserting intuitive temporal buffers between projects [6]. This manual approach to creating "safety time" is unreliable; an insufficient gap between projects risks system-wide congestion, while an overly generous one leads to poor resource utilization and reduced capital efficiency [7,8]. The central challenge, therefore, is to develop a data-driven, simulation-informed optimization procedure that determines project-specific inter-project buffer times and is explicitly evaluated under stochastic variability [9,10].

Academic research has explored advanced computational methods to address this. Discrete Event Simulation (DES) has been used to model factory dynamics, while metaheuristic algorithms have been

applied to optimize production sequences [11]. A disconnect persists, however, between these sophisticated optimization techniques and the practical cadence of production planning. While Simulation-Based Optimization (SBO) offers high rigorous validity, it often lacks the operational utility necessary for the routine generation of weekly schedules [12–14]. The operational reality of construction management rarely accommodates the computational expense or specialized expertise required to maintain complex optimization engines. Consequently, there is a critical need for a new class of decision-support artifacts that bridge this gap—providing the rigor of simulation without the computational burden [15].

To address this challenge, this study adopts a Design-Science Research (DSR) approach to design and evaluate a novel scheduling artifact: the Surrogate-Based Scheduling Framework. The objective is to equip planners with a scientific basis for determining project start times. The core logic of this artifact is that robust schedules must be derived from a realistic understanding of operational uncertainty. To achieve this, the framework utilizes a multi-level probabilistic demand model to generate a wide spectrum of potential production scenarios. Through extensive simulation, an optimization process identifies the ideal inter-project lag time needed to absorb variability. This complex relationship is then instantiated into a computationally simple planning rule via a Surrogate-Based Heuristic Derivation strategy. The final output is a practical decision-making tool that translates project-specific data into scientifically sized buffers, enabling the creation of resilient look-ahead plans without the need for real-time computational analysis.



Figure 1: Structure of the proposed framework.

2 Research Methodology

2.1 Research Approach

This study adopts the Design-Science Research (DSR) methodology [16]. DSR is a problem-solving paradigm focused on the creation and rigorous evaluation of innovative "artifacts" to address practical organizational challenges.

Consistent with the DSR process, this research follows a three-stage cycle:

1. Problem Identification (Relevance): Defining the need for a dynamic buffer allocation method that overcomes the limitations of static scheduling practices.
2. Artifact Design: Developing the "Offline Learning Framework." This framework operates through a structured learning cycle (Figure 1): (1) synthetic demand scenarios are generated to capture operational variability; (2) these scenarios are evaluated within a high-fidelity simulation model; (3) an optimization algorithm solves for the optimal Inter-Project Time Buffer (τ^*); and (4) the results are instantiated into a lightweight planning rule via surrogate-based heuristic derivation.
3. Evaluation (Rigor): Validating the utility of the artifact through a case study application (Section 3).

Unlike deterministic planning that assumes fixed task durations, this DSR artifact recognizes that robust schedules must be informed by the probabilistic reality of ETO production. The methodology shifts the decision point to the planning phase, allowing buffers to be sized scientifically before production begins.

2.2 Hierarchical Stochastic Demand Modelling

To ensure the scheduling artifact addresses the 'Relevance' requirement of DSR, the design process begins by modelling the environment's true variability. Therefore, a Hierarchical Stochastic Model synthesizes look-ahead windows that reflect the variability inherent in actual ETO demand. This scenario-based modelling logic is consistent with prior work that uses Monte Carlo / scenario sampling to generate diverse

realizations of uncertain future conditions before evaluating candidate strategies [17,18]. The model integrates four interdependent sub-components:

Project Occurrence Process: The arrival of incoming orders varies with market conditions. The number of projects N_{proj} within a planning horizon is modeled as a Poisson process:

$$P(N_{proj} = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (1)$$

where λ is the mean arrival rate, calibrated to reflect Low, Medium, or High workload intensities.

Project Scope: Each project has a defined work volume. The total quantity of units Q_i for Project *i* follows a discrete uniform distribution bounded by facility constraints:

$$Q_i \sim \mathcal{U}\{Q_{min}, Q_{max}\} \quad (2)$$

Product Mix (Complexity): Engineer-to-Order environments involve heterogeneous products with different complexity profiles. Let ρ denote the complexity ratio—the fraction of "complex" (high-variability) units within a project. The number of complex units is:

$$n_{c,i} = \text{Round}(\rho \cdot Q_i) \quad (3)$$

This parameter directly influences downstream processing time variability and resource contention, making it a critical input for buffer determination.

Delivery Constraints: External logistics (e.g., trailer availability, site readiness) constrain when projects can start. The allowable lead time δ_i is modeled stochastically to reflect real-world scheduling pressures:

$$D_i = T_{entry,i} + \text{Duration}_{est,i} + \delta_i \quad (4)$$

This temporal coupling ensures that the optimization considers both internal production variability and external logistical constraints.

2.3 Production System Simulation Environment

The simulation environment serves as the design testbed, allowing the artifact to be iteratively tested against non-linear dynamics before deployment. The stochastic production scenarios generated in Section 2.2 are evaluated within a computational simulation environment acting as a virtual testbed. While the specific modelling paradigm (e.g., Discrete Event Simulation, System Dynamics, or Agent-Based Modelling) may vary based on the facility's characteristics, the environment must faithfully replicate the non-linear dynamics of constrained production. The simulation framework generally operates on two critical layers to ensure validity:

Constraint-Based Physical Layer: The flow of work units is governed by rigorous physical and logical constraints rather than deterministic lead times. This includes finite buffer capacities (B_{Cap}) between workstations and specific resource availability windows (R_{avail}). By enforcing these limits, the simulation allows complex system behaviours—such as blocking (upstream work halting due to downstream congestion) and starvation (downstream idleness due to upstream delays)—to emerge endogenously from the system state rather than being pre-programmed.

Stochastic Process Execution: Task durations are modelled as stochastic variables drawn from probability distributions calibrated to historical performance data. To capture the heterogeneity of Engineer-to-Order (ETO) production, processing time P_t for a unit u is defined as a function of its specific complexity attributes:

$$P_t(u) = \begin{cases} f_{std}(\cdot) & \text{if } u \text{ is Standard Class} \\ f_{cplx}(\cdot) & \text{if } u \text{ is Complex Class} \end{cases} \quad (5)$$

where $f(.)$ represents a probability density function (e.g., Triangular, Log-Normal) appropriate for the specific task. This structure ensures that the optimization process specifically accounts for the volatility introduced by product mix variability, a primary driver of schedule unreliability in modular construction.

2.4 Optimization Formulation

This component represents the search process of the design cycle, identifying optimal configurations τ^* that maximize the utility function $J(\tau)$. The central problem is to determine the optimal Inter-Project Time Buffer vector $\tau^* = \{\tau_{1,2}, \tau_{2,3}, \dots, \tau_{N-1,N}\}$ that balances competing objectives. Here, $\tau_{i,i+1}$ denotes the scheduled lag between the planned completion of Project J_i and the start of Project J_{i+1}.

To avoid a highly fragmented ("scattered") schedule, the optimization procedure minimizes a weighted multi-objective cost function $J(\tau)$, which trades off process efficiency, production throughput, and overall schedule compactness:

$$J(\tau) = w_1 \cdot \widehat{CT} - w_2 \cdot \widehat{TH} + w_3 \cdot \widehat{MS} \quad (6)$$

where $\widehat{CT}$, $\widehat{TH}$, and $\widehat{MS}$ are the min-max normalized values of Cycle Time (CT), System Throughput (TH), and Makespan (MS), respectively. For a generic metric M, the normalized quantity is defined as:

$$\widehat{M} = \frac{M(\tau) - M_{min}}{M_{max} - M_{min}} \quad (7)$$

Within this framework, minimizing CT promotes efficient flow at the individual project level, while maximizing TH (achieved via the negative term) incentivizes higher production volume. The Makespan term (MS) serves as a penalty against schedule scattering, compelling the optimizer to select the tightest feasible buffers that avoid overlap without creating unnecessary idle time. The weights w_1, w_2, w_3 are user-specified and satisfy $w_1 + w_2 + w_3 = 1.0$.

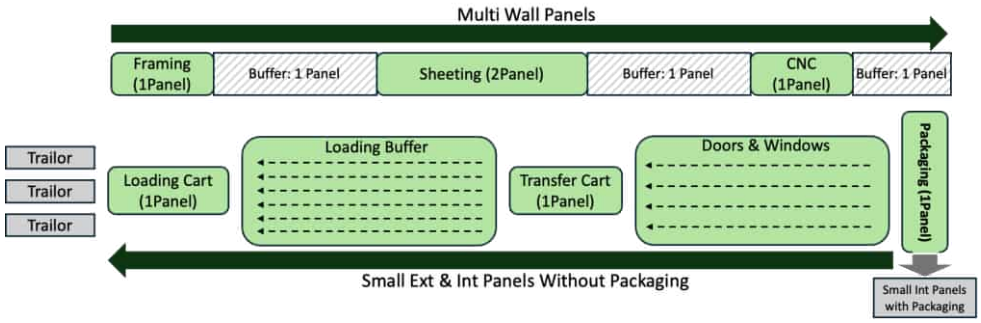


Figure 2: Schematic layout of the wall-assembly line.

The optimization is carried out under the domain constraint $0 \leq \tau_{i,i+1} \leq \tau_{max}$, where τ_{max} is the maximum commercially acceptable idle period (for example, one shift), ensuring that the solution is operationally realistic. The start-time relationship is given by:

$$Start_{i+1} = Finish_i^{planned} + \tau_{i,i+1} \quad (8)$$

By solving this optimization across all synthetic scenarios, the algorithm produces a rich dataset of optimal buffer configurations, each linked to the project characteristics that generated it.

2.5 Surrogate Heuristic Derivation

The final step in the design cycle is to instantiate the complex optimization results into a usable tool. As running the full simulation-optimization loop for every scheduling update is computationally prohibitive for routine operations. To bridge the gap between high-fidelity analysis and practical daily planning, the framework synthesizes the complex relationships found by the optimization engine into a lightweight Planning Heuristic via surrogate modelling. This step effectively "crystallizes" the intelligence of the simulation into a static format.

We posit a functional relationship between the project state vector *X* and the optimal inter-project buffer τ^*. Crucially, *X* comprises not only independent attributes of the incoming project (e.g., work content Q_i) but also transition metrics that capture the shift in operational intensity between sequential projects (e.g., Δ Complexity or Δ Volume between J_i and J_{i+1}). This ensures the heuristic accounts for sequence-dependent setup risks.

During the design phase, the optimization algorithm generates a training dataset $D = \{(X_n, \tau_n^*)\}_{n=1}^{M}$ spanning *M* diverse scenarios. A regression analysis is performed on *D* to approximate the function *f*. While the specific architecture of the surrogate model can range from non-linear approximators to decision trees depending on data complexity, a linear approximation is often preferred for its interpretability in industrial contexts. The general form of the heuristic is defined as:

$$\tau^* \approx \beta_0 + \sum_{j=1}^{k} \beta_j \cdot x_j \quad (9)$$

where, β_0, β_j are the learned coefficients representing the weight of each factor. x_j represents the Significant Predictors extracted from the project state vector *X*. These features are selected based on their correlation with production variability (e.g., total work content, resource intensity variability, or sequence-dependent complexity shifts).

The resulting output is a transparent, closed-form equation that eliminates the need for repeated simulation. This allows planners to implement the scientifically derived policy using standard spreadsheet tools or elementary calculation tables. By simply inputting the scheduled project attributesx_j, the heuristic instantaneously provides the optimal "safety time" τ^* required to maintain flow, transforming a computationally expensive optimization problem into a practical arithmetic task.

3 Artifact Evaluation: Case Study Implementation

To validate the proposed scheduling framework, a comprehensive case study was conducted within a real-world manufacturing environment. This section details the experimental setup, bridging the theoretical methodology with the specific physical constraints of an operating facility.

3.1 Evaluation Domain: Edmonton Assembly Line

The validation domain is a wood-frame wall panel assembly line in Edmonton, Alberta, organized as a flexible flow shop. As schematically illustrated in **Error! Reference source not found.**, the process flows sequentially from Framing through Sheeting and Nailing to Packaging. At the Packaging station, the flow splits: completed interior panels route directly to storage, while exterior panels proceed to four parallel

Door & Window (DW) installation lanes before reaching the Loading buffer. The layout is characterized by severe spatial constraints, featuring minimal buffer capacities (often limiting WIP to a single panel) between stations—a typical challenge in "volumetric" construction where large product dimensions restrict storage.

This specific facility was selected because it presents a convergence of internal variability and external logistical constraints that render static scheduling ineffective. The production environment is highly stochastic, with heterogeneous project specifications driving variable processing times. Crucially, the system is constrained by outbound logistics: panels must be loaded onto project-dedicated trailers, which have limited daily availability (avg. 3/day). In this tight spatial environment, traditional scheduling strategies that fail to account for variability frequently lead to immediate blocking when trailers are unavailable, causing cascading flow stagnation. This susceptibility to gridlock makes the line an ideal testbed for validating the proposed buffer optimization approach.

3.2 Configuration of the Simulation Testbed

To represent the stochastic production dynamics of the Edmonton facility, a hybrid simulation model was developed in Python using SimPy. The model combines a Discrete Event Simulation (DES) layer for material flow and capacity constraints with an Agent-Based Simulation (ABS) layer for workforce behavior. Physical constraints such as finite buffer capacities and the single-table Packaging limit are explicitly modeled, allowing blocking and starvation to emerge naturally from congestion. The ABS layer captures crew collaboration, non-linear productivity effects, and task-priority rules, with complex exterior panels prioritized over standard interior units.

The model was parameterized using empirical field data from the partner facility, including 100–150 processing-time records per station, supported by expert input to confirm logic and variability ranges. Processing times were differentiated by panel characteristics, such as interior versus exterior units. Triangular distributions were used as a practical representation of uncertainty in this data-constrained setting because they preserve minimum, most-likely, and maximum values and are easy to calibrate. The model was validated against a historical dataset of 84 Multi-Wall projects, achieving R^2 values of 0.95 for Framing and 0.89 for Doors & Windows, with MAPE between 16% and 18%.

Key assumptions include fixed observed routing, finite buffer capacities, rule-based crew collaboration and prioritization, station-specific triangular processing-time distributions, and schedule control through inter-project buffers rather than dynamic resequencing.

3.3 Evaluation Protocol

The framework was implemented through the 'Surrogate-Based Derivation' phase specified in the methodology. Leveraging the Hierarchical Stochastic Model, thirty distinct production scenarios were generated to encompass a broad range of workload intensities, spanning from low-demand periods (Poisson $\lambda = 8$) to high-intensity surges ($\lambda = 25$). The choice of 30 scenarios reflects a balance between coverage and computational tractability for this conceptual design stage. Rather than serving as an exhaustive sample, these scenarios were constructed to span low-, medium-, and high-intensity workload regimes and to expose the framework to diverse operating conditions. Moreover, because each scenario yields multiple adjacent project-pair decisions, the resulting surrogate was calibrated on approximately 800 optimized decision points rather than on only 30 scenario-level observations. While this was sufficient to derive and interpret a stable first-order heuristic, future work should examine convergence of the surrogate coefficients and scheduling performance as the number of generated scenarios increases.

For each scenario, the Simulation-Based Optimization engine determined the optimal Inter-Project Time Buffer (τ^*). The decision variable was the elapsed time (in minutes) between the planned finish of Project J_i and the release of Project J_{i+1}. The Genetic Algorithm minimized a weighted cost function that integrated Cycle Time, System Throughput, and Makespan. For this study, the weighting configuration was set to w_1=0.35 (Cycle Time), w_2=0.35 (Throughput), and w_3=0.30 (Makespan). Equal priority was assigned to the efficiency metrics (CT and TH), while the Makespan term served primarily as a 'control variable' to prevent the optimizer from suggesting an overly scattered schedule. It is important to note that this cost function is modular; in practice, facility managers can adjust these weights or introduce new metrics to reflect shifting operational priorities (e.g., prioritizing maximum volume during peak seasons).

3.4 Artifact Instantiation: Heuristic Calibration

To translate the simulation-based optimization results into a practical planning tool, a surrogate model was calibrated using the ~800 optimal decision points generated by the simulation. Linear Regression was selected for its interpretability and stability, with the aim of relating high-level project attributes to the optimal inter-project buffer (τ^*).

To identify the most informative predictors, a Theory-Guided Sparse Model Discovery approach was applied. The candidate library included raw production

metrics (e.g., Multi-Wall count, Small Panel count) and derived workload-intensity terms. Using LASSO with bootstrap-based stability selection, the process filtered out weak predictors and showed that raw unit counts alone were insufficient. Instead, the model consistently retained variables linked to bottleneck workload intensity, particularly the peak boundary intensity between adjacent projects and the directional change in workload.

Based on these results, the final regression was defined as:

$$\tau^* = \alpha + \beta\, M_i + \gamma\, \Delta M_i^+ + \epsilon \qquad 10$$

Where:

$M_i = \max(\text{mode}_{i-1}, \text{mode}_i)$ signifies the Peak Workload Intensity across the two adjacent projects. $\Delta M_i^+ = \max(\text{mode}_i - \text{mode}_{i-1}, 0)$ signifies the Anticipatory Workload Increase, triggered only when the incoming project is more demanding than the outgoing one.

The model was calibrated against the simulation-generated dataset, producing the following coefficients: $\alpha = -46.57$, $\beta = 0.186$, and $\gamma \approx -0.084$.

The coefficients illuminate distinct operational mechanisms. The positive weight on M_i ($\beta = 0.186$) reinforces that the primary constraint is the magnitude of the bottleneck; heavier project pairs require larger buffers. However, the negative coefficient for the anticipatory term ($\gamma \approx -0.084$) introduces a crucial insight: the "Clean-Slate Effect."

Mathematically, this term diminishes the required buffer when progressing from a lighter to a heavier project. Operationally, this suggests that a lighter outgoing project clears the system efficiently, establishing a stable, low-congestion environment for the subsequent demanding project. Conversely, a heavy outgoing project leaves residual variability in its wake, necessitating a larger safety margin for the next task regardless of its intensity.

The calibrated heuristic achieved an In-Sample Mean Absolute Error (MAE) of 32.46 minutes and a Root Mean Square Error (RMSE) of 37.40 minutes, with an R^2 of 0.657. Although an MAE of approximately 32 minutes indicates a noticeable gap from the theoretical optimum, its practical acceptability in this study is supported by the downstream validation results rather than by the error magnitude alone. When applied to the validation-week schedule, the calibrated heuristic still produced substantial system-level improvements, including a 37.7% reduction in average cycle time while maintaining statistically equivalent throughput. This suggests that the surrogate captured the dominant scheduling signal required for look-ahead decision support, even if individual buffer predictions were not exact. The R^2 value confirms that nearly two-thirds of the variability in optimal buffer sizing is captured by this transparent "Peak Workload" heuristic, establishing its validity as a dependable planning rule. However, the present study did not explicitly quantify how surrogate prediction error propagates into secondary operational risk metrics such as blocking frequency, due-date deviation, or extreme-case instability under peak conditions; this should be examined in future work.

It is acknowledged that the underlying production system exhibits inherent non-linear dynamics due to finite buffering and complex resource interactions. However, at this preliminary stage, high-fidelity feature data (e.g., granular task-level specifications) is often unavailable during the look-ahead planning phase. Consequently, a Linear Regression model was selected to demonstrate the potential of the methodology using only high-level metrics available to planners (e.g., panel counts, expected execution times). While an R^2 of 0.657 indicates that roughly one-third of the variance remains unexplained—suggesting the need for future non-linear modelling (e.g., Random Forests) as data granularity improves—the current model successfully captures the dominant 'Peak Workload' trend required to stabilize the schedule.

4 Results and Discussion

To assess the operational utility of the designed artifact, the methodology was evaluated using a calibrated simulation testbed initialized with a real 14-project production schedule from the partner facility. While a comprehensive live field implementation remains a subject for future research, this validation horizon was selected to align with the standard 1–2 week look-ahead planning window used by facility management. This experiment specifically tests the preliminary robustness of the heuristic in handling a realistic batch of heterogeneous projects compared to the facility's current static planning approach. The reported execution timelines, cycle times, and throughput values are simulation-derived performance measures obtained by applying the baseline and heuristic-controlled schedules to the calibrated testbed.

The comparative assessment demonstrates a marked enhancement in flow efficiency. The application of heuristic-derived buffers decreased the average Cycle Time (Lead Time) by 37.7%, declining from a baseline of 328.65 minutes to 204.70 minutes. This substantial reduction signifies a considerable drop in Work-in-Process (WIP) accumulation throughout the production floor. Notably, this improvement in flow speed was realized while maintaining the system's production rate; the heuristic-controlled throughput (3.617 panels/hour) was statistically equivalent to the baseline (3.668 panels/hour). Following the fundamental principles of Factory Physics [19], a reduction in Cycle Time (*CT*)

coupled with constant Throughput (*TH*) mathematically necessitates a reduction in Work-in-Process (*WIP*) (Little's Law: $WIP = TH \times CT$). In the context of volumetric construction, where floor space is the primary constraint, this reduction in on-line inventory is critical. It implies that the facility achieves the same output with significantly less physical clutter, reducing material handling risks and freeing up capital, thereby validating the heuristic's value beyond simple output metrics. The key simulation outputs considered in this evaluation are project execution timelines, average cycle time, and throughput; these are reported quantitatively in the text and visually compared in Figure 3. The Baseline schedule (Blue) displays extended project durations accompanied by significant temporal overlaps. These overlaps characterize periods of intense congestion where multiple projects contend for constrained buffer capacity, producing the "deadlocks" endemic to the facility's existing operations.

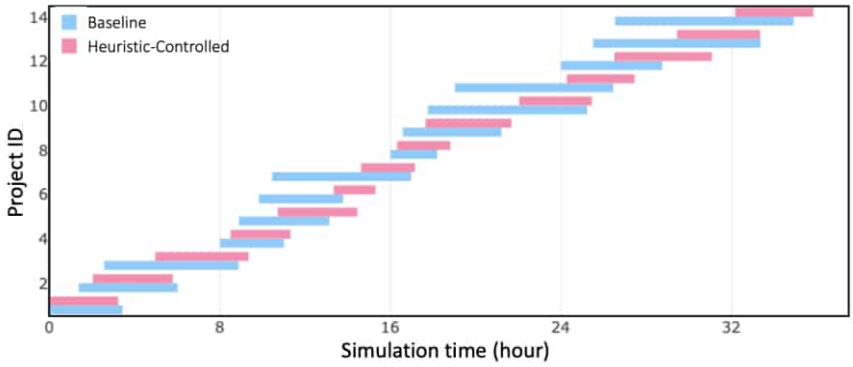


Figure 3: Comparative project execution timeline for the 14-project validation week: Baseline schedule vs. Heuristic-Controlled schedule.

By contrast, the Optimized schedule (Pink) exhibits more compressed execution windows with regular, orderly spacing. By introducing scientifically calibrated lags informed by the bottleneck intensity (M_i), the heuristic efficiently "harmonizes" the flow. Remarkably, the total Makespan of the production week remained essentially unchanged. This counterintuitive finding demonstrates that the "safety buffers" introduced by the heuristic do not postpone overall batch completion. Rather, they redirect unproductive "congestion time" (waiting within the line due to blockage) into productive "waiting time" (waiting outside the line prior to commencement). In volumetric construction, where floor space represents the limiting factor, the 38% improvement in Cycle Time bears strategic significance. By synchronizing production velocity with the stochastic properties of the bottleneck, the framework eliminates the cascading delays that characterize static schedules. The findings establish that disciplined planning can stabilize operations and reduce WIP liability while maintaining the facility's total weekly output capacity.

5 Conclusion

This study addresses the scheduling challenge in High-Mix Low-Volume (HMLV) modular construction, where product variability frequently disrupts static plans. Using a Design-Science Research (DSR) approach, it develops and evaluates a Surrogate-Based Scheduling Framework that translates complex simulation-optimization results into a practical daily planning rule. The framework follows a structured design cycle. A Hierarchical Stochastic Demand Model generates diverse synthetic production scenarios reflecting the variability of Engineer-to-Order projects. These scenarios are evaluated in a calibrated simulation testbed to identify the optimal inter-project buffers required to absorb system variability. The resulting decisions are then converted into a simplified surrogate-based heuristic, enabling planners to size safety margins based on project workload intensity without real-time optimization. Evaluation on a real wood-frame wall panel assembly line showed that the derived “Peak Workload” heuristic reduced average Cycle Time by 37.7% while maintaining statistically equivalent throughput. Under Little’s Law, this implies a substantial reduction in Work-in-Process (WIP), which is especially valuable in space-constrained prefabrication environments where congestion limits operational efficiency. The study also highlights several limitations. The current surrogate uses a linear regression model ($R^2 \approx 0.657$) to preserve interpretability under limited planning-stage information, but future work should explore richer features and non-linear models to better capture buffer-sizing behavior. In

addition, validation was limited to a one-week look-ahead horizon, so longer-term testing is needed to confirm robustness and generalizability. Finally, the optimization used fixed objective weights; future research should support dynamic weight adjustment so managers can adapt scheduling priorities to changing operational conditions.

6 Acknowledgement

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance grant ALLRP 581074–22, Discovery Grant RGPIN-2020-04420, and NSERC Alliance—Alberta Innovates Advance Program 232403286. All findings and conclusions expressed in this paper are those of the authors and do not reflect those of the contributors.

7 References

[1] Abdul Nabi M, El-adaway IH. Modular construction: Determining decision-making factors and future research needs. Journal of Management in Engineering 2020;36:04020085.

[2] Gosling J, Naim MM. Engineer-to-order supply chain management: A literature review and research agenda. Int J Prod Econ 2009;122:741–54.

[3] Salama T, Salah A, Moselhi O. INTEGRATED SCHEDULING OF MODULAR CONSTRUCTION USING "MOD-SCHEDULER." Proceedings of the 7th International Canadian Society for Civil Engineering Construction Conference (CSCE 2019), Montreal, Canada, 2019.

[4] Pourrahimian E, Salhab D, Fayek AR, Hamzeh F. Fuzzy-Chaos Framework for Analyzing and Quantifying Uncertainty in Labour Productivity in Complex Construction Environments. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 42, IAARC Publications; 2025, p. 571–9.

[5] Salama T, Salah A, Moselhi O. Alternative scheduling and planning processes for hybrid offsite construction. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 33, IAARC Publications; 2016, p. 1.

[6] Eltahan A, Lee G, Hamzeh F. Exploring Cognitive Load and Task Complexity in Dynamic Tracking Tasks: Insights for Construction Workflows. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 42, IAARC Publications; 2025, p. 532–9.

[7] Ballard G, Tommelein I. Current process benchmark for the last planner system. Lean Construction Journal 2016;89:57–89.

[8] She B, Chen B, Hall NG. Buffer sizing in critical chain project management by network decomposition. Omega (Westport) 2021;102:102382.

[9] Ma G, Wang A, Li N, Gu L, Ai Q. Improved critical chain project management framework for scheduling construction projects. J Constr Eng Manag 2014;140:04014055.

[10] Moghimi N, Arshang O, Hamzeh F. Digital twins in offsite construction: Current implementations, challenges, and future pathways. Autom Constr 2026;183:106820.

[11] Wu C, Jiang R, Li X. Integration of BIM and computer simulations in modular construction, A case study. Modular and Offsite Construction (MOC) Summit Proceedings 2016.

[12] Matsumoto K, Yamaguchi A, Oka T, Yasumoto M, Hara S, Iida M, et al. Simulation-based reinforcement learning approach towards construction machine automation. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 37, IAARC Publications; 2020, p. 457–64.

[13] Araújo GR, Gomes R, Gomes MG, Guedes MC, Ferrão P. Surrogate models for efficient multi-objective optimization of building performance. Energies (Basel) 2023;16:4030.

[14] Zhang L, Plathottam S, Reyna J, Merket N, Sayers K, Yang X, et al. High-resolution hourly surrogate modeling framework for physics-based large-scale building stock modeling. Sustain Cities Soc 2021;75:103292.

[15] Soman RK, Molina-Solana M, Whyte JK. Linked-Data based Constraint-Checking (LDCC) to support look-ahead planning in construction. Autom Constr 2020;120:103369.

[16] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Quarterly 2004:75–105.

[17] Moghimi N, Kashani H. Probabilistic Modeling to Develop the Optimal Seismic Resilience Enhancement Strategy for Industrial Facilities. Journal of Infrastructure Systems 2026;32:04025041. https://doi.org/10.1061/JITSE4.ISENG-2754.

[18] Moghimi N, Kashani H. A probabilistic framework to design the financing scheme for resilience enhancement of infrastructure systems. Reliab Eng Syst Saf 2025;262:111179.

[19] Hopp WJ, Spearman ML. Factory physics. Waveland Press; 2011.

AI Agent–based Component Rescheduling Framework for Industrialized Construction

Huayu Zhong[1], Hui Lu[2], Ke Chen[1]

[1]School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China
[2]School of Economics and Management, China University of Geosciences, Wuhan, China
hyzhong@hust.edu.cn, luhui@cug.edu.cn, chenkecm@hust.edu.cn

Abstract –

Industrialized construction relies on the tight integration of production, logistics, and on-site assembly, making the process highly sensitive to dynamic disturbances. To achieve rescheduling with minimum human intervention, this study proposes an intelligent rescheduling framework based on AI agents. First, we establish a unified Mixed-Integer Linear Programming (MILP) model to provide a standardized mathematical foundation for the rescheduling process. Second, a collaborative architecture comprising five functional AI agents is developed. By utilizing Large Language Models (LLMs) as their cognitive cores, these agents automatically transform natural-language disturbance reports into structured model patches and executable code. To ensure technical precision, a Retrieval-Augmented Generation (RAG) mechanism is integrated to ground the agents in domain-specific knowledge. Experimental results demonstrate that the framework can effectively bridge the gap between unstructured field reporting and mathematical optimization, enabling automated responses to disruptions.

Keywords –

Industrialized Construction; Dynamic Rescheduling; AI Agents; Framework

1 Introduction

Industrialized construction shifts traditional cast-in-situ construction methods toward a highly coordinated process of component production, transportation, and assembly. This transformation imposes spatiotemporal synchronization constraints across factories, logistics, and on-site resources[1]. Guided by just-in-time principles, buffer resources throughout the supply chain should be minimized, rendering the supply chain highly sensitive to various disturbances[2]. Consequently, local delays can propagate through the chain[3], triggering cascading effects such as vehicle idling[4] and on-site work stoppages[5]. Without mechanisms to generate actionable instructions, these chain reactions lead to overall performance degradation[6]. Therefore, prefabricated component scheduling requires not only robust static optimization but also agile operational rescheduling to mitigate disruptions.

Existing research in this domain spans from centralized optimization to distributed collaboration. While centralized approaches pursue global optimality, they often suffer from computational latency and lack the flexibility required to handle frequent disturbances[7-8]. To overcome these limitations, research has increasingly adopted Multi-Agent Systems (MAS) [9]. By modeling resources or tasks as autonomous entities with independent decision-making capabilities, MAS offer an adaptive alternative rooted in collaboration[10]. However, real-world disturbances are typically communicated via unstructured natural language such as site logs and verbal change orders. Traditional MAS, restricted by hard-coded decision logic[11], lack the ability to interpret ambiguous information and autonomously construct mathematical models. As a result, the pipeline from disturbance perception to decision generation remains heavily reliant on human intervention.

The emergence of Large Language Models (LLMs) provides a transformative pathway to overcome these hurdles. The primary advantage of LLMs lies not in replacing specialized optimization solvers, but in their capacity to serve as semantic translators[12]. This capability has catalyzed the rise of AI agents—autonomous entities that utilize LLMs as a cognitive core to perceive, reason, and act. An AI agent does not merely follow rigid rules; it interprets the intent behind a disturbance and manipulates computational tools to synthesize solutions.

Accordingly, this study proposes an intelligent rescheduling framework driven by AI agents. We first establish a unified Mixed-Integer Linear Programming (MILP) model to provide a standardized, solvable mathematical foundation for rescheduling[13]. Building on this, the framework employs AI agents as semantic-to-code bridges. By incorporating domain-specific

constraints through Retrieval-Augmented Generation (RAG), these agents automatically map unstructured disturbances into structured model modifications and executable solver code[14]. A collaborative workflow among specialized agents helps to achieve a fully automated loop from semantic disruption reporting to executable rescheduling commands.

2 Problem Description and Model Construction

This study specifically addresses the dynamic rescheduling of precast concrete components under disturbances. Grounded in the concept of interrupt management, the rescheduling strategy aims to minimize deviations between the revised schedule and the original baseline plan[15]. The optimization objective functions as a multi-criteria priority, balancing factors including (1) reducing costs associated with early or delayed component arrivals and (2) minimizing temporal shifts and resource reallocation overhead caused by schedule adjustments.

Coordination between the production facility and the construction site is maintained through the dynamic balancing of on-site inventory levels I_t. This framework adopts an event-driven rescheduling mechanism: upon the detection of a disturbance, a specialized Patch Agent translates the unstructured information into localized updates to automatically generate a reschedule solution.

Based on the problem description, the collaborative scheduling model is constructed as follows.

$$min\ Z = c_{late} * \sum_i l_i + c_{early} * \sum_t e_i + k_{time} * \sum_i shift_i^{time} + k_{line} * \sum_i \varphi_i^{line} \quad (1)$$

Equation (1) formulates the minimization objective, where c_{late} and c_{early} represent the unit penalty costs for the late arrival period l_i and early arrival period e_i of component i, respectively. To ensure schedule stability, the objective also penalizes deviations: k_{time} and k_{line} denote the penalty coefficients for temporal shifts and production line reallocations. Here, $shift_i^{time}$ represents the absolute deviation of the casting start time relative to the baseline, and φ_i^{line} is a binary variable indicating whether the production line assignment for component i has changed.

$$C_{i,o} = S_{i,o} + p_{i,o}, \forall\, i \in I \quad (2)$$

$$S_{i,od} \geq C_{i,oc} + \theta_i, \forall\, i \in I \quad (3)$$

$$S_{j,oc} \geq C_{i,od} - M_{big} * (1 - u_{i,j,l}^{line}) - M_{big} * (2 - x_{i,l} - x_{j,l}), \forall i, j \in I, i \neq j, \forall\, l \in L \quad (4)$$

$$S_{i,oc} \geq C_{j,od} - M_{big} * u_{i,j,l}^{line} - M_{big} * (2 - x_{i,l} - x_{j,l}), \forall i, j \in I, i \neq j, \forall\, l \in L \quad (5)$$

Equations (2-5) define the physical production logic. In Equation (2), $S_{i,o}$ and $C_{i,o}$ denote the start and completion times of process o (containing casting o_c and demolding o_d), and $p_{i,o}$ represents the processing duration. Equation (3) strictly enforces the curing constraint, requiring that the demolding start time $S_{i,od}$ must occur after the casting completion $C_{i,oc}$ plus the minimum curing time θ_i. Resource conflicts are resolved in Equations (4-5): $x_{i,l}$ is a binary variable indicating if component i is assigned to production line l, and $u_{i,j,l}^{line}$ manages the processing sequence between components i and j to prevent temporal overlap. M_{big} is a sufficiently large positive constant.

$$a_i = C_{i,od} + \tau_i, \forall\, i \in I \quad (6)$$

$$l_i \geq a_i - L_i, \forall\, i \in I \quad (7)$$

$$e_i \geq E_i - a_i, \forall\, i \in I \quad (8)$$

$$\sum_t z_{i,t} = 1, \forall\, i \in I \quad (9)$$

$$\sum_t (t\Delta) * z_{i,t} < a_i \leq \sum_t \big((t+1)\Delta\big) * z_{i,t}, \forall\, i \in I \quad (10)$$

$$I_t = I_{t-1} + \sum_i z_{i,t} - D_t, \forall\, t \in T \quad (11)$$

$$0 \leq I_t \leq W_{site}, \forall\, t \in T \quad (12)$$

Equations (6-12) bridge production and on-site assembly. Equation (6) calculates the site arrival time a_i based on the fixed transportation duration τ_i. Equations (7-8) quantify the deviation from the ideal arrival time window $[E_i, L_i]$. To map continuous arrival times to discrete time slots, Equations (9-10) introduce the binary variable $z_{i,t}$, which equals 1 if component i arrives within the time interval t of length Δ. The dynamic equilibrium of on-site inventory is governed by Equations (11-12), where I_t denotes the inventory level at time t, D_t represents the number of consumables used for assembly, and W_{site} is the maximum capacity of the on-site storage yard.

$$shift_i^{time} \geq S_{i,oc} - S_{i,oc}^0, \forall\, i \in I \quad (13)$$

$$shift_i^{time} \geq S_{i,oc}^0 - S_{i,oc}, \forall\, i \in I \quad (14)$$

Finally, Equations (13-14) define the absolute temporal deviation $shift_i^{time}$ by comparing the rescheduled casting start time $S_{i,oc}$ against the original baseline schedule $S_{i,oc}^0$. In summary, the collaborative scheduling model maintains physical and process

feasibility while minimizing deviations from the baseline plan. This provides a solvable MILP formulation for subsequent local rescheduling under event-driven mechanisms.

3 Methods

3.1 Overall Architecture of the Framework

Figure 1 shows the overall architecture of the proposed framework. Within this framework, LLMs are enabling advanced semantic comprehension and causal reasoning upon agent invocation. These agents map dynamic events to specific model patches and executable scripts, synthesizing structured scheduling instructions. To ensure domain alignment, RAG performs semantic similarity searches across specialized databases to retrieve contextually relevant knowledge fragments. This process guides LLMs in generating outputs that strictly adhere to established industry norms and technical constraints[16]. The MAS comprises sense agent (A_{sense}), patch agent (A_{patch}), code agent (A_{code}), scheduling agent ($A_{schedule}$), and dispatch-feedback agent ($A_{dispatch}$). These agents collaborate to form a standardized pipeline, spanning disturbance perception, model updating, code generation, optimization, and instruction dispatching.

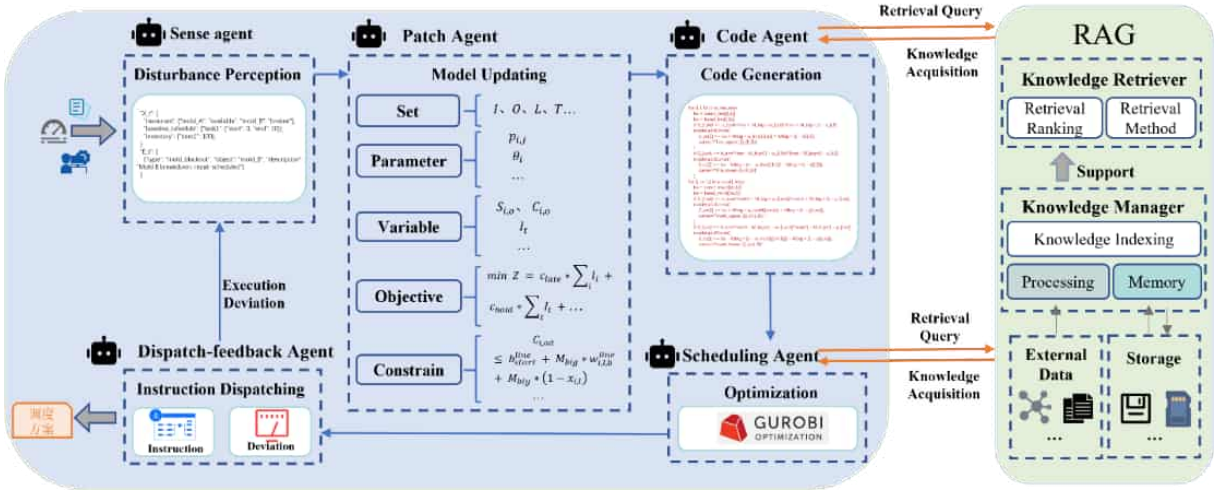


Figure 1. Framework Architecture

Information is transmitted between agents in standardized message formats, collectively completing the management from disturbance perception to instruction dispatching. For any message sent from agent i to agent j, the format is uniformly defined as follows:

$$m_t^{i\to j} = \langle type, payload, t\rangle \tag{15}$$

Here, $type$ specifies the message classification, while $payload$ contains the substantive data encoded as a JSON object. The variable t represents the timestamp, evolving over discrete time horizons $t = 0, 1, 2, \ldots$. Adopting this standardized protocol facilitates modular decoupling and ensures efficient inter-module communication.

3.2 Roles and Functions of AI Agents

The framework comprises five agents, each assigned specific functions and input/output parameters. The rescheduling process is initiated by A_{sense}, which aggregates multi-source data (e.g., operator logs, documents) into production states X_t and disturbance events E_t via function f_{sense}.These structured inputs are transmitted to A_{patch}, as shown in Equations (16-17). A_{patch} maps the disturbances into a mathematical model patch Π_t (Equation 18), which defines parameter variations and constraints, thereby updating the MILP model to M_t (Equation 19).

$$(X_t, E_t) = f_{sense}(NL_t^{req}, docs_t) \tag{16}$$

$$m_t^{sense\to patch} = \langle event_report, \{X_t, E_t\}, t\rangle \tag{17}$$

$$\Pi_t = \langle \Delta\text{Param}, \text{Blackouts}, \text{NewTasks}, \text{Cancels}, \text{WindowShifts}, \text{Freezes}\rangle = f_{patch}(X_t, E_t) \tag{18}$$

$$M_t = P(M_{t-1}, \Pi_t) \tag{19}$$

$$m_t^{patch\to code} = \langle model_update, \{M_t\}, t\rangle \tag{20}$$

Following the model update, A_{code} translates the intermediate model M_t into solver-specific executable code $code_t$, accompanied by a static analysis report R_{mod} to ensure syntax validity (Equation 21). This code

drives $A_{schedule}$, which invokes the optimization solver to generate the optimal plan sol_t and diagnostic metrics $Diag_t$ (Equation 23). Finally, $A_{dispatch}$ operationalizes the plan into production instructions and monitors execution feedback $feedback_{t+1}$ (Equation 25), which serves as the input for the subsequent rescheduling iteration (Equation 26).

$$(code_t, R_{mod}) = f_{code}(M_t) \quad (21)$$

$$m_t^{code \to schedule} = \langle MILP_code, \{code_t, R_{mod}\}, t \rangle \quad (22)$$

$$(sol_t, Diag_t) = f_{schedule}(code_t) \quad (23)$$

$$m_t^{schedule \to dispatch} = \langle plan, \{sol_t, Diag_t\}, t \rangle \quad (24)$$

$$(orders_t, feedback_{t+1}) = f_{dispatch}(sol_t) \quad (25)$$

$$m_{t+1}^{dispatch \to sense} = \langle \text{exec_feedback}, \{feedback_{t+1}\}, t+1 \rangle \quad (26)$$

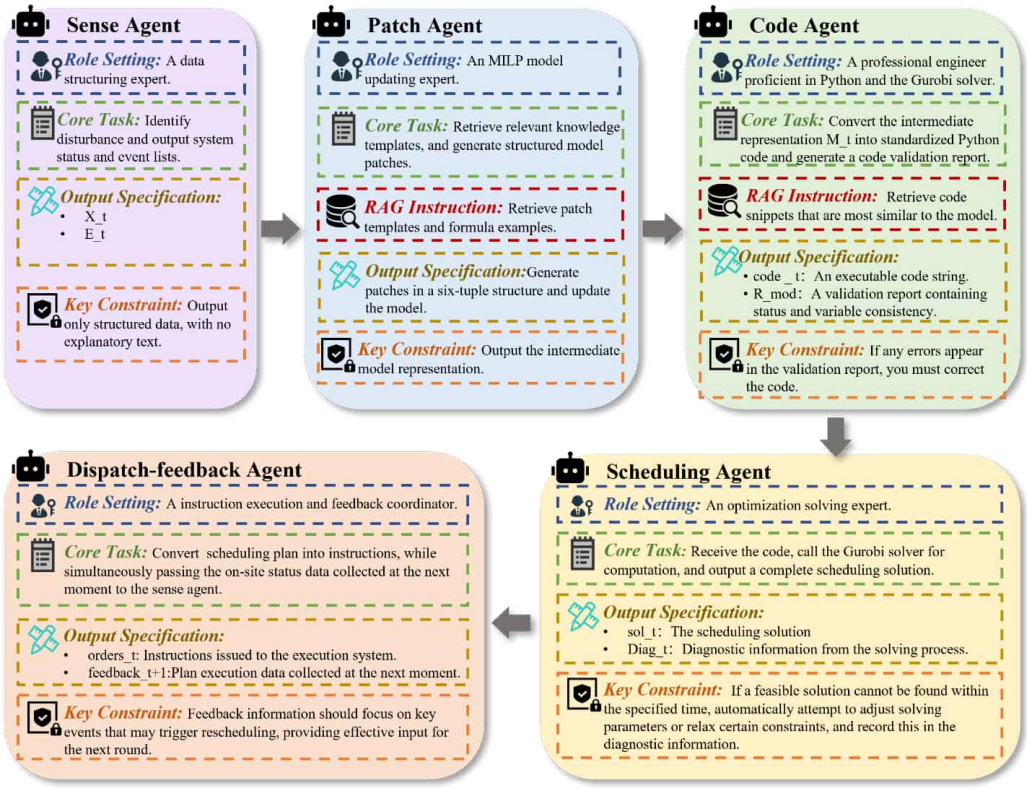


Figure 2. Prompt Structure for Each AI Agent

Prompt engineering enables LLMs to adapt to specific domains through techniques such as task decomposition and example-based guidance[17]. This study designed structured prompts that explicitly specify each agent's role, primary task, output format, and key constraints (see Figure 2). Notably, the prompts for A_{patch} and A_{code} incorporate specific retrieval instructions to leverage external domain knowledge.

3.3 RAG-Based Domain Knowledge Enhancement Mechanism

To ensure factual accuracy and executability, the framework incorporates RAG for A_{patch} and A_{code}. By retrieving relevant domain knowledge via vectorized indexing and injecting it into prompts, RAG mitigates LLM hallucinations and enforces contextual consistency.

3.3.1 Construction of External Knowledge Bases for A_{patch}

The A_{patch} knowledge base maps disturbances to corresponding MILP updates via scenario-patch-MILP pairs. Each entry defines transformation rules for constraints, parameters, and objective functions.

(1) Production Line Unavailability

For disturbances involving production line unavailability, the knowledge base instructs A_{patch} to generate $Blackouts$ patches. A binary auxiliary variable $w_{i,l,b}^{line}$ is introduced to depict the non-overlap logic

between the production line unavailable time $B_l^{line} = [b_{start}^{line}, b_{end}^{line}]$ and production tasks. Constraints (Equations 27-28), active only when $x_{i,l} = 1$, ensure that any tasks assigned to line l do not overlap with the blackout period.

$$C_{i,od} \leq b_{start}^{line} + M_{big} * w_{i,l,b}^{line} + M_{big} * (1 - x_{i,l}), \quad \forall i \in I, \forall l \in L, \forall b \in B_l^{line} \tag{27}$$

$$S_{i,oc} \geq b_{end}^{line} - M_{big} * (1 - w_{i,l,b}^{line}) - M_{big} * (1 - x_{i,l}), \quad \forall i \in I, \forall l \in L, \forall b \in B_l^{line} \tag{28}$$

(2) Emergency Insertion

When an emergency order arrives, it is translated into a $NewTasks$ patch via three update rules: (1) defining the new task set U to expand the total collection to $I^+ = I \cup U$ with initialized decision variables; (2) enforcing the required process and resource constraints for the inserted tasks; and (3) assigning a significantly higher delay penalty w_{i^*} to prioritize urgent insertions. The updated objective function is shown in Equation (29), and the inventory balance constraint is simultaneously extended to I^+, with the updated expression shown in Equation (30).

$$min\ Z = w_{i^*} * \sum_{i^* \in U} l_{i^*} + c_{late} * \sum_{i \in I} l_i + c_{early} * \sum_t e_i + k_{time} * \sum_{i \in I^+} shift_i^{time} + k_{line} * \sum_{i \in I^+} \varphi_i^{line} \tag{29}$$

$$I_t = I_{t-1} + \sum_{i \in I^+} z_{i,t} - D_t, \forall\, t \in T \tag{30}$$

While the preceding discussion primarily addresses disturbances at the production stage, the framework maintains the flexibility to map on-site construction disruptions to specific parameter updates in Equations (1-14). For instance, when encountering exogenous factors such as extreme weather or labor fluctuations, the knowledge base directs A_{patch} to generate a parameter patch ΔParam, reducing the assembly consumption rate D_t for the affected duration and triggers WindowShifts patches to recalibrate the ideal arrival time windows for various components. By continuously indexing disturbances into model parameter updates, the framework ensures its scalability in on-site construction.

3.3.2 Construction of External Knowledge Bases for A_{code}

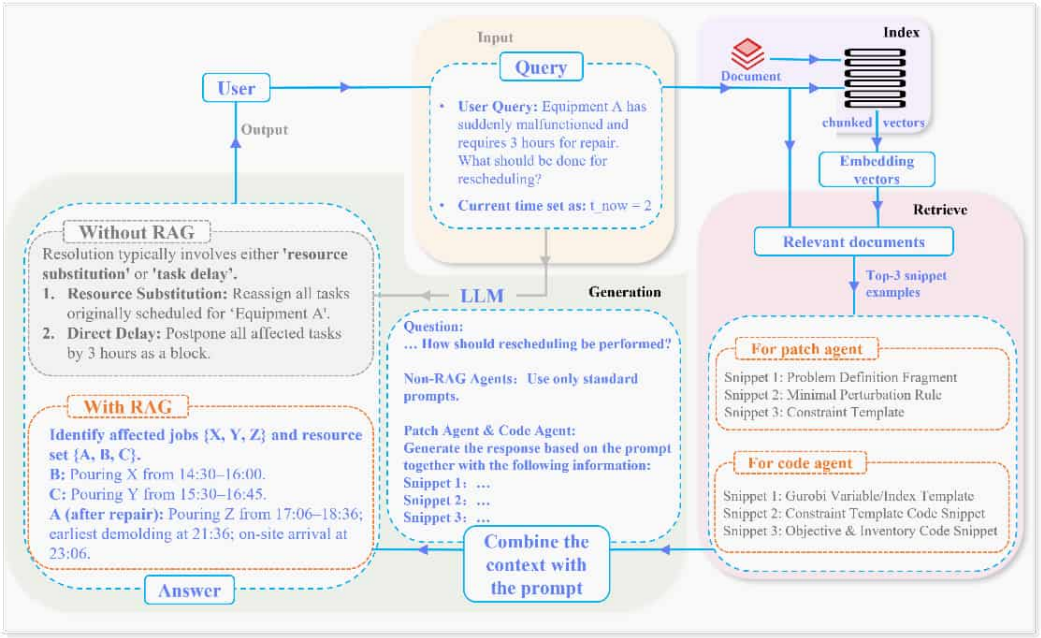


Figure 3. A representative example of the RAG process

The external knowledge base of A_{code} supports the conversion of MILP formulations into executable code by covering common MILP model structures, templates for variable and index definitions, and exemplars for constraint construction and solver interfaces. Given Gurobi's speed and stability for large-scale MILP[18], this study selects Gurobi as the solver for MILP models. To enhance the LLMs' ability to convert formulas into code, the process involves: (1) obtaining normalized MILP expressions from the A_{patch} external knowledge base; (2) driving the LLMs to generate model code using standard prompts without RAG technology; (3) validating executability via the Gurobi solver. This approach avoids semantically correct but non-executable issues that may

arise from relying solely on textual similarity[19-20]. Samples passing this verification are incorporated into the A_{code} knowledge base.

During the reasoning phase, RAG is introduced to reduce the hallucinations in LLM-generated responses and strengthen adherence to domain constraints. When A_{patch} or A_{code} is triggered, the framework retrieves the most contextually relevant template from its corresponding knowledge base. This retrieved knowledge is not memorized by the collaborative scheduling model but dynamically injected as critical contextual information into the original prompt, enhancing generation quality without fine-tuning[17]. The RAG application workflow is illustrated in Figure 3.

Despite the generative capabilities of LLM, the resulting code may encounter execution failures during the solving phase, such as syntax errors, undefined variables, or solver-induced infeasibility. To enhance execution robustness, this framework integrates automated fault diagnosis and repair mechanisms in A_{code} and $A_{schedule}$. In instances of code execution failure, the error message is fed back to A_{code} for context-aware rewriting. Furthermore, when the solver fails to identify a feasible solution, $A_{schedule}$ identifies conflicting constraints and passes the semantic feedback to A_{patch}, which facilitates the dynamic adjustment of non-hard constraints to restore feasibility.

In scenarios where a disturbance is not represented in the existing knowledge base, the framework extracts the original event E_t, the intermediate model M_t, and the diagnostic log $Diag_t$ to generate an anomaly report for expert review. Following manual intervention and verification, the unrecorded disturbance is vectorized and added to the knowledge base. In this way, the knowledge base is continuously expanded through both operational failure records and expert-in-the-loop corrections.

4 Case Study

4.1 Experimental Setup

The experimental environment is implemented in Python 3.9, integrating LangChain for agent orchestration. We employ GPT-4 as the inference model for the AI agents and Gurobi 10.0 as the MILP solver. The case scenario is set in a prefabrication factory containing two parallel production lines (l_1, l_2). The production plan encompasses tasks for facade i_1, composite slab i_2, balcony i_3, air conditioning panel i_4, and staircase i_5. Among these, facades, composite slabs, and staircases are critical path elements, with their late arrival penalty set at a higher rate of 10 units/hour[21]. The early arrival penalty coefficient is set at 50% of the late penalty to reflect that while early arrivals occupy inventory, their consequences are less severe than work stoppages due to material shortages.

The production includes casting, curing, and demolding, with casting and demolding occupying production-line capacity. After completing the casting process, components are transferred via an automated conveyor to the curing chamber. During this period, components are subject only to the minimum curing time constraint and do not occupy production line resources. That is, within the interval $(C_{i,oc}, S_{i,od})$, production line resources can be released for casting or demolding other components. Transportation time is set as $\tau_i = 1.5$ h. Time deviation penalties k_{time} and production line change penalties k_{line} are set at 5 units per hour and 10 units per occurrence, respectively. Other key process parameters are shown in Table 1. For case simplification, only one piece is produced for each component.

Table 1. Component Production Parameters and Logistics Constraints Setting

Component	$p_{i,oc}$ (h)	θ_i (h)	$p_{i,od}$ (h)	c_{late} (units/h)	c_{early} (units/h)
i_1	0.5	10	1.0	10	5
i_2	0.4	10	1.0	10	5
i_3	0.4	10	0.5	8	4
i_4	0.3	10	0.3	8	4
i_5	1.0	10	1.5	10	5

Before production begins, the framework invokes $A_{schedule}$ to construct the MILP model using the aforementioned parameters. All component castings must be intensively completed within the first few hours after t=0 to reserve sufficient curing time. Detailed data for the initial production schedule is shown in Table 2.

Table 2. Initial production schedule

Component	Line	$S_{i,oc} - C_{i,oc}$	$S_{i,od} - C_{i,od}$	a_i	$[E_i, L_i]$
i_1	l_1	0.0-0.5	10.5-11.5	13.0	[12, 15]
i_5	l_1	0.5-1.5	11.5-13.0	14.5	[12, 15]
i_2	l_2	0.0-0.4	10.4-11.4	12.9	[12, 15]
i_3	l_2	0.4-0.8	11.4-11.9	13.4	[12, 15]
i_4	l_2	0.8-1.1	11.9-12.2	13.7	[12, 15]

4.2 Disturbance Scenario Simulation

Production begins at t=0. At t=0.5, the framework receives the natural language log: “Hydraulic system failure on Production Line l_1. Full line shutdown required for 4 hours. Urgent action needed; prioritize staircase delivery.”.

The framework then initiates a collaborative response according to the workflow described in Section 3. A_{sense} parses the log, identifies the affected resource (l_1), the

outage duration, and the impacted task (i_5 casting), and outputs a structured report. A_{patch} generates a mathematical patch Π_t using retrieved knowledge about production line unavailability. The patch freezes completed tasks i_1, imposes fault period constraints on l_1, and reactivates stability penalty terms $k_{time} * \sum_i shift_i^{time} + k_{line} * \sum_i \varphi_i^{line}$ in the objective function. A_{code} converts the patched model into executable code, which is subsequently solved by $A_{schedule}$. Given the high lateness penalty for i_5, $A_{schedule}$ reallocates its casting operation by switching production lines and scheduling it after i_3 and before i_4. Finally, $A_{dispatch}$ is triggered to issue maintenance instructions to l_1 and revised production instructions to l_2.

4.3 Discussion

Table 3 compares a baseline response strategy (Strategy A) with the proposed cross-line rescheduling approach (Strategy B). The results demonstrate that, upon a production-line failure, the proposed framework can parse the disturbance semantics, patch the MILP formulation, and generate an updated production plan.

Strategy B represents an optimized trade-off. By accepting a fixed changeover penalty and small delays for low-priority component i_4, the framework prevents severe delays for critical component i_5. This decision stems from mathematical optimization of the updated model, highlighting the framework's advantage over simple rule-based responses. The end-to-end process completes automatically in under one minute, validating the framework's ability to bridge the semantic-to-model gap and support fully automated rescheduling in practical settings.

Table 3. Full Cost Comparison of Rescheduling Strategies

Cost Item	Strategy A (Wait)	Strategy B (Switch)
$k_{line} * \sum_i \varphi_i^{line}$	0	10.0
$k_{time} * \sum_i shift_i^{time}$	20	6.5
Late arrival penalty of i_5	35.0	0.0
Late arrival penalty of i_4	0.0	1.6
Z	55.0	18.1

5 Conclusion and Future Work

This study proposes an intelligent rescheduling framework that integrates the semantic reasoning of AI agents with the optimization capabilities of MILP. By positioning LLMs as the core cognitive processors, the framework delivers two primary contributions. First, it introduces an automated semantic-to-constraint translation mechanism that autonomously converts unstructured disturbance reports into formal mathematical constraints, thereby accelerating the transition from perception to solution. Second, the framework facilitates localized model refinement, which preserves computational efficiency by bypassing the need for a full model reconstruction.

However, current validation is limited to small-scale disturbance scenarios involving five components. This initial stage serves primarily as a proof-of-concept to demonstrate the feasibility of employing LLM-centric cognitive cores in scheduling workflows. As project scales expand, exact solvers often encounter exponential computational complexity. Therefore, future work should evaluate the framework against large-scale datasets featuring multi-node concurrent disturbances.

Acknowledgement

This study is supported by the National Natural Science Foundation of China (72571252), and the Science and Technology R&D Program of CSCEC International Operations (CSCI-2023-Z-5). Any opinions and findings do not necessarily reflect the views of the grantor.

References

[1]H. Zhong, K. Chen. Strategic planning and control in industrialized construction. *Automation in Construction*, 177: 15, 2025.

[2]J. Zhang, X. Song, E. Diaz. Critical chain project buffer sizing based on resource constraints. *International Journal of Production Research*, 55(3): 671-683, 2017.

[3]M. Hussein, A. E. E. Eltoukhy, A. Karam, I. A. Shaban, T. Zayed. Modelling in off-site construction supply chain management : a review and future directions for sustainable modular integrated construction. *Journal of Cleaner Production*, 310: 32, 2021.

[4]H. Liu, S. Wang, T. Yang, Z. Chen. Optimized transportation scheduling for precast concrete components considering heterogeneous vehicle-size matching. *Advanced Engineering Informatics*, 62: 15, 2024.

[5]A. Mohammadi, M. Tavakolan. Modeling the effects of production pressure on safety performance in construction projects using system dynamics. *Journal of Safety Research*, 71: 273-284, 2019.

[6]Y. Song, J. Wang, J. Lu, X. Si. Research on

collaborative scheduling of multiple projects of prefabricated building based on the niche genetic-raccoon family optimization algorithm. *Alexandria Engineering Journal*, 64: 1015-1033, 2023.

[7]T. van der Beek, J. T. van Essen, J. Pruyn, K. Aardal. Exact solution methods for the resource constrained project scheduling problem with a flexible project structure. *European Journal of Operational Research*, 322(1): 56-69, 2025.

[8]Z. Wang, H. Xu. Scheduling redundancy optimization for precast production to enhance disruption management in prefabricated construction. *Computers & Industrial Engineering*, 192: 15, 2024.

[9]K. Sanogo, A. M. Benhafssa, M. Sahnoun, B. Bettayeb, M. Abderrahim, A. Bekrar. A multi-agent system simulation based approach for collision avoidance in integrated job-shop scheduling problem with transportation tasks. *Journal of Manufacturing Systems*, 68: 209-226, 2023.

[10]G. M. Ali, A. Bouferguene, M. Al-Hussein. Crane mat layout optimization based on agent - based greedy and reinforcement-learning approach. *Journal of Construction Engineering and Management*, 149(8): 17, 2023.

[11]X. Zhang, C. Zhang, J. Sun, J. Xiao, Y. Yang, Y. Luo. EduPlanner : LLM-based multiagent systems for customized and intelligent instructional design. *IEEE Transactions On Learning Technologies*, 18: 416-427, 2025.

[12]Y. Ma, S. Zheng, Z. Yang, P. Zheng, J. Leng, J. Hong. Leveraging large language models in next generation intelligent manufacturing : retrospect and prospect. *Journal of Manufacturing Systems*, 82: 809-840, 2025.

[13]L. Zhao, W. Cheng, L. Meng, C. Zhang, Y. Ren, B. Zhang, P. Duan. MILP modeling and optimization of flexible job shop scheduling problem with preventive maintenance. *Computers & Industrial Engineering*, 201: 16, 2025.

[14]Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. D. Yuxi Bic, J. Sun, Meng, Wang, H. Wang. Retrieval-augmented generation for large language models: a survey. arXiv: 2312.10997, 2024.

[15]K. Shen, X. Li, X. Cao, Z. Zhang. Prefabricated construction process optimization based on rework risk. *Journal of Construction Engineering and Management*, 148(6): 14, 2022.

[16]P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, D. Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv: 2005.11401, 2020.

[17]Y. Xie, J. Liu, R. Wang, Z. Wang, K. Yu, Z. Song. Rapid generation method of process routes based on multi-agent collaboration with LLMs. *Advanced Engineering Informatics*, 68: 18, 2025.

[18]B. Zhang, P. Luo, G. Yang, B. Soong, C. Yuen. OR-LLM-agent: automating modeling and solving of operations research optimization problems with reasoning LLM. arXiv:2503.10009, 2025.

[19]J. W. L. Shi, W. Solihin, J. K. W. Yeoh. Fine-tuning a large language model for automated code compliance of building regulations. *Advanced Engineering Informatics*, 68: 103676, 2025.

[20]M. Peng, Z. Chen, J. Yang, J. Huang, Z. Shi, Q. Liu, X. Li, L. Gao. Automatic MILP model construction for multi-robot task allocation and scheduling based on large language models. arXiv:2503.13813, 2025.

[21]J. Yin, R. Huang, H. Sun, S. Cai. Multi-objective optimization for coordinated production and transportation in prefabricated construction with on-site lifting requirements. *Computers & Industrial Engineering*, 189: 15, 2024.

Multi-Horizon Forecasting of Indoor $PM_{2.5}$ Using Machine Learning in a Digital Twin Environment

Zihao Zheng[1], Karunakar Reddy Mannem[2], Borja Garcia de Soto[1]

[1]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
[2]Center for Research Computing, New York University Abu Dhabi (NYUAD), Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
zihao.zheng@nyu.edu, krm7@nyu.edu, garcia.de.soto@nyu.edu

Abstract –

Indoor air quality management remains largely reactive due to limited multi-horizon forecasting capabilities and deployment challenges under real-world constraints. This paper presents a Digital Twin (DT)-enabled framework for multi-horizon indoor $PM_{2.5}$ forecasting from 1 to 6 hours ahead using machine learning and multi-source environmental data from a university campus in Abu Dhabi, UAE. Indoor sensor measurements are combined with outdoor monitoring data and ERA5 wind variables used with a realistic five-day availability delay. CatBoost and Transformer models are evaluated using seasonal cross-validation, with CatBoost demonstrating advantages at short-term horizons while Transformer marginally performs better at the long-term horizon. SHAP analysis reveals that current indoor $PM_{2.5}$ is the dominant predictor for short-term forecasts. An ablation study demonstrates that incorporating lagged wind data yields statistically significant improvements in non-winter seasons.

Keywords –

Indoor air quality; CatBoost; Transformer; Reanalysis data

1 Introduction

Indoor air quality (IAQ) has gained increasing attention globally, particularly as people spend approximately 90% of their time in indoor environments and ambient $PM_{2.5}$ concentrations remain a persistent public health concern in many urban environments [1]. Among various indoor pollutants, particulate matter with a diameter of less than 2.5 micrometers ($PM_{2.5}$) poses significant health risks, including diabetes and prenatal disorders [2]. Indoor $PM_{2.5}$ primarily enters indoor environments through ventilation and infiltration processes, or is generated internally by human activities such as printing and photocopying, cleaning and maintenance activities (e.g., use of disinfectant sprays), frequent door opening that resuspends settled dust, and movement of occupants [2]. Many studies have proposed models to predict indoor PM2.5, but far fewer have dealt with what happens when those models must run continuously with delayed external data, mixed sensor streams, and decision-support needs in a live DT.

Digital Twin (DT) technology offers a transformative paradigm for data-driven IAQ management by creating virtual replicas of physical indoor spaces that enable real-time monitoring, prediction, and optimization. Current $PM_{2.5}$ monitoring systems in DT applications are primarily reactive, alerting occupants after pollutant concentrations exceed threshold levels. With advanced predictive analytics, DTs enable proactive measures for indoor $PM_{2.5}$ risk prevention and support preventive health management [3]. However, the application of predictive models to indoor environments faces unique challenges, particularly the need for real-time predictions with integration of available near real-time data sources such as indoor sensors, outdoor atmospheric monitoring stations, and satellite reanalysis data. What is still limited is the deployment of multi-horizon forecasting that can give operators useful lead time while relying only on data that are available at prediction time. Besides, wind-related variables from ERA5, the global meteorological reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [4], which are widely used in indoor and outdoor $PM_{2.5}$ studies for reliable wind data [5], [6], are subject to an inherent latency of approximately five days before becoming available. To the best of the authors' knowledge, the influence of this latency on indoor PM2.5 prediction has not been fully investigated, representing an operational constraint in real-time prediction deployment in DT applications.

This paper presents a framework that integrates near real-time multi-source data to forecast indoor $PM_{2.5}$ within a DT managed environment. Environmental data from indoor sensors ($PM_{2.5}$, humidity, and pressure), outdoor atmospheric monitoring stations (PM_1, $PM_{2.5}$, PM_{10}, temperature, humidity, pressure, and dew point), and lagged ERA5 meteorological reanalysis data (U and

V wind components, wind gust) are integrated to train CatBoost, Transformer, and persistence models for predictions up to six hours ahead. The contribution is not a new forecasting algorithm; it is a deployment-oriented forecasting workflow that accounts for data latency, tests whether delayed wind information is still useful, and shows how multi-horizon predictions can be pushed into a live DT interface.

2 Previous Studies

2.1 Hourly Indoor $PM_{2.5}$ Prediction

Recent studies have employed diverse machine learning architectures for indoor $PM_{2.5}$ forecasting. Recurrent architectures including LSTM, BiLSTM, and GRU have been widely adopted for capturing temporal dependencies in time-series air quality data [7]. Gradient boosting methods such as XGBoost and LightGBM demonstrate strong performance through ensemble learning approaches [6], [8]. Sparse attention–based Transformer models efficiently capture long-term dependencies in $PM_{2.5}$ time series and demonstrate improved performance for both short-term and long-term air quality forecasting [9]. Feature selection for indoor $PM_{2.5}$ prediction typically incorporates indoor environmental parameters ($PM_{2.5}$, temperature, humidity), outdoor atmospheric conditions (meteorological variables, ambient pollutants), and contextual factors (occupancy, season, time), resulting in a total of 10 to 21 features used for prediction [10], [11]. Even so, most published studies emphasize model accuracy, whereas deployment-oriented questions such as data latency, inference timing, and how much lead time is practically useful receive much less attention.

Current studies on indoor $PM_{2.5}$ forecasting primarily focus on model development and performance validation, with limited attention to real-world deployment challenges and operational constraints.

2.2 Predictive Digital Twin Application

DT technology has emerged as a promising platform for IAQ management, integrating IoT sensors, Building Information Modeling (BIM), and machine learning for real-time monitoring and visualization. Hu et al. [12] proposed a CNN-BiLSTM-MHA model for IAQ failure prediction integrated with a DT platform, demonstrating real-time alert capabilities but emphasizing anomaly detection over continuous predictive forecasting. Khosh-Amadi et al. [13] proposed an AIR-DT framework integrating IoT sensors, BIM, spatial analytics, and short-term IAQ prediction to enable real-time, asthma-oriented indoor air quality monitoring and intervention. Qian et al. [14] presented an intelligent IAQ management system based on DT platforms that integrates multi-source data from IoT and BIM, achieving rapid prediction and control responses within 30 minutes of threshold exceedance.

Despite these developments, critical gaps remain in integrating heterogeneous near real-time data sources from multiple domains (indoor, outdoor, meteorological) and implementing multi-horizon forecasting models that provide actionable lead times for building management interventions.

3 Framework of Digital Twin Enabled Predictive Analysis

The proposed framework comprises three layers encompassing seven processes, shown in Figure 1.

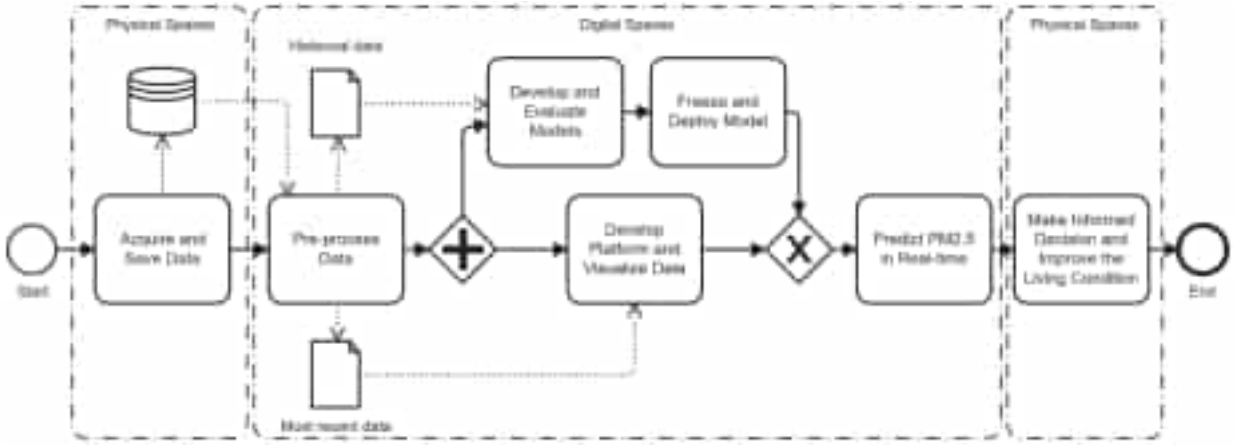


Figure 1: BPMN diagram for Digital Twin-enabled IAQ improvement with indoor $PM_{2.5}$ prediction

Data Acquisition Layer: acquire and save data from indoor sensors, outdoor stations, and ERA5 reanalysis; pre-process data through outlier removal following Tukey 1.5 × IQR rule, data imputation using KNN

method [15], and hourly aggregation and temporal alignment.

Model Development Layer: develop and evaluate models using CatBoost and Transformer through seasonal cross-validation and Diebold-Mariano (DM) tests [16], then freeze and deploy models in the backend DT system.

Broadcasting of Real-time Prediction Layer: develop platform and visualize data, predict $PM_{2.5}$ in real-time, make informed decisions and improve living conditions, enabling bilateral information flow in DT.

4 Implementation

4.1 Data Acquisition

Three categories of environmental data were collected to capture the multi-scale influences on indoor $PM_{2.5}$ concentrations. Indoor environmental parameters were measured using sensors deployed at NYU Abu Dhabi, recording $PM_{2.5}$, humidity, and pressure at 1-minute intervals [17]. Outdoor atmospheric conditions were monitored using two complementary systems. Outdoor air quality stations provided measurements of particulate matter (PM_1, $PM_{2.5}$, PM_{10}), temperature, relative humidity, and atmospheric pressure at 1-minute intervals, and meteorological stations supplied dew point observations at 15-minute intervals.

The realistic 5-day lag constraint of the ERA5 data source was considered in this study. Hourly measurements of 10-m wind gust, as well as zonal (U) and meridional (V) wind components, were extracted from ERA5 for the New York University Abu Dhabi campus location. The complete dataset spans from April 6, 2024, to November 16, 2025, covering approximately 19 months of observations. All variables used in this study are summarized in Table 1.

As part of data preprocessing, 3.2% of indoor sensor readings, 4.0% of outdoor measurements, and 0.7% of meteorological station data were classified as outliers and excluded from further analysis. All variables were subsequently resampled and temporally aligned to an hourly resolution using mean aggregation. Missing values were handled using k-nearest neighbors (KNN) imputation, which filled 3.6% of indoor sensor data, 2.9% of outdoor data, and 22.9% of meteorological station data. Although the meteorological station dataset (dew point) exhibited a higher proportion of missing values, this feature was retained due to its moderately positive Pearson correlation of 0.28 with indoor $PM_{2.5}$. The final dataset comprises 14,129 hourly observations with 13 input features.

An overview of the indoor $PM_{2.5}$ after the data pre-processing is shown in Figure 2. The monthly statistics reveal substantial temporal variability, with mean $PM_{2.5}$ concentrations ranging from 4.1 μg/m³ in May 2024 to 10.2 μg/m³ in October 2025. Lower concentrations were generally observed during winter and spring months (December-May), while elevated levels occurred during summer and fall months (June-November), a pattern consistent with the influence of intense heat, enhanced secondary aerosol formation, frequent dust events, and reduced atmospheric dispersion during warmer periods in Abu Dhabi. Accordingly, a seasonally stratified cross-validation strategy was implemented to evaluate model performance under varying climatic conditions.

Table 1: Parameters collected from different sources

Abbreviation	Parameter	Data source
Ind. PM2.5	Indoor $PM_{2.5}$	Indoor*
Ind. Humi.	Indoor Humidity	Indoor
Ind. Press.	Indoor Pressure	Indoor
Outd. PM1	Outdoor PM_1	Outdoor**
Outd. PM2.5	Outdoor $PM_{2.5}$	Outdoor
Outd. PM10	Outdoor PM_{10}	Outdoor
Outd. Temp.	Outdoor Temperature	Outdoor
Outd. Humi.	Outdoor Humidity	Outdoor
Outd. Press.	Outdoor Pressure	Outdoor
Dew Point	Dew Point	Meteo. station***
Wind Gust	Instantaneous 10-meter wind gust	ERA5
U Wind	10-meter U wind component	ERA5
V Wind	10-meter V wind component	ERA5

** Sensors provided and maintained by the S.M.A.R.T. Construction Research Group at NYUAD*
*** Sensors provided and maintained by CITIES at NYUAD*
**** Sensors provided and maintained by ACCESS at NYUAD*

4.2 Model Development

4.2.1 Model Selection and Configuration

CatBoost gradient boosting and Transformer neural network architectures were selected for systematic comparison based on their respective strengths in ensemble learning with feature interaction capabilities [8] and attention-based temporal modeling [9]. CatBoost is a strong choice for heterogeneous tabular inputs with nonlinear interactions and relatively limited feature engineering, while the Transformer is well suited to learning longer temporal dependencies from lagged sequences. Preliminary experiments with other models, such as Random Forest and BiLSTM, showed almost 50% higher RMSE and were therefore excluded from this study. Additionally, a persistence baseline model, which assumes future $PM_{2.5}$ concentrations remain equal to the current observed value, was included for statistical comparison as a benchmark baseline [10].

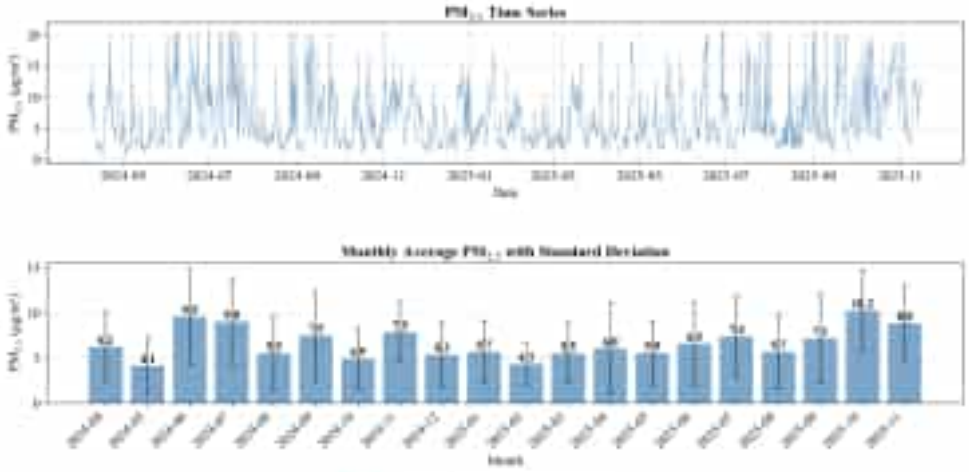


Figure 2: The overview and monthly average of the indoor $PM_{2.5}$

As previously indicated, the ERA5 API exhibits a five-day latency in data acquisition, which poses a limitation for real-time prediction in DT applications. Correlation analysis reveals that Wind Gust, U Wind, and V Wind exhibit moderate Pearson correlations of -0.27, -0.12, and 0.06, respectively, with indoor $PM_{2.5}$ when using contemporaneous data, but this operational lag constraint reduces these correlations to -0.04, -0.03, and -0.01. Nevertheless, the lagged wind features were retained as machine learning models may capture long-term and non-linear relationships that traditional linear regression cannot easily detect, potentially revealing residual atmospheric circulation patterns relevant to $PM_{2.5}$ dynamics. NaN wind data at the beginning of the dataset due to the five-day latency were filled using the earliest available wind observations.

Lag features are applied to all components to capture their temporal dependencies and periodicities [8]. Specifically, for each prediction horizon h, the input feature set includes observations from the current hour and the preceding six hours (at times t through $t-6$), as well as the corresponding hourly observation from 24 hours prior to the target prediction time (at $t+h-24$). All features were normalized using min-max scaling to ensure comparability across features and to support stable model training.

To evaluate model performance under different climatic conditions, a rolling-window cross-validation strategy was employed. Four folds (i.e., temporal splits) were constructed, as summarized in Table 2. For each fold, data from the eight months preceding the target season were used for model training. The data within the target season were then chronologically divided into two equal parts, with the first half used for validation and the second half reserved for testing. Across all folds, the approximate train/validation/test ratio is 72%/14%/14%.

The Transformer model was implemented using PyTorch, while the CatBoost model was implemented using the CatBoost Python library. Configurations of both models are listed in Table 3.

Table 2: Dataset splitting for cross validation

Split	Targeting Season	Time Frame
1	Spring	March, 2025 - May, 2025
2	Summer	June, 2025 - Aug., 2025
3	Fall	Sep., 2025 - Nov., 2025
4	Winter	Dec., 2024 - Feb., 2025

Table 3: Model configurations

CatBoost	Transformer
Max iterations: 2000	Max epochs: 200
Early stopping: yes	Early stopping: yes
Loss function: MAE	Loss function: MAE + Huber
Eval metric: MAPE	(delta=2)
Learning rate: 0.05	Eval metric: MAPE
Depth: 3	Batch size: 32
L2 leaf reg: 5	Learning rate: 0.0005
Bagging temperature:	Weight decay: 0.0001
0.8	Embedding dimension: 64
Random strength: 0	Number of attention heads: 8
	Transformer layers: 3
	Dropout rate: 0.1

4.2.2 Results of Model Training

Performance metrics were computed for each cross-validation fold and for each prediction horizon (H1-H6). Model performance was evaluated across RMSE, MAE, MAPE, and R^2 at six prediction horizons, as shown in Figure 3-Figure 6.

1. Performance comparison of horizons

The results indicate a consistent decline in prediction accuracy as the forecast horizon extends from H1 to H6, a trend observed across all models and validation splits.

For CatBoost, mean performance metrics averaged over all splits indicate RMSE values of 0.9412, 1.3176, 1.6461, 1.9479, 2.2253, and 2.4647 μg/m³ (Figure 3), and MAE values of 0.4854, 0.8147, 1.0878, 1.3006, 1.5342, and 1.6996 μg/m³ across the six horizons (Figure 4). Corresponding mean MAPE increased from 8.3% to 28.6% (Figure 5), while mean R^2 decreased from 0.9420 to 0.6332 (Figure 6).

Despite this degradation, CatBoost and Transformer models at the H6 demonstrate predictive ability with average R^2 values of 0.6332 and 0.6278, surpassing the persistence model of 0.5797 by 9.2% and 8.3% respectively.

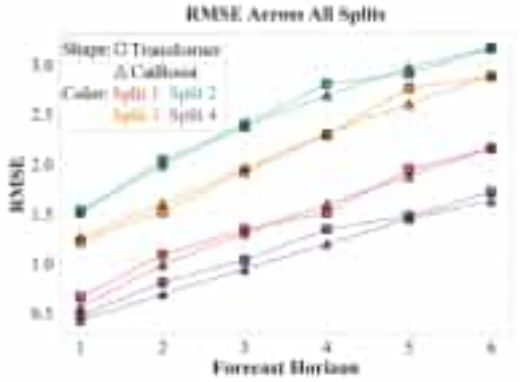


Figure 3: RMSE across all splits and 6 horizons

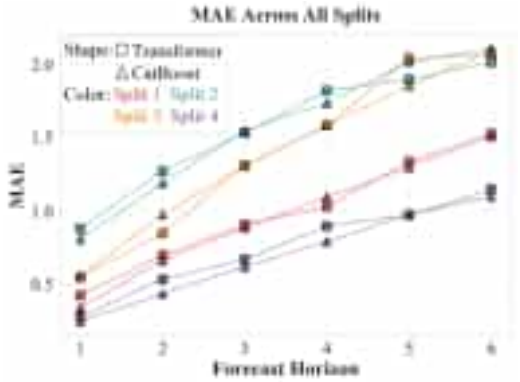


Figure 4: MAE across all splits and 6 horizons

2. Performance comparison of models

CatBoost demonstrated lower predictive error (RMSE, MAE, and MAPE) at short-term horizons compared to both Transformer and persistence models. At H1, CatBoost achieved an average MAE of 0.4854 μg/m³ across the four splits, compared to 0.5297 μg/m³ for Transformer and 0.5062 μg/m³ for persistence, representing improvements of 9.1% and 4.3% respectively. At H6, this reverses, with Transformer achieving 1.6870 μg/m³ while CatBoost achieving 1.6996 μg/m³. Both machine learning models outperformed the 1.7516 μg/m³ of persistence baseline by 3.1% and 3.8% respectively. CatBoost's gradient boosting architecture effectively captures the temporal dynamics and feature interactions governing indoor $PM_{2.5}$ evolution, particularly at shorter prediction horizons while Transformer marginally surpasses CatBoost at long-term prediction at H6.

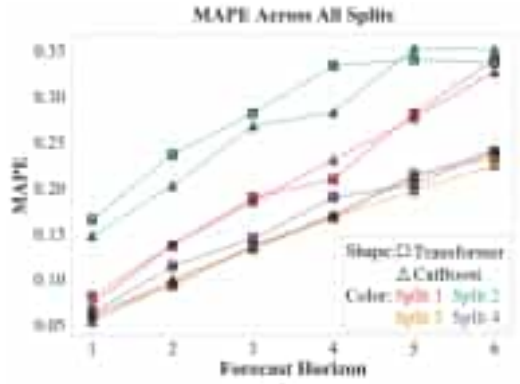


Figure 5: MAPE across all splits and 6 horizons

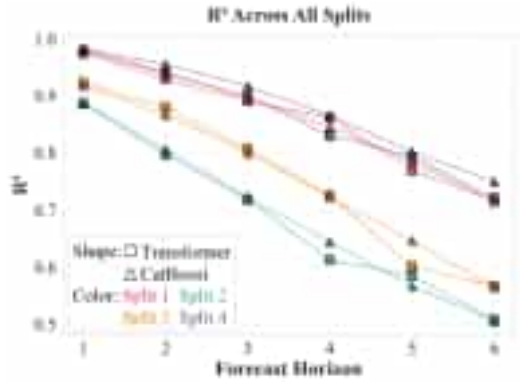


Figure 6: R^2 across all splits and 6 horizons

3. Performance comparison of seasonal split

Seasonal variations significantly influenced model performance across the four cross-validation splits, with Splits 1 (spring, in red) and 4 (winter, in purple) demonstrating notably better performance than Splits 2 (summer, in green) and 3 (fall, in orange). For Split 4, CatBoost achieved the lowest errors with MAE of 0.2448 μg/m³ at H1 and 1.0967 μg/m³ at H6, while for Split 1 CatBoost showed similarly favorable results with MAE of 0.3405 μg/m³ at H1 and 1.5067 μg/m³ at H6. In contrast, Splits 2 and 3 exhibited higher errors, with CatBoost's MAE at H6 reaching 2.1083 μg/m³ and 2.0868 μg/m³, representing increases of 92.2% and 90.3%, respectively, compared with Split 4.

This seasonal performance pattern is consistent with

the $PM_{2.5}$ concentration trends shown in Figure 2. Spring and winter are characterized by lower and more stable pollutant levels, resulting in smoother temporal dynamics that are easier for data-driven models to learn. In contrast, summer and fall exhibit higher mean concentrations and greater short-term variability, driven by dust events, heat-induced atmospheric stagnation, and changing wind patterns in the UAE. These factors increase the volatility of indoor $PM_{2.5}$, thereby amplifying forecast uncertainty and contributing to the observed degradation in model performance for Splits 2 and 3.

4.2.3 Statistical Validation

One-sided DM test was employed to evaluate the statistical significance of performance differences between models using absolute error. Table 4 presents the DM statistics and p-values for all pairwise comparisons across prediction horizons. For example, at the two-hour forecast horizon (H2), the DM statistic for CatBoost vs. Transformer is -1.91 with a p-value of 0.0281. Because the statistic is negative and the p-value is below 0.05, this indicates that CatBoost performs significantly better than the Transformer at this horizon across all splits, and the observed improvement is unlikely to be due to random variation.

Table 4: DM test results on different horizons

Horizon	CatBoost vs Persistence	Transformer vs Persistence	CatBoost vs Transformer
1	-3.47 (0.0003)	-1.53 (0.0635)	-0.81 (0.2096)
2	-3.86 (0.0001)	-1.28 (0.0997)	-1.91 (0.0281)
3	-4.07 (0.0000)	-3.08 (0.0010)	-0.61 (0.2714)
4	-4.02 (0.0000)	-2.75 (0.0029)	-1.00 (0.1577)
5	-3.61 (0.0002)	-2.48 (0.0065)	-1.11 (0.1333)
6	-3.20 (0.0007)	-3.02 (0.0013)	-0.19 (0.4242)

Note: Data pair format - DM statistic (p value). Negative DM statistic indicates superior performance of the first model. $p < 0.01$ indicates strong significance; $p < 0.05$ indicates significance; $p < 0.10$ indicates marginal significance.

The CatBoost model significantly outperformed the persistence baseline across all horizons ($p < 0.01$) while the Transformer model significantly outperformed the persistence baseline from H3-H6 ($p < 0.01$) and marginally outperformed persistence baseline at H1 and H2 ($p < 0.1$). Both models showed their ability to capture temporal dynamics beyond simple autocorrelation. CatBoost exhibited the stronger superiority over persistence, with DM statistics ranging from -3.20 to -4.07 ($p < 0.01$). Comparing the two machine learning models, CatBoost demonstrated marginal significance over Transformer at horizon H2 ($p < 0.05$). On the other horizons, the performance differences were not statistically significant ($p > 0.10$), indicating comparable predictive capabilities at these horizons.

4.2.4 Importance Analysis

The model interpretability and the importance of the features are supported by the SHAP (Shapley Additive exPlans). Figure 7 shows an example of feature importance and directional impact for indoor $PM_{2.5}$ at H1 of split 4 of CatBoost model. It shows that indoor $PM_{2.5}$ at the current emerges as the dominant predictor, with high feature values strongly driving higher predictions, while outdoor PM_1 ranks second, reflecting the close coupling between indoor and outdoor particulate matter. Contribution of wind gust feature emerges after indoor $PM_{2.5}$ and other outdoor features.

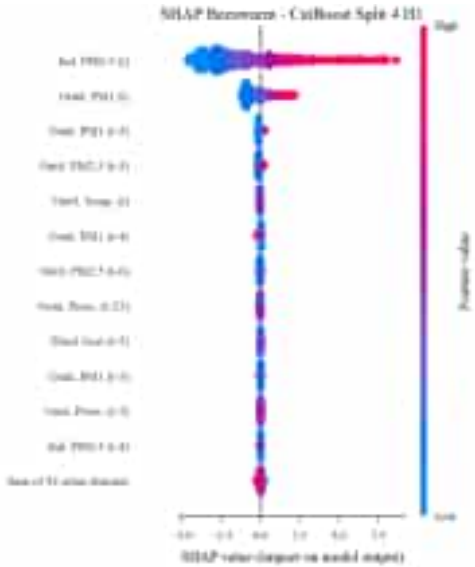


Figure 7: SHAP Beeswarm of CatBoost (H1)

4.3 Broadcasting of Real-time Prediction

A implementation of the real-time forecasting module is shown in Figure 8 and is accessible on the prototype online platform [18]. The data ingestion layer continuously acquires the most recent environmental measurements. Wind inputs are retrieved from the ERA5 reanalysis through the ECMWF API and incorporated into the forecasting workflow under a five-day availability delay, thereby ensuring consistency with operational data-access constraints. The forecasting service executes multi-horizon inference using the trained CatBoost models and publishes prediction outputs to the DT interface for visualization.

Although the models are developed using the data

collected at NYUAD campus, the workflow and three-layer framework can be transferred to other buildings that have comparable indoor sensing, outdoor measurements, and meteorological inputs.

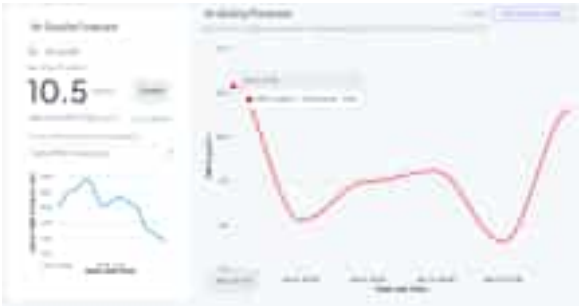


Figure 8: Screenshot of real-time forecasting platform

5 Ablation Study: Lagged Wind Feature

To validate whether using the lagged wind features provides meaningful predictive value, an ablation study was conducted using the CatBoost model aggregating all horizons. Table 5 shows the DM test results comparing the "Lagged ERA5" and "No ERA5" configurations over the four splits. It indicates that predictions with "Lagged ERA5" wind features are strongly significant in Split 1 (spring, $p < 0.01$) and 2 (summer, $p < 0.01$), marginally significant in Split 3 (fall, $p < 0.1$), while not significant in Split 4 (winter).

Table 5: DM test results of "Lagged ERA5" vs. "No ERA5" configurations

Split 1	Split 2	Split 3	Split 4
-3.42 (0.0003)	-3.31 (0.0005)	-1.51 (0.0658)	3.66 (0.9999)

Table 6 compares MAE performance across all prediction horizons for the "Lagged ERA5" and "No ERA5" configurations in Split 1. Overall, incorporating lagged ERA5 wind variables yields modest performance gains, with MAE variations remaining within ±2% across forecast horizons H2–H5.

Table 6: MAE (μg/m³) comparison of "Lagged ERA5" and "No ERA5" configurations in Split 1

Horizon	Lagged ERA5	No ERA5	Improvement
1	0.3405	0.3387	-0.5%
2	0.6594	0.6605	0.2%
3	0.8844	0.8863	0.2%
4	1.0954	1.1091	1.3%
5	1.2960	1.3070	0.9%
6	1.5067	1.5149	0.5%

6 Limitations

The five-day latency in ERA5 wind data yields only modest improvements (under 2%), suggesting a need for more recent wind data sources to enhance predictive performance. Reliance on a single reanalysis dataset (ERA5) and a single imputation method (KNN) may introduce dataset- or method-specific biases. In addition, all predictors used here are observational rather than directly manipulatable. Occupancy-related activities and HVAC operational variables were not available, even though such inputs would be valuable in a DT setting because they are closer to actionable control decisions. Finally, the framework was evaluated on a single campus in Abu Dhabi with a specific building configuration, and its generalizability to other climates and building configurations remains untested.

7 Conclusion

This paper presents a DT-enabled framework for multi-horizon indoor $PM_{2.5}$ forecasting using multi-source environmental data. CatBoost and Transformer models were evaluated using seasonal cross-validation, with CatBoost achieving average MAE of 0.4854 μg/m³ ($R^2 = 0.9420$) at one hour and 1.6996 μg/m³ ($R^2 = 0.6332$) at the six-hour horizon. Both models significantly outperformed persistence baselines, with CatBoost demonstrating statistically significant advantages at several intermediate horizons. SHAP analysis shows that current indoor $PM_{2.5}$ dominates short-term predictions, with its importance declining from 0.408 to 0.300 in CatBoost. An ablation analysis comparing "Lagged ERA5" and "No ERA5" configurations confirmed that incorporating lagged wind features yields statistically significant improvements, strengthening the case for deployment-ready forecasting.

Overall, the proposed framework provides a scalable foundation for data-driven indoor air quality management in DT–enabled environments. Future work should examine alternative real-time meteorological data sources, extend validation across diverse indoor settings and climates, and integrate building operational and occupancy information to further enhance predictive performance.

Acknowledgement

This research was partially supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013. The authors also gratefully acknowledge CITIES and ACCESS for providing access to sensor data used in this paper.

References

[1] J. Cai, J. Chen, Y. Hu, S. Li, and Q. He, "Digital twin for healthy indoor environment: A vision for the post-pandemic era," *Front. Eng. Manag.*, vol. 10, no. 2, pp. 300–318, Jun. 2023, doi: 10.1007/s42524-022-0244-y.

[2] N. R. Martins and G. Carrilho da Graça, "Impact of PM2.5 in indoor urban environments: A review," *Sustainable Cities and Society*, vol. 42, pp. 259–275, Oct. 2018, doi: 10.1016/j.scs.2018.07.011.

[3] J. D. Castaño Molina, Z. Zheng, and B. García de Soto, "Predicting Indoor PM2.5 Levels Using Deep Learning for Enhanced Digital Twin Applications," in *Proceedings of the 2025 European Conference on Computing in Construction*, Porto, Portugal, Jul. 2025, pp. 1–8. doi: 10.35490/EC3.2025.277.

[4] H. Hersbach *et al.*, "ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)." Accessed: Jan. 09, 2026. [Online]. Available: doi: 10.24381/cds.adbb2d47

[5] Z. Wang, P. Chen, R. Wang, Z. An, and L. Qiu, "Estimation of PM2.5 concentrations with high spatiotemporal resolution in Beijing using the ERA5 dataset and machine learning models," *Advances in Space Research*, vol. 71, no. 8, pp. 3150–3165, Apr. 2023, doi: 10.1016/j.asr.2022.12.016.

[6] W. Yu, B. Nakisa, S. W. Loke, S. Stevanovic, Y. Guo, and M. N. Rastgoo, "Indoor PM2.5 forecasting and the association with outdoor air pollution: a modelling study based on sensor data in Australia," May 13, 2024, *arXiv*: arXiv:2405.07404. doi: 10.48550/arXiv.2405.07404.

[7] B. Eren, İ. Aksangür, and C. Erden, "Predicting next hour fine particulate matter (PM2.5) in the Istanbul Metropolitan City using deep learning algorithms with time windowing strategy," *Urban Climate*, vol. 48, p. 101418, Mar. 2023, doi: 10.1016/j.uclim.2023.101418.

[8] A. Singh, M. Islam, and N. Dinh, "Forecasting Indoor Air Quality Using Machine Learning Models," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2024, pp. 1–6. doi: 10.1109/ICCE59016.2024.10444387.

[9] Z. Zhang and S. Zhang, "Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks," *Int. J. Environ. Sci. Technol.*, vol. 20, no. 12, pp. 13535–13550, Dec. 2023, doi: 10.1007/s13762-023-04900-1.

[10] S. A. Asklany, D. Mohammed, I. K. Youssef, M. Nawaz, and W. Al Malwi, "Forecasting urban air quality in Paris using ensemble machine learning: A scalable framework for environmental management," *PLoS One*, vol. 20, no. 11, p. e0336897, Nov. 2025, doi: 10.1371/journal.pone.0336897.

[11] R. Das, A. Acharyya, and S. Majumdar, "Influences of Human Presence on the Indoor Air Quality of Educational Institutions: Concurrent Multipollutant Sensing Approach," *IEEE Open Journal of Instrumentation and Measurement*, vol. 4, pp. 1–8, 2025, doi: 10.1109/OJIM.2025.3583294.

[12] W. Hu, C. Chang, H. Wu, K. Lai, and Y. Cai, "Digital Twin-enabled failure prediction for indoor air quality: A CNN-BiLSTM model with multi-head attention mechanism," *Journal of Building Engineering*, vol. 117, p. 114755, Jan. 2026, doi: 10.1016/j.jobe.2025.114755.

[13] N. Khosh-Amadi, S. Talebi, F. Cheung, and M. Al-Adhami, "Development of a spatially contextualised and AI-enabled digital twin for asthma-specific indoor air quality management," *Journal of Building Engineering*, vol. 118, p. 115027, Jan. 2026, doi: 10.1016/j.jobe.2025.115027.

[14] Y. Qian, J. Leng, K. Zhou, and Y. Liu, "How to measure and control indoor air quality based on intelligent digital twin platforms: A case study in China," *Building and Environment*, vol. 253, p. 111349, Apr. 2024, doi: 10.1016/j.buildenv.2024.111349.

[15] Y. Kim *et al.*, "Positive matrix factorization outperforms machine learning in imputing missing PM2.5 and further identifying spatial patterns in multi-sites without external data," *Urban Climate*, vol. 62, p. 102552, Aug. 2025, doi: 10.1016/j.uclim.2025.102552.

[16] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 134–144, Jan. 2002, doi: 10.1198/073500102753410444.

[17] Z. Zheng, J. D. Castaño Molina, E. Mengiste, S. Prieto, and B. García de Soto, "Digital Twins for Healthier Spaces: A Scalable Framework for Monitoring Indoor Environmental Quality," in *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 2025, pp. 1053–1060. doi: 10.5194/isprs-annals-X-G-2025-1053-2025.

[18] Z. Zheng, J. D. Castaño Molina, and B. García de Soto, "S.M.A.R.T. Forecast." Accessed: Jan. 09, 2026. [Online]. Available: https://digitaltwin.smart-crg.com/spaces/smart_lab/forecast/

Automatic Geometry Extraction from Design Drawings and Design Compliance Checking for Steel Scaffolds

Bing Sun[1], Boyu Wang[2], Chao Yin[1], Borja García de Soto[2] and Jack C. P. Cheng[1]

[1]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[2]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
bsunaq@connect.ust.hk, bw3005@nyu.edu, cyinac@connect.ust.hk, garcia.de.soto@nyu.edu, cejcheng@ust.hk

Abstract

Scaffolding accidents account for significant fatalities in the construction industry, with deviations between erected structures and approved design drawings being a leading cause. However, automated verification of scaffold compliance with design specifications remains an open challenge due to the difficulty of extracting geometric information from engineering drawings and matching them to as-built point cloud data. This paper presents a three-stage framework for automated design compliance checking of steel scaffolds that integrates multi-view drawing analysis with 3D point cloud verification. The first step extracts geometry from scaffold drawings using computer vision algorithms. The second step involves extracting geometry from as-built scaffolds based on point cloud analysis method. The third step verifies the existence of components by comparing the extracted as-built information with the design information. Validation experiments were conducted on a point cloud dataset acquired from a scaffold system in a construction site in Hong Kong. The results provided 100% detection accuracy for tubes and safety-critical components (toe-boards, platforms and stairs), with drawing-based geometric reconstruction accuracy achieving sub-centimeter precision (mean error: 1.94 mm for tube length). This work represents the first automated system for verifying scaffold adherence to design specifications, addressing a critical gap between code compliance and design compliance in construction safety automation.

Keywords –

3D point cloud; Design compliance checking; Drawing analysis; Scaffold; Construction automation; Computer vision

1 Introduction

Scaffolding is a critical yet potentially hazardous temporary structure in construction, often associated with severe accidents due to structural failures. While adherence to safety standards and design specifications during scaffold assembly is crucial for mitigating accident risks, deviations between erected scaffolds and approved safety drawings are frequently observed. These deviations include unauthorized changes to the structural configuration and improper installation of key components. Consequently, the verification of erected scaffolds against approved safety drawings is a routine and essential practice. However, manual design compliance check of scaffolds is slow, inaccurate, qualitative and dangerous. Frequent compliance checking of scaffolds is necessary as scaffolds are often taken down and rebuilt as the project moves forward. To address this challenge, there is a growing focus on developing automatic systems for compliance checking for scaffolds. The advancements in technologies such as computer vision, robot dogs, and unmanned aircraft can assist or replace workers in conducting efficient and safe compliance checks. Adopting automatic compliance checking for scaffolding is essential for advancing engineering management [1].

Recent studies on scaffold compliance checking have concentrated on component detection, including toe-boards [2], guard-rails [2,3], and work platforms [3]. Wang [2] analyzed the geometric attributes of platform-related components, though the scope remained limited. Kim et al. [3] relied exclusively on point cloud quantity to identify elements, overlooking critical geometric dimensions. Conversely, Luo et al. [4] fused imagery with point cloud data, utilizing Graph Neural Networks (GNNs) to evaluate structural metrics such as upright verticality and bottom runner clearance. However, reliance on single-image analysis renders such methods susceptible to occlusion, limiting their application to

single-layer scaffolds. Currently, a comprehensive approach that leverages geometric characteristics to verify all scaffold component types is lacking. Furthermore, few studies address compliance checking against architectural drawings, despite the critical importance of ensuring consistency between the as-built scaffolding and the original design.

To address these limitations, this study aims to develop an automated framework for design compliance checking of scaffolds, specifically focusing on component presence verification. Unlike previous approaches, this research bridges the gap between 2D design documentation and 3D site data. The primary contributions are twofold: 1) a novel drawing analysis method is proposed to automatically extract component positions and geometric attributes from rasterized design drawings; and 2) a robust compliance checking algorithm is developed to verify component existence through geometric feature extraction.

2 Literature Review

Some studies proposed automatic metal scaffolding inspection methods through analyzing point cloud data. Wang [2] proposed an automated technique utilizing 3D point cloud data to extract scaffold platforms and verify the regulatory compliance of toe-boards and guard-rails. Most studies employ rule-based approaches which have limited general applicability, while a small portion uses learning-based methods due to the scarcity of data. Kim, et al. [3] developed an automated scaffold inspection methodology integrating terrestrial laser scanner data acquisition and RandLA-Net-based 3D segmentation. However, the practical implementation of these systems faces two significant hurdles: the scarcity of large-scale, labeled scaffold datasets for model training, and the substantial degradation of prediction performance when processing sparse point cloud data [5]. To address the limitations regarding point cloud scarcity and non-uniformity in large-scale scaffold monitoring, Kim, et al. [5] proposed a reconstruction method that utilizes synthetic datasets for semantic segmentation and an upsampling adversarial network to generate 3D CAD models. Some researchers apply advanced robotic technologies to automate data collection. Chung, et al. [6] developed a fully automated point cloud acquisition system using a quadruped robot that autonomously executes a three-stage workflow of initial site exploration, SLAM-based scan plan generation, and scanning execution without requiring prior information. Due to the height restrictions of robot dogs, the collected point cloud data contains missing and omitted parts. Johansen, et al. [7] employed specialized UAVs to autonomously capture point cloud data from the building exterior during inactive hours using optimized flight paths derived from digital safety plans. Furthermore, there have been few studies on the geometric quality assessment of scaffolding. Therefore, this study proposes a method to extract geometric features from point cloud data.

3 Methodology

The proposed framework consists of three integrated modules (Figure 1): (1) Multi-view Drawing Analysis, which extracts 3D geometric information from 2D raster drawings, (2) Component Detection, which extracts 3D geometric information from point cloud data, (3) Design Compliance Checking, which verifies component existence by comparing as-built component information against extracted design information.

3.1 Multi-view Drawing Analysis

This module extracts 2D information from multiple views and then matches and reasons to obtain 3D coordinates and geometry. Table 1 shows the data extracted from the drawings, and the different elements (tube, toe-board, platform, and stair) are described in the following subsections.

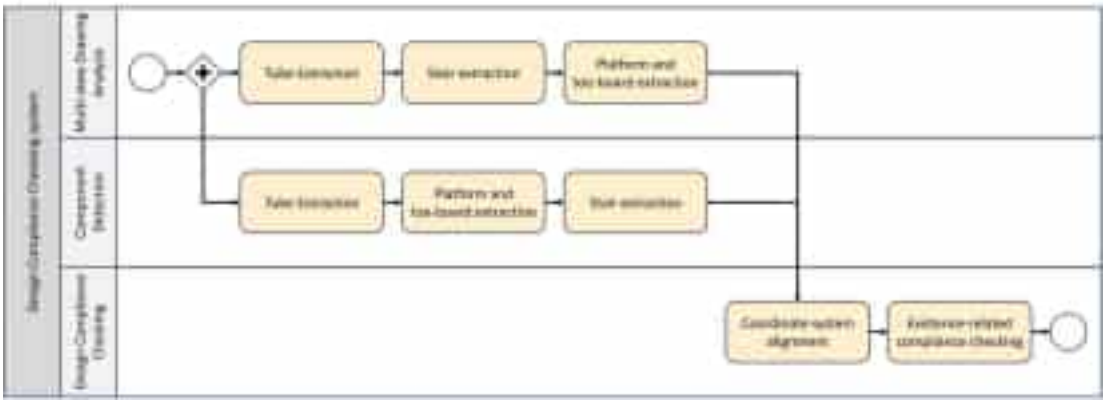


Figure 1. Proposed framework for design compliance checking of scaffolds

3.1.1 Tube Extraction

Scaffolding is defined as a temporary framework structure composed of tubes, with other types of components functioning as attachments to these tubes. Based on chromatic features, various components are extracted from the drawings, encompassing tubes, boards, and stairs. The Canny edge detector and Hough transform are utilized to identify and extract line segments representing the tubes. Given that a single tube in scaffold drawings may be represented by either a rectangle or a line segment, the detected segments are merged. Subsequently, the method of least squares is applied to perform linear fitting and generate definitive line segments. Finally, the starting and ending coordinates of each line segment are extracted to serve as the basis for subsequent analysis. When poor drawing quality compromises line segment extraction accuracy, the lengths and positions of the segments are calibrated by matching numerical annotations to the corresponding segments based on relative positions.

To integrate data extracted from different views, a unified coordinate system must be established. Scaffold drawings typically lack the axis lines and labels previous studies relied on to establish coordinate systems. Optical Character Recognition (OCR) techniques are employed to extract annotations from each view and filter for numerical annotations representing dimensions. These dimensional annotations are then matched with the corresponding tubes based on their coordinates. The scale ratio is calculated by dividing the annotated physical length by the pixel length of the tubes, which subsequently allows for the calculation of the actual tube lengths and 3D coordinates. Since image detection algorithms typically define the top-left corner of the image as the coordinate origin, a coordinate transformation is necessary to map image coordinates to the world coordinate system. When the world coordinate system is projected to form individual views, it is equivalent to retaining only two dimensions from the three-dimensional coordinates. An inverse operation is performed to restore the coordinates from all views to the world coordinate system with one dimension remains indeterminate at this stage. The origin of the world coordinate system is defined by the minimum coordinate values found among the start and end points of all detected tubes.

Existing two-view matching algorithms, utilizing either feature engineering or deep learning, prove ineffective for the three-dimensional reconstruction of scaffolding from engineering drawings. This ineffectiveness stems from the orthogonal projection of such drawings, where a 90-degree angular disparity often causes a structural member in one view to degenerate into a single node in another. To address this geometric ambiguity, the ‘Two-View Matching with Third-View Validation’ technique is employed to recover the missing spatial dimensions of individual tubes. The algorithm matches the endpoints of extracted line segments across two selected views and subsequently validates these candidates against a third view. However, this methodology inherently generates redundant line segments, thereby necessitating manual intervention for their removal.

3.1.2 Stair Extraction

The graphical representation of stair stringers and railings is morphologically similar to that of tubes. The location of the staircase and the quantity of steps are derived based on the position of the stringers. In the frontal view of the staircase, stringers are rendered as paired line segments perpendicular to the ground or as rectangles with their major axes aligned vertically. Conversely, in the lateral view of the staircase, the stringers manifest as inclined line segments or rectangles with tilted major axes due to projection and occlusion effects. The steps are depicted as short straight lines or rectangles that bridge the stringers and are horizontally relative to the ground. The Hough transform was utilized to extract stringers from both front and side views. The orientation of the staircase was determined based on the quantity of these detected stringers, which subsequently determined whether the identified geometric features represented the step length or width. Finally, the precise spatial position and dimensions of the staircase were derived using the coordinate transformation and 3D reconstruction methodologies previously applied to the tubes.

3.1.3 Platform and Toe-board Extraction

Scaffolding platforms are horizontal planar surfaces designed to support loads of personnel, tools and materials, represented in the plan view as shaded, colored or hatched rectangles. By extracting the minimum bounding rectangle from the top view, the actual x and y coordinates of the platforms within the world coordinate system are derived using the same transformation logic as tube detection. These rules stipulate that: 1) platforms must be located at the head and foot of staircases; 2) upper platforms require a minimum of four horizontal tubes for structural support; and (3) these supporting tubes must form a ground-parallel rectangular configuration.

Toe-boards are vertical plates positioned along platform perimeters to prevent slips and falls. Their extraction follows the platform methodology but utilizes front and side views. The final spatial coordinates of the toe-boards are determined by referencing the boundary coordinates of the underlying platforms.

Table 1. Component categories and checking items of the inspected scaffolds

Checking item	Extracted data from drawings				Extracted data from point cloud data			
	Tube	Toe-board	Platform	Stair	Tube	Toe-board	Platform	Stair
Number	√	√	√	√	√	√	√	√
Position	√	√	√	√	√	√	√	√
Orientation	√	√			√	√		
Length	√	√	√		√	√	√	
Width		√	√			√	√	
Thickness						√	√	
Radius					√			
Straightness					√			
Number of steps				√				√

3.2 Component Detection

3.2.1 Tube Extraction

Scaffolding tubes are typically composed of circular steel tubes, which are classified into various types based on their position and function, such as upright tubes, transverse horizontal tubes, and cross bracing. The Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm considering principal direction and cylindricity is employed to perform the initial clustering of the point cloud. The principal direction is initially estimated using Principal Component Analysis (PCA) and subsequently refined through line fitting via the Maximum Likelihood Estimation Sample Consensus (MLESAC) algorithm within the vicinity of the principal axis [8]. Due to high noise levels, non-uniform density, and the presence of numerous irregularly shaped connectors within the scaffolding point cloud, HDBSCAN tends to segment the data into multiple fragments. To address this, these fragments are merged based on collinearity and spatial distance to generate nearly complete point cloud clusters for individual tubes.

Constrained by scanner positioning and occlusions, the resulting scaffolding point clouds often appear as partial cylinders. Thus, the principal direction calculated directly via PCA exhibits significant deviation and renders direct cylinder fitting methods ineffective. Therefore, an iterative slicing approach to determine the centerline is adopted [9]. First, the point cloud is sliced equidistantly along the initial principal direction calculated by PCA. Circles are fitted to each slice using Least Squares to extract centers and diameters. Slices with highly reliable fitted circles within a preset diameter threshold are selected. Then, Random Sample Consensus (RANSAC) is employed to fit a straight line through their centers. This process is repeated iteratively until the angular difference between two consecutive centerlines falls below a specified threshold, at which point the fitted line is defined as the final centerline. The final diameter is determined by averaging the diameters of the valid slices.

3.2.2 Platform and Toe-board Extraction

The positions and geometric dimensions of the platforms and toe boards are determined by extracting planes from the point cloud. For platform detection, local normal vectors are calculated based on the eight nearest neighbors of each data point to filter for approximately horizontal points within the entire point cloud dataset. A Z-value histogram with a bin width of 0.1 m is computed for all data points. Bins positioned above ground level, containing a point count exceeding three times the average of all bins, are identified as potential working platform locations. Consecutive bins are merged, and the M-estimator Sample Consensus (MSAC) algorithm is employed to fit the planes. The extracted platforms are projected onto the xy-plane, and DBSCAN clustering is utilized to separate distinct planes located at the same elevation but spatially disconnected. However, in practical scaffolding scenarios characterized by high noise levels, erroneous planes may be generated from connected surfaces at identical heights, such as stair treads or the upper surfaces of tubes affected by noise. To mitigate this, a grid coverage-based platform segmentation method is proposed. First, the minimum bounding rectangle (MBR) of the extracted plane is computed and discretized into 5 cm rectangular grids. Coverage rate for each row and column is determined based on the point count within the grid. If two consecutive rows or columns fail to meet the coverage threshold, indicating a potential gap, then a split point is established to segment the point cloud. This segmentation process is iterated until all platform fragments meet the coverage criteria. The principal directions of these fragments are fitted using RANSAC to determine their MBRs. After excluding outliers via histogram statistics, the length and width of each platform fragment are calculated. Fragments meeting the defined minimum size thresholds are retained as

extracted platforms.

Finally, toe boards are searched for within a range of 10 cm inward and outward from the platform boundaries. Planar fragments are extracted using a method similar to that used for the platforms. The proposed grid coverage-based segmentation method is also applied to eliminate the interference of guardrails located above the toe boards.

3.2.3 Stair Extraction

Based on the positions of the uprights and the platform, the potential region of the staircase is identified. The point cloud within this area is subsequently segmented and subjected to planar projection. The Probabilistic Hough Transform is utilized to detect short line segments parallel to the ground, while DBSCAN is employed to cluster similar height values to preliminarily determine the number of steps. Because staircase steps typically share uniform dimensional specifications with identical lengths and widths, these dimensions are applied as constraints for filtering during the extraction process. Finally, the position of the staircase is defined by the geometric centroid of the identified steps.

3.3 Design Compliance Checking

3.3.1 Coordinate System Alignment

To enable subsequent position-based component matching for compliance checking, the drawing coordinate system is aligned with the scanner coordinate system using the coordinates of the uprights. The upright detection pipeline begins by extracting horizontal slices and segmenting points via region growing to fit circles based on radius and root-mean-squared error constraints [2]. The spatial distribution of the resulting centers is converted into a binary density map, facilitating the construction of a convex hull that identifies the precise centroid coordinates of the uprights [2]. To determine the alignment, two upright centroids from the drawing analysis and two from the point cloud are selected. Assuming a correspondence between these pairs, the Singular Value Decomposition (SVD) method is employed to solve for the rigid body transformation matrix. Transformed point cloud upright centroids that fall within a specified distance threshold of the drawing upright centroids are classified as inliers. An Exhaustive Search is performed to iterate through all possible combinations. The transformation matrix maximizing the inlier count is selected as the optimal solution. If several solutions have equal inliers, the one with the matrix with the lowest Root Mean Square Error (RMSE) is selected. Finally, this optimal transformation matrix is applied to the raw 3D point cloud to achieve global alignment.

3.3.2 Existence-related Compliance Checking

Component presence is validated by contrasting the position and geometric attributes derived from the drawings against those from the point cloud. Platforms and toe-boards are matched via centroid distance, with subsequent assessments of their length, width, and orientation. Since the uprights were already filtered by diameter during the extraction process, they are matched based on the orientation and position of their centerlines. The presence of the component is confirmed by a successful match. Finally, stairs are validated by directly comparing their position and the number of steps. The component categories and checking items of the inspected scaffolds are presented in Table 1.

4 Case study

4.1 Data collection

A scaffold system located at a construction site in Hong Kong was selected for this case study. The scaffold system consisted of 42 tubes, 4 toe-boards, 1 platform and 1 staircase with 8 steps. The design drawings for the scaffold system are stored as color raster images, as illustrated in Figure 2. Point cloud data of a scaffold system was acquired using a high-resolution SHARE SLAM S20 handheld 3D LiDAR scanner. During data acquisition, the scanner was maneuvered slowly around the scaffold, continuously emitting laser pulses and recording the reflected signals. The built-in SLAM algorithm within the SHARE PointCloud Studio software was then utilized to generate a high-density point cloud, which was exported in plain text format for subsequent analysis. Finally, manual pre-processing was performed using CloudCompare to remove extraneous environmental data. The resulting pre-processed point cloud is illustrated in Figure 3.

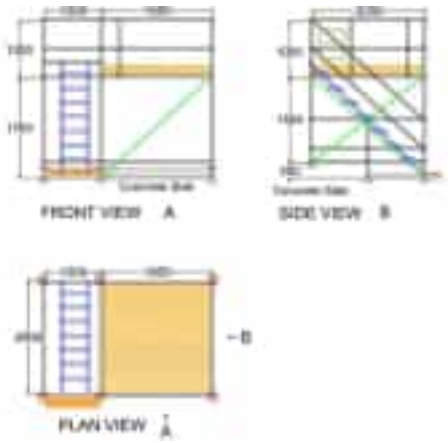


Figure 2. Scaffolding design drawing

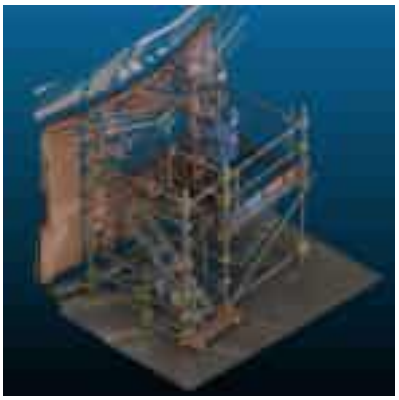


Figure 3. Collected and preprocessed scaffolding point cloud

4.2 Results

4.2.1 Geometry Extraction from Design Drawings

In the drawing analysis, the discrepancies between the measured values and the ground truth values for scaffold tubes were minimal. Figure 4 shows the spatial distribution of errors between the extracted scaffold tube measurements and the ground truth in three dimensions. Each tube is visualized in its 3D location, and its color indicates the magnitude of error according to the corresponding color bar (green = low discrepancy, red = high discrepancy). The position difference heatmap (left) highlights where centerline or endpoint localization deviates from the reference, the length difference heatmap (middle) shows segment-wise deviations in tube length, and the angle difference heatmap (right) summarizes orientation errors. For all three metrics, most tubes appear in green tones, indicating small discrepancies throughout the structure. Using a scale of 4 mm per pixel, the average discrepancies were 8.334 mm for position, 1.935 mm for length, and 0 for angle, confirming minimal deviation from ground truth and high geometric consistency in the drawing-based scaffold analysis. Overall, discrepancies remain low across the scaffold geometry, indicating close agreement between the measured results and the ground truth. However, the extraction errors for platforms and toe-boards were slightly higher (Table 2). Position errors ranged from 8.77 mm (platform) to 49.56 mm (toe-board 3), meaning toe-boards were harder to localize accurately than the platform. Length errors were similar for all components (−72 to −74 mm). Width errors were small and consistent for toe-boards (−8.00 mm), but much larger for the platform (−74.00 mm). The angle error was 0° for every component, showing that orientation was correctly preserved. These differences mainly come from the use of color segmentation, where filled regions start at the tube edges, while the drawing dimensions are referenced to the tube centerlines. Therefore, the deviations are systematic and acceptable for scaffold modeling.

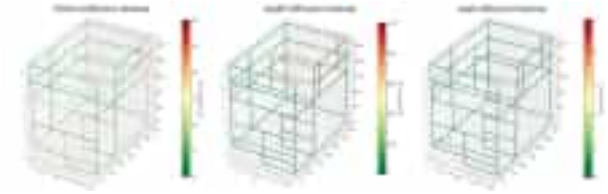


Figure 4. 3D discrepancy heatmaps comparing the measured scaffold tube parameters against ground truth

Table 2. Discrepancies between measured and ground truth values for scaffold toe-board and platform components

Component	Position difference (mm)	Length difference (mm)	Width difference (mm)	Angle difference (deg.)
Toe-board 1	41.42	-72.00	-8.00	0
Toe-board 2	47.97	-74.00	-8.00	0
Toe-board 3	49.56	-72.00	-8.00	0
Toe-board 4	47.97	-74.00	-8.00	0
Platform	8.77	-72.00	-74.00	0

4.2.2 Design Compliance Checking

Table 3 presents the results of the existence-related compliance checking for the case study. A total of 42 tubes, 1 staircase, 1 platform and 4 toe-boards were successfully detected. Compared with the design drawings, an additional stair step was constructed in the as-built scaffold.

Table 3. Results of existence-related design compliance checking of the case study

	Quantity in drawing	Quantity in point cloud data	Detected quantity
Tube	42	42	42
Platform	1	1	1
Toe-board	4	4	4
Stair	1	1	1
Steps of stair	7	8	8

5 Discussion

During tube detection, due to total or partial occlusion and the co-linear superposition of boundaries, segment endpoints or the vertices of overlapping rectangles become indistinguishable, rendering actual geometric attributes indeterminate. This inherent metric ambiguity affects both human and computational systems, necessitating expert knowledge or top-down semantic

constraints for its resolution. As a result, certain short line segments situated within the interior of the scaffold structure may be erroneously extracted as longer segments, leading to the generation of redundant tubes during the 3D reconstruction process. In this study, we manually remove redundant tubes. Future research can incorporate stronger structural and semantic priors into the tube reconstruction pipeline, including scaffold connectivity rules, topological constraints, and the functional roles of the components.

A limitation of this study is that the method was validated on only one small dataset containing 42 tubes. Although the overall framework is not inherently constrained by scaffold size, scaling to larger and more complex scaffolds is expected to introduce additional challenges, including heavier occlusion, denser tube arrangements, repeated structural patterns, and error propagation from local line extraction to global 3D reconstruction. These factors may increase ambiguity in tube interpretation and lead to more redundant tube instances. In future work, we will evaluate the method on larger and more diverse datasets.

In addition, the quality of the Point Cloud Data (PCD) requires further improvement. Primarily, constraints related to scanner elevation and layout, along with the inherent complexity of the scaffolding structure, lead to non-negligible incompleteness in the captured point clouds. Specifically, the numerous support tubes beneath the platforms often generate excessive noise in the undersurface data. Despite deploying handheld mobile scanners carried up the stairs manually, the point clouds of the platform's upper surface remain fragmented due to occlusions caused by other equipment located on the platform. While the proposed grid coverage-based platform segmentation method mitigates the effects of point cloud sparsity and uneven distribution to a certain extent, the accuracy of platform dimension extraction remains compromised. Moreover, regarding scaffolding positioned adjacent to walls, spatial inaccessibility results in blind spots on the back of the structure. Similarly, lower-level components often obstruct the line of sight to higher scaffolding tiers. To address these limitations, future work should employ more flexible devices, such as Unmanned Aerial Vehicles (UAVs), to scan the target range from various angles and heights. Furthermore, developing appropriate point cloud completion methods to increase data density and completeness without altering the geometric characteristics of the components serves as a potential solution.

6 Conclusion

This paper proposed a framework for automated design compliance checking of steel scaffolds, addressing a critical gap in construction safety automation. By integrating multi-view drawing analysis with 3D point cloud verification, the proposed method enables quantitative comparison between as-built conditions and approved design specifications. Specifically, this study has two main contributions: 1) proposes a multi-view drawing analysis method that can effectively extract information and enables 3D reconstruction, and 2) enables effective scaffold design compliance checking. While current results are promising, limitations include occlusion-related false negatives, assumption of approximate positional correctness, and single case study validation. Future work should integrate multi-modal data sources (photogrammetry, BIM) and validate across diverse scaffold configurations.

Acknowledgement

The authors would like to thank the RGC General Research Fund (16212425) for supporting this research.

References

[1] Xiaochun Luo, Mingyong Qin, Zeyu Gao, Hui Yan, and Xincong Yang. Ground abstract structure concepts of scaffolding systems for automatic compliance checking based on reasoning segmentation. Expert Systems with Applications, 270:126563, 2025.

[2] Qian Wang. Automatic checks from 3D point cloud data for safety regulation compliance for scaffold work platforms. Automation in Construction, 104:38–51, 2019.

[3] Jeehoon Kim, Juhyeon Kim, Nahye Koo, and Hyoungkwan Kim. Automating scaffold safety inspections using semantic analysis of 3d point clouds. Automation in Construction, 166:105603, 2024.

[4] Xiaochun Luo, Yutong Tang, Yongqi Wei, Chengqian Li, and Qi Fang. Gnn-based spatial relationship modeling for automated scaffold component function recognition and intelligent compliance checking. Automation in Construction, 180:106557, 2025.

[5] Juhyeon Kim, Jeehoon Kim, Yohan Kim, and Hyoungkwan Kim. 3D reconstruction of large-scale scaffolds with synthetic data generation and an upsampling adversarial network. Automation in Construction, 156:105108, 2023.

[6] Duho Chung, Juhyeon Kim, Sunwoong Paik, Seunghun Im, and Hyoungkwan Kim. Automated system of scaffold point cloud data acquisition using a robot dog. Automation in Construction, 170:105944, 2025.

[7] K.W. Johansen, J. Teizer, and C. Schultz. Automated rule-based safety inspection and compliance checking of temporary guardrail systems in construction. Automation in Construction, 168:105849, 2024.

[8] Yusheng Xu, Sebastian Tuttas, Ludwig Hoegner, and Uwe Stilla. Reconstruction of scaffolds from a photogrammetric point cloud of construction sites using a novel 3D local feature descriptor. Automation in Construction, 85:76–95, 2018. ISSN 0926-5805.

[9] Liu Yang, Jack C.P. Cheng, and Qian Wang. Semi-automated generation of parametric BIM for steel structures based on terrestrial laser scanning data. Automation in Construction, 112:103037, 2020. ISSN 0926-5805.

Drone-Based Façade Defect Detection via Decision-Level Fusion of Thermal Infrared and Visible Images

Jiaqi Li[1], Qingrui Yue[1,2], Zezhi Ding[1], Haofeng Yan[3], Nan Jin[2], Xincong Yang[1,*]

[1]School of Intelligent Civil and Ocean Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen, 518055, China
[2]Safe Urban Development Institute of Science and Technology (Shenzhen), Shenzhen, 518024, China
[3]School of Civil Engineering, Harbin Institute of Technology, Harbin, 150090, China

Abstract -

Urban façade deterioration, especially tile delamination and related surface damage, poses significant safety risks in dense urban environments. Conventional inspections based on manual visual assessment and acoustic sounding are labor-intensive, hazardous, and subjective. This paper presents an autonomous UAV-based façade inspection framework that combines oblique-photogrammetry-driven path planning with decision-level fusion of infrared and visible imagery for defect identification. A 3D model of the target building is reconstructed to extract geometric features for generating safe flight trajectories at a fixed standoff distance of 10 m, supporting near-complete and spatially consistent dual-modal data acquisition under the planned inspection route. Modality-specific defects are then segmented from RGB and thermal data using deep-learning models, and the fused defect map is produced through decision-level fusion of surface appearance cues and subsurface thermal signatures. A field deployment on a building demonstrates the feasibility and efficiency of the framework, completing the inspection of approximately 4,200 m² of façade area using 208 waypoints within 50 minutes of total on-site operation. The results show that the proposed method provides a practical and safer solution for façade condition assessment and maintenance prioritization.

Keywords -

Façade inspection, Drone path planning, Infrared thermography, Decision-level fusion, Subsurface delamination

1 Introduction

Building façades function as the first line of defense for infrastructure, enduring continuous environmental stress throughout their lifecycle. Synergistic effects of material deterioration, construction imperfections, and severe weather can compromise façade integrity, serviceability, and public safety. In China, statistical analysis of façade failures reveals that tile detachment is a predominant hazard, frequently resulting in casualties and economic losses [1]. The risk is amplified in dense urban environments where buildings directly face crowded streets. Consequently, implementing regular and rigorous inspection protocols for aging tiled façades is imperative for public safety maintenance.

Conventionally, façade health monitoring relies on manual visual inspection and acoustic methods. Ground-based visual surveys, often aided by optical telescopes or handheld thermography, attempt to identify external anomalies. However, these methods suffer from low efficiency and are prone to human error due to the inspectors' subjective judgment [2]. Alternatively, hammer sounding—which detects subsurface delaminations via acoustic response—is widely practiced but inherently inefficient. It requires direct physical access via scaffolds or gondolas, exposing personnel to significant safety hazards associated with working at heights.

Recent strides in unmanned aerial vehicle (UAV) technology have significantly improved infrastructure monitoring [3,4]. Equipped with high-resolution sensors, UAVs offer a promising alternative for accessing hard-to-reach vertical surfaces. Despite this potential, a robust methodology for reliably detecting subsurface delaminations through automated means remains underdeveloped. Addressing this gap, this study proposes a novel drone-based inspection framework that integrates infrared thermography and RGB imagery to enhance the precision and efficiency of façade defect detection.

To address these challenges, this paper presents an autonomous inspection framework that integrates

adaptive flight path planning with decision-level fusion of thermal infrared and visible images for façade assessment. The proposed method first leverages oblique photogrammetry to reconstruct a 3D model of the target building, which supports the generation of safe and feasible flight trajectories for data acquisition. Subsequently, synchronized visible light images and thermal infrared data are analyzed using a deep learning-based segmentation pipeline. A decision-level fusion strategy is then applied to combine surface appearance cues with subsurface thermal anomalies, thereby reducing false detections. The feasibility of the framework is validated through a field deployment at a large commercial building, demonstrating efficient on-site operation and reliable defect detection compared with conventional inspection practices.

2 Methodology

2.1 Overview of the Proposed Framework

Figure 1 presents the schematic workflow of the proposed drone-based inspection system, which integrates autonomous path planning with dual-modal defect recognition. The methodology follows a coarse-to-fine strategy. Initially, oblique aerial imagery is acquired to construct a photogrammetric 3D model of the building. This geometric model serves as the environmental baseline for planning a flight path tailored for close-range inspection. During the inspection phase, the drone autonomously captures paired thermal infrared (IR) and visible light (RGB) imagery along the planned route. These images are processed by independent deep learning models to detect surface anomalies (e.g., peeling, cracks) and subsurface delamination. To maximize detection reliability, a decision-level fusion strategy is applied, synthesizing the outputs from both sensors into a unified and more reliable façade health assessment.

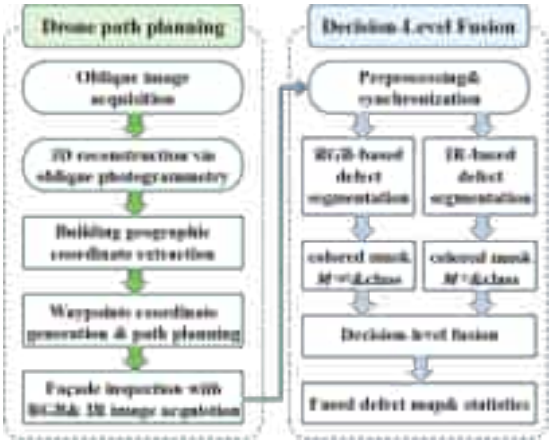


Figure 1. The overall framework of the proposed drone-based façade inspection system

2.2 Drone Path Planning Method

Drones equipped with flight control systems can execute complex, predefined trajectories [5]. However, manual path planning for complex building façades is inefficient and prone to collision risks. This study proposes an automated method for generating optimized inspection routes based on the extraction of architectural envelope key points from a preliminary 3D model.

2.2.1 Oblique 3D Modeling and Geographic Feature Extraction

The prerequisite for precise path planning is an accurate representation of the building geometry. Initially, a UAV imaging system captures the target building from multiple perspectives. This multi-view acquisition strategy reduces visual occlusions and provides sufficient image coverage for high-quality photogrammetric reconstruction.

Subsequently, the collected imagery is processed through a photogrammetric workflow, including image distortion correction, feature matching, bundle adjustment, dense point cloud generation, and mesh reconstruction. This process, often implemented using Multi-View Stereo algorithms in software such as Agisoft Metashape or ContextCapture [6,7], produces a high-precision oblique photogrammetric 3D model of the target building, as illustrated in Figure 2.

Based on the georeferenced 3D model, the geographic coordinates, including longitude, latitude, and altitude, of the building’s key corner points are extracted. These coordinates provide the geometric constraints for generating the subsequent inspection waypoints.


Figure 2. Process of oblique photogrammetry 3D reconstruction and key point extraction

2.2.2 Waypoint Generation Mechanism

To ensure complete coverage of the façade, the

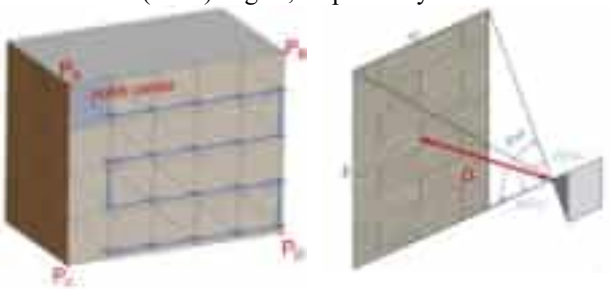
continuous wall surface must be discretized into a set of discrete inspection waypoints.

First, the physical dimensions of the single-image footprint are determined. Let the four corner vertices of the target façade be denoted as $\mathrm{P_A}(\lambda_A,\phi_A,h_A),\mathrm{P_B}(\lambda_B,\phi_B,h_B),\mathrm{P_C}(\lambda_C,\phi_C,h_C),\mathrm{P_D}(\lambda_D,\phi_D,h_D)$ (Figure 3(a)). Modeling the camera as a pinhole system, the physical coverage dimensions, namely the width w and height h, at a given inspection distance D can be determined from the camera's horizontal and vertical fields of view, respectively, as illustrated in Figure 3(b):

$$w = 2D \cdot \tan\left(\frac{\theta_h}{2}\right) \tag{1}$$

$$h = 2D \cdot \tan\left(\frac{\theta_v}{2}\right) \tag{2}$$

where θ_h and θ_v represent the horizontal and vertical field of view (FOV) angles, respectively.



(a) Flight trajectory positioning (b) Field of View (FOV) calculation model

Figure 3. Geometric parameters for path planning

Since the extracted building key points utilize the Geographic Coordinate System (GCS, i.e., WGS84), whereas the geometric subdivision necessitates metric computation, a coordinate transformation is required. The geographic coordinates (λ,ϕ,h) are converted to the Earth-Centered, Earth-Fixed (ECEF) Cartesian coordinates (X,Y,Z) using the following standard transformation equations:

$$X = (N+h)\cos(\phi)\cos(\lambda) \tag{3}$$

$$Y = (N+h)\cos(\phi)\sin(\lambda) \tag{4}$$

$$Z = \left[\left(1-e^2\right)N+h\right]\sin(\phi) \tag{5}$$

with the radius of curvature in the prime vertical N defined as:

$$N = \frac{a}{\sqrt{1-e^2\sin^2(\phi)}} \tag{6}$$

where a denotes the Earth's semi-major axis ($\approx 6,378,137\mathrm{m}$) and e represents the Earth's eccentricity (≈ 0.081819).

In the ECEF frame, the Euclidean distances between the corner points (e.g., $\mathrm{P_A}$ to $\mathrm{P_B}$ and $\mathrm{P_C}$) are calculated to determine the total width and height of the façade:

$$d_{AB} = \sqrt{\left(X_B-X_A\right)^2+\left(Y_B-Y_A\right)^2+\left(Z_B-Z_A\right)^2} \tag{7}$$

$$d_{AC} = \sqrt{\left(X_C-X_A\right)^2+\left(Y_C-Y_A\right)^2+\left(Z_C-Z_A\right)^2} \tag{8}$$

Based on the single-image coverage (w, h) and the total façade dimensions, the façade is discretized into an $m \times n$ grid of inspection areas. To ensure sufficient image overlap between adjacent viewpoints, the horizontal and vertical step sizes are defined as $w(1-O_w)$ and $h(1-O_h)$, respectively, where O_w and O_h denote the prescribed horizontal and vertical overlap ratios. Accordingly, the required numbers of inspection columns and rows are calculated as:

$$m = 1+\left\lceil \frac{d_{AB}-w}{w(1-O_w)} \right\rceil \tag{9}$$

$$n = 1+\left\lceil \frac{d_{AC}-h}{h(1-O_h)} \right\rceil \tag{10}$$

The center point $(X_{\text{center}},Y_{\text{center}},Z_{\text{center}})$ of each grid cell is then calculated and translated outward along the normal vector $\vec{N}=(N_x,N_y,N_z)$ of the plane $\mathrm{P_AP_BP_C}$ by the standoff distance D. This operation yields the precise 3D coordinates of each camera waypoint in the Cartesian frame:

$$X_{\text{shifted}} = X_{\text{center}} + D \cdot N_x \tag{11}$$

$$Y_{\text{shifted}} = Y_{\text{center}} + D \cdot N_y \tag{12}$$

$$Z_{\text{shifted}} = Z_{\text{center}} + D \cdot N_z \tag{13}$$

Finally, an inverse coordinate transformation is applied to convert these Cartesian waypoints back to the Geographic Coordinate System, allowing them to be uploaded to the drone's flight controller. The longitude λ is calculated directly, while the latitude ϕ and height h are derived using Bowring's closed-form approximation [8]:

$$\lambda = \arctan\left(\frac{Y}{X}\right) \tag{14}$$

$$\phi = \arctan\left(\frac{Z+e'^2\cdot b\cdot\sin^3(\theta)}{p-e^2\cdot a\cdot\cos^3(\theta)}\right) \tag{15}$$

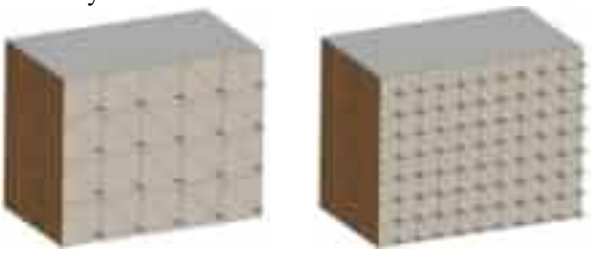
$$h = \frac{p}{\cos(\phi)} - N \tag{16}$$

To facilitate this calculation, the auxiliary parameters are defined as follows:

$$e'^2 = e^2 / (1-e^2),\ p = \sqrt{X^2 + Y^2},\ b = a\sqrt{1-e^2} \quad (17)$$

$$\theta = \arctan(\frac{Z \cdot a}{p \cdot b}) \quad (18)$$

The standoff distance D is a critical trade-off parameter. As illustrated in Figure 4, a smaller D yields smaller (higher spatial resolution) Ground Sampling Distance (GSD) and richer detail but limits the single-image field of view. This necessitates a denser grid of waypoints, resulting in longer flight times. Conversely, a larger D improves efficiency but may compromise the detection of fine cracks. Accordingly, D=10 m was selected for the field deployment as a practical compromise between image resolution and inspection efficiency.



(a) Sparse sampling at a larger distance (b)Dense sampling at a closer distance

Figure 4. Impact of standoff distance on waypoint density

2.2.3 Executable Flight Path Generation

Once the discrete waypoints are generated, the final step is to generate an executable inspection trajectory by determining an appropriate traversal sequence. The objective is to visit all waypoints while minimizing total flight time and energy consumption, subject to practical constraints such as battery endurance and a predefined safety standoff from the façade.

This problem can be addressed using coverage patterns and, when needed, graph-based routing. Geometric coverage strategies such as the Boustrophedon (back-and-forth) or spiral patterns provide simple and consistent coverage with controllable overlap [9]. When obstacle-aware local routing is required, shortest-path solvers (e.g., Dijkstra or A*) can be used to connect adjacent waypoint segments under additional constraints. Figure 5 compares the trajectories produced by different strategies. For façade inspection, the Boustrophedon pattern is adopted in this study due to its implementation simplicity and stable overlap management.

Finally, the optimized trajectory is converted into a DJI Waypoint Mission by exporting the ordered waypoint list together with mission parameters (e.g., altitude, speed, heading/yaw, gimbal angle, and camera triggering). This enables the planned route to be directly uploaded to the DJI flight controller for autonomous execution.

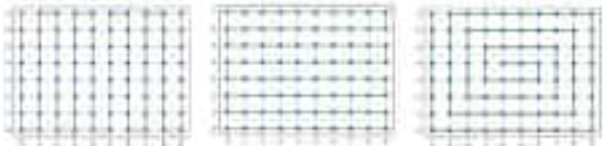


Figure 5. Flight paths generated by different methods

2.3 Façade Defect Detection and Fusion

This section details the proposed dual-modal defect detection framework, which employs deep learning models to process RGB and thermal imagery independently, followed by a decision-level fusion strategy to synthesize the results (as shown in Figure 6).

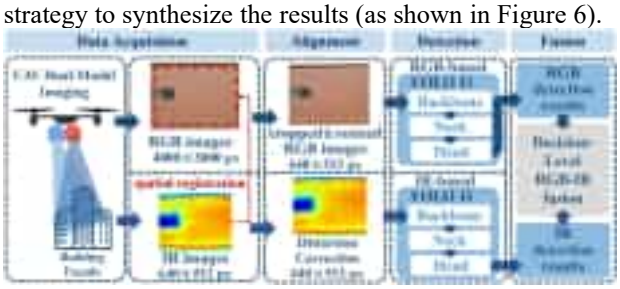


Figure 6. The proposed dual-modal defect detection framework

2.3.1 Surface Anomaly Segmentation via RGB Imagery

RGB imagery is effective in capturing high-frequency texture information and surface chromatic features. However, latent subsurface delaminations typically lack distinct visual characteristics in the early stages. They become observable in the visible spectrum only when the defect evolves into external manifestations, such as peeling or cracking. Consequently, RGB data is primarily utilized in this study to detect surface-level damage (e.g., tile cracking and peeling) associated with advanced deterioration.

To automate defect extraction, YOLO11-seg was adopted from the Ultralytics YOLO family [10,11] because it provides a strong balance between segmentation accuracy and inference efficiency. In this study, it was used as the modality-specific segmentation backbone for both the RGB and thermal branches. As a one-stage segmentation model, YOLO11-seg performs feature extraction, multi-scale feature aggregation, and pixel-level mask prediction in an end-to-end manner, enabling robust defect segmentation across different scales.

2.3.2 Subsurface Delamination Detection via Infrared Thermography

Infrared thermography is employed to reveal non-

visible subsurface defects based on differences in thermal inertia and heat-transfer behavior. In delaminated areas, an air gap between the tile and the substrate increases thermal resistance and disrupts heat conduction. Under solar excitation or during the cooling phase, the detached façade surface exhibits a different heating/cooling rate compared with sound areas, producing a measurable temperature contrast ΔT on the surface that can be captured by the thermal sensor [12].

In this study, the thermal images were converted into pseudo-color representations and used as inputs to a dedicated model. The model was fine-tuned on the infrared dataset to segment thermally anomalous regions associated with subsurface delamination. These thermal anomalies, typically manifested as localized hot or cold regions, are often not visible in RGB imagery.

2.3.3 Decision-Level Fusion of Detection Results

To obtain a holistic and physically interpretable assessment of façade condition, we adopt a decision-level fusion strategy [13] that operates on the independent segmentation outputs of the RGB and IR branches. Unlike pixel-level or feature-level fusion, the proposed strategy integrates detection results after modality-specific inference. This ensures that the distinct physical semantics of each sensor are preserved: visible imagery captures superficial appearance anomalies (M_{surf}), such as cracks and peeling, whereas infrared thermography highlights subsurface structural defects (M_{sub}) through thermal variations.

A key objective of the fusion is to identify compound defects, i.e., surface damage spatially correlated with subsurface delamination, since such regions often indicate elevated structural risk. To account for minor registration errors and boundary uncertainties between the dual-modal images, a morphological dilation is first applied to the subsurface defect mask:

$$\widetilde{M}_{sub} = \text{Dilation}(M_{sub}, r) \tag{19}$$

where r denotes the radius of the structuring element and serves as a small spatial tolerance for overlap matching, rather than a substantial geometric expansion of the defect region. In the present dataset, the residual RGB-IR misalignment after registration was empirically found to remain within approximately 5 infrared pixels. Accordingly, r was selected as a small value within this tolerance range, so that true cross-modal overlap would not be missed due to slight spatial deviation.

Based on the spatial relationship between the surface mask M_{surf} and the dilated subsurface mask $\widetilde{M}_{sub}$, the final fused result map M_{final} is constructed. For any given pixel p on the façade, the defect category is determined according to the following hierarchical logic:

$$M_{final}(p) = \begin{cases} \text{Compound Defect}, & \text{if } p \in (M_{surf} \cap \widetilde{M}_{sub}) \\ \text{Surface Defect}, & \text{if } p \in (M_{surf} \setminus \widetilde{M}_{sub}) \\ \text{Subsurface Defect}, & \text{if } p \in (M_{sub} \setminus M_{surf}) \\ \text{Background}, & \text{otherwise} \end{cases} \tag{20}$$

This hierarchical rule prioritizes compound defects by assigning the label Compound Defect to regions where visible surface anomalies overlap with subsurface delamination, while regions detected by only one modality are retained as Surface Defect or Subsurface Defect. In this way, the fusion result links superficial manifestations to potential internal risks and provides a more interpretable representation of façade condition. From an engineering perspective, cracks or peeling that spatially coincide with thermal anomalies may indicate an elevated likelihood of underlying delamination. In addition, because the RGB and IR models are trained independently, the proposed decision-level fusion decouples optimization across modalities, which simplifies implementation and facilitates the integration of heterogeneous sensing results.

3 Field Experiments and Analysis

3.1 Experimental Setup and Data Acquisition

To validate the feasibility and robustness of the proposed framework, a field experiment was conducted on a commercial building. The DJI Mavic 3T drone was employed as the flight platform. This UAV is equipped with a synchronized wide-angle RGB camera and a thermal infrared camera, enabling the simultaneous acquisition of dual-modal imagery.

The flight mission was executed following the path planning method described in Section 2.2. Figure 7 illustrates the planned trajectory. The captured RGB-IR image pairs are presented in Figure 8. It can be observed that the images collected autonomously along the planned route exhibit high spatial consistency and sufficient overlap, which establishes a high-quality dataset for subsequent algorithmic analysis.

The field deployment was conducted in November under clear-sky conditions, with stable daylight and relatively low outdoor wind speed, which were favorable for both RGB and thermal data acquisition. The thermal campaign was carried out in two rounds corresponding to the façade heating and cooling stages on November 29, 2023, in Chengdu, China, for a façade oriented about 40° south of east. At around 10:30 a.m., the solar azimuth was close to the façade normal direction, providing favorable direct solar excitation. At that time, the incident solar irradiance was

approximately 750 W/m², the ambient temperature was about 14°C, and the convective heat-transfer coefficient was estimated to be 8-10 W/(m²·K). By contrast, at around 5:00 p.m., the façade had entered its self-shadowed condition and direct solar heating was no longer present. The ambient temperature was about 15°C, and the convective heat-transfer coefficient was approximately 10-12 W/(m²·K). As a result, the acquired thermal infrared images covered both the heating and cooling phases of the façade, during which delamination-related thermal signatures were relatively distinct. These conditions provided thermal images with favorable signal-to-noise characteristics for subsequent model training and defect recognition. It should be noted, however, that the quality of defect-related visual and thermal cues may vary under different environmental conditions: RGB imagery is more sensitive to illumination changes, whereas thermal infrared imagery is more strongly influenced by solar loading, wind, and the thermal properties of façade materials.



Figure 7. Visualization of the executed flight path in the 3D model

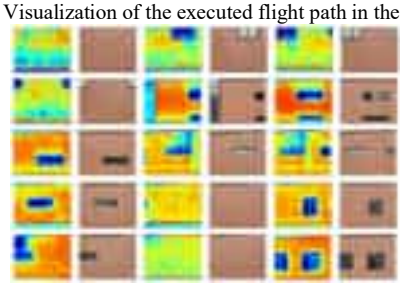


Figure 8. Samples of synchronized RGB and IR image pairs collected during the flight

Since the decision-level fusion in this study depends on the spatial correspondence between visible and thermal defect regions, an explicit cross-modal alignment procedure was applied before defect analysis. The visible and thermal cameras mounted on the UAV gimbal have a fixed relative pose, which provides the geometric basis for stable registration. Both sensors were first calibrated using a calibration board, and all images were corrected for lens distortion prior to alignment. Because the UAV acquired façade images at a fixed standoff distance along the planned route and the imaging parameters were kept unchanged throughout the mission, the relative projection relationship between the two sensors remained approximately constant across the dataset. Therefore, after distortion correction, a single affine transformation matrix was adopted for all RGB-IR image pairs.

The affine transformation parameters were estimated from five representative RGB-IR image pairs with clearly identifiable shared features, and the final matrix was obtained by averaging the estimated parameters. To evaluate the registration quality, 20 additional image pairs were randomly selected for verification. The results showed that the alignment achieved an average registration error of approximately 2 pixels, with all observed errors remaining within 5 pixels. In addition, because the RGB and thermal sensors have different native resolutions, the visible images were resampled to the thermal image size (639×509 pixels) before registration and model inference. This resolution unification improved the consistency of cross-modal spatial matching and also simplified the subsequent defect-detection pipeline.

3.2 Model Implementation and Training Details

The dataset combined the current field campaign with previously accumulated inspection data from the same façade. The waypoint route and two-round acquisition process contributed 416 paired images, yielding a final raw dataset of 766 RGB and 766 original IR images. For the thermal branch, each original infrared image was converted into pseudo-color images using three different palettes. To avoid data leakage, the train/validation/test split was performed at the level of the original RGB-IR pairs before pseudo-color expansion, using a ratio of 7:2:1. The RGB branch focused on visible surface defects, whereas the thermal branch focused on thermally anomalous regions associated with subsurface delamination.

The experimental workflow consisted of data annotation, model training, and performance evaluation. In the data annotation step, the collected images were labeled using LabelMe. Annotation masks were created by experienced civil engineers, who manually annotated tile cracks and peeling in RGB images, and thermally anomalous regions associated with subsurface delamination in infrared images.

For modality-specific defect extraction, two dedicated segmentation models were fine-tuned using the Ultralytics YOLO11 framework, one for the RGB branch and the other for the thermal branch. Training was conducted with an input resolution of 640×640 pixels, a batch size of 16, the AdamW optimizer, and early stopping with a patience of 100 epochs. To improve robustness, data augmentation including color perturbation, rotation, translation, scaling, and flipping was applied to the training set. The best-performing checkpoint was automatically selected based on validation performance. The evolution of the validation

metrics during training is shown in Figure 9.

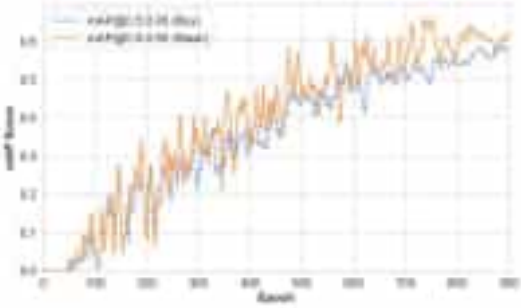


Figure 9. Evolution of validation box and mask mAP@0.5:0.95 over training epochs

3.3 Quantitative and Qualitative Analysis

Quantitative evaluation on the test set demonstrates the robust segmentation capability of the trained YOLO11n-seg model. For the segmentation masks, the overall precision and recall reached 0.795 and 0.869, respectively, with an mAP@0.5 of 0.829 and an mAP@0.5:0.95 of 0.619. Class-wise results indicate that the framework maintained effective performance across subsurface delamination, peeling, and crack categories, although subsurface delamination and crack remained more challenging than peeling. These results confirm that the framework can effectively segment both surface-visible defects and thermally expressed subsurface anomalies. The adopted YOLO11n-seg backbone also showed favorable inference efficiency, with an average per-image processing time of 0.8ms for preprocessing, 0.5ms for inference, and 1.2ms for postprocessing, supporting rapid batch processing and timely on-site review after UAV data acquisition.

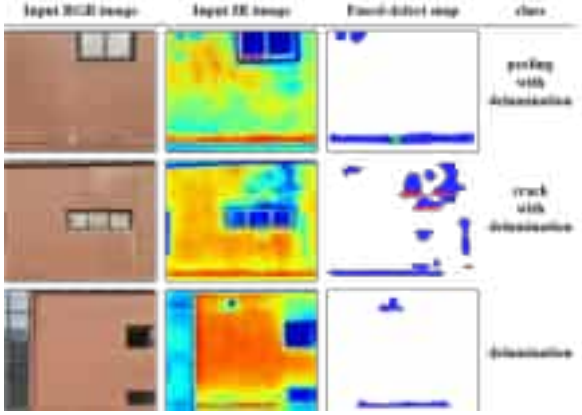


Figure 10. Visualization of defect segmentation results

Qualitative identification results are presented in Figure 10. The proposed framework effectively characterizes damage types in both modalities, identifying tile cracks and peeling in RGB imagery while simultaneously detecting thermally anomalous regions associated with subsurface delamination in infrared imagery. Following the decision-level fusion strategy, the modality-specific masks are combined to generate a comprehensive façade condition map.

In addition to qualitative visualization, the practical value of the decision-level fusion was assessed at the façade level. RGB-only results identified visible surface defects such as cracks and peeling, whereas IR-only results identified thermally anomalous regions associated with subsurface delamination. Based on the spatial-overlap rules, the fusion process further identified 18 compound-defect regions, including crack-with-delamination and peeling-with-delamination cases, which could not be explicitly characterized by either modality alone and therefore provide more actionable information for maintenance prioritization.

A brief sensitivity check, performed by varying the dilation radius from 2 to 8 pixels, showed that the compound-defect results remained largely stable, indicating limited sensitivity to this tolerance parameter under the present registration accuracy.

Field verification was conducted according to defect visibility and accessibility. Visible defects such as cracks and peeling were manually reviewed, accessible low-rise suspected delamination areas were examined by hammer sounding, and higher or less accessible regions were checked indirectly using a telephoto camera, particularly where bulging or outward deformation was visible. The majority of detected defects were practically confirmed, although some inaccessible and visually flat regions could not be fully verified. Therefore, the current validation should be regarded as a practical engineering verification rather than a fully exhaustive ground-truth survey.

4 Discussion

The proposed framework represents a practical improvement over conventional façade inspection and maintenance practices. The field experiments demonstrated that digital 3D façade models can be effectively translated into executable UAV inspection missions, reducing dependence on pilot experience and improving the consistency of image acquisition. In addition, the combination of visible and thermal infrared imagery helps overcome the limitations of single-sensor inspection: visible imagery captures surface defects such as cracks and peeling, whereas thermal infrared imagery provides complementary information related to subsurface delamination. The experimental results show that the proposed dual-modal framework can exploit this complementarity effectively, while the decision-level fusion strategy further improves the interpretability of façade condition by linking visible deterioration with potential internal risks.

The transferability of the framework should be understood at both the detector and fusion levels.

Detector performance may vary with façade material, surface texture, and environmental conditions, whereas the rule-based fusion logic is expected to remain broadly applicable as long as relevant visual and thermal cues are observable. Nevertheless, the current study was validated on a tiled planar façade, and broader validation on other façade types and more complex geometries is still needed. The current path-planning algorithm is mainly suited to planar façades, although irregular envelopes may be approximated as multiple locally planar patches for practical inspection. In addition, thermal-anomaly visibility depends on solar-loading and wind conditions. Future work will focus on broader validation, adaptive planning for complex 3D geometries, and deeper cross-modal fusion.

5 Conclusions

This study presents a façade inspection framework that integrates efficient UAV-based data acquisition with deep-learning-based defect detection for façade safety assessment. The proposed workflow was developed and validated through a field deployment on a commercial building, including oblique photogrammetry-based 3D modeling, route generation, and defect recognition based on modality-specific segmentation and decision-level fusion. The results demonstrate that the proposed approach enables efficient and safe inspections while providing a more informative assessment of both surface-visible defects and subsurface delamination risks.

The main contributions of this research are summarized as follows:

(1) Automated flight route generation: A UAV inspection route generation method was developed based on architectural key points extracted from oblique photogrammetry models, enabling consistent and automated façade data acquisition through waypoint-based missions.

(2) Dual-modal defect detection and interpretable fusion. A defect-identification framework was established using dedicated YOLO11-seg models for visible and thermal infrared imagery, followed by decision-level fusion of the modality-specific outputs. By associating surface appearance anomalies with subsurface thermal signatures, the framework provides a more comprehensive and practically interpretable assessment of façade condition, offering useful support for intelligent infrastructure inspection and maintenance.

6 Acknowledgement

This research is supported by Key Technologies R&D Program (Grant No. 2024YFC3808602), Shenzhen Science and Technology Programs (Grant No. KJZD20230923114310021, Grant No. JCYJ20241202123722029).

References

[1] P. Wang, J. Xiao, Z. Duan, et al. Intelligent development trend of building enclosure damage detection. J. Archit. Civ. Eng., 39(4):24-37, 2022.

[2] J. D. Silvestre and J. De Brito. Ceramic tiling in building façades: Inspection and pathological characterization using an expert system. Constr. Build. Mater., 25(4):1560-1571, 2011.

[3] X. Wang, C. Demartino, Y. Narazaki, et al. Rapid seismic risk assessment of bridges using UAV aerial photogrammetry. Eng. Struct., 279:115589, 2023.

[4] S. Jordan, J. Moore, S. Hovet, et al. State-of-the-art technologies for UAV inspections. IET Radar Sonar Navig., 12(2):151-164, 2018.

[5] S. Ivić, B. Crnković, L. Grbčić, et al. Multi-UAV trajectory planning for 3D visual inspection of complex structures. Autom. Constr., 147:104709, 2023.

[6] A. Y. Yiğit and M. Uysal. Automatic crack detection and structural inspection of cultural heritage buildings using UAV photogrammetry and digital twin technology. J. Build. Eng., 109952, 2024.

[7] M. Štroner, R. Urban, J. Seidl, T. Reindl, and J. Brouček. Photogrammetry using UAV-mounted GNSS RTK: Georeferencing strategies without GCPs. Remote Sens., 13:1336, 2021.

[8] B. R. Bowring. Transformation from spatial to geographical coordinates. Survey Review, 23(181):323-327, 1976.

[9] T. M. Cabreira, L. B. Brisolara, and R. F. J. Paulo. Survey on coverage path planning with unmanned aerial vehicles. Drones, 3(1):4, 2019.

[10] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr., 5(4):1680-1716, 2023.

[11] Ultralytics. "Ultralytics YOLO11," 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[12] I. Garrido, E. Barreira, R. M. S. F. Almeida, et al. Introduction of active thermography and automatic defect segmentation in the thermographic inspection of specimens of ceramic tiling for building façades. Infrared Phys. Technol., 121:104012, 2022.

[13] X. Yang, R. Guo, and H. Li. Comparison of multimodal RGB-thermal fusion techniques for exterior wall multi-defect detection. J. Infrastruct. Intell. Resil., 2(2):100029, 2023.

Intelligent Automation and Optimization of Structural Design in BIM-Based Workflows: Case Studies

Tafraout Salim[1] and Bourahla Nouredine[2]

[1] Ecole Nationale Supérieure des Travaux Publics ENSTP, LTPiTE Laboratory, Algiers, Algeria
[2] Ecole Nationale Polytechnique, Civil Engineering Department, LGSDS Laboratory, Algiers, Algeria
s.tafraout@enstp.edu.dz, nouredine.bourahla@g.enp.edu.dz

Abstract

This paper presents an AI-driven framework designed for automating and optimizing structural systems within a Building Information Modeling (BIM) workflow. The process begins by examining architectural layouts from IFC files to automatically define a valid search space for structural elements, ensuring that the solutions adhere to a set of structural design constraints. Central to this approach is a multi-objective optimization engine driven by a genetic algorithm (GA). It commences with an initial population of structural configurations, which are then validated against criteria for architectural, structural, and seismic design. Versatility and robustness of the framework are demonstrated through three challenging case studies. For reinforced concrete (RC) wall-slab systems and cold-formed steel (CFS) structures, a GA optimally determines the primary load-bearing elements to satisfy structural and seismic requirements. For RC dual systems (RCDS), a novel hybrid optimization strategy was implemented, beginning with a hill-climbing algorithm that performs a targeted search to determine optimal column placements. Following this, a genetic algorithm is employed to refine the positioning and configuration of shear walls within the established column grid, ultimately enhancing the performance of the dual system. The results indicate that the framework consistently produces high-performance, compliant structural designs rapidly. By automating intricate initial design tasks, this approach not only accelerates the design process but also enhances collaboration between architects and engineers, demonstrating significant potential for widespread adoption in the Architecture, Engineering, and Construction (AEC) industry.

Keywords –
Automation Design; Hybrid Optimization; Genetic Algorithm; BIM; Seismic Design; AEC Industry.

1 Introduction

The integration of artificial intelligence (AI) is catalyzing a profound metamorphosis within the Architecture, Engineering, and Construction (AEC) industry. Due to the iterative and multidisciplinary nature of design and construction processes, this sector is particularly well-suited for the application of AI-driven generative and optimization techniques. A comprehensive review of generative AI in structural engineering identifies several pivotal research domains, including data representation, intelligent generation algorithms, design evaluation metrics, and the synthesis of generation and optimization workflows [1, 2].

Architects are increasingly using AI to augment the design process [3, 4, 5]. Machine learning algorithms enable rapid iterations for complex geometries, facilitating the creation of more sophisticated architectural forms [6, 7]. Moreover, AI is optimizing mechanical, electrical, and plumbing (MEP) systems, thereby enhancing energy performance during the early stages of design [8, 9]. Within Building Information Modeling (BIM) environments, these advanced tools contribute to improved project oversight through automated monitoring and analysis [10].

Recent advancements in structural engineering have introduced innovative methods for automation in design and optimization that transcend traditional methodologies. A significant approach is represented by AIBIM, which incorporates Generative AI in conjunction with Building Information Modeling. This methodology employs a Physics-based Conditional Diffusion Model (PCDM) to generate high-quality structural designs, taking into account real-world constraints such as building height and seismic intensity [11]. Furthermore, the utilization of Generative Adversarial Networks (GANs) has been demonstrated in the automation of layout design for steel frame-braced structures, leveraging existing drawings from experienced engineers to enhance design efficiency [12].

Of particular interest, newly developed frameworks, rule-based approaches employing AI algorithms show high potential for BIM integration [13-15]; however, some more case studies for comparison and performance evaluation are required. In continuation of this previous work, the authors present a synthesis of these newly formulated frameworks in a single system-tool outlining the design processes for the three distinct types of structures made from various materials. The first type discussed is wall-slab reinforced concrete (RC) structures [13]. The second type focuses on cold-formed steel (CFS) structures [14]. Lastly, the RC dual systems, as detailed in [15].

The efficiency and robustness of these frameworks have been validated through testing BIM models of 17 architectural buildings, incorporating various configurations and levels of complexity. The unified GA/HCA pipeline automates the process of converting IFC to structural layouts across common systems.

As novelty, our proposed framework advances beyond our prior individual system studies [13-15] through three key innovations: (1) a unified IFC-to-layout pipeline integrating all three structural systems within a single framework; (2) standardized multi-objective fitness functions enabling direct performance comparison across RC, CFS, and dual systems; and (3) comprehensive validation across diverse BIM models demonstrating framework generality and robustness. These advancements transform fragmented single-system tools into a production-ready, multi-material seismic design platform.

2 AI-based structural design framework

The proposed framework addresses three common seismic-resistant structural systems for low to medium-rise buildings: Reinforced Concrete (RC) wall-slab systems, Cold-formed steel (CFS) shear wall panels, and RC dual systems, which combine moment frames and shear walls. These systems represent the load-resisting configurations that are commonly found worldwide.

The approaches reviewed for optimizing structural design start with the implementation of an architectural 3D-BIM model. This model aids in identifying suitable locations for structural elements while taking into account architectural considerations such as openings, slab limits, and partition walls. This ensures a continuous load path from the roof to the foundations. The optimization problem is inherently multi-objective, aiming to select the most effective structural configuration from a range of alternatives. Once common architectural constraints are established, the methodologies diverge based on the type of structure, utilizing specialized algorithms and design focuses. For reinforced concrete (RC) wall-slab systems, a single genetic algorithm is employed to generate an optimized structure from a given set of elements. In contrast, cold-formed steel (CFS) structures use a tailored genetic algorithm to optimize the layout of shear walls, while also evaluating various weights for the objective functions. For dual systems, a hybrid protocol begins with a hill-climbing algorithm to determine an optimal frame, followed by a genetic algorithm to effectively distribute the shear walls. All of these approaches ensure that the optimal solution complies with established structural rules and meets stability criteria in accordance with relevant codes of practice.

3 Structural constraints and design rules

In contemporary architectural practice, it is imperative that structural systems afford architects enhanced flexibility in spatial planning while simultaneously optimizing surface utilization. Engineers are tasked with the challenge of devising efficient structural solutions that align with the architectural design, ensuring that strength and stability are not compromised. Typically, buildings employ self-supporting vertical elements, such as columns and walls, complemented by horizontal components like beams and slabs, to manage lateral forces resulting from seismic and wind loads, effectively transferring these forces to the foundation. This section outlines the structural criteria for the developed Generative Algorithm (GA) and Hill-Climbing Algorithm (HCA) within an automated design framework. Focusing on prevalent structural systems for common multi-storey buildings, the GA optimizes the placement of discrete elements, such as shear walls or CFS panels, while the HCA is utilized to optimize the column positions along the axes in dual systems.

3.1 Reinforced concrete wall-slab system

The design of Reinforced Concrete (RC) Wall-Slab systems adeptly confronts various construction challenges by ensuring a rigid connection between the slabs and RC walls. This integration results in a monolithic structure that significantly enhances stability through inherent redundancy. The walls function as load-bearing elements, effectively resisting both vertical loads and lateral forces. The essential criteria for the genetic algorithm employed in the multi-objective optimization process are summarized as follows [13]:

3.2 Individual shear wall lengths

Seismic design codes, such as the Algerian RPA2024 and Eurocode EC8, delineate the parameters for reinforced concrete (RC) shear walls, specifically in terms of their thickness and length. These parameters

critically influence the walls' stiffness, strength, and modes of failure. The minimum thickness of these walls is determined by the free height of the floor, thus ensuring adequate structural integrity. Within the context of the developed framework, the permissible lengths for these shear walls range from 0.8 to 5.5 meters, which aligns with the requirements for low- to medium-height buildings. This framework aims to optimize the performance and reliability of shear walls under seismic loading conditions.

3.3 Total length of RC shear walls

The total length of reinforced concrete (RC) shear walls can be defined as the ratio of their total cross-sectional area to the surface area of the corresponding floor. To ascertain the requisite shear wall length for ensuring structural stability and strength, an iterative parametric analysis, in conjunction with established earthquake engineering recommendations, was employed. The minimum length is calculated based on the floor area (Sf) and the overall height of the building (H) using the following methodology:

$$L_{w,m} = 0.007H.\ S_f \text{ for } H < 30 \text{ m} \quad (1)$$

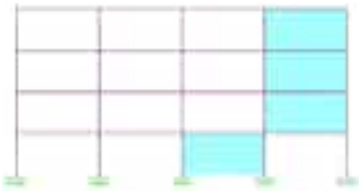
$$L_{w,m} = 0.013H.\ S_f \text{ for } 30 \text{ m} < H < 100 \text{ m} \quad (2)$$

3.4 RC shear wall layouts

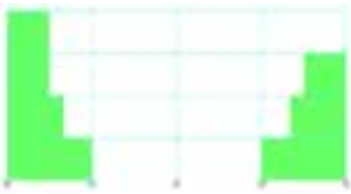
The configuration of reinforced concrete (RC) walls is essential for optimizing a structure's torsional response in the face of seismic activity. Two primary criteria are delineated in this context: Firstly, shear walls should be strategically positioned in proximity to the center of stiffness and aligned with the center of mass across both horizontal axes. Secondly, these walls should be situated as distantly as possible from the center of the floor to increase torsional stiffness. This specific arrangement plays a crucial role in preserving plan uniformity and facilitating non-torsional flexibility, thereby enhancing the overall seismic performance of the structure.

3.4.1 Other criteria

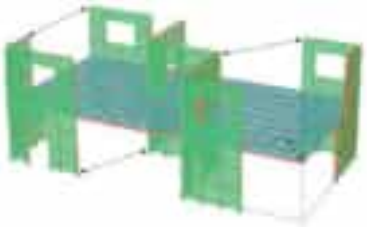
- Load path continuity: It is imperative to maintain load path continuity to ensure the effective vertical transfer of lateral forces to the foundation system, as illustrated in Figure 1-a.
- Variation in shear-wall length with height: An optimal structural design necessitates a progressive reduction in the stiffness of shear walls, which is achieved by decreasing their length from the base to the top of the structure, as depicted in Figure 1-b.
- Geometric slab conditions: Established guidelines regarding geometric slab conditions highlight critical parameters, such as the span distance between two reinforced concrete walls and the maximum allowable free length of a cantilever, which is typically restricted to a maximum of 2.5 meters, as shown in Figure 1-c.



a) Vertical continuity



b) Length variation with height



c) Geometric condition for slab

Figures 1. RC Wall-Slab design criteria

3.5 Cold-formed steel structures

A cold-formed steel (CFS) structure represents a sophisticated form of light-frame construction characterized by the assembly of vertical and horizontal components comprised of repetitive framing members (Figure 2). The integration of rigid connections between the slabs and walls significantly contributes to the overall stability and redundancy of the system. Vertical loads are efficiently transferred from the floor joists to lipped channel "stud" members, which are anchored to the foundation utilizing unlipped channel "track" sections. Furthermore, lateral forces resultant from seismic or wind loads are effectively transmitted to the foundation through the incorporation of shear

wall panels or bracing systems, presuming the presence of rigid diaphragms and strict adherence to capacity design criteria. This structural configuration underscores the importance of careful design considerations in ensuring the resilience and performance of cold-formed steel construction in various loading scenarios.



Figure 2. Examples of CFS structure

The shear wall system (SWP) serves as the principal lateral load-resisting framework, employing cold-formed steel (CFS) C-shaped framing affixed to either steel or wood sheathing through mechanical fastening. This system is engineered to exhibit inelastic behavior at the connection points until failure occurs.

The nominal strength of the system will guide the preliminary design process concerning the requisite lengths for each floor, subsequently followed by comprehensive cross-section evaluations.

Fundamental criteria necessary for the effective implementation of this system in the framework of GA multi-objective optimization are delineated below [14]:

3.5.1 Joist directions

The orientation of joists within a structural slab is predominantly influenced by the spans involved as well as the configuration of wall openings. It is recommended that joists be aligned with the longer span; in instances where the dimensions are comparable, they should be oriented parallel to the walls that exhibit fewer openings.

3.5.2 Length of individual wall

The minimum length of a single wall panel is dictated by the spacing between two consecutive studs, with a specified minimum distance of L_{wmin} = 60 cm. Furthermore, the height-to-width (h/w) ratio of an SWP is constrained by a maximum allowable ratio of 4:1. It is posited that the optimal length of an SWP is represented by L_{wopt} = h, where "h" denotes the height of the storey.

3.5.3 Total Length of SWPs

Shear wall panels (SWPs) must exhibit the capacity to withstand lateral loads induced by wind or seismic events from both directions. The minimum total length of shear walls required "L_{wtmin}" at each level and for each directional influence is determined by calculating the level shear lateral load (Vi) and dividing it by the nominal strength of the shear wall per unit length ($R_{nom.}$). This calculation is essential for ensuring structural integrity and stability in the face of lateral forces.

3.5.4 Other criteria

In alignment with the principles established for the RC-SW system, it is imperative to incorporate two critical structural criteria into the objective function on this structural configuration: the continuity of the vertical load path and the spatial distribution of the shear wall panels (SWP) in plan view. The integration of these criteria facilitates a reduction in overall stiffness while mitigating the risk of vertical discontinuities, thereby minimizing the torsional response of the building. This approach not only enhances structural performance but also contributes to the overall stability and resilience of the system.

3.6 Reinforced concrete dual systems

The automated design process for the structural system initiates with the identification of optimal layouts for moment-resisting frames (MRFs), a task that presents significant challenges due to the various configurations available for column placements. The subsequent phase encompasses the strategic arrangement of shear walls within the established MRF framework [15].

The process starts by overlaying architectural floor plans to facilitate the alignment of wall partitions and non-structural components, such as doors and windows, thereby aiding in the identification of potential column locations. To optimize this arrangement, a Hill-Climbing Algorithm (HCA) is selected to define the axis positions of the columns. This is particularly suitable for a given set of predefined potential axes. It iteratively moves from one solution to its neighbors by applying slight changes to the positions of the axes. At each iteration, neighbors of the current solution are generated and evaluated. The best one, based on the

given criteria, is then selected to replace the current one for the next iteration. This allows for rapid convergence to optimal axis positions.

This automated methodology, founded upon the geometric data provided by the Industry Foundation Classes (IFC), yields feasible column positions, as presented in Figures 3 and 4.

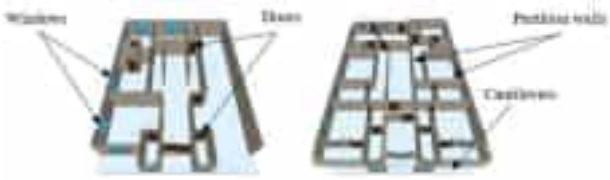


Figure 3. 3D-BIM model (IFC)

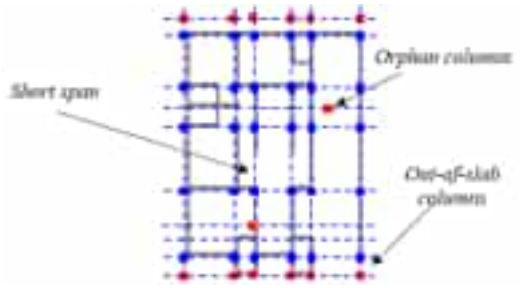


Figure 4. Potential column positions

Once optimal MRF grids have been established, the positioning of shear walls to facilitate the creation of reinforced concrete dual systems (RCDS) must be meticulously aligned with this predetermined grid framework.

It is imperative to enforce a minimum distance constraint to avert the formation of short links between load-bearing elements. Furthermore, in compliance with prevailing seismic codes, the moment-resisting frame (MRF) is required to endure a minimum of 25% of the base shear. The lateral loads acting on the structure are simultaneously resisted by the shear walls (SWs) and the MRF, with their capacity to resist loads being determined by their respective stiffness characteristics.

The seismic design classification of the RCDS as per Eurocode 8 (EC8) is contingent upon the proportion of shear resistance contributed by both the MRF and the shear walls. In scenarios where either the MRF or the shear walls account for more than 50% of the total shear resistance of the structure, stringent control measures must be implemented. This is achieved through the imposition of minimum and maximum length constraints for the shear walls in both principal directions. Optimization of the RCDS layout is facilitated through a scoring approach predicated on a defined fitness function.

In addition to shear wall dimensions, other critical parameters, such as the total slab area and the effective covered area, which accounts for the surface area directly supported by the shear walls while considering overlaps with adjacent structural elements and overhangs from the slab edges, are systematically calculated for each configuration. Scoring formulas are developed and systematically integrated into a multi-objective optimization function, thereby regulating the ratio of shear walls' contributions to the gravity load-bearing capacity within the RCDS framework.

4 GA-based process for generating structural design

In an architectural BIM platform, the IFC protocol facilitates the export of the model's geometry and elements. The algorithm defines slab boundaries and wall partition layouts for each floor, including openings. Structural elements are generated iteratively based on genetic actions. The Genetic Algorithm (GA) optimizes the structural layout after a set number of iterations, effectively meeting constraints and maximizing the objective function.

1. Random generation of the initial population: The algorithm commences by generating an initial population that comprises a diverse set of structural elements. These elements are randomly allocated within the architectural partitions to establish a foundational layout for subsequent iterations.
2. Selection: The selection phase operates on the principle of fitness ranking, wherein the layouts of the subsequent population are chosen based on the performance metrics of all individuals within the current population. This evaluation encompasses layouts produced through crossover and mutation processes, ensuring that the most promising designs are carried forward.
3. Crossover: This function facilitates the unification of structural characteristics from two or more existing configurations to produce a novel layout. This process ensures that the resultant solution retains desirable attributes from the parent layouts while adhering to established structural constraints, thereby enhancing the viability of the new configuration.



Figure 5. Crossover function

4. Mutation: The mutation function serves to enhance the variability within the population by implementing alterations to the structural layout. This may involve the insertion of the new elements, the removal of an existing one, or the modification of the dimensions of the current elements, such as shortening or extending the length of the SW. These changes expand the range of possible designs, enhancing innovative structural solutions.

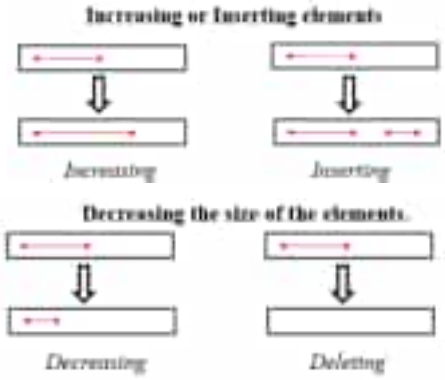


Figure 6. Mutation function

A multi-objective optimization function has been established to direct the population towards an optimal solution, grounded in the established criteria pertinent to each structural type. This optimization function encompasses a diverse set of criteria, each designated with a score ranging from 0 to 1.0, reflecting the validity of the criterion's condition.

The fitness value of a given layout solution is computed through a weighted linear combination of scores across the various criteria. Initially, uniform weights of 1.0 were applied, yielding acceptable outcomes. Subsequently, a sensitivity analysis was conducted to fine-tune these weights in accordance with the significance of each criterion.

The termination process occurs following the completion of the primary evolutionary loop, typically when the objective function reaches a predetermined threshold. However, in the context of multi-objective functions, defining a specific value poses significant challenges. Consequently, a limited number of iterations, generally ranging from 50 to 100, has been implemented, resulting in satisfactory outcomes.

5 Performance criteria of the multi-objective function

The performance parameters associated with the multi-objective function were rigorously evaluated through the analysis of seventeen (17) architectural BIM models, specifically to assess the efficacy of GA in structural design. The selection of samples was methodical, reflecting a diversity in dimensions, forms, placement of openings, and configurations of partitions. The input extracted from the IFC models was instrumental for the GA, which subsequently generated optimized structural solutions through a series of iterative processes. Generated structural designs conformed to established architectural standards, catering to the specified requirements for interior partitions, positioning of openings, and the delineation of wall load paths.

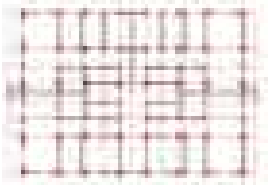
For RC-SW and Dual designs, three-dimensional numerical models were developed utilizing BIM-compatible finite element (FE) software to rigorously evaluate the structural performance of layouts generated through the GA. The assessment was conducted to ascertain compliance with established structural design constraints. The findings reveal that all design criteria were satisfactorily fulfilled, particularly regarding stability conditions and structural strength [13-15]. For the SWP design, the structural performance was evaluated by juxtaposing the GA-generated layout with a conventional structural layout [14]. Figures 7a, 7b, and 7c present examples of generated designs for each type of structure.



a) RC-SW design



b) SWP design



c) Dual system design

Figure 7. Various generated structural designs

Table 5.1 summarizes key metrics that quantify performance, robustness, and efficiency claims across these 17 BIM models:

Table 1: Performance metrics across 17 BIM models

Criterion	RC W-S	CFS SWP	RC Dual
Const. Satisfaction	100%	100%	100%
Avg. Conv. Iterations	50 - 100		
Tors. Irreg. Reduction	~35%	~35%	~30%
Overstrength Factor	> 2.0		

These results confirm the framework's robustness: 100% of seismic code compliance across all systems, consistent convergence within design limits, significant torsional improvements, and ductile overstrength meeting seismic performance objectives, where more than 90% of the structural members of tested models remain in the elastic branch [13-18].

The genetic algorithm demonstrated notable robustness and efficacy in the generation of layouts, particularly for the RC-SW system. This was substantiated by varying key design constraints and adjusting the weight scales within the objective function, which indicated a significant efficiency in managing the design criteria [16-17].

Additionally, a recent analysis reviewed the models based on the hill-climbing algorithm (HCA) for the development of moment-resisting frames (MRFs) structural designs. The structures were subjected to pushover analysis, which facilitated the identification of capacity curves and performance points tailored to specific seismic demands. Notably, despite the utilization of general objective functions, the evaluated layouts exhibited strong performance and adhered to seismic standards. The analysis demonstrated that over 90% of structural members did not exhibit the formation of plastic hinges, with the MRFs sustaining performance within the Immediate Occupancy (IO) to Life Safety (LS) range, indicative of substantial overstrength. These insights suggest promising avenues for future research aimed at enhancing the algorithm through the integration of a Performance-Based Design (PBD) approach, thereby improving the control of structural performance [18].

6 Conclusion

The integration of automation and optimization within the Architecture, Engineering, and Construction (AEC) industry signifies a pivotal transformation in the paradigms of building design and construction methodologies. The deployment of advanced computational tools, artificial intelligence, and data analytics serves to enhance the efficiency, precision, and innovative capacity inherent in structural design processes.

This paper presents a comprehensive review of a newly developed framework proposed by the authors, which employs Genetic Algorithms (GA) and Hill-Climbing Algorithm (HCA) to facilitate the automatic generation of optimal structural systems predicated on specified architectural configurations in a three-dimensional Building Information Modeling (3D-BIM) format. The core premise of this methodology is the establishment of a multi-objective optimization process. This process entails the stochastic generation of a vast array of potential structural configurations that adhere to architectural constraints, alongside a defined set of general structural and seismic design principles. Through successive iterations, the optimization process converges towards solutions that increasingly align with the predetermined criteria.

The proposed method has been systematically applied to three prevalent structural designs:

Reinforced Concrete Shear Wall and Slab (RC SW-Slb), cold-formed steel Panels (SWP), and dual systems. The outcomes relative to the performance, efficiency, and robustness of the generated structures indicate substantial advancements in the formulation of optimized structural configurations for the examined design samples, all of which conform to the specified design criteria for a given architectural layout.

References

[1] W. Liao, X. Lu, Y. Fei, Y. Gu, Y. Huang. Generative AI design for building structures. *Automation in Construction, 157, 105187*, 2024. https://doi.org/10.1016/j.autcon.2023.105187

[2] A. Nikitin, S. Sinenko. Features of the use AI in generative design of building and structures. *Journal of mechanics of continua and mathematical sciences, Vol – 19 No – 5,* 2024. https://doi.org/10.26782/jmcms.2024.05.00001

[3] Y. Li, H. Chen, P. Yu, L. Yang. A Review of Artificial Intelligence in Enhancing Architectural Design Efficiency. *Journal of Applied Science, 15, 1476*, 2025. https://doi.org/10.3390/app15031476

[4] S. W. Dharmatanna, E. S. Wijaya. The Study of AI integrated Simulation in Building Information Modelling (BIM) Use at Architectural Design Studio. *Journal of Artificial Intelligence in Architecture, vol. 4, no. 1, pp. 23–34*, 2025. https://doi.org/10.24002/jarina.v4i1.9415

[5] J. Anshul, K. Ananda Babu. Application of Artificial Intelligence (AI) in the sustainable design of green buildings. *Journal of Environmental Pollution and Environmental Modelling, Vol. 8(1): 59-65*, 2025.

[6] A. Ayman, Y. Mansour, H. Eldaly. Applying machine learning algorithms to architectural parameters for form generation. *Automation in Construction, Volume 166, 105624*, 2024. https://doi.org/10.1016/j.autcon.2024.105624

[7] D. D. Shuvo, I. Mobasshira, S. Md. Habibur Rahman, A. Shakil, K. Moumita. Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. *Heliyon, Volume 10, Issue 5, e26888,* 2024. https://doi.org/10.1016/j.heliyon.2024.e26888

[8] R. Deng, X. Li, Y. Tian. A Review: The Application of Path Optimization Algorithms in Building Mechanical, Electrical, and Plumbing Pipe Design. *Buildings, 15(12),2093,* 2025, https://doi.org/10.3390/buildings15122093

[9] D.N. Tahat, A. Mansoori, K. Tahat, R. Alfaisal, H. Yousuf, S.A. Salloum. Integrating Generative AI into Predictive Modeling for Energy Efficiency Optimization in Building Design. Generative AI in Creative Industries. *Studies in Computational Intelligence, vol 1208. Springer,* 2025). https://doi.org/10.1007/978-3-031-89175-5_36

[10] G. Wei, D. Zhou, X. Yuan. Target Detection Technology of Mechanical, Electrical, and Plumbing Components Based on CV. *Applied Science, 15, 2803*, 2025, https://doi.org/10.3390/app15052803

[11] Z. He, Yu-H. Wang , J. Zhang. Generative AIBIM: An automatic and intelligent structural design pipeline integrating BIM and generative AI. *Information Fusion (114) 102654*, 2025, https://doi.org/10.1016/j.inffus.2024.102654

[12] B. Fu, Y. Gao, W. Wang. Dual generative adversarial networks for automated component layout design of steel frame-brace structures. *Automation in Construction, 146, 104661*, 2023, https://doi.org/10.1016/j.autcon.2022.104661

[13] S. Tafraout, N. Bourahla, Y. Bourahla, A. Mebarki. Automatic structural design of RC wall-slab buildings using a genetic algorithm with application in BIM environment. *Automation in Construction 106 102901*, 2019, https://doi.org/10.1016/j.autcon.2019.102901

[14] N. Bourahla, S. Tafraout, Y. Bourahla, A. Serier El-Hirtsi, A. Skoudarli. GA based design automation and optimization of earthquake resisting CFS structures in a BIM environment. *Structures 43, 1334–1341*, 2022, https://doi.org/10.1016/j.istruc.2022.07.041

[15] N. Bourahla, S. Larfi, K. Souaci, Y. Bourahla, S. Tafraout. Intelligent automation and optimization of reinforced concrete dual systems for earthquake resisting buildings in a BIM environment. *Journal of Building Engineering 76, 107111*, 2023, https://doi.org/10.1016/j.jobe.2023.107111

[16] N. Bourahla, S. Tafraout, Y Bourahla. Performance evaluation of an intelligent structural design process in a BIM environment: case of RC wall slab resistance systems. *In Proceeding of the conference RACEI,* pages 604-610, India, 2019.

[17] N. Bourahla, S. Tafraout, Y. Bourahla. Intelligent structural design in BIM platforms: Optimization of RC wall slab systems. *European Conference on Product and Process Modeling, ECPPM,* Russia, 2021, http://dx.doi.org/10.1201/9781003191476-36

[18] S. Tafraout, N. Bourahla. Seismic performance of automatically generated ductile moment resisting frames. *Conference of the Civil and Environmental Engineering for Resilient, Smart and Sustainable Solutions CEES*, Saudi Arabia, 2024, https://doi.org/10.21741/9781644903414-23.

Hybrid simulation in construction: Trends and design framework

Xuming Zhu[1] and Jinchi Han[1]

[1]Department of the Built Environment, National University of Singapore, Singapore
E0688540@u.nus.edu, hanjinchi@bucea.edu.cn

Abstract

Due to construction work's complexity and uncertainty, single simulation methods such as system dynamics (SD), discrete-event simulation (DES), and agent-based modelling (ABM) face significant limitations when applied independently. Hybrid simulation offers a promising solution, yet lacks both a comprehensive review and a structured framework for construction applications. This paper addresses these gaps by providing a thorough review identifying trends and improvement areas, and presenting a detailed framework integrating SD, DES, and ABM tailored to the construction industry. The study also identifies future research directions, including case studies to validate the framework's effectiveness and demonstrate its practical application in real-world construction projects.

Keywords

Hybrid simulation; Construction; System dynamics; Discrete-event simulation; Agent-based modelling; Framework

1 Introduction

In the construction industry, simulation has become an essential tool for planning, analyzing, and optimizing complex projects [1]. The dynamic nature of construction sites, with its multiple variables such as time, cost, labor, materials, and site-specific constraints, makes it challenging to manage efficiently using traditional analytical methods. Simulations provide a means to model construction processes, forecast outcomes, and assess the impact of various strategies across different scenarios. They aid in improving safety, productivity, and quality by providing a virtual platform to test and optimize construction operations prior to real-world implementation [2].

Three fundamental simulation approaches, namely system dynamics (SD), discrete-event simulation (DES), and agent-based modelling (ABM), are widely employed in construction [3]. SD focuses on capturing the behavior of complex systems over time by modelling the interrelationships among components, often through causal loop diagrams and stock-flow models. DES, in contrast, models systems as sequences of discrete events that alter the system state, such as the initiation or completion of tasks, the allocation or release of resources, or the arrival of materials and equipment on site. ABM models systems through autonomous agents whose state-dependent actions and interactions determine both individual behavior and the emergent evolution of the system. In this context, it is important to recognize that SD, DES, and ABM are simulation paradigms, conceptual approaches that define how a system is abstracted and represented at a logical level, and are fundamentally distinct from physics-based simulation platforms that replicate physical-world behaviors within three-dimensional virtual environments. For instance, NVIDIA Isaac Sim provides a GPU-accelerated environment for simulating robotic kinematics, rigid-body dynamics, and sensor responses such as LiDAR and camera feeds, primarily for the purpose of testing and training AI-driven robotic systems [4]. Similarly, Unity, originally developed as a game engine, has been increasingly adopted for robotics and industrial simulation due to its real-time physics engine and photorealistic rendering capabilities, enabling applications such as digital twin development and robot navigation testing [5]. While these platforms offer valuable capabilities for replicating physical interactions at the individual object or equipment level, they do not address how to conceptualize and structure the broader system-level processes and interdependencies that characterize construction projects. This study focuses on the paradigm level, as the integration of these conceptual approaches remains a more pressing methodological challenge in construction simulation compared to the selection of specific implementation platforms. While these methods contribute valuable insights, no single approach can fully capture the multifaceted complexity of real-world construction projects. Reliance on a single method either narrows the scope of analysis or produces oversimplified representations, thereby limiting both validity and practical applicability [3].

To overcome these limitations, an integration

framework is essential to coordinate simulation paradigms, prevent incompatible sub-models, and enable cross-project replication [3]. However, current frameworks for hybrid simulation remain fragmented, hindered by inconsistent terminology and a tendency to operate at high levels of abstraction [6], [7], treating each method as a unified whole. This oversimplification fails to reflect the reality of hybrid modelling, where numerous combinations of interaction are possible, creating significant challenges for model development. Consequently, these frameworks offer limited practical guidance for researchers attempting to construct models, as they lack the details needed to handle the complexity of hybrid modelling.

Although numerous studies have reviewed single simulation methods in construction, research on hybrid simulation in construction remains limited. The few existing studies [2], [8], [9] primarily summarize frameworks developed in other industries rather than proposing approaches tailored to the construction-specific characteristics: temporary nature, site-specific constraints, high uncertainty, and diverse activities requiring different simulation approaches. Moreover, no comprehensive review systematically maps hybrid simulation applications, development trends, and thematic focus in construction, limiting researchers' ability to track field evolution and identify future directions. Therefore, this study aims to address these gaps by providing a comprehensive review and a holistic framework for implementing hybrid simulation approaches in the construction domain.

The remainder of this paper is organized as follows: Section 2 describes the literature collection; Section 3 conducts a bibliometric analysis; Section 4 proposes a framework for hybrid simulation in construction; conclusion and future work are provided in Section 5.

2 Literature collection

This study adopts a comprehensive approach to search and collect literature, guided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [10] (Figure 1). To collect papers on hybrid simulation in construction, databases including “Scopus”, “Web of Science”, and “Google Scholar” were used to search for the papers from 2000-2024. Three search term sets (hybrid simulation-related, method-related, and construction-related) are shown in Table 1. Four method combinations were used: “SD AND DES”, “SD AND (ABM OR ABS)”, “DES AND (ABM OR ABS)”, and “SD AND DES AND (ABM OR ABS)”, where agent-based simulation (ABS) is an ABM variation. Terms from the first and third sets were paired with each method combination, covering all possible combinations (such as “Hybrid simulation” AND “SD AND DES” AND “Construction"). Papers underwent two-stage screening: initial review of titles, abstracts, and metadata, followed by full-text examination. This method maximizes inclusion of relevant papers, with automation tools removing duplicates and ineligible studies.

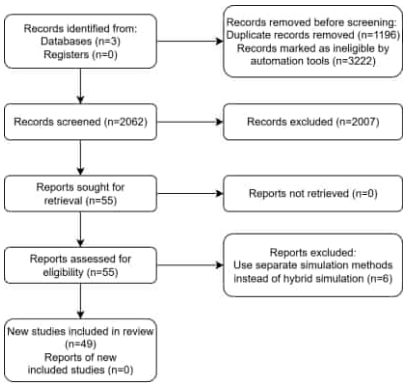


Figure 1. PRISMA flow diagram

The next step involved a manual screening process to exclude irrelevant papers, followed by retrieval of the remaining studies for a more detailed eligibility assessment. Finally, a total of 49 papers were selected (listed in Table 2) for the analysis.

Table 1 Search terms used to collect papers

Hybrid simulation-related search terms	Specific method-related search terms	Construction-related search terms
Hybrid simulation	System dynamics/SD	Construction
Mixed simulation	Discrete event simulation/DES	Civil engineering
Multi-method simulation	Agent-based modelling/ABM	Building
Integrated simulation	Agent-based simulation/ABS	Infrastructure
		Project management

Table 2 Database of papers included

Review & framework	[11] [12] [1] [3] [2]		
Stage	**SD-DES**	**SD-ABM**	**DES-ABM**
Pre-constr.	[13] [14]	[15]	-
Constr.	[8] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31]	[32] [33] [34] [35] [36] [37]	[32] [38] [39] [40] [41] [42] [43]
Post-constr.	[44]	-	[45] [46] [47]
Cross-stage	[48] [49]	[50] [51] [52]	[53] [54]

3 Bibliometric analysis

As identified in Section 1, no comprehensive review has systematically mapped hybrid simulation applications, development trends, and thematic focus in construction. To address this gap, this section presents a bibliometric analysis of the collected literature.

3.1 Bibliometric analysis on publications

This bibliometric analysis examines the publication trends from 2007 to 2024. Figure 2 illustrates the yearly distribution of publications, highlighting a peak in 2018 with a total of 9 publications. A polynomial-fitted dashed trend line represents the overall trajectory, indicating a gradual increase in publications leading up to 2018, followed by fluctuations in subsequent years. Notably, during the five-year period from 2015 to 2019, nearly half of the total publications (24 out of 49) were produced, underscoring a concentrated period of research activity.

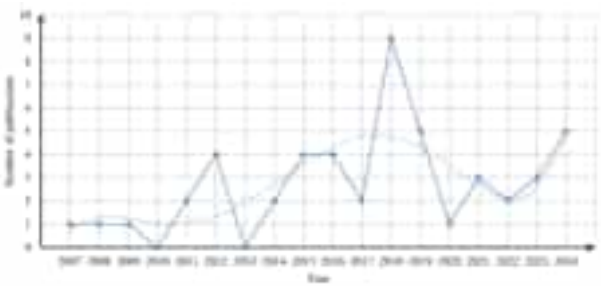


Figure 2. Yearly number of publications

A comprehensive analysis of keywords is conducted. Of 210 unique keywords, 40 appear more than once, signifying recurring research topics. The ten most frequent keywords are: SD, DES, hybrid simulation, simulation, ABM, ABS, construction management, productivity, construction industry, and construction.

3.2 Bibliometric analysis on journals

Bibliometric analysis of journals identifies key publication venues, tracks influential studies, and reveals field impact and productivity. An analysis of 49 selected papers shows that they are distributed across 30 journals or conferences, with only 9 journals featuring more than one publication. "Automation in Construction" leads in publication count, followed by "Construction Innovation" and "Journal of Construction Engineering and Management". 9 journals collectively account for 57.1% (28 out of 49) of total publications.

3.3 Bibliometric analysis on authors

Bibliometric analysis on authors documents contributions. Among the 49 selected papers, 131 authors were identified, of whom 24 published more than one paper. There are 6 distinct research clusters and 5 overall research groups, as some clusters exhibit strong interconnections. The largest group comprises 8 authors with 9 papers, where SangHyun Lee and Simaan AbouRizk occupy key positions. The second largest group consists of 5 authors with 4 papers, centered on Farnad Nasirzadeh. The third largest includes 4 authors with 6 publications, centered on Tarek Zayed.

4 Framework for hybrid simulation

As identified in Section 1, existing hybrid simulation frameworks lack consideration of construction-specific characteristics. To address this gap, this section presents an integrated framework (Figure 3) that serves as a practical decision-support tool for hybrid simulation in construction. The hybrid simulation design framework consists of four phases that guide researchers and industry practitioners systematically from problem definition to implementation.

4.1 Problem definition

The framework's initial phase develops comprehensive understanding of the construction problem, like other hybrid simulation frameworks [55], helping researchers articulate problems with modelling precision and determine whether complexity warrants a hybrid approach. Precise problem definition requires four elements: a problem statement identifying concrete issues and consequences; primary performance focus (productivity, quality, or safety); construction context (temporal and spatial); and scope boundaries balancing comprehensiveness with feasibility. Researchers then analyze characteristics along three dimensions: continuous (feedback loops, flows, accumulations), discrete (events, queuing, sequential operations), and behavioral (autonomous decision-making), each suggesting different modelling requirements. This characterization determines whether hybrid simulation is necessary: single methods suffice when problems align with one dimension, while hybrid approaches are essential when problems span multiple dimensions. If hybrid simulation is needed, proceed to Phase 2 for method selection; otherwise, exit for single methods.

4.2 Model selection

To guide method selection, we systematically analyzed 49 reviewed papers to identify patterns in hybrid simulation applications. Papers were classified by construction phase (pre-construction, construction, post-construction, cross-stage) and hybrid approach. While some papers incorporated all three methods (SD, DES, ABM), they typically applied them in paired

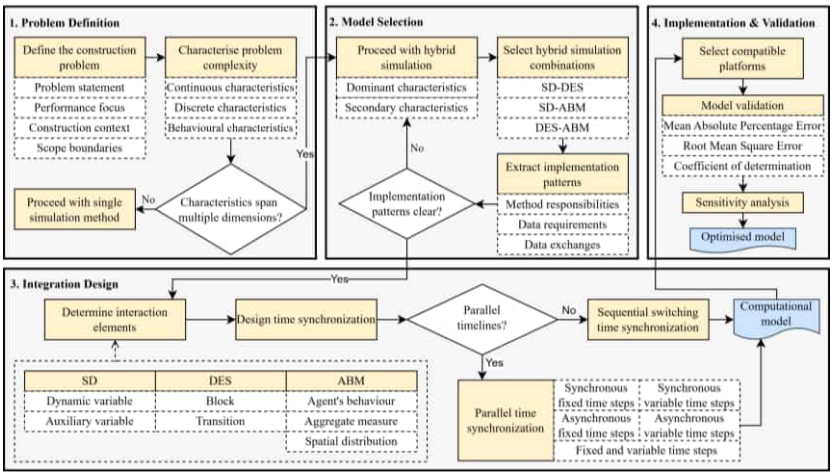


Figure 3. Framework for hybrid simulation design

combinations rather than fully integrated systems. Consequently, we focused on three hybrid subcategories: SD-DES, SD-ABM, and DES-ABM. This systematic classification, calculated from Table 2, revealed clear patterns in how researchers have matched hybrid approaches to specific construction challenges. Construction phase dominates with 30 studies (61%), while pre-construction (3 studies) and post-construction (4 studies) remain underexplored. Cross-stage applications represent 7 studies (14%), addressing strategic planning spanning multiple phases. SD-DES combinations lead with 22 studies, reflecting alignment between system-level dynamics and operational processes. DES-ABM follows with 12 studies addressing process modelling and agent behavior, while SD-ABM accounts for 10 studies focusing on system-wide factors interacting with individual behaviors.

Researchers evaluate hybrid simulation methods based on Phase 1 problem analysis. The dominant characteristic determines the primary method, whilst secondary characteristics guide complementary method selection. Once methods are selected, researchers verify that method responsibilities are clearly defined, required data exists for both methods, and sufficient implementation guidance is available. Meeting these criteria produces a conceptual model for Phase 3 integration design. Gaps require refinement or reconsideration of earlier framework steps.

4.3 Integration design

Phase 3 addresses how the selected methods will exchange information and coordinate their execution. This paper references the hybrid SD-ABM model presented by Nguyen et al. [55]. For each method (SD, DES, and ABM), key elements are identified and summarized, acting as interaction points between different methods. For SD, the key elements are dynamic variables (stocks and flows that accumulate or change over time through feedback) and auxiliary variables (intermediaries that establish relationships without accumulating). In DES, the primary elements are blocks, whose characteristics can influence and be influenced by other methods, and transitions, which are either deterministic (leading to a single block) or stochastic (providing probabilistic pathways to multiple blocks). For ABM, the essential elements include individual agent behavior (rules and decisions guiding actions), aggregate measures (collective outcomes from agent interactions), and spatial distribution (agent positioning and movement within space).

After designing the information exchange elements, the next step is synchronization of time advancement. The summarized categories are presented. The first category is sequential switching, where dynamic switching occurs along a single timeline. In this case, only one method is active at a time, and a new method can only begin once the previous one has ended. For other hybrid simulation models, the timelines are always parallel. In synchronous fixed time steps category, the interchanges of information happen at the same time with equal time intervals. While in synchronous variable time steps category, the data exchanges occur simultaneously but at varying intervals. Asynchronous fixed time steps and asynchronous variable time steps follow a similar pattern to synchronous ones, but in an asynchronous way. Finally, the data exchange can combine both fixed and variable time steps, meaning that whenever a model is updated, it immediately sends information to the other model. Phase 3 results in a computational model ready for implementation.

4.4 Implementation & validation

Phase 4 transforms this design into an executable simulation through systematic implementation and

validation. Implementation begins with choosing between single platforms supporting all methods (simplified integration) or multiple specialized platforms (greater flexibility but complex bridging) [3]. Development follows an incremental approach: build method components independently, verify individual functionality, then implement integration mechanisms. Hybrid model validation addresses three levels. Component validation tests each method against benchmarks [36]. Interface validation ensures data exchanges maintain integrity through transformations and synchronization [44]. Integrated validation examines overall model behavior against historical construction data, revealing emergent properties from method interactions [40].

Accuracy metrics include mean absolute percentage error (MAPE) and root mean square deviation (RMSE), and R-squared for overall fit [17]. Acceptable thresholds depend on application. Sensitivity analysis systematically varies key parameters to identify critical inputs and assess model stability [51]. Well-constructed models demonstrate logical sensitivity to critical parameters while maintaining robustness against reasonable variations in less certain inputs. A final optimized model is obtained once it passes both the accuracy and sensitivity tests.

4.5 Case study of construction equipment repair yard inventory management

To demonstrate the practical applicability of the proposed framework, a hybrid simulation case study involving repair yard inventory management for vibroflotation equipment (hereafter referred to as vibros) in foundation construction is mapped onto the four phases, as illustrated in Figure 4.

In Phase 1, the problem concerns managing spare parts inventory for a fleet of vibros, each comprising interdependent components. Component failures cause equipment downtime and costly emergency procurement, with productivity and cost efficiency as the primary performance focus. The problem spans all three dimensions: continuous dynamics in inventory stock and flows; discrete events in the vibro operational process; and behavioral characteristics from individual vibros and their components, each with unique degradation patterns and interdependencies. A hybrid simulation approach is therefore warranted.

In Phase 2, the DES component serves as the main operational workflow, modelling vibro assignment, spare parts demand prediction, normal ordering, and monitoring. During monitoring, preventive maintenance leads the vibro directly to the repair yard, while corrective maintenance first evaluates whether emergency ordering is necessary. ABM represents individual vibros (Vibro 1 to Vibro n) as agents with status parameters (working, under maintenance, or idle), and each component within a vibro is also modelled as an agent. A component correlation matrix captures part interdependencies and feeds into the DES demand prediction module. SD models the ordering strategy through a stock-flow structure (orders into inventory, demands as outflow), governed by parameters including supplier reliability, safety stock, order quantity, lead time, emergency order quantity, and consumption quantity. Optimization objectives include fill rate, turnover ratio, service level, and cost metrics (normal ordering, emergency ordering, holding, and penalty costs).



Figure 4. Hybrid simulation implementation of construction equipment repair yard inventory management

In Phase 3, the interaction points are defined as follows: ABM transmits component correlation matrices and vibro status to the DES demand prediction module; DES updates the SD model through ordering events (order inflow) and repair completion (consumption outflow); and SD inventory levels and safety stock status influence whether normal or emergency ordering is activated in DES. Historical maintenance and repair records initialize the ABM agents, and a feedback mechanism from the DES "Send to repair yard" block updates these records, enabling dynamic calibration as the simulation progresses. The DES component operates on an event-driven basis while SD advances continuously, with information exchanged at each relevant state change.

In Phase 4, the model will be built in Python as a single-platform approach to simplify integration across the three methods. Development follows an incremental strategy, building and verifying each component

independently before implementing integration mechanisms. Validation will address three levels: component, interface, and integrated, with accuracy assessed using metrics such as MAPE, RMSE, and R-squared against historical construction data. Sensitivity analysis will vary key parameters to identify critical inputs and assess model stability. The validation of this case study is currently ongoing, as data collection from the repair yard is still in progress.

5 Conclusion and future work

This study addresses critical gaps in construction hybrid simulation research: the absence of comprehensive reviews and lack of frameworks guiding SD, DES, and ABM integration. Following PRISMA guidelines, bibliometric analysis of 49 papers revealed declining research from 2019-2021, reversing upward from 2022, with publications spanning interdisciplinary venues and forming 6 distinct authorship clusters.

The developed four-phase framework revealed that SD-DES combinations dominate current research, with focus concentrated on active construction phases. The framework's modular structure enables continuous adaptation as technologies emerge. An illustrative case study on repair yard inventory management for vibroflotation equipment has been presented to demonstrate the framework's applicability, and future work will focus on completing its validation as field data collection progresses, further demonstrating practical value in real-world construction applications.

References

[1] A. Khodabandelu and J. Park, "Agent-based modeling and simulation in construction," *Autom. Constr.*, vol. 131, p. 103882, Nov. 2021, doi: 10.1016/j.autcon.2021.103882.

[2] O. Bokor, L. Florez, A. Osborne, and B. J. Gledson, "Overview of construction simulation approaches to model construction processes," *Organ. Technol. Manag. Constr. Int. J.*, vol. 11, no. 1, pp. 1853–1861, Mar. 2019, doi: 10.2478/otmcj-2018-0018.

[3] S. C. Brailsford, T. Eldabi, M. Kunc, N. Mustafee, and A. F. Osorio, "Hybrid simulation modelling in operational research: A state-of-the-art review," *Eur. J. Oper. Res.*, vol. 278, no. 3, pp. 721–737, Nov. 2019, doi: 10.1016/j.ejor.2018.10.025.

[4] L. Xu, D. Veeramani, and Z. Zhu, "Automated physics-based modeling of construction equipment through data fusion," *Autom. Constr.*, vol. 168, p. 105880, Dec. 2024, doi: 10.1016/j.autcon.2024.105880.

[5] S. Backman, D. Lindmark, K. Bodin, M. Servin, J. Mörk, and H. Löfgren, "Continuous Control of an Underground Loader Using Deep Reinforcement Learning," *Machines*, vol. 9, no. 10, p. 216, Sep. 2021, doi: 10.3390/machines9100216.

[6] J. S. Morgan, S. Howick, and V. Belton, "A toolkit of designs for mixing Discrete Event Simulation and System Dynamics," *Eur. J. Oper. Res.*, vol. 257, no. 3, pp. 907–918, Mar. 2017, doi: 10.1016/j.ejor.2016.08.016.

[7] G. Wallentin and C. Neuwirth, "Dynamic hybrid modelling: Switching between AB and SD designs of a predator-prey model," *Ecol. Model.*, vol. 345, pp. 165–175, Feb. 2017, doi: 10.1016/j.ecolmodel.2016.11.007.

[8] S. Moradi, F. Nasirzadeh, and F. Golkhoo, "A hybrid SD–DES simulation approach to model construction projects," *Constr. Innov.*, vol. 15, no. 1, pp. 66–83, Jan. 2015, doi: 10.1108/CI-10-2013-0045.

[9] M. Fakhimi, N. Mustafee, and T. Eldabi, "Hybrid Simulation in Construction," in *2023 Winter Simulation Conference (WSC)*, San Antonio, TX, USA: IEEE, Dec. 2023, pp. 1397–1408. doi: 10.1109/WSC60868.2023.10407401.

[10] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[11] N. Kedir, N. Siraj, and A. R. Fayek, "Application of System Dynamics in Construction Engineering and Management: Content Analysis and Systematic Literature Review," *Adv. Civ. Eng.*, vol. 2023, no. 1, p. 1058063, 2023, doi: 10.1155/2023/1058063.

[12] V. Zamani, E. Yavari, and H. Taghaddos, "A science mapping lens on discrete event simulation applications in construction engineering and management," *Autom. Constr.*, vol. 166, p. 105625, Oct. 2024, doi: 10.1016/j.autcon.2024.105625.

[13] O. Bokor, L. Florez, B. Gledson, and A. Osborne, "Input for Hybrid Simulation Modelling Construction Operations," in *Proceedings of the Creative Construction Conference 2019*, Budapest University of Technology and Economics, 2019, pp. 571–577. doi: 10.3311/CCC2019-078.

[14] S. RazaviAlavi and S. AbouRizk, "A hybrid simulation approach for quantitatively analyzing the impact of facility size on construction projects," *Autom. Constr.*, vol. 60, pp. 39–48, Dec. 2015, doi: 10.1016/j.autcon.2015.09.006.

[15] H. Jo, H. Lee, Y. Suh, J. Kim, and Y. Park, "A dynamic feasibility analysis of public investment projects: An integrated approach using system dynamics and agent-based modeling," *Int. J. Proj. Manag.*, vol. 33, no. 8, pp. 1863–1876, Nov. 2015, doi: 10.1016/j.ijproman.2015.07.002.

[16] F. Peña-Mora, S. Han, S. Lee, and M. Park, "Strategic-Operational Construction Management:

Hybrid System Dynamics and Discrete Event Approach," *J. Constr. Eng. Manag.*, vol. 134, no. 9, pp. 701–710, Sep. 2008, doi: 10.1061/(ASCE)0733-9364(2008)134:9(701).

[17] X. Xu, J. Wang, C. Z. Li, W. Huang, and N. Xia, "Schedule risk analysis of infrastructure projects: A hybrid dynamic approach," *Autom. Constr.*, vol. 95, pp. 20–34, Nov. 2018, doi: 10.1016/j.autcon.2018.07.026.

[18] C. Z. Li, J. Hong, C. Fan, X. Xu, and G. Q. Shen, "Schedule delay analysis of prefabricated housing production: A hybrid dynamic approach," *J. Clean. Prod.*, vol. 195, pp. 1533–1545, Sep. 2018, doi: 10.1016/j.jclepro.2017.09.066.

[19] S. Moradi, F. Nasirzadeh, and F. Golkhoo, "Modeling Labor Productivity in Construction Projects using Hybrid SD-DES Approach," *Sci. Iran.*, vol. 24, no. 6, pp. 2752–2761, Aug. 2017, doi: 10.24200/sci.2017.4171.

[20] A. Alvanchi, S. Lee, and S. AbouRizk, "Modeling Framework and Architecture of Hybrid System Dynamics and Discrete Event Simulation for Construction," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 26, no. 2, pp. 77–91, Feb. 2011, doi: 10.1111/j.1467-8667.2010.00650.x.

[21] N. Gerami Seresht and A. R. Fayek, "Modeling earthmoving operations in real time using hybrid fuzzy simulation," *Can. J. Civ. Eng.*, vol. 49, no. 4, pp. 627–635, Apr. 2022, doi: 10.1139/cjce-2020-0790.

[22] S. Lee, S. Han, and F. Peña-Mora, "Integrating Construction Operation and Context in Large-Scale Construction Using Hybrid Computer Simulation," *J. Comput. Civ. Eng.*, vol. 23, no. 2, pp. 75–83, Apr. 2009, doi: 10.1061/(ASCE)0887-3801(2009)23:2(75).

[23] T. T. Dang, B. Schoesser, M. Thewes, and M. Koenig, "Evaluation of productivities influenced by disturbances and different soil compositions in microtunnelling using process simulation," *Tunn. Undergr. Space Technol.*, vol. 76, pp. 10–20, Jun. 2018, doi: 10.1016/j.tust.2018.03.002.

[24] A. Alvanchi, S. Lee, and S. AbouRizk, "Dynamics of Working Hours in Construction," *J. Constr. Eng. Manag.*, vol. 138, no. 1, pp. 66–77, Jan. 2012, doi: 10.1061/(ASCE)CO.1943-7862.0000384.

[25] A. Alvanchi, S. Lee, and S. AbouRizk, "Dynamics of workforce skill evolution in construction projects," *Can. J. Civ. Eng.*, vol. 39, no. 9, pp. 1005–1017, Sep. 2012, doi: 10.1139/l2012-087.

[26] H. Alzraiee, T. Zayed, and O. Moselhi, "Dynamic planning of construction activities using hybrid simulation," *Autom. Constr.*, vol. 49, pp. 176–192, Jan. 2015, doi: 10.1016/j.autcon.2014.08.011.

[27] S. Abdelmageed, S. Abdelkhalek, M. Hussien, and T. Zayed, "A hybrid simulation model for modules installation in modular integrated construction projects," *Int. J. Constr. Manag.*, vol. 24, no. 13, pp. 1407–1418, Oct. 2024, doi: 10.1080/15623599.2023.2219939.

[28] H. Alzraiee, O. Moselhi, and T. Zayed, "A Hybrid Framework for Modeling Construction Operations Using Discrete Event Simulation and System Dynamics," pp. 1063–1073, Jul. 2012, doi: 10.1061/9780784412329.107.

[29] X. Shi, K. Fu, Y. Qiao, J. Santos, and Z. Yuan, "Accident cause analysis for lifting prefabricated components of assembled buildings based on hybrid simulation technology from human factor perspective," *Eng. Constr. Archit. Manag.*, Nov. 2024, doi: 10.1108/ECAM-04-2024-0431.

[30] H. Taghaddos, M. H. Heydari, and A. Asgari, "A hybrid simulation approach for site layout planning in construction projects," *Constr. Innov.*, vol. 21, no. 3, pp. 417–440, 2021, doi: 10.1108/CI-05-2020-0069.

[31] S. Moharrami, M. Taghaddos, S. RazaviAlavi, and S. AbouRizk, "A hybrid simulation approach for microtunneling construction planning," *Constr. Innov.*, vol. 21, no. 3, pp. 363–378, Jul. 2021, doi: 10.1108/CI-05-2020-0068.

[32] Y. M. Goh and M. J. Askar Ali, "A hybrid simulation approach for integrating safety behavior into construction planning: An earthmoving case study," *Accid. Anal. Prev.*, vol. 93, pp. 310–318, Aug. 2016, doi: 10.1016/j.aap.2015.09.015.

[33] M. Khanzadi, F. Nasirzadeh, M. Mir, and P. Nojedehi, "Prediction and improvement of labor productivity using hybrid system dynamics and agent-based modeling approach," *Constr. Innov.*, vol. 18, no. 1, pp. 2–19, 2018, doi: 10.1108/CI-06-2015-0034.

[34] F. Nasirzadeh, M. Khanzadi, and M. Mir, "A hybrid simulation framework for modelling construction projects using agent-based modelling and system dynamics: an application to model construction workers' safety behavior," *Int. J. Constr. Manag.*, vol. 18, no. 2, pp. 132–143, Mar. 2018, doi: 10.1080/15623599.2017.1285485.

[35] H. Liang, K.-Y. Lin, and S. Zhang, "Understanding the Social Contagion Effect of Safety Violations within a Construction Crew: A Hybrid Approach Using System Dynamics and Agent-Based Modeling," *Int. J. Environ. Res. Public. Health*, vol. 15, no. 12, Art. no. 12, Dec. 2018, doi: 10.3390/ijerph15122696.

[36] C. Wu, C. Chen, R. Jiang, P. Wu, B. Xu, and J. Wang, "Understanding laborers' behavioral diversities in multinational construction projects

using integrated simulation approach," *Eng. Constr. Archit. Manag.*, vol. 26, no. 9, pp. 2120–2146, 2019, doi: 10.1108/ECAM-07-2018-0281.

[37] Q. Fang, X. Chen, D. Castro-Lacouture, and C. Li, "Intervention and management of construction workers' unsafe behavior: A simulation digital twin model," *Adv. Eng. Inform.*, vol. 58, p. 102182, Oct. 2023, doi: 10.1016/j.aei.2023.102182.

[38] A. Conrads, M. Scheffer, H. Mattern, M. König, and M. Thewes, "Assessing maintenance strategies for cutting tool replacements in mechanized tunneling using process simulation," *J. Simul.*, vol. 11, no. 1, pp. 51–61, Feb. 2017, doi: 10.1057/s41273-016-0046-5.

[39] M. Jung, M. Park, H.-S. Lee, and S. Chi, "Multimethod Supply Chain Simulation Model for High-Rise Building Construction Projects," *J. Comput. Civ. Eng.*, vol. 32, no. 3, p. 04018007, May 2018, doi: 10.1061/(asce)cp.1943-5487.0000751.

[40] S. Xiang, M. Arashpour, and R. Wakefield, "Hybrid Simulation Modeling of Hoist Down-Peak Operations in Construction Sites," presented at the 33th International Symposium on Automation and Robotics in Construction, Auburn, AL, USA, Jul. 2016. doi: 10.22260/ISARC2016/0020.

[41] B. Matejevic, M. Zlatanovic, and D. Cvetkovic, "The Simulation Model for Predicting the Productivity of the Reinforced Concrete Slabs Concreting Process," *Teh. Vjesn.-Tech. Gaz.*, vol. 25, no. 6, pp. 1672–1679, Dec. 2018, doi: 10.17559/TV-20170627195003.

[42] Y. Wei, Z. Lei, and M. S. Altaf, "Simulation-based comparison of push- and pull-based planning in panelized construction," *Autom. Constr.*, vol. 158, p. 105228, Feb. 2024, doi: 10.1016/j.autcon.2023.105228.

[43] A. Attajer and B. Mecheri, "Multi-Agent Simulation Approach for Modular Integrated Construction Supply Chain," *Appl. Sci.*, vol. 14, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/app14125286.

[44] S. Hwang, M. Park, H.-S. Lee, and S. Lee, "Hybrid Simulation Framework for Immediate Facility Restoration Planning after a Catastrophic Disaster," *J. Constr. Eng. Manag.*, vol. 142, no. 8, Aug. 2016, doi: 10.1061/(ASCE)CO.1943-7862.0001146.

[45] S. Abdelkhalek and T. Zayed, "Simulation-based planning of concrete bridge deck inspection with non-destructive technologies," *Autom. Constr.*, vol. 119, p. 103337, Nov. 2020, doi: 10.1016/j.autcon.2020.103337.

[46] B. Zhang, W. K. Chan, and S. V. Ukkusuri, "Agent-based discrete-event hybrid space modeling approach for transportation evacuation simulation," in *Proceedings of the 2011 Winter Simulation Conference (WSC)*, Phoenix, AZ, USA: IEEE, Dec. 2011, pp. 199–209. doi: 10.1109/WSC.2011.6147751.

[47] B. Zhang, W. K. V. Chan, and S. V. Ukkusuri, "On the modelling of transportation evacuation: an agent-based discrete-event hybrid-space approach," *J. Simul.*, vol. 8, no. 4, pp. 259–270, Nov. 2014, doi: 10.1057/jos.2014.3.

[48] L. Rabelo, H. Eskandari, T. Shaalan, and M. Helal, "Value chain analysis using hybrid simulation and AHP," *Int. J. Prod. Econ.*, vol. 105, no. 2, pp. 536–547, Feb. 2007, doi: 10.1016/j.ijpe.2006.05.011.

[49] H. Alzraiee, T. Zayed, and O. Moselhi, "Methodology for synchronizing Discrete Event Simulation and System Dynamics models," in *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*, Berlin, Germany: IEEE, Dec. 2012, pp. 1–11. doi: 10.1109/WSC.2012.6464997.

[50] Z. Ding, W. Gong, S. Li, and Z. Wu, "System Dynamics versus Agent-Based Modeling: A Review of Complexity Simulation in Construction Waste Management," *Sustainability*, vol. 10, no. 7, 2018, doi: 10.3390/su10072484.

[51] A. Mostafavi, D. Abraham, and D. DeLaurentis, "Ex-Ante Policy Analysis in Civil Infrastructure Systems," *J. Comput. Civ. Eng.*, vol. 28, no. 5, p. A4014006, Sep. 2014, doi: 10.1061/(ASCE)CP.1943-5487.0000350.

[52] D. Páez-Pérez and M. Sánchez-Silva, "A dynamic principal-agent framework for modeling the performance of infrastructure," *Eur. J. Oper. Res.*, vol. 254, no. 2, pp. 576–594, Oct. 2016, doi: 10.1016/j.ejor.2016.03.027.

[53] M. Hussein, A. Karam, A. E. E. Eltoukhy, A. Darko, and T. Zayed, "Optimized multimodal logistics planning of modular integrated construction using hybrid multi-agent and metamodeling," *Autom. Constr.*, vol. 145, p. 104637, Jan. 2023, doi: 10.1016/j.autcon.2022.104637.

[54] M. Assaf, S. Assaf, W. Correa, R. Lemouchi, and Y. Mohamed, "A Hybrid Simulation-Based Optimization Framework For Managing Modular Bridge Construction Projects: A Cable-Stayed Bridge Case Study," in *2023 Winter Simulation Conference (WSC)*, Dec. 2023, pp. 3094–3105. doi: 10.1109/WSC60868.2023.10407539.

[55] L. K. N. Nguyen, S. Howick, and I. Megiddo, "A framework for conceptualising hybrid system dynamics and agent-based simulation models," *Eur. J. Oper. Res.*, vol. 315, no. 3, pp. 1153–1166, Jun. 2024, doi: 10.1016/j.ejor.2024.01.027.

A VLM-Driven Workflow for Generating Context-Specific Fire Hazards for VR Construction Safety Training

Kexin Liu[1], Mohamed Sabek[1], Gaang Lee[1], Max Kinateder[2], Vicente A. Gonzalez[3]

[1]Infrastructure Human Tech (IHT) Lab, Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, Alberta, Canada
[2]Construction Research Centre, National Research Council Canada, Ottawa, Ontario, Canada
[3]Infrastructure and Human Tech (IHT) Lab, Strategic Projects Insight Centre in Engineering (SPICE), Department of Civil and Environmental Engineering, University of Alberta, Canada
kexin10@ualberta.ca, sabek@ualberta.ca, gaang@ualberta.ca, Max.Kinateder@nrc-cnrc.gc.ca, vagonzal@ualberta.ca

Abstract –

While Virtual Reality (VR) provides a promising platform for construction safety training, generating virtual environments (VEs) that are tailored to specific project and task contexts remains a bottleneck. Current workflows, ranging from manual modeling to semi-automated BIM-to-VR pipelines, are often labor-intensive and result in static, idealized VEs that fail to capture the dynamic reality of active construction sites. This paper addresses these limitations by proposing a cyclical framework driven by Vision-Language Models (VLMs), designed to both generate procedural fire hazard scenarios and systematically capture expert knowledge for continuous model improvement. By focusing on fire safety, a domain requiring a deep understanding of spatial relationships between objects and activities, the workflow integrates VLMs to extract tacit risk context that traditional object detection misses. Human-in-the-loop corrections then serve as ground-truth data for future model fine-tuning. As an initial proof of concept, we developed a semi-automated Unity prototype that connects unstructured site imagery with 3D asset generation through a hybrid generative pipeline. The prototype illustrates the potential of the proposed workflow to transform raw construction imagery into privacy-preserving, context-specific digital twin training environments. By reducing the manual effort required for VR content creation, this approach lays the foundation for scalable, self-evolving AI-driven safety training systems.

Keywords –

Construction safety, Fire hazards, Virtual Reality, Vision-Language Models

1 Introduction

Construction fire safety remains a persistent challenge, driven by the complex, dynamic nature of job sites and the interactions among workers, equipment, and changing site conditions [1]. VR is a promising solution by providing controlled environments for behavioural research and immersive platforms for safety training [2]. However, the creation and maintenance of VEs for training remains a major bottleneck, limiting both research validity and the scalability of VR-based training systems in practice [3].

This bottleneck is driven by two intersecting challenges. First, high-quality VR training scenarios still rely heavily on the tacit knowledge from senior safety personnel [4]. Existing literature, standards, and reported incidents provide important general principles, but they rarely describe hazards at the level of detail needed to build convincing, site- and task-specific VR scenarios. Additionally, each construction project is unique, requiring flexible, context-specific judgement and decision-making. As a result, expert input remains essential at two crucial stages: (1) to accurately identify complex, context-dependent hazards from raw site data [5]; and (2) to validate the realism of the resulting virtual simulation [4]. Despite this importance, this tacit knowledge of experts is rarely formalized or reused in a structured way [6]. At the same time, the construction industry faces a worldwide aging workforce and accelerating labor shortages [7-9]. As experienced personnel retire, this specialized domain expertise is at risk of disappearing. Without a mechanism to capture and digitize this visual reasoning, the industry risks losing the capacity to generate and verify realistic safety scenarios.

Second, even with expert knowledge available, the technical workflows for turning this hazard-relevant knowledge into VR content remain inefficient. Current

digital twins incrementally update elements or zones based on scans, sensor data, or manual inputs, which is effective for progress monitoring and quality control [10]. However, for safety training, it is still time-consuming and labour-intensive to propagate these updates into VR environments [11]. Most importantly, within the large volume of new data, only a subset of the updated information is needed for safety training. What matters are fire hazard–relevant changes, such as new ignition sources or flammable materials introduced by new construction activities, rather than a perfect geometric copy of every site detail. This suggests an alternative strategy in which, instead of fully twinning the whole site, we focus on twinning hazard-relevant information while still reflecting the local context where workers operate.

Building on these challenges, this paper presents a semi-automated, VLM-driven workflow that twins hazard-specific information into VR while capturing and storing expert knowledge. As a proof-of-concept, we developed an initial prototype implementation within the popular game engine Unity to illustrate the practical feasibility of the proposed framework.

2 Related Work

2.1 Construction VR Scene Generation

VR is increasingly used in the architecture, engineering, and construction (AEC) sector for design reviews, project planning, and safety training, providing more transparent communication and improved learning outcomes [12]. However, achieving high ecological validity in VR scenes, meaning they accurately reflect task-relevant conditions in specific construction environments, remains challenging due to their complexity [3]. These environments contain many objects, changing site conditions, and heterogeneous data sources such as BIM, point clouds, and GIS, all of which require significant cleaning and optimization before becoming VR-ready [13].

Currently, there are two primary workflows to generate virtual environments. Semi-automated BIM-to-VR pipelines are the most common but frequently suffer from loss of BIM semantics during export and require aggressive geometric optimisation to meet the rendering constraints of standalone VR headsets [14]. Manual scene authoring supports highly customized simulations but is slow and requires advanced modelling and game-engine expertise [15]. Crucially, neither workflow includes context-specific hazards, such as oily rags left near heaters or other flammable materials. Standard BIM models focus on permanent, “as-designed” elements [1] and rarely capture short-lived site conditions like waste, temporary storage, or poor housekeeping. In scan- or sensor-based workflows, these transient objects are often filtered out as noise [10]. However, such temporary combustibles and unsafe arrangements are a common source of fire risk on active sites and are therefore essential components of realistic safety training scenarios.

Recent advances in generative AI offer a promising alternative. Text-to-3D and image-to-3D models, such as Hunyuan3D 2.0, Microsoft TRELLIS [16], and Meta SAM 3D [17], can synthesize high-quality 3D objects directly from natural-language descriptions or site imagery, reducing reliance on fragmented asset libraries and bypassing many manual modelling steps. However, despite these advances, the AEC domain still lacks a practical workflow that applies generative models to VR scene generation, especially for safety-critical contexts such as hazard recognition research and training.

2.2 Potential of VLMs for Construction Hazard Recognition Training

The integration of advanced computational models into construction safety management is an active and expanding area of research. Traditional computer vision (CV) techniques, often relying on convolutional neural networks (CNNs), have demonstrated success in object detection tasks, such as identifying personnel, machinery, and personal protective equipment [18]. However, these methods often struggle to interpret context and relationships among objects, which is paramount for safety analysis. A ladder, for example, is not inherently a hazard; it becomes one when it is unsecured, placed on an unstable surface, or used near an open electrical panel.

This is the gap that VLMs are beginning to fill. VLMs such as LLaVA (Large Language and Vision Assistant) [19] and GEMMA 3 [20], and those based on the GPT-4 architecture [21], combine the perceptual capabilities of CV with the sophisticated reasoning of large language models [22]. This multimodal approach allows VLMs to move beyond mere object identification to complex scene contextual processing and complex pattern recognition. Recent studies have explored using VLMs to analyze static site images for safety-rule violations by processing both the visual data and textual prompts containing regulatory guidelines [18,23]. This allows them to identify not only the presence of a potential hazard, but the reason it violates a safety protocol.

Despite these advances, current studies have largely focused on demonstrating the hazard detection capabilities of VLMs [22], while the translation of VLM outputs into reusable, context-rich VR training assets remains underexplored. In particular, limited work has examined how contextual information embedded in construction hazard scenes can be structured into a format that can be directly interpreted and rendered in a virtual environment. Additionally, the use of VLMs in safety-critical settings requires caution. VLMs may generate plausible but incorrect hazard descriptions or

safety explanations, commonly referred to as "hallucinations." They may also struggle with fine-grained spatial reasoning, such as estimating 3D proximity or depth from 2D imagery [27]. Therefore, VLM-generated outputs may require further structuring and validation by human experts before being integrated into VR-based safety training environments.

3 Research Method

We propose a five-stage, human-in-the-loop workflow designed to transform unstructured construction site imagery into structured, context-aware fire hazard libraries for VR. As illustrated in Figure 1, the framework operates iteratively. It not only generates hazard scenarios but also uses expert validation to continuously fine-tune the underlying VLM. This helps the system adapt to specific project constraints and reduces the rate of hallucinations that can contribute to unnecessary information or misleading hazard detections over time. The following sections detail each stage.

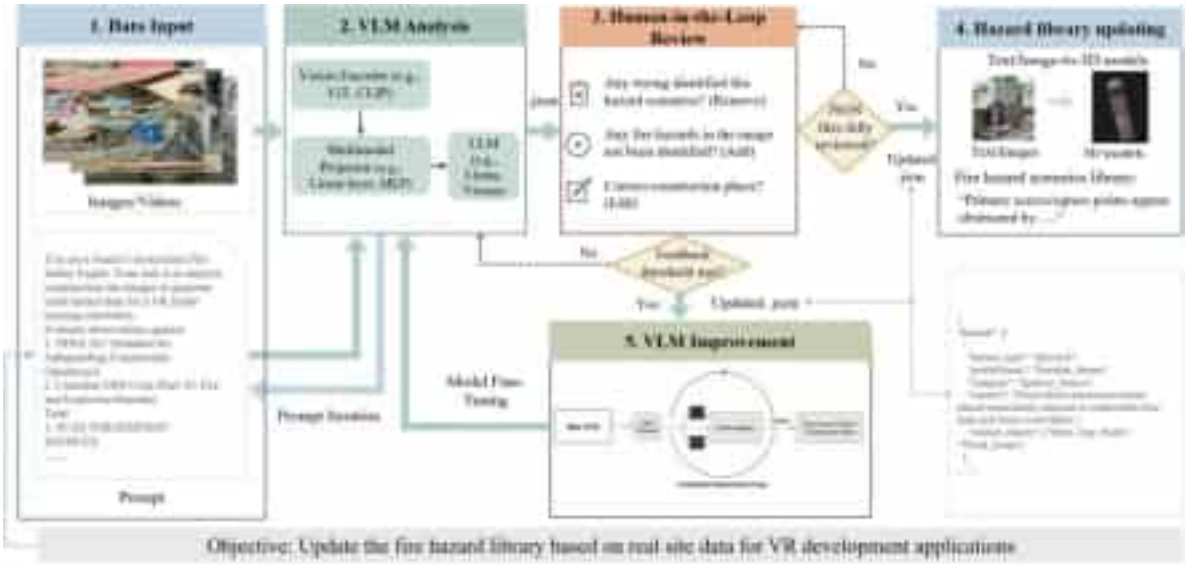


Figure 1. The proposed VLM-driven workflow for virtual fire hazard library update

3.1 Data Input

The workflow begins with the ingestion of raw visual data (images or video frames) captured from active construction sites. To ensure practical scalability and minimize additional hardware costs, the workflow can integrate with existing site CCTV surveillance systems.

To accurately translate these 2D images into valid 3D VR components, we employed a goal-oriented prompt engineering strategy [24]. The prompt design was reverse engineered from the specific data requirements of the VR Scene Library. Specifically, the system requires not only the identification of physical objects (e.g., oil drums or welding equipment) and their spatial context, but also the recognition of abstract violations. Abstract hazards, such as blocked emergency exits or missing fire extinguishers, are critical for safety training but lack a single physical object to detect. Therefore, the prompt is structured to extract both explicit physical assets and implicit procedural non-compliance.

After defining the expected output, we implemented a multi-stage prompting architecture to achieve reliable extraction. First, we applied Role-Prompting by instructing the model to adopt the persona of a "Senior Construction Fire Safety Expert"[25]. This constrains the model's focus to regulatory compliance rather than generic image description. Second, we used Chain-of-Thought (CoT) [26] to ensure the model follows a logical checklist and to mitigate hallucinations. This step-by-step reasoning is critical for visual analysis [22], as it prevents the model from guessing and ensures that every identified hazard is based on clear visual evidence. Finally, we required the output in a strict data format (JSON). This structured format ensures the data is machine-readable, allowing for direct parsing into the user interface and facilitating the automated construction of fine-tuning datasets.

3.2 VLM-based Hazard Analysis

In this stage, the VLM processes the visual input alongside the prompt to generate a structured description of potential hazards. The model identifies specific hazard instances and extracts critical metadata required for VR instantiation. This includes:

- Hazard classification, distinguishing between

physical hazards that must be represented as 3D entities in the virtual environment (e.g., heaters and wood debris) and abstract hazards that represent procedural or regulatory violations (e.g., blocked egress routes, improper material storage).

- Prefab associations, where each physical hazard is mapped to a corresponding VR asset ("prefab") to enable automatic placement within the simulation.
- Contextual metadata, which captures the hazard's relevance, associated risks, and the environmental cues driving the classification (e.g., proximity of a heater to combustible traps, presence of fuel containers near ignition sources).
- Relational linkages, documenting interactions between hazards, such as ignition sources located within NFPA 241's minimum clearance distance from Class A or B fuels, to support scenario logic and rule-based triggers in the simulation.
- Compliance attributes, representing inferred violations of NFPA 241 or the Canadian OHS Code (Part 10), enabling automated generation of safety feedback and assessment scoring.

3.3 Human-in-the-Loop Review

Given the safety-critical nature of this domain, relying on VLM output alone is insufficient due to the potential for semantic errors. We therefore introduce a Human-in-the-Loop review process [27]. A domain expert reviews the JSON output against the original image to verify the presence and classification of the hazard. The expert can correct the output by removing false positives or amending inaccurate context descriptions and can augment it by adding hazards that the VLM has missed. The review is iterative, and the JSON file is edited until the expert confirms that all images in the batch have their hazards fully marked and correctly represented. Only cases where the expert modifies the VLM output are counted as feedback and stored in a feedback repository. The final, validated JSON is stored in an updated JSON file and used both to refine prompts, hazard templates, and configuration rules for the VLM and to populate a feedback repository. VLM updates are triggered once the number of such corrections exceeds a predefined feedback threshold [27]. The specific threshold is configurable depending on project requirements and update cadence; for example, it could correspond to the completion of a scheduled image batch, such as a weekly data collection cycle. In this way, this process acts as a form of "human coaching", helping future hazard descriptions and scene suggestions better align with expert judgement.

3.4 Hazard Library Update

The validated data from Section 3.3 is used to support two parallel processes: generating 3D models (corresponding to the physical field in the VLM output) and populating the Abstract Scene Library (corresponding to the abstract field). This library serves as the semantic back-end for the VR environment. This eliminates the need to repeatedly search for site photographs, locate appropriate 3D objects, or request additional expert validation, thereby streamlining the construction of accurate, context-aware VR fire-safety training scenes.

3.5 Continuous VLM Improvement

A critical component of the proposed workflow is the establishment of a "data flywheel" (Stage 5), which ensures the VLM's accuracy and domain specificity improve over time. This stage operationalizes the feedback gathered during the Human-in-the-Loop Review (Stage 3). The "Updated.json" file generated by the human expert is not only the finalized data for the VR library but also serves as a high-quality, expert-validated ground truth dataset. This curated dataset becomes the training data for iteratively refining the VLM. By comparing the VLM's initial JSON output with the expert-corrected JSON, we can identify systematic errors in the model's reasoning, such as misclassifying construction phases, overlooking "abstract" hazards, or providing incomplete contextual descriptions.

As depicted in the workflow (Figure 1), this refinement process can be efficiently implemented using parameter-efficient fine-tuning (PEFT) techniques, such as Low-Rank Adaptation (LoRA). LoRA allows fine-tuning the VLM on this new, domain-specific dataset by updating only a small subset of the model's parameters. This approach is computationally much more efficient than fully retraining the entire model, making it feasible to run periodic improvements as more curated data is collected. The ultimate goal of this continuous improvement loop is to enhance VLM's performance in construction-specific hazard analysis, thereby reducing the manual correction burden on human experts in future iterations and increasing overall pipeline automation.

4 Implementation

This section presents a proof-of-concept prototype developed as the primary artifact to illustrate the proposed workflow concept. Accordingly, systematic parameterization, full module integration, and empirical evaluation across a larger image corpus will be addressed in future work. Unity (2022.3.46f1) was selected for its mature VR development capabilities and its ability to integrate external AI modules through standard APIs. Although the workflow can be implemented in other engines, Unity provides a stable and practical environment for rapid prototyping. Within Unity, we

implemented a custom user interface (UI) that facilitates user interaction, displays intermediate results, and coordinates communication with external system components. As shown in Figure 2, Unity serves as the central layer, linking data processing, model inference, and VR scene assembly into a cohesive pipeline.

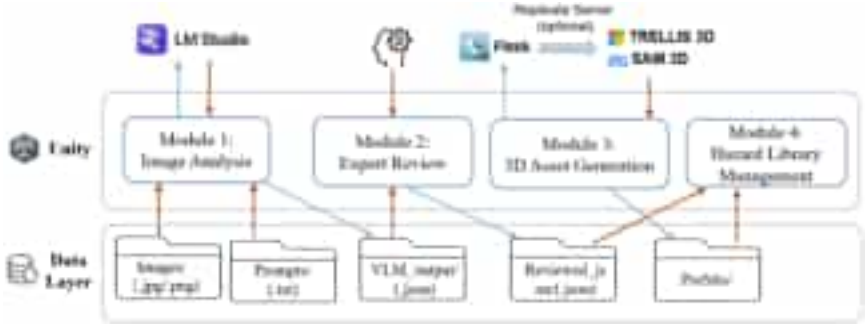


Figure 2. Schematic overview of the Unity plug-in implementing the proposed workflow

4.1 Image Analysis Module

The Image Analysis Module serves as the primary entry point for the fire hazard analysis and hazard library generation workflow. The implementation is compatible with widely used local LLM servers, LM Studio (www.lmstudio.ai/). We selected LM Studio to facilitate the efficient local deployment of quantized models via a standardized REST API, minimizing integration complexity. This enables the use of various open-source VLMs, such as LLaVA or GEMMA 3, which are well-suited for this type of detailed visual analysis (as shown in Figure 3). It is designed with local VLM backends, ensuring data privacy and control.



Figure 3. Locally hosted VLM models in LM Studio

For a proof-of-concept demonstration, the prototype was tested using a small set of publicly available construction-site images. These images were used for feasibility illustration rather than as a formal evaluation dataset. As shown in Figure 4, in the configuration tab, the user selects the VLM model, target image, and domain-specific prompt (from Section 3.1) from the local file system. In our prototype, it is put under the Unity project asset file for easy management. Upon clicking "Analyze", the plug-in converts the image to a base64 string and serializes it into a single API request. This request is sent via an HTTP POST request to the local VLM's API endpoint (e.g., LM Studio's default server at http://127.0.0.1:1234). The plug-in receives the raw JSON text response from the VLM and displays it in the "Results Preview" field. This JSON data is then passed to the "Expert Review" module (Section 4.2), which deserializes the JSON and uses its contents to populate the interactive fields for human-in-the-loop curation. This setup provides a self-contained, integrated toolset for a VR developer to manage the entire data-to-asset pipeline without leaving the Unity editor.



Figure 4. Image Analysis Module UI

4.2 Expert Review Module

This module streamlines the Human-in-the-Loop review stage described in Section 3.3 by providing an interface for experts to inspect and modify all identified fire-hazard scenarios. As shown in Figure 5, it starts with loading the JSON file generated by the Image Analysis module. It presents the extracted fields, such as prefab name, category, and related contextual objects, in a structured and readable format. Experts can edit any field by selecting "Edit," which switches the interface into an editable state, allowing them to correct or update the information. If a scenario has been misidentified, the expert can remove it using the "Remove" option. Conversely, if the VLM-based detection missed a relevant fire hazard, the expert can add a new one using the "Add New Hazard" function. After completing their review, experts can confirm the revisions by "Confirm All Changes." The system then generates an updated JSON file as described in Section 3.3, ensuring that all corrections and additions are incorporated into the downstream modules.



Figure 5. Expert Review Module UI

4.3 3D Asset Generation Module

Module 3 employs a hybrid asset generation pipeline. The system first queries the internal Hazard Library (Module 4) for pre-validated prefabs matching the detected hazard scenarios. If no match is found, the system triggers an on-demand generative workflow, as shown in Figure 6.

The generative process supports both text-to-3D and image-to-3D pathways by choosing different models. In this prototype, we tested the image-to-3D route using TRELLIS [16] and SAM-3D [17]. We explored two deployment strategies to integrate these generative models into the Unity environment. In the local deployment configuration, the models are hosted on the workstation (Intel Core i9-13900KF, 64 GB RAM, NVIDIA GeForce RTX 4090 24GB), and Unity communicates with them through a lightweight Python microservice built with Flask (https://flask.palletsprojects.com), which passes Unity's generation requests to the locally running 3D model. This option avoids per-generation costs but requires a high-performance GPU and sufficient memory.

Alternatively, the cloud-based configuration uses Replicate's model-hosting service (https://replicate.com), where Unity sends requests directly to the cloud and receives generated assets without dedicated hardware. This approach is significantly more accessible and can run on a standard laptop, although each generation incurs a small cost (typically $0.02–$0.07 depending on the model and version). Depending on the available computing resources and budget constraints, either configuration can be used to support text-based or image-based 3D asset generation within the pipeline.

This module is not yet fully developed due to the complexity of integrating Image-to-3D models into the plug-in, but this will be completed in future work. To validate the workflow concept, Figure 7 shows the 3D models generated by SAM 3D through raw construction site images.



Figure 6. 3D Asset Generation Module UI



Figure 7. Example of a 3D construction object generation from raw site images

4.4 Library Management Module

This module aims to display all existing prefabs within the Unity project alongside the fire-hazard scenarios generated or revised in the updated JSON file. As shown in Figure 8, it functions as a central repository, allowing users to review previously identified hazards and track how scenarios have evolved over time. By maintaining a structured record of hazards and their associated assets, the module supports efficient development of VR environments that better reflect real site conditions. Due to current limitations in the 3D asset-generation model, this component is not yet fully implemented. Its underlying functions will be further developed and validated in future work.



Figure 8. Library Management Module UI

5 Conclusion and Future Work

This study makes three main contributions. First, it introduces a hazard-focused twinning concept that updates only fire hazard–relevant elements in VR, rather than mirroring the entire construction environment. This provides an alternative to full digital twins for training use cases where task-relevant ecological validity is more important than exhaustive geometric fidelity. Second, it proposes a novel, semi-automated pipeline connecting real construction data, VLM reasoning, and procedural 3D objects generation. By integrating VLM analysis with a human-in-the-loop verification step, the framework not only automates the creation of context-aware fire hazard scenarios but also systematically digitizes the tacit knowledge of senior safety experts. Third, it presents an initial Unity-based prototype that illustrates how the workflow can be implemented in practice and used to generate fire hazard-relevant models from real construction imagery.

Despite these contributions, several limitations remain. The system's performance is sensitive to the quality and perspective of input imagery. Occluded objects or poor lighting can hinder VLM semantic extraction. Additionally, while the framework extracts spatial context from 2D images, it currently lacks precise depth estimation. In some cases, objects may appear closer or farther away than they actually are in the original scene. Furthermore, while the feedback loop for model fine-tuning is theoretically defined, the long-term impact of expert corrections on model performance has not yet been empirically validated in this study. Finally, the seamless integration of different models within Unity still requires additional development efforts, and systematic quantitative evaluation of hazard identification performance (e.g., precision and recall metrics) is beyond the scope of this proof-of-concept study and is identified as a priority for future empirical validation.

Future research includes strengthening the core components of the proposed workflow. First, improving depth reasoning from 2D inputs, whether through lightweight monocular depth models [28] or multi-view image capture, will help reduce spatial ambiguities during VLM interpretation. Additional effort is also needed to formally evaluate the feedback loop by measuring how expert corrections affect model accuracy across successive fine-tuning cycles. Such an assessment would clarify whether the system can reliably accumulate and operationalize tacit safety knowledge over time. Finally, the framework's scope could be broadened beyond fire safety to address other critical challenges, such as fall protection and visual clutter analysis, thereby validating the generalizability of VLM-driven simulation across the construction safety domain.

Acknowledgement

This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) via the Discovery Grant program (Grant No. RGPIN-2023-04239). The authors gratefully acknowledge NSERC funds, which have facilitated the development of this research.

References

[1] S. Zhang, K. Sulankivi, M. Kiviniemi, I. Romo, C. M. Eastman, and J. Teizer. BIM-based fall hazard identification and prevention in construction safety planning. *Safety Science*. 72:31–45, 2015.

[2] S. S. Man, H. Wen, and B. C. L. So. Are virtual reality applications effective for construction safety training and education? A systematic review and meta-analysis. *Journal of Safety Research*. 88:230–243, 2024.

[3] X. Li, W. Yi, H.-L. Chi, X. Wang, and A. P. C. Chan. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in Construction*. 86:150–162, 2018.

[4] V. Getuli, P. Capone, A. Bruttini, and S. Isaac. BIM-based immersive Virtual Reality for construction workspace planning: A safety-oriented approach. *Automation in Construction*. 114:103160, 2020.

[5] M. Namian, A. Albert, and J. Feng. Effect of Distraction on Hazard Recognition and Safety Risk Perception. *Journal of Construction Engineering and Management*. 144(4)2018.

[6] I. S. Abotaleb *et al.* A framework to integrate virtual reality into international standard safety trainings. *Engineering, Construction and Architectural Management*. 32(4):2320–2341, 2025.

[7] Statistics Canada. Labour shortage trends in Canada. Accessed: Jan. 01, 2026. Online: https://www.statcan.gc.ca/en/subjects-start/labour_/labour-shortage-trends-canada

[8] Claire McAnaw Gallagher. The Construction Industry: Characteristics of the Employed, 2003–2020., U.S. Bureau of Labor Statistics. Accessed: Jan. 01, 2026. Online: https://www.bls.gov/spotlight/2022/the-construction-industry-labor-force-2003-to-2020/home.htm

[9] European Labour Authority. Labour shortages and surpluses in Europe 2024. Accessed: Jan. 01, 2026. Online: https://www.ela.europa.eu/en/publications/labour-shortages-and-surpluses-europe-2024

[10] V. V. Tuhaise, J. H. M. Tah, and F. H. Abanda. Technologies for digital twin applications in construction. *Automation in Construction*. 152:104931, 2023.

[11] N. Akindele, R. Taiwo, H. Sarvari, B. I. Oluleye, I. A. Awodele, and T. O. Olaniran. A state-of-the-art analysis of virtual reality applications in construction health and safety. *Results in Engineering*. 23:102382, 2024.

[12] Y. Zhang, H. Liu, S.-C. Kang, and M. Al-Hussein. Virtual reality applications for the built environment: Research trends and opportunities. *Automation in Construction*. 118:103311, 2020.

[13] M. Johansson and M. Roupé. Real-world applications of BIM and immersive VR in construction. *Automation in Construction*. 158:105233, 2024.

[14] S. Alizadehsalehi, A. Hadavi, and J. C. Huang. From BIM to extended reality in AEC industry. *Automation in Construction*. 116:103254, 2020.

[15] V. Getuli, P. Capone, A. Bruttini, and S. Isaac. BIM-based immersive Virtual Reality for construction workspace planning: A safety-oriented approach. *Automation in Construction*. 114:103160, 2020.

[16] J. Xiang *et al.* Structured 3D Latents for Scalable and Versatile 3D Generation. 2025.

[17] SAM 3D Team *et al.* SAM 3D: 3Dfy Anything in Images. 2025.

[18] M. Adil, G. Lee, V. A. Gonzalez, and Q. Mei. Using Vision Language Models for Safety Hazard Identification in Construction., 2025.

[19] H. L. C. W. Q. L. Y. J. Liu. Improved Baselines with Visual Instruction Tuning., in *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, :26296–26306.

[20] Gemma Team *et al.* Gemma 3 Technical Report. 2025. Online: http://arxiv.org/abs/2503.19786

[21] OpenAI *et al.* GPT-4 Technical Report. 2024. Online: http://arxiv.org/abs/2303.08774

[22] X. Chen and Z. Zou. Are Large Pre-trained Vision Language Models Effective Construction Safety Inspectors? 2025. Online: http://arxiv.org/abs/2508.11011

[23] Y. Wang, H. Luo, and W. Fang. An integrated approach for automatic safety inspection in construction: Domain knowledge with multimodal large language model. *Advanced Engineering Informatics*. 652025.

[24] Q. Ma, W. Peng, C. Yang, H. Shen, K. Koedinger, and T. Wu. What Should We Engineer in Prompts? Training Humans in Requirement-Driven LLM Use. *ACM Transactions on Computer-Human Interaction*. 32(4):1–27, 2025.

[25] E. Saravia. Prompt Engineering Guide. *https://github.com/dair-ai/Prompt-Engineering-Guide*. 2022.

[26] J. Wei *et al.* Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. 2023. Online: http://arxiv.org/abs/2201.11903

[27] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*. 56(4):3005–3054, 2023.

[28] D. Meng *et al.* Conditional DETR for Fast Training Convergence. in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, :3631–3640.

An Automatic Update Framework for As-designed Pipeline BIM Model Based on Laser Scanning Point Cloud

Xinru Wang[1], Bin Yang[1] and Tianjia Lu[1]

[1] College of Civil Engineering, Tongji University, China
2430797@tongji.edu.cn, yangbin@tongji.edu.cn, 2454442@tongji.edu.cn

Abstract –

Building Information Modeling (BIM) is widely used in construction management, particularly for mechanical, electrical, and plumbing (MEP) systems, where accurate pipeline information is critical for efficient operation and maintenance. In practice, however, as-built pipelines often deviate from the as-designed BIM models, and manually updating these models is time-consuming. This study proposes an automatic pipeline for BIM updating, which leverages laser-scanned point clouds together with as-designed BIM priors. The framework integrates point cloud semantic segmentation, parameter extraction, and spatial topology analysis to establish component correspondences and generate an updated BIM model. A case study in a complex MEP environment demonstrates that this framework can robustly update pipeline BIMs, reducing modeling time from 3-5 days to approximately 3 hours, thereby improving the efficiency of BIM-based MEP management.

Keywords –

Point cloud; Building Information Modeling; As-designed model; Automatic modeling

1 Introduction

BIM has become a key enabler for efficient lifecycle management in the architecture, engineering and construction industry, especially when integrated with Internet of Things (IoT) for real-time data acquisition and facility operation optimization. In smart construction, enriched BIM models are increasingly required to support performance monitoring and maintenance of MEP systems [1]. However, as-designed BIM models often deviate from as-built conditions due to construction tolerances and on-site modifications, which significantly degrades the reliability of BIM-based MEP management.

In practice, BIM models are often generated manually from component specifications and installation records, causing low efficiency in complex MEP environments. Laser scanning can capture dense point clouds that accurately represent as-built environment [2]. However, raw point clouds lack semantic and topological information and cannot be directly used for engineering applications [3]. Although automated scan-to-BIM approaches have been developed [4-5], many rely primarily on local geometric features, which limits their robustness in reconstructing complex MEP systems.

To address these limitations, this paper proposes an automatic framework for updating as-designed pipeline BIM models using laser-scanned point clouds. The main contributions are summarized below. First, a semantic segmentation strategy is developed by integrating DBSCAN with deep learning to improve recognition of multi-scale pipeline components in cluttered MEP environments. Second, a logic-based geometric refinement method is proposed by utilizing as-designed BIM priors to correct coordinate deviations caused by noise and occlusion. Finally, the research bridges the semantic gap between unstructured as-built data and structured as-designed components by applying graph isomorphism, thereby enabling a fully closed-loop parametric update. Overall, the integrated workflow minimizes manual intervention and maximizes MEP modeling fidelity at the same time.

2 Related Work

2.1 Point Cloud Segmentation

Point cloud segmentation is critical for automated scan-to-BIM workflows but remains challenging in complex MEP environments [6-7]. Deep learning approaches such as PointNet++ and transformer-based networks [8] have improved segmentation performance. Synthetic or annotated datasets have also been explored to address data scarcity [9-10]. However, robustness is still limited by scale variation and occlusion, motivating hybrid methods that combine spatial preprocessing with learning-based segmentation.

2.2 Geometry Parameter Identification

Following segmentation, parameter identification extracts geometric information from point clouds for

parametric BIM reconstruction. Early methods mainly relied on manual feature extraction and geometric fitting techniques such as RANSAC [11]. Although effective for regular shapes, these approaches depend on prior knowledge and manual tuning, limiting their applicability to complex MEP components with irregular geometry. Recent studies have therefore shifted toward deep learning-based parameter extraction, which learns structural representations directly from data and can handle both regular and irregular components without explicit feature engineering [12]. This data-driven paradigm improves the automation and robustness of parametric BIM reconstruction in complex MEP systems.

2.3 Automatic Modeling

Automatic modeling aims to convert geometric and semantic data into parametric BIMs. However, accurately reconstructing as-built pipelines remains challenging. Early Scan-to-BIM frameworks focused on detecting spatial discrepancies [13]. Subsequent studies [14-15] leveraged 3D point clouds for automated geometric modeling of cylindrical MEP elements. Recently, combining semantic segmentation with geometric extraction has enabled transformation from point cloud to BIM via visual programming (e.g., Dynamo) [4]. Furthermore, hybrid methods fusing image semantics with LiDAR precision enhance robustness in complex environments [16]. These studies collectively provide a solid foundation for automated pipeline reconstruction.

Despite recent progress in Scan-to-BIM research, several limitations still restrict its practical application in complex MEP environments. First, many existing approaches rely heavily on local geometric features, which reduces robustness in scenes with severe occlusion and dense pipeline layouts. Second, prior information from as-designed BIM models is not sufficiently utilized. Third, current methods still face challenges in achieving automated synchronization between as-designed and as-built pipeline models.

3 Methodology

This study proposes a framework to automatically update as-designed pipeline BIM model with laser-scanned as-built point clouds. As shown in Figure 1, the framework takes data acquired from a laser scanner as inputs. After point cloud preprocessing, DBSCAN and PointNet++ perform semantic segmentation. On this basis, parameter extraction with logical reasoning derives the geometric parameters of the as-built pipelines, while spatial topology analysis is used to establish the mapping between point-cloud instances and BIM components. Finally, as-built parameters are used to automatically update the as-designed BIM model, thereby generating an updated as-built pipeline BIM. The details of each step are presented in the following sections.

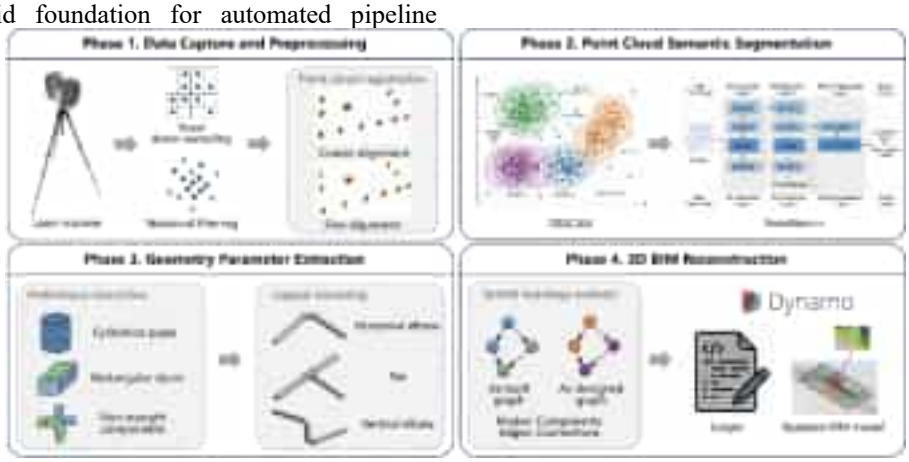


Figure 1. Proposed methodology framework

3.1 Data Capture and Preprocessing

High-quality point clouds are essential for accurate BIM updating, as they directly influence all subsequent steps. In MEP scenarios, dense and accurate laser scans enable reliable identification of pipeline geometry and connections, improving the robustness of BIM updating [17]. Point cloud quality primarily depends on the data acquisition strategy and preprocessing procedures, which together determine point density and noise level.

This study uses a terrestrial laser scanner (Trimble TX5) to obtain the as-built data of the MEP system, its detailed technical specifications are summarized in Table 1. Multi-station laser scanning is employed to ensure adequate coverage and to mitigate the impact of occlusions in indoor MEP environments. The raw point clouds are first preprocessed using voxel down-sampling with a grid size of 0.01 to reduce data redundancy while

preserving key geometric features, and then refined through statistical outlier removal to eliminate noise and irrelevant points. This process yields sparse and low-noise point clouds suitable for subsequent segmentation and parameter extraction.

Table 1. Characteristics of the Trimble TX5 Scanner

Laser Type	Parameters
Distance range	0.6-120 m
Accuracy	±2 mm@25 m
Angle range	Horizontal: 360°; vertical: 300°
Acquisition speed	Up to 976,000 pts/s

The coordinate system of the as-designed model is used as the reference for the workflow. Walls and columns, assumed to be consistent with the design, are selected as registration primitives to align the as-built point clouds. A combined registration strategy using Fast Global Registration (FGR) for coarse alignment and Iterative Closest Point (ICP) for fine alignment transforms the as-built pipeline point clouds into the design coordinate system, providing a reliable basis for further analysis [18].

3.2 Point Cloud Semantic Segmentation

After obtaining the as-built pipeline point cloud in unified coordinate system, it is necessary to segment and identify different pipeline components for parametric modeling. PointNet++ [19] is a widely used semantic segmentation network that samples centroids uniformly and aggregates local features through set abstraction layers. However, in complex MEP environments with large scale variation, large-diameter pipes dominate the sampled points, while small components are under-sampled. This causes the loss of critical local geometry for small parts and reduces segmentation accuracy. To address this limitation, this study proposes a two-stage semantic segmentation method that integrates DBSCAN with deep learning to better capture pipeline components across scales.

3.2.1 Density-Based Pipeline Clustering

To reduce scene complexity and improve sampling balance, this study first applies the DBSCAN algorithm [20] to segment the global point cloud into independent pipeline instances. DBSCAN identifies clusters by locating points that satisfy a minimum density criterion, controlled by the neighborhood radius (ϵ) and the minimum number of points. In MEP systems, pipelines are densely arranged but usually separated by small spatial gaps. By setting ϵ according to the minimum design clearance, DBSCAN clusters points from the same physical run into a single instance while discarding isolated noise. This separation allows each pipeline segment to be processed independently in the subsequent stage, avoiding feature interference from adjacent and unrelated structures.

3.2.2 PointNet++-Based Component Semantic Segmentation

Once individual pipeline instances are extracted via DBSCAN, the resulting segments are fed into PointNet++ model for fine-grained component recognition. This step classifies them into eight key categories: exhaust pipes, exhaust pipe elbows, exhaust pipe tees, air pipes, air pipe elbows, air pipe tees, water pipes and water pipe tees, covering the main elements in MEP layouts.

We construct a dedicated training set by collecting pipeline point clouds from multiple construction sites and manually composing them into diverse spatial layouts. Random translations, rotations and gaussian noise are applied to enhance robustness to sensor errors and scan-angle variations. The final dataset, summarized in Table 2, balances large pipes and small fittings.

Table 2. Dataset description

Label	Category	Train Set	Test Set
1	Exhaust pipe	158	40
2	Exhaust pipe elbow	120	32
3	Exhaust pipe tee	96	24
4	Air pipe	160	46
5	Air pipe elbow	138	40
6	Air pipe tee	96	24
7	Water pipe	180	48
8	Water pipe tee	122	32

In the segmentation stage, each isolated and normalized component is uniformly resampled to 8192 points, ensuring that even small parts retain sufficient point density. The network employs Multi-Scale Grouping (MSG) in its set abstraction layers to capture features across multiple radii, enabling it to learn both the geometry of straight pipes and the complex shapes of elbows and tees. Through feature propagation, local geometric cues are integrated with global context to assign a semantic label to every point. This instance-level strategy significantly improves overall performance and produces accurate component boundaries in dense MEP scenes.

3.3 Geometry Parameter Extraction

3.3.1 Preliminary extraction

For cylindrical pipe segments, RANSAC cylinder fitting algorithm is employed to robustly estimate the centerline and radius by minimizing the orthogonal distance between points and the cylinder surface. Once the axis direction is obtained, the pipe length is computed

by projecting all points onto the axis and measuring the distance between the two extreme projections.

For rectangular pipes, RANSAC-based plane fitting algorithm is applied to extract the four main faces. The width and height are derived from the distances between opposing parallel planes, while the pipe orientation is inferred from the plane normal vectors.

For non-straight components such as elbows and tees, an indirect strategy is adopted. Insertion points and orientations are inferred from the intersections of centerlines of adjacent straight segments, ensuring topological consistency between fittings and pipes during reconstruction.

3.3.2 Correction Based on Logical Reasoning

However, geometric parameters extracted directly from point clouds are often affected by noise, occlusions and fitting errors, resulting in joint gaps. To address this, a logical reasoning scheme adjusts the as-built parameters based on engineering constraints and as-designed data. As shown in Figure 2, the reasoning rules are defined for three types of connections.

If two pipes are connected by a horizontal elbow, we first extend the central axis of the two pipes. Then we calculate the intersection of the two center axes, which can be thought of as the center coordinates of the elbow. Next, we correct the endpoint coordinates of the two pipes by analyzing the known radii in Figure 2(a). The final coordinate formulas are shown below, where r represents the standard elbow radius from as-designed model, x and y represent the calculated horizontal elbow center coordinates, a_5, a_6, b_1 and b_3 represent the original endpoint coordinates.

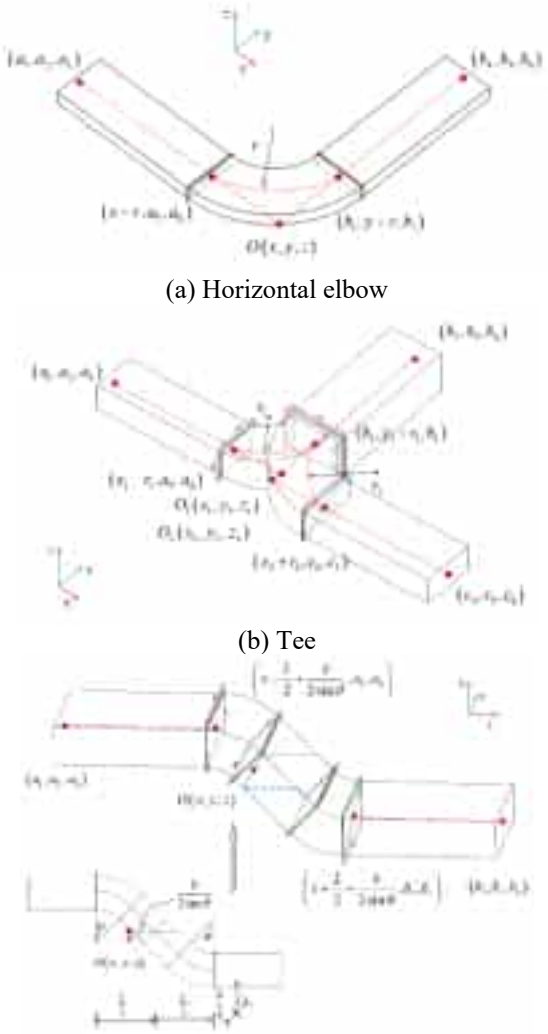
$$(x - r, a_5, a_6) \quad (1)$$
$$(b_1, y + r, b_3) \quad (2)$$

If two pipes are connected by a tee, we first find the three pipe endpoints around the tee, and distinguish the three pipes into two main pipes on a straight line, and a separate branch pipe. Then, we find the two intersection points of the central axis of the branch pipe and the central axis of each main pipe respectively. Next, we match the two radii with intersection and pipe, so the pipe endpoint coordinates are corrected by analyzing the known radii as shown in Figure 2(b). The inference formulas are as follows, r_1 and r_2 are two known radii of the tee in the as-designed model, x_1, x_2 and y_1 are the calculated center coordinates of the tee, a_5, a_6, b_1, b_2, c_2 and c_3 represent the endpoint coordinates.

$$(x_1 - r_1, a_5, a_6) \quad (3)$$
$$(b_1, y_1 + r_1, b_3) \quad (4)$$
$$(x_2 + r_2, c_2, c_3) \quad (5)$$

If two pipes are connected by upper and lower elbows, we first calculate the horizontal plane between the upper and lower pipes by deriving the central axis of the two pipes. Then we exploit the RANSAC algorithm to extract the cuboidal extent of the elbow horizontal section based on the horizontal plane. Next, we extract the line perpendicular to the tube axis at the edges of the horizontal plane and take their midpoints as standard points as Figure 2(c) shows. Afterwards, we correct the endpoint coordinates by the formulas. L, h and θ represent the horizontal distance, height and curve angle of the upper and lower elbows respectively in the as-designed model, x represents the calculated standard coordinate of elbow and a_5, a_6, b_2, b_3 represent the original coordinates.

$$\left(x - \frac{L}{2} + \frac{h}{2sin\theta}, a_5, a_6\right) \quad (6)$$
$$\left(x + \frac{L}{2} + \frac{h}{2sin\theta}, b_2, b_3\right) \quad (7)$$



(a) Horizontal elbow

(b) Tee

(c) Vertical elbow

Figure 2. Logical reasoning mathematical analysis

The logical reasoning module processes all detected connections in the pipeline network. For each connection node, the module first retrieves the corresponding connector type and its standard size from the as-designed BIM model, then applies the corresponding reasoning rules, and updates the endpoint coordinates of the pipeline accordingly. When the same pipeline involves multiple corrections, the corrections will be carried out sequentially, and the final coordinates are calculated according to the weighted average of the confidence of each connecting node. This iterative optimization process transmits geometric constraints to the entire network, effectively eliminating the cumulative error.

3.4 Graph Isomorphism and Modeling

3.4.1 Spatial Topology Analysis

To establish correspondence between as-built point cloud components and as-designed BIM elements, both datasets are represented as directed graphs. In each graph, vertices denote pipeline components, while directed edges encode physical connection from pipes to connectors based on spatial adjacency. The construction of the as-built graph is based on the segmented point cloud instances. The as-designed graph is generated from the BIM model by extracting component attributes and predefined connection relationships.

Graph matching is performed through a topology guided search strategy. First, candidate node pairs are initialized based on component type consistency and similarity of geometric attributes such as diameter, length and orientation. To reduce ambiguity in repetitive pipeline layouts, node degree and local connection patterns are further used as pruning constraints.

Starting from distinctive seed components, the matching process incrementally propagates along connected branches of the pipeline network. During traversal, each potential correspondence is verified by jointly evaluating geometric compatibility and topological consistency. Backtracking is applied when conflicts occur, ensuring a one-to-one mapping between the as-built graph and the design graph. The resulting correspondence enables automatic transfer of geometric parameters to the as-designed BIM model.

3.4.2 3D BIM Reconstruction

In the final stage, we use parametric programming in Dynamo to update the pipeline BIM model so it matches the as-built data. Dynamo scripts import the updated start and end coordinates, as well as the size and type of each pipe, elbow, and tee. Using the mapping between point cloud instances and as-designed BIM components, the system then automatically updates the corresponding parameters in the prefabricated family library.

4 Results and Discussion

To evaluate the feasibility of the proposed framework, we conducted a case study on standard floor A1# of a shopping center. The environment features a complex MEP layout with narrow spaces and densely distributed pipelines. The scene includes eight types of pipeline components, such as air pipes, exhaust pipes, water pipes and their corresponding elbows and tees. Its complexity is further increased by severe occlusion: 16 pipes are blocked by others, 5 pipes overlap, and 8 connectors are heavily covered. This challenging scenario provides a rigorous test platform for assessing the framework's capability to update the pipeline BIM model based on the as-built point cloud and the as-designed model.

4.1 Point Cloud Collection and Processing

As shown in Figure 3, the acquired point cloud provides comprehensive coverage of the pipeline system despite severe occlusion. After registration, the point cloud achieved accurate alignment with the designed model. The RMSE decreased from 0.16m to 9.97×10^{-3}m, demonstrating effective spatial consistency. This precise alignment establishes a reliable basis for subsequent processes. The registration results are illustrated in Figure 4.



(a) Engineering practice (b) As-built point cloud

Figure 3. As-built pipelines



(a) FGR (b) ICP with FGR

Figure 4. Point cloud registration result

4.2 Point Cloud Segmentation

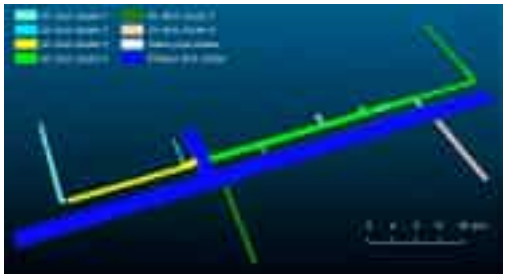
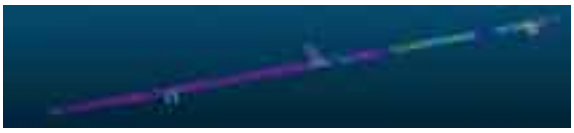


Figure 5. DBSCAN initial segmentation result

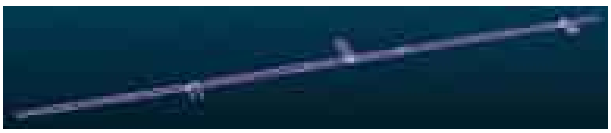
We integrated DBSCAN clustering with PointNet++ for semantic segmentation. With a radius of 0.025m and a minimum of 20 points, DBSCAN partitioned the point cloud into 8 clusters and removed 1,185 noise points (Figure 5). These clusters were subsequently processed by PointNet++, achieving an overall accuracy of 95.85%.



(a) Ground truth



(b) PointNet++ without DBSCAN



(c) PointNet++ with DBSCAN

Figure 6. Water pipeline point cloud segmentation results

Figure 6 compares the water pipeline segmentation results in three cases: the actual scene (Figure 6(a)), PointNet++ alone (Figure 6(b)), and the combined DBSCAN-PointNet++ method (Figure 6c). Figure 7 further demonstrates successful instance segmentation of exhaust pipes, elbows and tees.



(a) Pipes



(b) Elbows



(c) Tees

Figure 7.Exhaust pipeline point cloud segmentation results

Table 3. PointNet++ segmentation results with DBSCAN and without DBSCAN

Type	Without DBSCAN			With DBSCAN		
	Precision (%)	Recall (%)	IoU (%)	Precision (%)	Recall (%)	IoU (%)
Exhaust pipe	94.35	93.62	88.65	97.48	94.13	91.89
Exhaust pipe elbow	48.94	94.67	47.63	89.72	91.83	83.09
Exhaust pipe tee	71.58	82.89	62.36	70.80	73.93	56.66
Air pipe	94.69	95.20	90.37	98.90	98.92	97.84
Air pipe elbow	62.19	46.70	36.37	80.28	74.85	63.22
Air pipe tee	40.82	90.59	39.16	59.72	94.11	57.57
Water pipe	96.55	90.92	9.89	99.31	97.18	96.53
Water tee	52.31	82.90	11.54	53.85	83.85	48.79
Average	70.18	65.81	48.25	81.26	88.60	74.45

Using DBSCAN as a preprocessing step proves particularly effective in complex MEP scenes with significant scale variations, where PointNet++ alone often fails to capture data from small pipes. As shown in Table 3, the configuration with DBSCAN improves performance across all categories. For example, precision for air pipe elbows increases from 62.19% to 80.28% and recall for water tees rises from 82.90% to 83.85%. These results demonstrate the robustness of the proposed two-stage semantic segmentation strategy.

4.3 Geometry Parameter Extraction

Geometric parameters were first extracted via the RANSAC algorithm. However, initial extraction exhibited noticeable spatial errors. These were substantially reduced by applying the proposed logical reasoning module, which corrects pipe endpoints using known connector parameters from the as-designed model. Table 4 shows the resulting average coordinate errors.

Table 4. Average Coordinate Error of Different Pipes

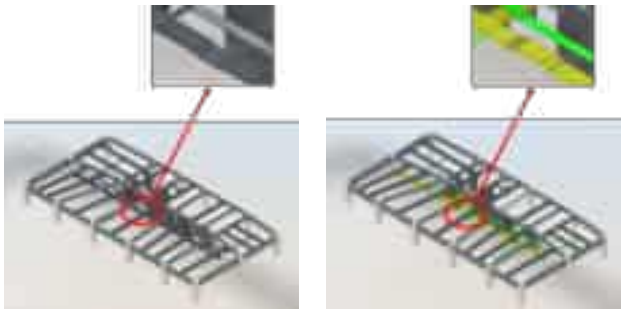
Types		Average spatial coordinate error		
		x (m)	y (m)	z (m)
Cuboidal exhaust pipes	Initial	0.114	0.022	0.002
	Corrected	0.009	0.002	0.002
Cuboidal air pipes	Initial	0.027	0.026	0.003
	Corrected	0.010	0.011	0.003
Cylindrical water pipes	Initial	0.023	0.025	0.003
	Corrected	0.003	0.003	0.003

The results indicate that using RANSAC alone is

often insufficient due to point cloud noise and occlusion, leading to misalignment between pipes and connectors. By mathematically inferring correct endpoints from connection relationships, the logical reasoning module effectively mitigates this issue. The result of errors generally below 0.01 m demonstrates the advantage of incorporating prior knowledge from the as-designed model over purely geometric methods, ensuring both complete and accurate connections.

4.4 Graph Isomorphism and Modeling

To automate the modeling process, spatial topology analysis was performed. Two directed graphs were constructed: one representing the logical structure of the as-built point cloud and the other corresponding to the as-designed BIM model. Graph isomorphism analysis established a one-to-one mapping between point cloud instances and BIM components. Based on this mapping, parametric programming was used to automatically update the geometric parameters. Figure 8 visually compares the as-designed BIM model (Figure 8(a)) with the updated BIM model (Figure 8(b)), demonstrating the success of reconstruction.



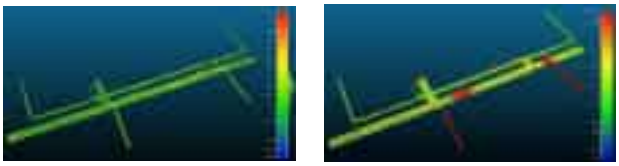
(a) As-designed (b) Updated

Figure 8. Pipeline BIM model

This step bridges the gap between geometric data and semantic BIM objects. By representing the pipeline network as a directed graph, the framework resolves the correspondence problem, enabling precise association between each BIM element and its corresponding extracted parameters. This automated mapping streamlines the transition from raw data to an information-rich BIM model, which is essential for the framework's practical application.

4.5 Precision and Efficiency Analysis

Our framework was evaluated in terms of both modeling accuracy and time efficiency. Deviation analysis between the updated BIM model and the laser-scan point cloud is presented in Figure 9(a). Errors followed a normal distribution, with over 90% of points exhibiting deviations under 0.03 m and an average deviation of only 0.0046 m. Figure 9(b) further compares the output with the original as-designed model, clearly indicating areas where on-site modifications were made. A significant gain in efficiency was achieved. Whereas traditional manual modeling for a scene of this complexity typically takes 3-5 days, our framework completed the update in only 3 hours, corresponding to a time reduction of nearly 90%. This demonstrates that the proposed framework offers an accurate and efficient solution for updating as-designed pipeline BIM models.



(a) Deviation distribution (b) Designed vs updated

Figure 9. Model comparison

5 Conclusion and Future Work

This study proposes and validates an automated framework for bridging as-designed BIM models with as-built point cloud data. By integrating two-stage semantic segmentation, parameter extraction with logical reasoning and graph isomorphism analysis, the framework addresses key challenges in complex MEP environments, including severe occlusion, multi-scale components and automated semantic association. Validation using a real-world shopping center case demonstrates that the proposed method can robustly update high-fidelity BIM models that accurately capture on-site conditions.

The results indicate that the hybrid segmentation strategy substantially enhances the detection of small-scale pipelines in cluttered scenes, effectively mitigating the loss of detail associated with conventional approaches. Moreover, by incorporating prior design knowledge through logical reasoning instead of relying solely on geometric fitting, the updated models preserve logical connectivity and topological correctness even when the raw data are imperfect. In addition, graph isomorphism analysis enables automated mapping between point-cloud instances and BIM elements, streamlining the workflow and reducing the time and labor required for manual modeling, while still meeting the accuracy requirements of facility management.

It is important to emphasize that in this study, the term automatic mainly refers to the automated updating of pipeline BIM components after data input and parameter initialization. Although several preprocessing parameters are defined according to engineering experience, the framework eliminates manual modeling operations and enables efficient generation of an updated BIM model.

Despite these contributions, several limitations warrant further research. First, the framework targets

standard pipeline geometries, future work should expand the component library to cover more complex MEP elements, such as valves and pumps. Second, while terrestrial laser scanning is highly accurate, adding complementary data sources such as photogrammetry or handheld SLAM could improve completeness in narrow or occluded areas. Finally, extending the algorithms toward processing in near real time would support quality control during construction by enabling immediate detection and correction of deviations.

References

[1] Y. Chen and X. Luo. Indoor scan-to-BIM workflows: Progress, challenges, and future directions (2014–2024). *Automation in Construction*, 181:106578, 2026.

[2] E. Valero, F. Bosché, and M. Bueno. Laser scanning for BIM. *Journal of Information Technology in Construction (ITcon)*, 27(23):486–495, 2022.

[3] S. Zbirovský and V. Nežerka. Open-source automatic pipeline for efficient conversion of large-scale point clouds to IFC format. *Automation in Construction*, 177:106303, 2025.

[4] B. Wang, C. Yin, H. Luo, J. C. P. Cheng, and Q. Wang. Fully automated generation of parametric BIM for MEP scenes based on terrestrial laser scanning data. *Automation in Construction*, 125:103615, 2021.

[5] X. Li, V. J. L. Gan, K. Li, and M. Li. High-precision 3D BIM reconstruction for mechanical, electrical and plumbing components using terrestrial laser scanning and LiDAR point clouds. *Journal of Building Engineering*, 113661, 2025.

[6] J. Patil and M. Kalantari. Automatic Scan-to-BIM—The Impact of Semantic Segmentation Accuracy. *Buildings*, 15(7):1126, 2025.

[7] S. Jing, G. Cha, M. B. Maru, B. Yu, and S. Park. Improved building MEP systems semantic segmentation in point clouds using a novel multi-class dataset and local--global vector transformer network. *Journal of Building Engineering*, 96:110311, 2024.

[8] R. Zhang, Y. Wu, W. Jin, and X. Meng. Deep-learning-based point cloud semantic segmentation: A survey. *Electronics*, 12(17):3642, 2023.

[9] M. Mérizette, N. Audebert, P. Kervella, and J. Verdun. 3DSES: an indoor LiDAR point cloud segmentation dataset with real and pseudo-labels from a 3D model. *arXiv preprint arXiv*:2501.17534, 2025.

[10] H. Yue, Q. Wang, H. Huang, X. Xia, H. Fang, and J. C. P. Cheng. Enhancing semantic segmentation of MEP scenes with deep learning and BIM-generated synthetic point clouds. *Advanced Engineering Informatics*, 68:103723, 2025.

[11] R. Schnabel, R. Wahl, and R. Klein. Efficient RANSAC for point-cloud shape detection. *In Computer Graphics Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[12] H. Yue, Q. Wang, Y. Yan, and G. Huang. Deep learning-based point cloud completion for MEP components. *Automation in Construction*, 175:106218, 2025.

[13] L. Centarti, N. E. Güidi, and B. Messina. Automated Scan-To-BIM Methodology for Accurate 3D Modeling of Embedded MEP Systems. In *Representation Across Boundaries: New Links with AI, AI-GEN, and XR Tools for Cultural Heritage and Innovative Design*, pages 905–924. Springer, Cham, 2026.

[14] V. S. Kalasapudi, Y. Turkan, and P. Tang. Toward automated spatial change analysis of MEP components using 3D point clouds and as-designed BIM models. In *2014 2nd International Conference on 3D Vision*, volume 2, pages 145–152. IEEE, 2014.

[15] F. Bosché, M. Ahmed, Y. Turkan, C. T. Haas, and R. Haas. The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction*, 49:201–213, 2015.

[16] B. Wang, Q. Wang, J. C. P. Cheng, C. Song, and C. Yin. Vision-assisted BIM reconstruction from 3D LiDAR point clouds for MEP scenes. *Automation in Construction*, 133:103997, 2022.

[17] B. Wang, Z. Chen, M. Li, Q. Wang, C. Yin, and J. C. P. Cheng. Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes. *Automation in Construction*, 162:105384, 2024.

[18] Y. Zha, M. Chen, and S. Lu. Research on point cloud registration of industrial parts based on FGR-ICP algorithm. In *Journal of Physics: Conference Series*, vol. 1941, no. 1, p. 012014, 2021.

[19] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017.

[20] E. Schubert, J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

Optimal Transport-based Affective User Requirements Realignment for Automated Spatial Design Revision

Jin-Bin Im[1], Rong-Lu Hong[1], Chang-Hyun Choi[2], Jung-Eun Ha[1] and Ju-Hyung Kim[1]

[1]Department of Architectural Engineering, Hanyang University, Republic of Korea
[2]Research and Development Department, Freaktech Corp., Republic of Korea
jbim0202@hanyang.ac.kr, abraham0814@hanyang.ac.kr, rkfaorl144@gmail.com, cuttigirl@naver.com, kcr97jhk@hanyang.ac.kr

Abstract

Modern residential occupants require their affective intents to be reflected in spatial design to achieve satisfaction. However, current generative architectural design studies focus on morphological and functional requirements, overlooking affective intent. This study proposes an affective user requirements realignment framework for automated, minimally disruptive spatial design revision based on Optimal Transport. The proposed BIM-integrated framework (i) represents spatial experience as a multidimensional affective distribution derived from 1,638 annotated responses across 54 parametric configurations, (ii) predicts the distribution from discrete design parameters using a lightweight Transformer regressor, and (iii) selects a minimally disruptive revision via entropic-regularized Wasserstein distance to a user-specified target affect across 6,480 candidate configurations. The selected parameter set is automatically applied through a Revit C# add-in to support closed-loop design review. Results demonstrated consistent convergence toward target affects (reduction: 0.036–0.290) with positive directional alignment across all ten affective dimensions (mean cosine similarity = 0.502), demonstrating the feasibility of affect-preserving revision in BIM workflows as a proof-of-concept.

Keywords –

Design Review, User-Centered Design, Automation, Affective Design, User Requirements.

1 Introduction

Residential architecture has moved beyond meeting morphological and functional needs to accommodate the psychological and emotional expectations [1]. Homes are now considered as personal environments that shape routines, well-being, and identity [2], which increases the pressure to improve the collection and integration of emotional requirements into design workflows. However, most residents lack architectural vocabulary and spatial literacy; therefore, they express their needs through narratives and subjective impressions rather than parametric terms [3]. Feedback such as “the room feels too small” conveys a strong experiential meaning, but is difficult to translate into quantifiable interventions, creating a persistent interpretive gap and risking the loss of user intent [4,5]. Although environmental psychology and neuroarchitecture link built features to emotional responses, real spaces involve tightly interdependent relationships between geometry, materiality, lighting, and atmosphere, making single-variable interpretations unreliable [6,7].

Recent advances in generative Artificial Intelligence (AI) have improved early stage option generation and communication [8]. However, most studies emphasize generation, rather than the more practically critical problem of revision. Revisions require interpreting feedback, preserving design intent, respecting constraints, and maintaining coupled spatial relationships [9]. This is economically important because early design phases account for a large share of the total design/documentation effort; thus, emotional misalignments discovered later can trigger costly rework or discourage necessary changes.

Accordingly, this study focused on affective design revisions in Building Information Modeling (BIM) environments by formulating emotional requirements as explicit affective targets. Rework has been shown to contribute to over 50% of total cost growth in construction project [10], and economic analysis of BIM-based high-rise project have directly quantified design errors as causing 5–20% rework overhead [11]. Those phenomena underscore the economic importance of addressing design intent misalignments early in the review process. To ensure clarity and reproducibility in design review contexts, user input is operationalized as the selection of a predefined affective dimension, allowing the revision process to concentrate on distribution-aware optimization and minimal design modification, rather than linguistic interpretation.

2 Related Studies

2.1 AI Application in Architectural Design Alternatives

The integration of generative AI into architectural workflows transforms traditional decision-making processes. Notably, generative design introduces an AI-powered approach that combines engineering design principles with cloud computing to automatically generate and optimize options based on parameters such as materials, dimensions, and cost constraints. In architectural designs, this approach has been extended to generate reference imagery [8], photorealistic interior visuals [12], and exterior façade designs guided by layout sketches and prompts [13].

The deep learning models employed in these efforts include generative adversarial networks (GANs), diffusion models, and transformers. GANs and diffusion models are particularly effective in image-to-image generation tasks such as diagram alterations and style transfer [14], with diffusion models suited to high-resolution rendering via platforms such as Dream Studio [15]. Transformer-based models, such as generative pre-trained Transformer, specialize in text processing for code generation, design documentation, and javascript object notation BIM object creation [16].

Generative modeling techniques are frequently used in BIM-based research. Studies including computer aided design-to-BIM automation [17] and conversational BIM agents coupled-GAN [18] have demonstrated the efficiency of generating certain objects or resources in BIM-based workflows. These studies highlight the need for multimodal inputs and a semantic pipeline to derive architectural variables for integration into structured BIM. However, generative models focus on generating outputs from zero-based scenarios; therefore, examining the efficiency of revising BIM-based workflows is required.

2.2 User Requirements Realignment

Architectural design is a tightly coupled system. Changing one element (e.g., window size, and ceiling height) can trigger cascading and unintended effects elsewhere; therefore, revision cannot be treated as a wholesale replacement without risking damage to the already successful aspects of the design [4]. Preserving design intent, including both explicit architectural decisions and implicit experiential coherence, is therefore essential because even small edits can disturb spatial harmony, function, and emotional responses [19]. Neuroarchitecture and environmental psychology further demonstrate that affective experiences emerge from multifeatured interactions (e.g., geometry, materiality, lighting, and enclosure), implying that selective change while maintaining stability in other dimensions is generally necessary [20].

This motivates distribution-aware revision. Instead of optimizing a single scalar target, the goal is to move the predicted affective distribution toward the user's intended distribution with minimal collateral displacement [21]. Information-theoretic divergences such as Kullback-Leibler and Jenson-Shannon [22] and overlap-based metrics such as Hellinger and Bhattacharyya [23,24] can quantify mismatch; however, they consider probability shifts as pointwise differences and do not encode how "far" mass moves between affect categories. This is a key limitation of minimal-change design control, in which only a subset of attributes should shift.

Optimal transport (OT) addresses this problem by endowing distributions with geometry. The Wasserstein (Earth Mover's) distance measures the minimum "work" required to transport probability mass between distributions given a ground-cost matrix capturing inter-label distances, making it well suited to small, selective, and cost-aware revisions [25]. Entropic regularization and Sinkhorn-based approximations enable fast and differentiable computation inside iterative optimization loops [26], with theoretical support for stability and sample behavior [27]. Overall, although classical divergences are lightweight, OT-based metrics model label-space transitions more directly, providing a stronger foundation for revision systems that adjust target affects while preserving the other components.

3 Proposed Methods

3.1 Research Framework

The framework began by allowing the user to select a target affective dimension encoded as a target affective distribution. A transformer regressor predicted the affective distributions for candidate spatial configurations, and an OT) module evaluated all 6,480 parametric design vectors (Ceiling Height:8 × WWR:10 × CCT:9 × Floor:3 × Color:3), by computing the Wasserstein distance (WD) from the target. A design with the minimum transport cost was selected, providing a distance-based ordering of alternatives and quantifying the minimal parameter change required to align the predicted affect with the user's intent.

3.2 Design of Experiment

3.2.1 Design Variable Selection

Emotional responses to architectural space originate from the combined effects of multiple spatial features, rather than isolated parameters. Previous studies in environmental psychology, interior architecture, and neuroarchitecture have consistently identified lighting

conditions, materiality, openness, and enclosure-related variables as dominant contributors to affective experiences in residential interiors. In particular, lighting attributes such as correlated color temperature (CCT) and luminance balance regulate arousal and comfort, whereas window-to-wall ratio (WWR) and ceiling height influence perceived openness and spaciousness. The material and surface color further modulate warmth, calmness, and overall valence.

Based on accumulated evidence, five interior design variables were selected for this study: ceiling height, WWR, CCT, room color, and floor material. These variables are strongly associated with emotional evaluations in residential spaces, making them suitable for controlled affective design experiments [28–30].

3.2.2 Affective Annotation

A systematic affective annotation procedure was conducted to construct a dataset linking spatial design parameters to human affective responses. The process began with the development of domain-grounded affective vocabulary. Drawing on environmental psychology, interior architecture, and neuroarchitecture, 22 affective adjectives were initially extracted from previous empirical studies [29–31].

Using four thesaurus expansion steps, synonyms and near-synonyms (e.g., happy-pleasant) were grouped, and semantically redundant terms were merged. Subsequently, the KJ clustering method was applied to identify conceptual clusters and remove ambiguous or overly broad terms. This process resulted in a final set of ten affective vocabularies closely aligned with the spatial experience: pleasant, exciting, cozy, spacious, attractive, simple, bright, open, calm, and comfortable.

3.2.3 Experimental Design

A parametric BIM environment was constructed using Autodesk Revit 2023 to generate the controlled interior design stimuli. The model represented a typical urban apartment living unit common in high-density residential developments with base dimensions of 4380 × 7440 × 2300 mm. The geometry was intentionally simplified to isolate the perceptually salient variables and minimize confounding architectural factors.

Full factorial combinations of the variables yielded 243 possible configurations. An L27 fractional factorial design was adopted to reduce redundancy, while preserving orthogonality. An additional 27 configurations were selected using the cyclic shift method to ensure uniform coverage of the multidimensional design space. In total, 54 interior design stimuli were labeled with their corresponding design vectors.

3.3 Inference Module: Affect-Spatial Elements Mapping

Each design configuration x_i was defined using five parameters: ceiling height, WWR, CCT, floor material, and room color. To ensure compatibility with the learning model $f_\theta(x)$, the data were encoded and normalized as follows. The continuous variables (ceiling height, WWR, CCT) were standardized using a StandardScaler, and the categorical variables (floor material, room color) were one-hot encoded with the first category dropped, yielding a 7-dimensional input vector.

For each configuration, the affective response vector α_i consisted of 10 ratings normalized to [0, 1]. To support OT-based computation, the ratings were further normalized into a probability distribution.

$$\alpha_i' = \frac{\alpha_i}{\sum_{k=1}^{K} \alpha_{i,k}}, K = 10, \sum_{k=1}^{K} \alpha_{i,k}' = 1 \quad (1)$$

A lightweight Transformer regressor was trained on all 54 design configurations to predict multidimensional affective responses from spatial design variables. Model selection and performance estimation were conducted using Leave-One-Out Cross-Validation (LOOCV), the statistically appropriate method for the design-level dataset size. The one-hot encoded design vector was projected onto a 64-dimensional embedding space. Learnable positional encoding was added and treated as a single-token sequence for the transformer processing. The embedding was then passed through a three-layer transformer encoder, where each layer included four-head self-attention, a 128-dimensional feed-forward network, and residual connections with layer normalization, enabling the model to learn nonlinear interactions among the design attributes.

Finally, a two-layer regression head (64→128→10) output continuous scores for the ten affective dimensions. Overall, the model mapped discrete design parameters onto a compact latent space, captured their interdependencies via stacked self-attention, and predicted the affective distribution of each configuration.

3.4 Optimizer Module: Entropic-Regularized Wasserstein Distance

The affective intent selected by the user determines the affective distribution of the target (P^*). When a user expresses a desire to increase or decrease a specific affective dimension, the corresponding component of the distribution is adjusted proportionally, followed by normalization.

$$P^* = \text{Normalize}(P + \Delta) \quad (2)$$

where Δ denotes the user-intended directional change. This normalized vector is the optimization objective of

an OT-based search.

For each spatial design vector x in the Space-Affect database, the transformer regression model $f_\theta(x)$ predicts its affective distribution.

$$Q_x = f_\theta \tag{3}$$

Thus, each candidate design produced a probability distribution over the same 10 affective dimensions as the target.

To quantify the discrepancy between the target affect distribution P^* and each predicted distribution Q_x, the WD was computed using the entropy-regularized Sinkhorn divergence [25].

$$W_\varepsilon(P^*, Q_x) = \min_{\Gamma \in U(P^*, Q_x)} \langle \Gamma, M \rangle + \varepsilon \mathrm{KL}(\Gamma \parallel P^{*} \otimes Q_x) \tag{4}$$

where $U(P^*, Q_x)$ denotes the set of admissible transport plans, M is the ground cost matrix and ε controls the entropy regularization strength. A uniform ground metric ($M_{ij} = 0$ and $M_{ij} = 1$ for $i \neq j$) was adopted as no validated psychometric distance exists for the architectural affective space and inter-affect correlations are uniformly high (mean r = 0.799), making the uniform metric the most conservative prior. The Frobenius inner product formulation enables linear-time Sinkhorn iterations and stable gradient propagation through $f_\theta(x)$.

Because the design space contained all 6,480 possible combinations, the optimizer evaluated the Sinkhorn distance for every candidate x.

$$\mathrm{x}^* = \arg\min_{x \in X} W_\varepsilon\,(P^8, Q_x) \tag{5}$$

where X denotes the full Cartesian product of all design variable attributes. This guarantees that the returned configuration is a globally optimal design vector with respect to the target affect.

3.5 Evaluation Metrics

3.5.1 Affect-Spatial Elements Mapping Evaluation

The proposed regression model was evaluated using metrics that captured the overall error magnitude, proportionality, and explanatory power. As a single metric can be misleading, particularly when targets differ in scale or when large errors are costly, we used five standard regression measures: mean absolute error (MAE) and R^2. MAE measures the average absolute error.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}| \tag{6}$$

It expresses errors as percentages, enabling comparisons across affective dimensions within different ranges. Finally, R^2 measures the proportion of variance explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{7}$$

Higher values indicate a stronger overall fit. Together, these metrics provide a balanced assessment of multidimensional affective regression performance.

3.5.2 Optimization Evaluation

Two complementary success criteria were defined: (i) positive WD reduction, the optimized design achieves a lower WD to the target distribution than the initial design; and (ii) positive directional alignment, cosine similarity > 0 between the intended and actual affective shift direction. WD reduction measures how effectively optimization decreases the discrepancy between the target affective distribution P^* and the predicted distribution Q_x.

$$\mathrm{Reduction} = \frac{W_\varepsilon(P^*, Q_{init}) - W_\varepsilon(P^*, Q_{final})}{W_\varepsilon(P^*, Q_{init})} \tag{8}$$

where Q_{init} and Q_{final} denote the initial and optimized affective states, respectively. Values closer to 1 indicate stronger affective convergence toward the target [25,26].

To assess the directional consistency with user intention, a perceptual alignment score was defined using the cosine similarity between the intended affective shift and the predicted shift.

$$\mathrm{Alignment} = \cos\big((P^* - P), (Q_x - P)\big) \tag{9}$$

A value close to 1 indicates strong alignment with the target affect, whereas negative values indicate a deviation from the intended direction [32].These metrics ensure that optimized revisions are both affectively closer to the target state and directionally consistent with user intent.

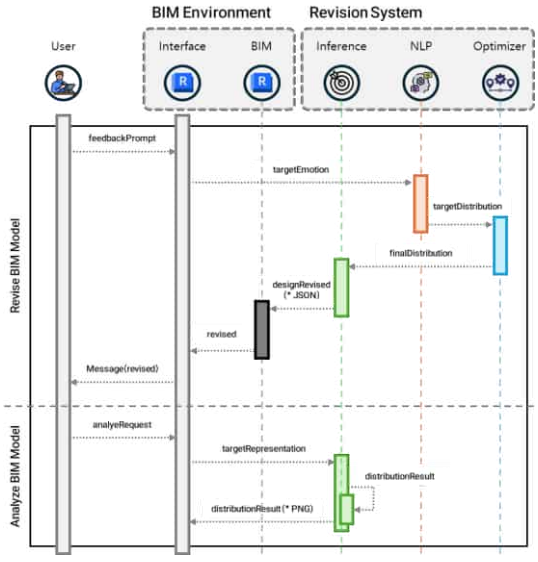


Figure 1. Proposed UML sequence diagram

3.6 BIM Add-in Application

The proposed system for affective design automation was fundamentally instantiated as a cohesive Revit add-in and entirely implemented using C# application programming interface (API) within the Visual Studio Environment. This integral implementation ensures seamless interaction with the BIM, minimizes data latency, and maximizes efficiency within the design workflow [18]. The comprehensive data flow and interlayer communication of this closed-loop system are precisely depicted using a Unified Modeling Language (UML) sequence diagram (Figure 1**Error! Reference source not found.**), which illustrates the sequential interactions among its modular components. This architectural approach directly integrates advanced computational modules into a commercial BIM platform.

4 Results

4.1 Inference Module Results

A total of 182 participants (male: 106, 58.2%, female: 76, 41.8%, age: 35.2±11.7, range 17–63) were recruited across architecture-related (n=70, 38.5%) and non-architecture (n=112, 61.5%) backgrounds, yielding 1,638 affective responses. The internal consistency of the ten affective scales was confirmed with Cronbach's α = 0.938. One-way analysis of variance confirmed statistically significant design condition effects across all ten affective dimensions (F=3.70–11.76, $p < .001$).

To justify the architectural choice of the Transformer regressor, 29 models across seven categories were systematically compared using LOOCV in Table 1. The Transformer (d = 32, nhead = 4, L = 1) achieved the highest overall LOOCV R^2 = 0.588 (MAE = 0.367), outperforming XGBoost (R^2 = 0.576), Gaussian Process-Matern (R^2 = 0.540), and MLP networks (R^2 = 0.352). Beyond predictive accuracy, tree-based XGBoost produced step-function prediction, collapsing to identical values for unseen parameter combinations. Since the OT optimizer searches intermediate values, smooth interpolation is structurally required, making the Transformer the appropriate model.

As shown in Table 1, the Transformer achieved LOOCV R^2 ranging from 0.483 (cozy) to 0.705 (open), with a mean of 0.588. Strong performance was observed for open (0.705), pleasant (0.685), spacious (0.640), attractive (0.636), and bright (0.593), indicating that these dimensions are well captured by the five spatial variables. Excitement (0.531), cozy (0.492), and calm (0.489) showed relatively lower explanatory power, attributable to high within-condition individual variance in these affective dimensions rather than model limitations.

Table 1. Top 10 model comparison using LOOCV.

Model	LOOCV R^2	MAE	Style
TF (d=32, h=4, L=1)	**0.588**	**0.367**	**Smooth**
XGBoost (d=3, n=100)	0.576	0.368	Step
GP (Matérn 1.5)	0.560	0.384	Smooth
RF (n=50, d=5)	0.546	0.381	Step
GBR (n=50, d=3)	0.544	0.385	Step
Ridge (α=1)	0.540	0.392	Linear
Lasso (α=0.01)	0.536	0.392	Linear
GP (Matérn 2.5)	0.534	0.389	Smooth
BayesianRidge	0.534	0.395	Linear
SVR (rbf, C=10)	0.503	0.415	Kernel

4.2 Optimizer Module Results

Across the ten affective categories, the optimization framework consistently reduced the WD to the target distribution, indicating that it can adjust the design variables to decrease the overall affective mismatch. As reported in Table 2, the reduction values ranged from 0.036 to 0.290, indicating meaningful convergence, even within a discrete and limited design space.

Table 2. Performance of the optimizer module

No.	Initial	Target	Final	Red.	Align.
1	0.106	0.255	0.118	0.078	0.388
2	0.092	0.243	0.129	0.247	0.941
3	0.085	0.237	0.108	0.143	0.437
4	0.099	0.249	0.132	0.215	0.699
5	0.126	0.272	0.168	0.290	0.732
6	0.095	0.246	0.111	0.100	0.346
7	0.109	0.258	0.132	0.148	0.373
8	0.082	0.235	0.118	0.232	0.734
9	0.106	0.255	0.115	0.064	0.221
10	0.100	0.250	0.107	0.036	0.150

Note: 1=Cozy, 2=Bright, 3=Excitement, 4=Open, 5=Simple, 6=Pleasant, 7=Spacious, 8=Attractive, 9=Calm, 10=Comfortable

The largest reductions (Red.) were achieved for simple (0.290), bright (0.247), and attractive (0.232), suggesting that these affects were highly responsive to the available physical variables. Spacious (0.148), excitement (0.143), and pleasant (0.100) exhibited moderate improvements, whereas cozy (0.078), calm (0.064), and comfortable (0.036) showed smaller but still positive reductions. The variation in reduction magnitude reflects differences in how sensitively each affective dimension responds to the five design variables.

The directional change was further examined using the alignment metric (Align.), which measures the cosine similarity between the intended affective shift direction and the actual distributional change. Positive alignment was achieved across all ten affective dimensions (mean = 0.502), with particularly strong alignment for bright

(0.941), attractive (0.734), simple (0.732), and open (0.699), confirming that the optimizer reliably moved the affective distribution in the intended direction. Moderate alignment was observed for cozy (0.388), spacious (0.373), and pleasant (0.346), while calm (0.221) and comfortable (0.150) exhibited weaker directional consistency, reflecting the limited sensitivity of these dimensions to the five parametric variables considered. Notably, all alignment scores were positive, indicating that the optimization consistently produced directionally appropriate adjustments without counterproductive shifts.

4.3 Revit API-based Design Revision

This section describes the sequence of steps, target emotion selection, inference, optimization, and the final application of design changes. The efficacy of the add-in was demonstrated by comparing pre- and post-revision affective distributions. The initial spatial parameters of the design case study are listed in Table 3.

Table 3 Initial and revised design parameters

Variables	Initial Design	Revised Design
Ceiling Height	2.7m	2.8m
WWR	30%	90%
CCT	2700K	3500K
Room Color	Gray	White
Floor Material	Tile	Tile

4.3.1 Initial Design Analysis

The add-in performed an objective analysis of the current design state. This involved utilizing a pretrained affective prediction model to estimate the distribution of the 10 affective dimensions induced by the initial spatial configuration. This initial analysis served as the baseline for the design and was used to benchmark the improvement in the target affective dimension after revision. At this stage, the system primarily generated the current affective profile without setting a specific optimization goal. An example of the initial design analysis is shown in Figure 2.

4.3.2 Design Revision Optimization

This module employed a retrieval-based optimization algorithm using the affective prediction model as its objective function to explore new combinations of the five spatial variables. The evaluation of the optimization used an OT-based method to calculate the distance between the affective distributions. In this case, the search successfully identified a revised combination of parameters that achieved the required increase in 'bright' while maintaining the overall coherence of the affective distribution. Based on these, revised design parameters were retrieved from the inference layer and are presented in **Error! Reference source not found.**.

4.3.3 Apply and Reset

Finally, the optimized spatial parameters were automatically applied to the BIM parameters. Following this automatic modification, the system captures the viewpoint image of the revised design. The user is then presented with this rendering, along with the predicted affective distributions, to make a final decision.

The primary validation of the system was a comparison of the initial and revised affective distributions. As shown in Figure 3, the 'bright' score significantly increased in the revised design compared to the initial states. This improvement validated the proposed methodology. Based on the visualization of the rendering and affective data, the user can choose to apply or reset the revised design.



Figure 2. Revit API UI for initial BIM Analysis



Figure 3. Revit API UI for Revised Design

5 Discussion

This study demonstrated the proof-of-concept that affective design revision can be formulated as a distribution-alignment problem rather than a single-variable optimization task. The use of OT provides a principled mechanism for reducing unintended affective shifts, with the degree of collateral affective displacement varying across dimensions. The revised results demonstrate consistent positive WD reduction and directional alignment across all ten affective dimensions, empirically supporting the framework's validity as a proof-of-concept.

The uniform ground metric was adopted because no

validated psychometric distance exists for the architectural affective space, inter-affect correlations are uniformly high (mean r = 0.799), and the uniform metric constitutes the most conservative prior. A sensitivity analysis comparing three cost matrix formulations yielded 70% agreement in optimal design selection, with the uniform metric performing consistently in all discordant cases. Future work may explore empirically grounded cost matrices derived from affective space or psychometric scaling.

This work differentiates itself from three prior streams. BIM optimization studies have targeted structural or cost-based objectives, whereas this work introduces multidimensional affective distributions as the optimization criterion within an automated revision loop. Affective design studies characterize space–emotion relationships without closing the loop into parametric revision, which this work achieves end-to-end. OT applications in affective computing have used Wasserstein distance as a training loss, whereas here OT serves as a design search criterion over a smooth, interpolable 6,480 candidate landscape, made possible by the Transformer's continuous interpolation for unseen parameter combinations. The current exhaustive search guarantees global optimality at negligible computational cost (< 1 ms); scaling to continuous spaces will require Bayesian optimization, differentiable Sinkhorn layers, or reinforcement learning-based exploration.

Perceptual validation through user studies and open-ended natural language feedback interpretation remains critical directions for future work, alongside assessments of design coherence and system usability.

6 Conclusion

This paper presents a BIM-integrated framework for automating affective design revision using optimal transport theory. By linking parametric spatial variables with predicted affective distributions, the proposed method enables designers to enhance selected emotional qualities while reducing unintended distributional displacement in most affective dimensions. The framework was implemented as a Revit add-in, demonstrating its technical feasibility within existing BIM-based design workflows.

The results suggest that optimal transport provides a promising, mathematically grounded foundation for affective design revision, as demonstrated in this proof-of-concept implementation. Although the current framework operates on a discrete set of five spatial variables and ten affective dimensions, it establishes a proof-of-concept foundation for future user-centered design automation. The proposed approach bridges the gap between user affective requirements and parametric design operations, thereby real-world practice validation is required.

Acknowledgement

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (grant number RS202400356697).

References

[1] J.C. Vischer. Towards an environmental psychology of workspace: How people are affected by environments for work. *Architectural Science Review*, 51(2):97–108, 2008.

[2] R. Gifford, D.W. Hine, W. Muller-Clemm, and K.T. Shaw. Why architects and laypersons judge buildings differently: Cognitive properties and physical bases. *Journal of Architectural and Planning Research*, (19)(2):131–148, 2002.

[3] R. Luck. Dialogue in participatory design. *Design Studies*, 24(6):523–535, 2003.

[4] S. Abbas, N. Okdeh, R. Roufayel, H. Kovacic, J.-M. Sabatier, Z. Fajloun, and Z. Abi Khattar. Neuroarchitecture: How the perception of our surroundings impacts the brain. *Biology (Basel)*, 13(4):220, 2024.

[5] K.T. Lee, J.B. Im, and J.H. Kim. Effects of human perception dimensions constituting individual temperament and experience on indoor space composition decision-making presented in a mixed-reality environment. *Architectural Science Review*, 1–16, 2025.

[6] J.B. Im, R.L Hong, M. Joo, E. Zhang, and J.H. Kim. Visual fidelity effects on occupants' performance, mental states, and emotions in mixed reality environments. *Architectural Science Review*, 1–21, 2025.

[7] M. Joo, J.B. Im, K.T. Kee, and J.H. Kim. Comparative analysis of work performance and physiological response based on office types and indoor space design. *Architectural Engineering and Design Management*, 1–17, 2025.

[8] J.-K. Lee, Y. Kim, E. Shin, S. Choo, and S.H. Cha. An AI-assisted approach for creating and archiving interior design references using 360-degree panoramic images. *Architectural Science Review*, (68)(6):506–518, 2024.

[9] K. Li, V.J.L. Gan, M. Li, M.Y. Gao, R.L.K. Tiong, and Y. Yang. Automated generative design and prefabrication of precast buildings using integrated BIM and graph convolutional neural network. *Developments in the Built Environment*, 18:100418, 2024.

[10] P.E.D. Love. Influence of Project Type and

Procurement Method on Rework Costs in Building Construction Projects. *Journal of Construction Engineering and Management*, 128(1):1–92, 2002.

[11] N. Ham, S. Moon, J.-H. Kim, and J.-J. Kim. Economic Analysis of Design Errors in BIM-Based High-Rise Construction Projects: Case Study of Haeundae L Project. *Journal of Construction Engineering and Management*, 144(6):05018006, 2002.

[12] J.K. Lee, H. Jeong, Y. Kim, and S.H. Cha. Creating spatial visualizations using fine-tuned interior design style models informed by user preferences. *Advanced Engineering Informatics*, 62:102686, 2024.

[13] H. Jo, J.-K. Lee, Y.-C. Lee, and S. Choo. Generative artificial intelligence and building design: Early photorealistic render visualization of façades using local identity-trained models. *Journal of Computational Design and Engineering*, 11(2):85–105, 2024.

[14] R.W. Gallega and Y. Sumi. Exploring the use of generative AI for material texturing in 3D interior design spaces. *Frontiers in Computer Science*, 6:1493937, 2024.

[15] J.A. Guridi, C. Cheyre, M. Goula, D. Santo, L. Humphreys, A. Souras, and A. Shankar. Image generative AI to design public spaces: A reflection of how AI could improve co-design of public parks. *Digital Government: Research and Practice*, 6(1):1–14, 2025.

[16] G. Zhang, C. Lu, and Q. Luo. Application of large language models in the AECO industry: Core technologies, application scenarios, and research challenges. *Buildings*, 15(11):1944, 2025.

[17] M. Shin, S. Park, B. Koo, and T.W. Kim. Automated CAD-to-BIM generation of restroom sanitary plumbing system. *Journal of Computational Design and Engineering*, 11(2):70–84, 2024.

[18] S. Jang, G. Lee, J. Oh, J. Lee, and B. Koo. Automated detailing of exterior walls using NADIA: Natural-language-based architectural detailing through interaction with AI. *Advanced Engineering Informatics*, 61:102532, 2024.

[19] G. Franz and J.M. Wiener. From Space Syntax to Space Semantics: A behaviorally and perceptually oriented methodology for the efficient description of the geometry and topology of environments. *Environment and Planning B: Planning and Design*, 35(4):574–592, 2008.

[20] A. Grasso-Cladera, M. Arenas-Perez, P. Wegertseder-Martinez, E. Vilina, J. Mattoli-Sanchez, and F.J. Parada. Neuroscientific insights into the built environment: A systematic review of empirical research on indoor environmental quality, physiological dynamics, and psychological well-being in real-life contexts. *International Journal of Environmental Research and Public Health*, 22(6):824, 2025.

[21] G. Pei, H. Li, Y. Lu, Y. Wang, S. Hua, and T. Li. Affective computing: Recent advances, challenges, and future trends. *Intelligent Computing*, 3:0076, 2024.

[22] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

[23] P. Bhattacharya, D. Saraswat, D. Savaliya, S. Sanghavi, A. Verma, V. Sakariya, S. Tanwar, R. Sharma, M.S. Raboaca, and D.L. Manea. Towards future internet: The metaverse perspective for diverse industrial applications. *Mathematics*, 11(4):941, 2023.

[24] I. Csiszár and P.C. Shields. Information theory and statistics: A tutorial. *Foundations and Trends in Communications and Information Theory*, 1(**4**):417–528, 2004.

[25] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11:355–607, 2019.

[26] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems 26*, pages 2292–2300, NV, USA, 2013.

[27] J. Feydy, T. Séjourné, F.-X. Vialard, S. Amari, A. Trouvé, and G. Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. *In The 22nd international conference on artificial intelligence and statistics*, pages 2681–2690, Okinawa, Japan, 2019.

[28] J.B. Im, M. Joo, K.T. Lee, and J.H. Kim. A mixed reality approach to identify cognitive performance and mental states in preferred vs. non-preferred individual and collaborative work environments. *Building and Environment*, 274:112806, 2025.

[29] S.H. Cha, C. Koo, T.W. Kim, and T. Hong. Spatial perception of ceiling height and type variation in immersive virtual environments. *Building and Environment*, 163:106285, 2019.

[30] J.H. Ma, E. Erdogmus, S. Kangisser, and E. Yang. A comparative analysis of the effectiveness of immersive virtual reality on end-user design review. *Building and Environment*, 267:112237, 2025.

[31] K.T. Lee, J.B. Im, S.J. Park, and J.H. Ki. Conceptual framework to support personalized indoor space design decision-making: A systematic literature review. *Buildings*, *12*(6):716, 2022.

[32] E. Cambria. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107, 2016.

Defect Specification Modeling Enabled Comprehensive Rating System with Individual and Collective Impact Analysis

Youheng Guo[1,2], Xuesong Shen[1], Khalegh Barati[1], and James Linke[2]

[1]School of Civil and Environmental Engineering, The University of New South Wales (UNSW Sydney), Sydney, NSW 2052, Australia
[2]GeoAI, Sydney, NSW 2019, Australia
youheng.guo@student.unsw.edu.au, x.shen@unsw.edu.au, khalegh.barati@unsw.edu.au, j.linke@geoai.au

Abstract

As the safety of civil infrastructure gains more attention nowadays, inspection plays an important role in evaluating structural conditions. Ranking the level of defect is one of the most direct ways to understand the status of the building. Traditional inspections are based on handwritten records from human inspectors, which tend to be subjective and lack consistency. Therefore, digitalizing the data is a promising solution for recording. Based on the captured data, surface defects can be visualized and parameterized. This research aims to provide an innovative framework to analyze the effect of the defects based on Defect Specification Modeling (DSM) and rank the severity of different sections. A comprehensive diagnosis is conducted by considering the type, dimension, and location of the defects. It can identify structurally dangerous areas and provide solid evidence for decision-making. A case study carried out on a cooling tower is presented to validate the proposed approach. The comparison between severity rankings based solely on defect area and a comprehensive defect index reveals the practicability of the proposed methodology.

Keywords –

Defect Specification Modeling; Defect Ranking; Decision Making

1 Introduction

Traditionally, inspections have been conducted by human inspectors. Although these inspectors possess extensive experience, the reliability of inspection records is difficult to maintain, as report quality varies between individuals, potentially causing inconsistencies within a project. Moreover, manual inspection processes are prone to introduce errors, making them inefficient and time-consuming. In recent years, advancements in Computer Vision (CV) and Light Detection and Ranging (LiDAR) technologies have facilitated the digitization of inspection records in formats such as photographs, mesh models, and point clouds. Additionally, defect Building Information Modeling (BIM) has been developed to visualize and parameterize structural defects. However, limited analysis has been conducted on these defect models. The research gap remains in effectively utilizing the generated defect BIM models to extract information for decision support and maintenance strategies. Furthermore, there are no detailed standards for assessing the impact of defects on the structural integrity of the target infrastructure.

This paper aims to utilize Digital Specification Modeling (DSM) to develop a systematic framework for leveraging the severity of existing defects while considering multiple factors defined in relevant standards. By analyzing the provided defect model, an index for different defect can be calculated, enabling the ranking of severity across different structural areas. Based on this result, maintenance strategies can be optimized for varying situations. The main contribution of this work is directing the traditional experience-based structural condition assessment to a more objective and automated process which can benefit decision-making.

2 Related Work

With the acquired defect information, conducting a thorough defect rating is essential for efficient decision-making and objective asset management over the existing structure. A well-structured rating system allows stakeholders to prioritize maintenance efforts, allocate resources effectively, and mitigate potential structural failures.

2.1 Defect Modeling in BIM

BIM is a process involving collaborating between

different construction stakeholders based on a 3D model. Borrowing the concept, DSM will enable structural analysis, defect ranking, and even assisting decision making.

Borin and Cavazzini [1] recreated the Three-Dimensional (3D) shape of concrete spalling pathology in the BIM model which acts as a decision support system for the built environment. To automate the onsite data collection process and enhance the reliability of diagnosis, Dias, et al. [2] discussed the application of new inspection techniques. They applied severity of degradation to evaluate the global performance of the given façade. Bolourian [3] created a public point cloud dataset for surface defect semantic segmentation and proposed a Deep Learning (DL) approach to detect different types of concrete surface defects in point clouds. Guo, et al. [4] proposed a Convolutional Neural Network (CNN) based approach to detect minor cracks on concrete surface. By facilitating the automated crack detection method, Guo, et al. [5] proposed an innovative approach to use photogrammetry to acquire spatial information to quantify the detected crack on image. Choi, et al. [6] classified the defects as cracks, surface deterioration, displacement, and deformation. Based on the extracted information, they proposed a semiautomated visual approach to filter building condition information and used Dynamo to convert the information into BIM. Yiğit and Uysal [7] generated a Level of Detail 3 (LOD3) Digital Twin (DT) by triangulating the high-resolution images from Unmanned Aerial Vehicle (UAV). Cracks are automatically detected with the aid of DL and converted into vectors. They think the proposed method is suitable for rapid evaluation in terms of Structural Health Monitoring (SHM). Bahreini and Hammad [8] introduced the Ontology for Concrete Surface Defects (OCSD) to provide comprehensive data related to concrete surface defects, encompassing inspection, diagnosis, defect concepts, repair, rehabilitation, and replacement.

2.2 Regulations for Defect Categorizing

Among different methods, United States Army Construction Engineering Research Laboratories (USACERL) appeared to be the most comprehensive standard. Uzarski [9] added descriptions for the USACERL Condition Index. But it is still based on inspectors' experiences when it comes to predicting the condition and repairing time.

Amani, et al. [10] overviewed Component Condition Index System (CCIS) and investigated from technology, measurement, assessment function, and component maintenance. The authors stated that Condition Index (CI) is fundamental for deciding time and cost to invest in repairing or replacing existing facilities. The CI scale rates the status of health from 0 to 100 or D to A based on rating scale theory. It is believed that condition description based on visual guidance is more efficient and it would become an integral infrastructure management system. Queensland-Government [11] suggested that the type, width, location, and orientation of cracks can significantly affect the structure durability. Therefore, accurately measuring cracks is essential for monitoring purposes. The severity is classified as hairline (less than 0.1mm), minor (0.1-0.3mm), moderate (0.3-0.6mm), and severe (more than 0.6mm). According to Queensland-Government [11], the conditions of structure are categorized into five states from good, fair, poor, very poor, to unsafe. The exposure classification is defined as relatively benign, mildly aggressive, aggressive, and most aggressive. Vicroads [12] noted that if the spall is less than 25 mm in depth or 15 cm in diameter, it can be categorized as a small spall.

Although researchers have been working on detecting existing defects and creating BIM based defect models which enhance both efficiency of capturing process and consistency of the data, several gaps remain in the prior studies. Most of the current defect models are used for visualization, while limited research has been allocated to exploit the captured defect data and develop a reproducible assessment system for scoring the condition of building components. In addition, Existing regulations can provide defect classification thresholds but lack a systematic method to quantify the defect attributes and rank the defect accordingly.

Therefore, this study aims to develop a structured defect ranking framework connecting spatial defect information and a transparent scoring method. By evaluating and converting different attributes, a more objective condition assessing method can be achieved and reproduced. Furthermore, the proposed framework can facilitate optimized maintenance and decision making for infrastructure asset management.

3 Research Methodology

This severity scoring framework is developed from structural maintenance and infrastructure inspection standards. By analyzing the provided defect model, a series of defect attributes are obtained and integrated into the proposed framework. The evaluation criterion is normalized ranging from 1 to 10 (least to most severe) which is consistent with infrastructure condition index system. The scoring discipline originates from engineering judgement and widely used inspection standards. Several factors influence the assessment results, including defect type, size, location, exposure, duration of existence and shape. Variations in prioritization can lead to different outcomes, affecting the severity ranking of different structural areas.

A flowchart outlining the assessment process for

individual defects is presented below (see Figure 1). Both the influence of individual defects and the combined effect of multiple factors should be considered as seen in Equation 1, where CS stands for Component Score, IDS stands for Individual Defect Score, GF stands for Group Factor, and CF stands for Combination Factor. Although detailed measurements are presented below, weight and scaling factors will need to be adjusted according to specific cases.

$$CS = \sum (IDS \times GF \times CF) \quad (1)$$

3.1 Individual Effect

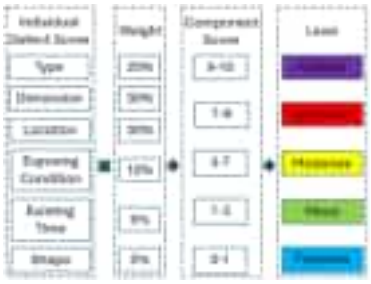


Figure 1. Proposed systematic workflow for individual defect ranking of component.

As seen in Figure 1, the component score is determined based on three key factors: defect ranking criteria, defect score (S_i), and weight (WC_i) (see Equation 2). The defect ranking criteria include defect type, dimensions, location, exposure conditions, and duration of existence. For each criterion, scores are normalized across different conditions using a ten-point scale, where a score of 10 represents the most severe extent, and a score of 1 represents the least severe level.

$$IDS = \sum_{i=1}^{n} S_i \cdot WC_i \quad (2)$$

3.1.1 Defect Type

Different types indicate different levels of potential risk. For example, cracks compromise the structural integrity of the surface, potentially leading to internal damage at early stage. While spalling typically affects a larger surface area and involves internal material exposure, it is more severe than cracking. Overall, if the defect type is more likely to pose danger to structural integrity, the severity score will correspondingly be higher.

3.1.2 Dimension

Defect dimension is one of the most direct factors that represent the level of degradation. For most defects, the affected area serves as the primary measurement criterion, while other dimensions, such as depth, width, and length, are also essential. Defects can be classified into minor, moderate, and major extents according to specific thresholds. Generally, larger defect dimensions correspond to higher severity scores.

3.1.3 Location

The location of identified defects is a significant factor influencing structural stability. Defects near critical components incur higher severity scores. The most critical areas (rated 8-10) include load-bearing components such as joints, columns, beams, and walls, where any flaw can potentially lead to structural collapse. Besides, defects at elevated locations pose significant risks, as falling debris may threaten human safety and the surrounding environment. Moderate-risk locations (rated 4-7) include slabs, floors, and ceilings. Defects in these areas may cause localized hazards but are unlikely to result in global structural failure. Non-load-bearing structures (rated 1-3), such as partition walls and cladding, are considered non-critical, as defects in these areas are less likely to cause severe structural problems.

3.1.4 Exposing Condition

Exposure conditions play a crucial role in the progression of defects. Structures fully exposed to harsh outdoor environments, such as costal erosion, high winds, severe storms, heavy snow, or intense heat, experience the highest severity levels (rated 9-10), as these extreme conditions significantly accelerate degradation and corrosion. Normal outdoor environments (rated 6-8) also have adverse effects on structural integrity, though to a lesser extent. Semi-enclosed areas, such as parking garages, patios, and sunrooms, are less affected by weather conditions and are assigned moderate severity scores (rated 3-5). Indoor environments remain relatively stable, experiencing minimal environmental impact, and therefore receive the lowest defect severity scores (rated 1-2).

3.1.5 Duration of Existence

The duration of a defect's existence is another critical factor influencing structural stability. It can be divided into three phases: emerging, developing, and existing. Developing defects (rated 8–10) are considered more significant than emerging ones, as their progression indicates increasing risk. Emerging defects (rated 4-7) may not be immediately critical. However, if a known defect is detected constantly growing, it poses a greater hazard compared to newly formed defects. In contrast, existing defects (rated 1-3) are the least concerning because they have remained stable over time without posing a major risk.

3.1.6 Shape

Moreover, the shape of a defect plays a crucial role in

surface integrity. For instance, in the case of spalling, the characteristics of its edges can vary. Defects with spiky or sharp edges (rated 6-10) tend to exhibit higher stress concentrations at the corners which can accelerate defect propagation. In contrast, defects with rounded or smooth edges (rated 1-5) are more stable, as the applied forces are more evenly distributed, reducing the likelihood of further deterioration.

3.2 Collective Effect

Apart from the impact of individual defects mentioned in the previous section, the collective effect of multiple defects should also be investigated, as defects often occur in clusters. When multiple defects appear within a single structural unit, various factors, such as intensity influence and center influence, should all be considered. The combined effect of these factors is illustrated in Figure 2.

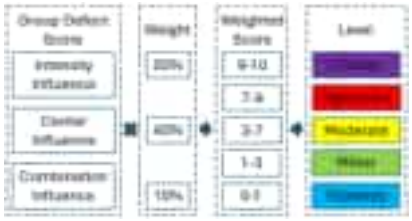


Figure 2. Proposed systematic workflow for group defect ranking of component.

3.2.1 Intensity Influence

Distance is the primary indicator of the combined impact of multiple defects. When defects are densely concentrated within a specific area, the local structural zone becomes more vulnerable, increasing the likelihood of sudden failure. Two approaches can be employed to quantify the level of defect concentration.

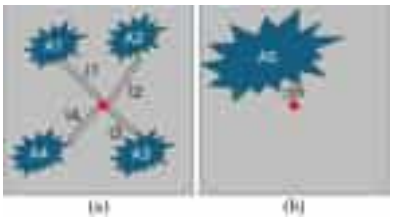


Figure 3. Combined impact factor: (a) adjusted weighted approach, and (b) finding the initial defect.

The first step involves calculating the adjusted weighted intensity factor (see Equation 3) based on a distance function, as illustrated in Figure 3 (a). In this equation, A_i represents the defect area, d_i denotes the distance between two defects, β is the scaling factor, and I_i is the intensity factor when a single defect is positioned at the center. Different initial defect locations may lead to varying results. Since this index is computed with respect to a single center, the outcome may be biased. To enhance scientific accuracy and minimize bias, the weighted intensity factor WI is introduced (see Equation 4). This factor is derived by averaging the individual intensity indices within the same structural component, as shown in Figure 3 (b).

$$I_i = \frac{\sum_1^i A_i e^{-\beta d_i}}{\sum_1^i A_i} \quad (3)$$

$$WI = \frac{\sum_1^i I_i}{i} \quad (4)$$

3.2.2 Center Influence

In addition to the influence of defect density, the group effect on the structural component itself should also be analyzed, as illustrated in Figure 4. In this figure, $A_1 - A_4$ represent defect areas, while $D_1 - D_4$ denote vectors extending from the geometric center of the structural component to the centers of individual defects. The red dot indicates the geometric center of the component.

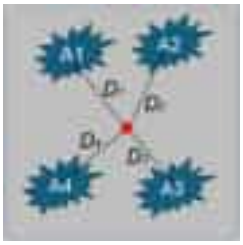


Figure 4. The center impact of the defects towards the center of the component.

The center impact of defects is a key parameter for evaluating their overall effect on a structural component. This index accounts for both the direction and distance of each defect relative to the component's geometric center. The magnitude of the resulting general vector represents the overall impact. The calculation of the center impact index is detailed in Equations 5-7, where k is the scale factor, A_i represents the defect area, D_i is the vector from the component center to defect center, W_i is the weighted vector, V_x is the influence vector in x direction, V_y is the influence vector in y direction, V_z is the influence vector in z direction, and V_i is the gross center influence vector.

$$W_i = D_i \cdot A_i \quad (5)$$

$$V_i = \frac{\sum_1^i k \cdot W_i}{\sum_1^i A_i} \quad (6)$$

$$|V_i| = \sqrt{{V_x}^2 + {V_y}^2 + {V_z}^2} \quad (7)$$

3.2.3 Combination Influence

When more than one kind of defect appears on the same component, combination factor is introduced to calculate the interaction effects between different types of defects. Different combinations of multiple defects can signify varying structural issues. For example, in the case of corrosion, cracking, and spalling, severity increases progressively from the infiltration of greenish deposits to cracks and then to a peeled concrete surface. The presence of both corrosion and cracks generally has a lower CF compared to the combination of cracks and spalling, which directly endangers structural integrity. If only one type of defect is involved in the component, CF is not applied and regarded as the minimum of 1.

4 Case Study

To evaluate the effectiveness of the proposed rating system, a case study is conducted on an aging cooling tower. The study begins with a brief introduction to the background of the cooling tower. High-resolution images were first captured by a UAV (see Figure 5), as well as LiDAR point cloud data using a terrestrial laser scanner (see Figure 6). Subsequently, a spalling BIM model (see Figure 7) is generated in Autodesk Revit using. Finally, a detailed analysis is performed based on the proposed methodology, utilizing the extracted spalling information to assess defect severity.



Figure 5. Captured image of spalling near the top ring of an aging cooling tower.

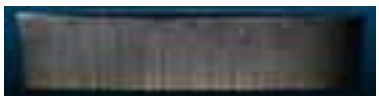


Figure 6. The captured LiDAR point clouds of defective area near the top ring of an aging cooling tower.



Figure 7. The BIM model of the aging cooling tower with spalling spread on the surface.

4.1 Project Background

Since the cooling tower became operational, high-temperature steam has condensed into moisture near the top ring, leading to the gradual corrosion of the concrete surface, as shown in Figure 5. After decades of cyclic exposure, the surface began to swell and eventually detach, resulting in concrete debris falling. These detached fragments pose a significant safety risk to workers on the ground. More critically, the exposure of reinforcement bars accelerates their deterioration, as continuous steam exposure enhances corrosion. Over time, this degradation may damage the stability of the upper structure. Therefore, monitoring the distribution of spalling and assessing its severity across different areas is important for ensuring structural integrity and worker safety.

4.2 Data Analysis

As outlined in Section 3, the individual effects are analyzed in the context of a specific case where spalling is the main defect, and the structure exhibits uniform properties. Additionally, the group effects, primarily including intensity influence and center influence, are quantified according to the conditions observed in the cooling tower. These calculations provide a comprehensive assessment of defect severity.

4.2.1 Individual Effect

The recorded spalling areas, measured in square meters, range from $0.03m^2$ to $4m^2$. These areas are categorized into three groups based on area: minor spalling ($0m^2$-$0.5m^2$), moderate spalling ($0.5m^2$-$1m^2$), and major spalling ($1m^2$-$4m^2$). In this case, a total of 157 minor defects, 54 moderate defects, and 25 major defects were identified. The severity scores assigned to these categories are 3 points for minor defects, 7 points for moderate defects, and 10 points for major defects.

The significance of location is primarily determined by its impact on load-bearing structures. In the case of the cooling tower, the entire structure is axisymmetric, meaning that all concrete panels serve an equivalent functional role. Consequently, height is the key variable influencing location-based importance. The height range is classified into three categories: minor effect (605m-608m), moderate effect (608m-612m), and major effect (612m-615m). Based on these criteria, 17 spalling locations are classified as non-critical (scoring 3 points), 85 locations are categorized as concerning (scoring 7 points), and 134 locations are identified as critical (scoring 10 points).

Regarding shape, no prominent spiky edges were observed among the existing spalling. Instead, most defects exhibit irregular forms. Some examples of these irregular and smooth-edged shapes are presented in Figure 8. Since the defects lack sharp edges, a severity score of 6 is assigned to irregularly shaped spalling, while round-shaped defects are given a lower severity score of 1.

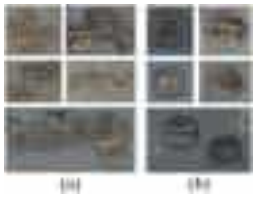


Figure 8. Two shapes of defects: (a) irregular/spiky shapes, and (b) smooth shapes.

As observed in Figure 5, the defects near the top ring primarily consist of spalling. Due to the lack of recorded data, the exact duration of defect existence remains unknown. Similarly, the exposure conditions are largely uniform because all defects are located on the concrete surface at similar heights. Therefore, the indexes for defect type, existing time, and exposure condition can be excluded from the analysis.

As discussed above, three main factors – spalling area, location, and shape are considered in the evaluation. The weighting of assessment criteria represents the relative impact of severity for each type of defect. The weights are determined based on practical inspection experience and engineering judgment in this study. Dimensions of defects are put in the first place as surface material loss is directly related to structural integrity. Location of defects is regarded as the second important factor as the appearance in critical areas can also pose significant safety risks to structural capacity. Shape of defect is allocated a lower weight because the influence is less compared to other two factors. Future studies will adopt more structured approaches or expert surveys to refine the scheme. The weight distribution for these factors is assigned as follows: 50% for area, 40% for location, and 10% for shape. The calculated factors for Panel 1 are provided as an example (see Table 1).

Table 1 Scaled area, location, and shape factors for Panel 1.

Spalling No.	1-1	1-2	1-3	1-4	1-5	1-6	1-7
Area Factor	10	3	3	7	3	7	7
Location Factor	10	10	10	10	10	7	7
Shape Factor	1	1	1	1	1	1	1
Scaled Area Factor	5	1.5	1.5	3.5	1.5	3.5	3.5
Scaled Location Factor	4	4	4	4	4	2.8	2.8
Scaled Shape Factor	0.1	0.1	0.1	0.1	0.1	0.1	0.1

4.2.2 Collective Effect

This section examines the impact caused by multiple defects, specifically focusing on intensity influence and center influence. Since only a single defect type is present, the combination factor is not considered. In this case, the structural unit used for calculating the group effect is a concrete panel. Take Panel 1 as an example, the centers of the identified spalling defects are marked in Figure 9 (a), while the geometric center of the panel is indicated in Figure 9 (b).



Figure 9. Panel 1: (a) numbered centers of spalling, and (b) the center of panel component.

Following Equation 3, the scaling factor β is set to 0.1. The intensity index for different spalling instances is presented in Table 2. As observed, variations in the spalling center result in changes in the intensity index, ranging from 0.7458 to 0.8599. To mitigate potential bias, the derived single intensity indexes are averaged according to Equation 5.

Table 2. Individual intensity factors for spalling on panel 1.

Spalling	Single Intensity Index
1-1	0.8599
1-2	0.7739
1-3	0.8028
1-4	0.8529
1-5	0.8101
1-6	0.7458
1-7	0.7808

For the center index, individual weighted vectors are calculated using Equation 5. Afterward, the scaling factor k is set to 0.2 to obtain the magnitude of the gross center influence index for each concrete panel, as calculated using Equation 6 and 7.

Since only spalling appears on the concrete surface, the combination factor is excluded from the analysis. The group effect factors are weighed as follows: 60% for the intensity influence index and 40% for the center influence index. The intensity influence, center influence index, and combined group effect factors for panel 1 to 10 are listed in Table 3.

Table 3. Calculated influence index for panels 1 to 10.

Panel	Intensity Influence Index	Center Influence Index	Combined Group Index
1	0.8037	0.7155	0.8
2	0.8060	0.8254	0.8
3	0.7568	0.5501	0.7
4	0.7509	0.4367	0.6
5	0.8303	0.4273	0.7
6	0.8111	0.4015	0.6
7	0.8037	0.3436	0.6
8	0.7808	0.5214	0.7
9	0.8368	0.6863	0.8
10	0.8181	0.6571	0.8

To validate the statistics derived, the distribution of spalling on each concrete panel is shown in Figure 10. As previously discussed, the more concentrated the spalling defects are in each area, the higher the value of the intensity influence index. For example, even with a similar amount of spalling, the intensity index of Panel 9 is higher than that of Panel 8 due to the more concentrated spalling in Panel 9. For the center influence index, the value increases as the spalling group moves farther away from the center of the panel. As seen from panel 25 to 27, the center influence index decreases as the defect groups move closer to the panel center.



Figure 10. The projected top ring panels of defects.

4.3 Panel Defect Index

Since both the individual defect scores and group factors are available, the panel defect index can be determined. Figure 11 presents a comparison between the panel defect index obtained using the proposed methodology and the index based solely on the spalling area. The top 10 highest indexes are highlighted in red, the 10 lowest indexes in green, and the moderate values in yellow. The differences in severity rankings indicate that the proposed method offers a more comprehensive assessment by incorporating multiple defect-related factors.



Figure 11. Comparison between the panel defect index and sum of spalling area.

To better understand the distribution of defects, spalling is modeled into different sizes to indicate the affected area (see Figure 12 (a)). The colors represent varying severities, ranging from green (minor) to yellow (moderate) and red (severe). Additionally, the categorized concrete panels based on the defect index are presented in Figure 12 (b). A combined visualization of both the panels and spalling is displayed in Figure 12 (c) to provide a comprehensive overview of the structural condition.



Figure 12. (a) The categorized spalling: green ($0m^2 – 0.5m^2$), yellow ($0.5m^2 – 1m^2$), and red ($1m^2 – 4m^2$), the categorized panel based on defect index: green (minor), yellow (moderate), and red (major), and (c) the categorized panel and spalling

4.4 Validation

The proposed framework is evaluated through comparative analysis and engineering consistency assessment. Compared with the conventional area-based approach, the proposed method derives a more comprehensive result by considering multiple attributes. The assessment is also consistent with engineering deductions, when higher concentration and more critical location can lead to higher scores. Future work is expected to perform a direct comparison with expert assessments.

5 Discussion

The proposed methodology is applied to an aging cooling tower to demonstrate its feasibility. A semantically rich defect specification model is developed, incorporating essential defect and component rating information. To enhance interpretability, the severity of defects is visualized using a color-coded system, facilitating intuitive assessment of structural conditions. This enhanced visualization supports decision-making and maintenance strategy development by leveraging the defect BIM model and parameterized information. The proposed rating system is designed to be adaptable across different projects. Although certain criteria involve limited engineering judgement, explicit scoring rules can maintain repeatability and consistency. One of the main challenges remains the absence of common standards for validating results which can affect the reliability of evaluations. Future research will focus on standardizing the assessment criteria to improve the consistency and robustness of the defect rating system. The integration of Knowledge Graph (KG) could facilitate a more thorough understanding of defect propagation and interdependence.

6 Conclusions

The increasing number of aging infrastructure assets can potentially pose dangers to structural safety and human users. To mitigate the bias introduced by traditional human-based inspections, this study presents a practical workflow - from rating the condition of defects to visualizing categorized defects within a BIM model. A comprehensive assessment criterion is proposed, considering both individual and grouped defect effects. The individual defect assessment includes factors such as type, location, dimension, exposure condition, existing time, and shape. Each factor is assigned a score based on its severity. By multiplying the weights of these factors and summing the scaled values, the overall score for individual defect can be calculated. This score is then aggregated at the component level. Next, group factors – including intensity, center, and combination influences - are evaluated. Integrating both individual and group factors provides a more reasonable defect score for each component.

Acknowledgments

This research is supported by funding from the Australian Research Council (ARC) Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), and the ARC Research Hub for Human-Robot Teaming for Sustainable and Resilient Construction (IH240100016).

References

[1] P. Borin and F. Cavazzini. Condition assessment of RC bridges: Integrating machine learning, photogrammetry and BIM. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:201–208, 2019.

[2] I. S. Dias, I. Flores-Colen, and A. Silva. Critical analysis about emerging technologies for building's façade inspection. *Buildings*, 11(2):53, 2021.

[3] N. Bolourian. *Point cloud-based deep learning and UAV path planning for surface defect detection of concrete bridges*. Concordia University, 2022.

[4] Y. Guo, Z. Wang, X. Shen, K. Barati, and J. Linke. Automatic detection and dimensional measurement of minor concrete cracks with convolutional neural network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:57–64, 2022.

[5] Y. Guo, X. Shen, J. Linke, Z. Wang, and K. Barati. Quantification of structural defects using pixel level spatial information from photogrammetry. *Sensors*, 23(13):5878, 2023.

[6] M. Choi, S. Kim, and S. Kim. Semi-automated visualization method for visual inspection of buildings on BIM using 3D point cloud. *Journal of Building Engineering*, 81:108017, 2024.

[7] A. Y. Yiğit and M. Uysal. Automatic crack detection and structural inspection of cultural heritage buildings using UAV photogrammetry and digital twin technology. *Journal of Building Engineering*:109952, 2024.

[8] F. Bahreini and A. Hammad. Developing an ontology for concrete surface defects to enhance inspection, diagnosis and repair information modeling. *Infrastructures*, 9(12):220, 2024.

[9] D. Uzarski. *Development of condition indexes for low volume railroad track*. 1993.

[10] N. Amani, M. Nasly, and R. A. Samat. Infrastructure component assessment using the condition index system: Literature review and discussion. *Journal of Construction Engineering and Project Management*, 2(1):27–34, 2012.

[11] Queensland Government. *Structures inspection manual*. Structures inspection manual. On-line: https://www.tmr.qld.gov.au/_/media/7c6ed28706904a0fb9664bddccc01fc2.pdf, Accessed: 31/12/2025.

[12] VicRoads. *Bridge maintenance and repair manual*. Bridge maintenance and repair manual. On-line: https://content.vic.gov.au/sites/default/files/2024-05/Bridge-Maintenance-and-Repair-Manual-%28v1.5%29.pdf, Accessed: 31/12/2025.

A Systematic Study on Instance-Level Bridge Point Cloud Segmentation under Cross-Type and Synthetic–Real Training Conditions

Jeong Kyu Lee[1], Ui Chan Lee[1], Jong Won Ma[1]
[1] School of Civil and Environmental Engineering, Yonsei University
smile6569@yonsei.ac.kr, air0223@yonsei.ac.kr, jongwon.ma@yonsei.ac.kr

Abstract -

Bridge point cloud instance segmentation is essential for digital twin construction and automated bridge management. However, deep learning–based approaches are limited by the scarcity of instance-level annotated real-world data and the domain gap between synthetic and real point clouds. In practical end-to-end Scan-to-Bridge Information Modeling(BrIM) pipelines, segmentation models must operate directly on raw point clouds without assuming prior bridge-specific domain knowledge or relying on domain-driven preprocessing.

To address these challenges, synthetic bridge point clouds for slab and PSC girder bridges were generated using physics-based LiDAR simulation, with a unified component definition applied to both synthetic and real data to construct consistent instance-level annotations. Background elements, including terrain and vegetation, were incorporated into the synthetic scenes to reduce the gap between realistic conditions. Experimental results show that synthetic-only training yields the lowest performance, followed by real-only training, while the best performance is achieved when synthetic and real data are jointly used.

For the Japan dataset, mIoU and mAP improved from 0.253 and 0.131 to 0.531 and 0.638, respectively. For the Cambridge dataset, although mIoU slightly decreased from 0.556 to 0.549, mAP increased from 0.617 to 0.688, indicating improved instance-level performance.

Instance-level segmentation is particularly important in bridge applications, as maintenance and long-term monitoring require component-wise analysis and consistent tracking of *maintenance-relevant structural elements or structural systems*. Given the limited visibility of abutments in real-world and their frequent exclusion in prior studies, we further evaluated performance under scenarios with deck and background components to reflect realistic end-to-end deployment conditions.

Keywords –

Scan-to-BrIM; Instance segmentation; Synthetic point clouds;

1 Introduction

Bridges are critical infrastructure that connects local communities and supports economic activities, making efficient management and maintenance essential. However, conventional bridge management heavily relies on visual inspections and manual operations, which are time-consuming, labor-intensive, and costly [1]. As a result, digital transformation of bridge management has gained increasing attention, with digital twins and Scan-to-Bridge Information Modeling (Scan-to-BrIM) emerging as promising solutions [2].

Within Scan-to-BrIM pipelines, bridge components must be identified from point clouds acquired via laser scanning or photogrammetry and subsequently reconstructed as structured geometric models. Point cloud segmentation therefore plays a central role in enabling downstream modeling and analysis tasks [3]. In particular, instance-level segmentation is required for practical bridge applications, as maintenance, inspection, and long-term monitoring depend on component-wise analysis and consistent tracking of maintenance-relevant structural elements or systems. For this purpose, recent deep learning–based approaches have shown clear advantages over traditional rule-based methods in handling complex bridge geometries and diverse structural configurations [4].

However, the effectiveness of deep learning–based bridge segmentation is fundamentally constrained by the availability and quality of training data. Generating accurate instance-level annotations for real-world bridge point clouds requires substantial domain expertise, manual effort, and time, while large-scale acquisition of real bridge data is often limited by logistical and financial constraints. To address these challenges, many studies have explored training strategies that combine synthetic and real bridge point cloud data, leveraging synthetic data to mitigate labeling costs and data scarcity while incorporating real data to improve realism and generalization.

The synthetic and real data generation pipeline adopted in this study builds upon established workflows in prior literature, while incorporating methodological refinements tailored to bridge point cloud instance segmentation.

Despite recent progress, many existing studies still rely on simplified problem formulations for bridge point

cloud instance segmentation, such as excluding certain structural components (e.g., abutments), merging partially visible elements into dominant instances like decks, or conducting segmentation on pre-filtered point clouds without background points. These simplifications embed strong, often implicit, prior assumptions about bridge geometry and data preprocessing, and therefore depart from realistic end-to-end Scan-to-BrIM workflows, where models must directly process raw, cluttered point clouds acquired from field measurements.

To address this gap, we propose a stepwise experimental framework that explicitly controls and progressively relaxes these assumptions. Rather than treating instance segmentation under a single idealized setting, our framework systematically evaluates model behavior across multiple levels of realism by varying the presence of background points and structurally ambiguous components. This design enables a controlled analysis of how training strategies generalize as the problem formulation transitions from idealized to practically deployable conditions.

2 Related work

2.1 Semantic Segmentation of Bridge Point Clouds

Bridge point cloud segmentation is a fundamental technology for Scan-to-BIM, automated bridge inspection, and digital twin construction, as it enables the identification and separation of structural components from raw 3D point clouds at both semantic and instance levels. Accurate segmentation is a prerequisite for downstream tasks such as structural assessment, deformation analysis, and maintenance planning.

Early studies on bridge point cloud segmentation primarily relied on handcrafted geometric features and rule-based heuristics, including curvature, normal vectors, region growing, and threshold-based clustering [5]. While these approaches achieved reasonable performance in controlled environments, they exhibited strong dependence on manually tuned parameters and limited robustness when applied to complex real-world scenes with occlusions, noise, and varying point densities.

With the rapid advancement of deep learning, data-driven approaches have become dominant in point cloud segmentation. Methods based on PointNet, PointNet++ [6-7], and sparse convolutional networks significantly improved segmentation accuracy by learning discriminative features directly from point cloud data. More recently, transformer-based instance segmentation frameworks have attracted increasing attention due to their ability to model long-range dependencies and global contextual relationships, leading to improved performance in complex and cluttered scenes [8].

Despite these advances, applying deep learning–based segmentation methods to bridge point clouds remains challenging. Bridge datasets often suffer from severe class imbalance, where small components such as abutments or parapets contain significantly fewer points than major components like decks or girders. In addition, acquiring large-scale, labeled bridge point clouds is costly and time-consuming, limiting the availability and diversity of training data [9]. These challenges restrict the generalization capability of existing models when deployed in real-world bridge environments.

2.2 Synthetic Bridge Point Cloud Generation

Deep learning–based bridge point cloud segmentation requires large amounts of labelled data. However, constructing component-level ground-truth annotations for real bridges is extremely costly and labor-intensive. To address this limitation, extensive research has explored the use of synthetic bridge point cloud data.

Early synthetic data generation methods typically sampled point clouds directly from 3D CAD or mesh-based bridge models [10]. Although these approaches enabled efficient data generation and precise labelling, they failed to capture critical characteristics of real-world point clouds, such as sensor noise, non-uniform sampling density, occlusion effects, and background clutter. Consequently, models trained on such data often exhibited limited generalization to real-world scenarios.

To improve realism, subsequent studies adopted physics-based LiDAR simulation or UAV–SfM pipelines to generate synthetic point clouds that more closely resemble real sensor measurements [11-12]. These methods incorporated factors such as scanning geometry, measurement noise, and partial occlusion, leading to improved data fidelity. Nevertheless, many existing approaches remain limited to a small number of bridge types or simplified environmental settings, often neglecting surrounding terrain, vegetation, and background structures.

As a result, a persistent domain gap between synthetic and real bridge point clouds has been widely reported and recognized as a major factor degrading model performance in real-world applications [13]. This limitation has motivated recent research to investigate training strategies that combine synthetic and real data or progressively adapt models across domains. However, systematic comparisons of different training data compositions under controlled experimental settings remain limited, particularly in the context of instance-level bridge segmentation.

3 Methodology

To ground our methodology in existing research

practice, this study first examines six representative prior works that utilize the same real-world bridge point cloud dataset [14-15-16-17-18-19]. From a training-strategy perspective, most previous studies focus on improving segmentation performance by combining synthetic and real data, applying data augmentation to real data, or leveraging pseudo-labeling techniques to exploit unlabeled real-world point clouds [15-16-17]. Other approaches emphasize the generation of synthetic bridge models designed to closely resemble real bridges in order to enhance training effectiveness [8].

Meanwhile, several studies effectively simplify the segmentation task by constraining scene composition or component definitions at the dataset level. For example, in Xia et al. [14], only pier, slab, and background classes are considered, resulting in a highly simplified scene configuration. In the Yang et al. [15] and Yang et al. [17], background and abutment components are excluded altogether, meaning that the input point clouds do not represent raw scanning conditions. Although the Xu et al. [16] partially includes background and abutments, most bridges are truncated around the mid-span region, thereby avoiding complex boundary conditions near abutments.

In the Lin et al. [18], loss functions are strongly tuned based on bridge-specific structural priors, introducing explicit domain assumptions into the learning process. Furthermore, in the Chen et al. [19], abutments and piers are merged into a single category, further reducing component-level granularity and simplifying the segmentation task.

In contrast, practical end-to-end Scan-to-BrIM pipelines operate directly on raw point clouds acquired from field measurements, where background clutter and partially occluded components are inherently present and cannot be reliably pre-filtered. To reflect this practical constraint, this study explicitly distinguishes three training scenarios—synthetic data only, real data only, and combined synthetic–real data—and conducts a comprehensive, stepwise performance comparison using raw point cloud inputs under a unified model architecture and evaluation protocol. Through this analysis, we systematically investigate how training data composition influences bridge point cloud instance segmentation performance in realistic, domain-agnostic settings. A summary of performance comparisons reported in previous studies is presented in Table 1.

Table 1. Segmentation of performance comparisons reported in prior studies

Dataset	mIoU (%)	mAP (%)	OA (%)	Acc (%)
Xia et al. [14]	94.72	-	-	-
Yang et al. [15]	91.84	-	95.56	96.74
Xu et al. [16]	41.4	63.3		
Yang et al. [17]	89.83	-	95.34	94.31
Lin et al. [18]	74.42	-	87.67	86.65
Chen et al. [19]	81.0	-	-	92

3.1 Datasets

To evaluate the performance of the proposed method, two real-world bridge point cloud datasets from Japan and Cambridge were used in this study. By leveraging datasets constructed in different countries, environments, and scanning conditions, we aim to comprehensively assess the generalization capability and practical applicability of the proposed framework. Examples of the real bridge point clouds are shown in Fig. 1.



Figure 1. Representative bridge point clouds from the Japan and Cambridge datasets.

3.1.1 Japan Dataset

Japan Dataset is a proprietary dataset collected in December 2023 from nine reinforced concrete bridges located in rural areas of Japan using a Matterport Pro3 terrestrial laser scanner[20]. The scanner provides a 360° horizontal field of view and a 295° vertical field of view, with a measurement accuracy of ±20 mm at a distance of 10 m and a maximum acquisition rate of 100,000 points per second.

All points were manually annotated into five structural component classes: Abutment, Girder, Deck, Parapet, and Background. For abutment components, parts of the frontal regions were reclassified as Background to account for geometric complexity and severe class imbalance.

For this dataset, following the original bridge IDs provided in the dataset, Bridges 2, 4, 5, 6, 8, and 9 were used for training, while Bridges 1, 3, and 8 were used for testing.

3.1.2 Cambridge Dataset

Cambridge Dataset is a publicly available benchmark dataset, consisting of highway bridge point clouds collected in the Cambridgeshire region of the United Kingdom using a FARO Focus 3D X330 terrestrial laser

scanner[21]. To ensure consistency number in Japan Dataset, a total of nine bridges (Bridges 2–10) were used, excluding Bridge 1 from the original set of ten bridges.

The bridges have total lengths ranging from 55 to 91 m, span lengths of 13 to 35 m, and heights between 5 and 8 m. Structurally, they include multiple piers and horizontally curved deck geometries. All points were annotated into six structural component classes: Abutment, Pier, Girder, Deck, Parapet, and Background.

For this dataset, following the original bridge IDs provided in the dataset. Bridges 3, 4, 5, 7, 8, and 10 were used for training, while Bridges 2, 6, and 9 were used for testing.

3.2 Synthetic and Real Data Generation Pipeline

This study proposes an integrated data generation and annotation pipeline that combines synthetic and real-world data to construct training datasets for 3D instance segmentation of bridge components. An overview of the proposed pipeline is illustrated in Fig. 2.

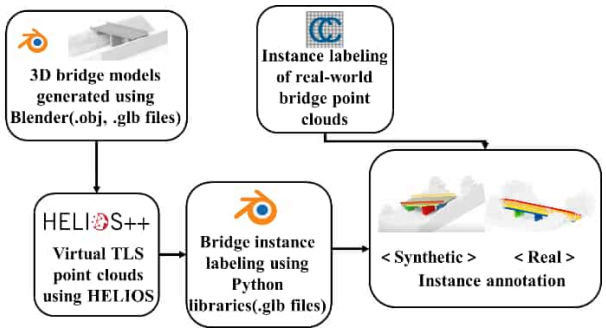


Figure 2. Overview of the proposed data generation and annotation pipeline for bridge component instance segmentation.

Bridge models of slab bridges and PSC girder bridges were first collected from the online repository www.bigdata-eng.com in IFC format and subsequently converted into OBJ files using Blender. Additional structural elements such as piers and surrounding terrain were manually modeled to construct complete 3D bridge scenes in Blender. The synthetic bridge dataset consists of 48 slab bridges and 9 PSC girder bridges, reflecting structural diversity in terms of bridge length (10–45 m for slab bridges and 30–50 m for girder bridges) and pier geometry (rectangular, square, and cylindrical). In addition, terrain-inclusive versions were generated for all bridges to account for realistic topographic conditions. Representative examples of slab and girder bridges are shown in Fig. 3. Virtual terrestrial laser scanning (TLS) was then performed on the generated 3D bridge models using HELIOS++ to produce synthetic bridge point clouds.

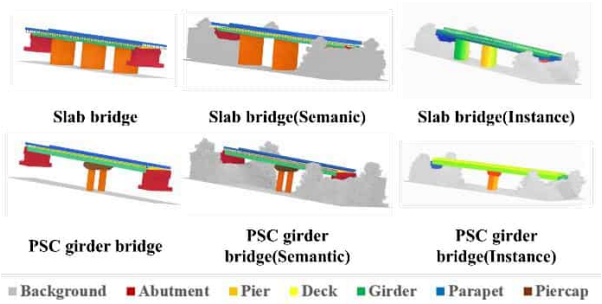


Figure 3. Representative synthetic bridge models used in this study: slab bridges and PSC girder bridges.

To emulate real-world TLS acquisition conditions, scanner placement, field of view, ray distribution, and range attenuation effects were explicitly modeled, and scanner configurations were defined based on the specifications of the RIEGL VZ-400. In particular, TLS scanners were automatically deployed at regular intervals on multiple horizontal planes derived from the bridge bounding box, including planes located above and below the bridge deck in the XY plane and an additional plane positioned 2 m below the structural center. This strategy ensured sufficient coverage of bridge geometries from multiple viewpoints. Scanners that were placed inside the bridge mesh were automatically excluded to prevent unrealistic scanning artifacts. As a result, synthetic point clouds with spatial distributions and point densities comparable to those obtained from real TLS surveys were efficiently generated. The scanner spacing was set to 10 m or 15 m depending on whether background elements were included. Terrain and vegetation were automatically placed to replicate realistic bridge environments. The synthetic dataset was ultimately divided into two variants, with and without vegetation and surrounding hills, while ground surfaces were included in both variants. Detailed HELIOS++ scanner settings are summarized in Table 2.

Following point cloud generation, a Python-based post-processing pipeline was applied to automatically assign component-level instance labels to the synthetic point clouds. Structural component information defined during the Blender modeling stage was preserved in GLB files, and a nearest-neighbor matching process between mesh surfaces and point clouds was used to consistently assign semantic labels and instance IDs to individual points. In contrast, real-world bridge TLS point clouds were manually annotated using CloudCompare.

Table 2. HELIOS++ TLS simulation settings used for synthetic bridge point cloud generation.

Parameter	Value
Pulse frequency	100,000 Hz
Scan frequency	120 Hz
Scan angle	100°

Head rotation speed	10°/s
Vertical FOV	−40° to 60°
Horizontal FOV	−180° to 180°
Trajectory interval	1.0 s

For all point cloud data, RGB values were unified to grayscale to minimize the influence of color information. A voxel down-sampling with a resolution of 0.05 m was applied, followed by spatial cropping into 10 m-sized volumes using a sliding-window strategy. Here, the overlap ratio is defined as the percentage of spatial overlap between adjacent cropped volumes. Overlap ratios of 0%, 12.5%, and 25% were applied, resulting in approximately 100k–300k points per scene. The final dataset includes both synthetic and real bridge point clouds and shares consistent class definitions and instance labeling schemes, enabling direct use for training and evaluating 3D point cloud instance segmentation models.

3.3 Experimental Setting / Evaluation Metrics

All experiments were conducted on a workstation equipped with an AMD Ryzen 7 5800X CPU, an NVIDIA RTX 3090 GPU with 24 GB VRAM, and 32 GB of system memory. The experimental environment was configured on Ubuntu 20.04, using Python 3.8 and CUDA 11.3.

In this study, Relation3D[22] was adopted as the baseline instance segmentation model, and all models were trained for 200 epochs. The training configuration of Relation3D followed the official settings for the ScanNetV2 dataset, except that the batch size was set to 10 and the number of data loader workers was set to 5. In addition, different overlap ratios (0%, 12.5%, and 25%) were applied during data preparation.

Intersection over Union (IoU) measures the degree of overlap between the predicted and ground-truth regions for a given class, indicating how accurately the predicted area matches the reference annotation. Based on IoU, Average Precision (AP) evaluates instance-level segmentation performance by summarizing the precision–recall relationship obtained from IoU-based matching between predicted and ground-truth instances. Mean Intersection over Union (mIoU) is then computed by averaging the IoU values across all classes, providing an overall measure of semantic segmentation accuracy. Finally, mean Average Precision (mAP) is obtained by averaging the AP values over all classes, reflecting the overall instance segmentation performance.

4 Experimental Results

The Japan and Cambridge datasets were used as training sets, while the synthetic dataset consisted of 48 slab bridges and 9 PSC girder bridges. The synthetic dataset was further divided into two variants depending on the presence of surrounding Background. Overlap ratios of 0%, 12.5%, and 25% were applied during data preparation to represent low, moderate, and high levels of spatial redundancy between adjacent subspaces. Specifically, 0% corresponds to a non-overlapping baseline, while 12.5% and 25% introduce progressively increasing overlap within a practical range before excessive redundancy occurs. This design enables a systematic analysis of the trade-off between computational cost and segmentation accuracy. For evaluation, the respective test sets of the Japan and Cambridge datasets were used.

From this perspective, this study compared three training scenarios based on instance segmentation that include vegetation and abutments: training with synthetic data only, training with real data only, and training with a combination of synthetic and real data. Experimental results show that training with synthetic data alone consistently yielded the lowest performance. For the Japan dataset (Table 3), under background-included settings, AP values for overlap ratios of 0%, 12.5%, and 25% were 0.057, 0.088, and 0.078, respectively. When background points were excluded, AP further decreased to 0.045, 0.040, and 0.032 under the same overlap conditions. In contrast, mIoU values were relatively higher under background-excluded settings, reaching 0.291, 0.279, and 0.242, respectively.

Table 3. Accuracies on the Japan dataset for synthetic-only training across overlap ratios (with/without background)

Background\Overlap	0%	12.5%	25%
	mIoU / mAP	mIoU / mAP	mIoU / mAP
Background(0)	0.257 /0.057	0.206 /0.088	0.210 /0.078
Background(x)	0.291 /0.045	0.279 /0.040	0.242 /0.032

Similar trends were observed in the Cambridge dataset (Table 4). When background points were included, AP values at overlap ratios of 0%, 12.5%, and 25% were 0.266, 0.279, and 0.266, respectively, whereas background-excluded settings resulted in mAP values of 0.213, 0.207, and 0.231 under the same conditions. The mIoU values were similar across all settings. Since real bridge environments require the separation of individual components under complex scenes where abutments, ground surfaces, terrain, and surrounding structures are mixed, training with realistic scenes that include

background elements is essential for instance segmentation.

Table 4. Accuracies on the Cambridge dataset for synthetic-only training across overlap ratios (with/without background)

Background\Overlap	0%	12.5%	25%
	mIoU / mAP	mIoU / mAP	mIoU / mAP
Background(0)	0.300 /0.266	0.291 /0.279	0.294 /0.266
Background(x)	0.298 /0.213	0.306 /0.207	0.299 /0.231

Training with real data only led to a clear performance improvement. However, the optimal overlap ratio varied across datasets. For the Japan dataset (Table 5), the highest performance was achieved at an overlap ratio of 12.5%, with an mIoU of 0.253 and an mAP of 0.131. In contrast, the Cambridge dataset (Table 6) achieved its best performance at an overlap ratio of 25%, yielding an mIoU of 0.556 and an mAP of 0.617. This difference is closely related to bridge scale and the crop-based data construction strategy. For structurally smaller Japanese bridges, increasing the overlap ratio can limit the model's ability to sufficiently learn component arrangements and instance-level relationships. In contrast, the longer and more complex Cambridge bridges provide richer spatial contexts, which are more favorable for learning instance boundaries.

Table 5. Accuracies on the Japan dataset under real-data-only training with different overlap ratios.

Overlap	mIoU	mAP
0%	0.075	0.055
12.5%	0.253	0.131
25%	0.181	0.112

Table 6. Accuracies on the Cambridge dataset under real-data-only training with different overlap ratios.

Overlap	mIoU	mAP
0%	0.462	0.485
12.5%	0.415	0.382
25%	0.556	0.617

The most significant performance improvements were observed when synthetic and real data were combined for training. For the Cambridge dataset, applying a 25% overlap ratio and a background-excluded setting for synthetic data resulted in an mIoU of 0.549 and an AP of 0.688. Similarly, for the Japan dataset, an overlap ratio of 12.5% with background-excluded synthetic data achieved an mIoU of 0.531 and an AP of 0.638 (Table 7). In both datasets, better performance was achieved when the synthetic data excluded background elements,

Table 7. Optimal overlap ratio selection for synthetic–real combined training and corresponding performance.

Background \Name(Overlap)	Cambridge (25%)	Japan (12.5%)
	mIoU / mAP	mIoU / mAP
Background(0)	0.530 / 0.668	0.488 / 0.543
Background(x)	0.549 / 0.688	0.531 / 0.638

which can be attributed to the fact that background characteristics were learned primarily from real data. Detailed performance results for each component are presented in Figs. 4 and 5.

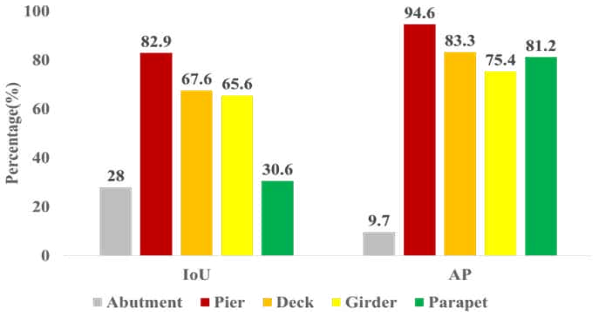


Figure 4. Component-wise instance segmentation performance on the Japan dataset.

These results indicate that synthetic data contributes beyond merely increasing the amount of training data, effectively supporting instance segmentation learning for components that are difficult to sufficiently observe in real TLS data. In particular, abutments, which are adjacent to the ground and structurally complex, often exhibit sparse point distributions and unclear boundaries with background elements, making instance segmentation challenging. Synthetic data provides clear geometric structures and instance boundaries for such components, enabling the model to learn object-level segmentation capabilities required in real end-to-end maintenance pipelines.

Furthermore, when the entire synthetic dataset was combined with real bridge data collected from different regions for training (Table 8), the Japan dataset showed improvements in mAP and mIoU from 0.131 and 0.253 to 0.238 and 0.279, respectively, compared to training with real data only. Similarly, for the Cambridge dataset, mAP and mIoU increased from 0.617 and 0.556 to 0.734 and 0.605. These results outperform the case where training was conducted using real data from the same regional domain only (mAP 0.688, mIoU 0.549), demonstrating that combining background-free synthetic data with multi-domain real data is also effective for improving domain generalization.

Table 8. Performance Comparison of Cross-Domain Real-Data Training with Fixed Synthetic Data

Overlap/Configuation	mIoU	mAP	Abutment	Pier	Deck	Girder	Parapet
			IoU/AP	IoU/AP	IoU/AP	IoU/AP	IoU/AP
Japan(12.5%)	0.279	0.238	0.069/0.002	- / -	0.374/0.138	0.460/0.446	0.214/0.365
Cambridge(25%)	0.605	0.734	0.551/0.223	0.726/0.964	0.659 / 0.834	0.689/0.772	0.402/0.876

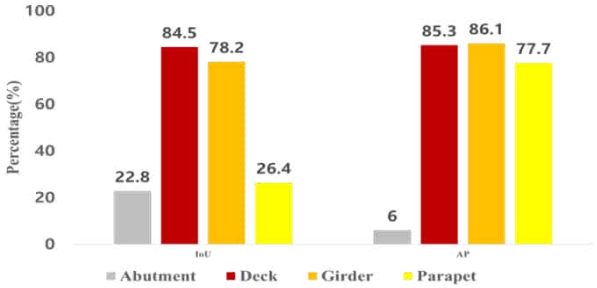


Figure 5. Component-wise instance segmentation performance on the Cambridge dataset.

Instance segmentation results for Bridge 8 in the Japan dataset and Bridge 6 in the Cambridge dataset are visualized in Fig. 6. In the visualization, background points are shown in gray, incorrectly predicted regions are highlighted in red, and correctly segmented instances are distinguished by different colors.

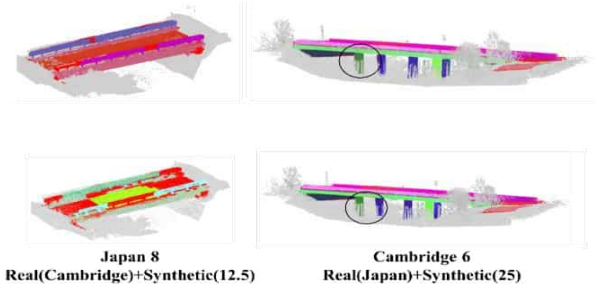


Figure 6. Instance Segmentation Results for Bridge 8 (Japan) and Bridge 6 (Cambridge)

5 Conclusion and Limitation

This study systematically analyzed the performance variations of 3D point cloud–based component-level instance segmentation for bridge maintenance and digital twin construction according to different training data compositions. The fundamental motivation for performing bridge instance segmentation is that, even when components belong to the same category, each component is treated as an independent management unit at an appropriate management level from a maintenance perspective. Multiple piers, parapets, and abutments, which are spatially separable structural units, may be classified into the same semantic class. However, in actual maintenance processes, each component exhibits different damage states, deterioration patterns, inspection histories, and repair records. Therefore, point-wise semantic classification alone is insufficient to satisfy these management requirements, and instance segmentation that identifies individual components as object-level entities is a necessary step in end-to-end Scan-to-BrIM and bridge digital twin pipelines.

Nevertheless, this study has several limitations. First, the experimental scope is limited to slab bridges and PSC girder bridges. Second, although the synthetic dataset reflects variations in component geometry, it is still based on a limited set of component configurations and structural details. Third, some components, such as abutments, still exhibit relatively low performance due to complex geometries and severe class imbalance.

In summary, this study clearly defines bridge instance segmentation as an essential object-level analysis step at the system-management level from a maintenance perspective and demonstrates the performance of training with scenes that include both bridge components and surrounding elements such as background and abutments. The results further show that combining synthetic data with multi-domain real data is an effective strategy for robustly separating individual components in realistic bridge scenes, providing practical guidelines for 3D point cloud–based Scan-to-BrIM and bridge digital twin construction.

Although this study includes both slab-type and girder-type bridges from publicly available datasets in Korea to incorporate structural diversity, the datasets used in this study remain limited. While the effectiveness of the proposed framework is demonstrated, further validation on a broader range of bridge types and real-world datasets is required. In addition, a more comprehensive exploration of training strategies—such as synthetic–real mixing ratios, pretraining versus joint training, and evaluation across multiple models—remains as future work to further enhance generalizability and robustness.

References

[1] J. O. Sobanjo, “Cost Estimating Under Uncertainty: Issues in Bridge Management” *Transportation Research Circular* 498 (2000).

[2] R. F. S. Cerquin, M. Alarcon, and R. M. Delgadillo, “BrIM and Digital Twin Integration for Structural Health Monitoring and Analysis of the Villena Rey Bridge via Laser Scanning,” *Appl. Sci.*, vol. 15, no. 21, Nov. 2025, doi: 10.3390/app152111741.

[3] M. Castellani, M. A., G.-M. E., A. F., and U. F., “UAV photogrammetry and laser scanning of

bridges: a new methodology and its application to a case study," *Procedia Struct. Integr.*, vol. 62, pp. 193–200, Jan. 2024, doi: 10.1016/j.prostr.2024.09.033.

[4] G.-Q. Zhang, B. Wang, J. Li, and Y.-L. Xu, "The application of deep learning in bridge health monitoring: a literature review," *Adv. Bridge Eng.*, vol. 3, no. 1, p. 22, Dec. 2022, doi: 10.1186/s43251-022-00078-7.

[5] C. Perea, J. Alcala, V. Yepes, F. Gonzalez-Vidosa, and A. Hospitaler, "Design of reinforced concrete bridge frames by heuristic optimization," *Adv. Eng. Softw.*, vol. 39, no. 8, pp. 676–688, Aug. 2008, doi: 10.1016/j.advengsoft.2007.07.007.

[6] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660. Accessed: Jan. 29, 2026. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Qi_PointNet_Deep_Learning_CVPR_2017_paper.html

[7] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jan. 29, 2026. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/d8bf84be3800d12f74d8b05e9b89836f-Abstract.html

[8] A. U. Rahman and V. Hoskere, "Instance segmentation of reinforced concrete bridge point clouds with transformers trained exclusively on synthetic data," *Autom. Constr.*, vol. 173, p. 106067, May 2025, doi: 10.1016/j.autcon.2025.106067.

[9] R. Lu, I. Brilakis, and C. R. Middleton, "Detection of Structural Components in Point Clouds of Existing RC Bridges," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 34, no. 3, pp. 191–212, 2019, doi: 10.1111/mice.12407.

[10] A. R. Colligan, T. T. Robinson, D. C. Nolan, and Y. Hua, "Point cloud dataset creation for machine learning on CAD models," *Comput.-Aided Des. Appl.*, vol. 18, no. 4, pp. 760–771, Jan. 2021, doi: 10.14733/cadaps.2021.760-771.

[11] L. Winiwarter *et al.*, "Virtual laser scanning with HELIOS++: A novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning," *Remote Sens. Environ.*, vol. 269, p. 112772, Feb. 2022, doi: 10.1016/j.rse.2021.112772.

[12] G. Lenda, N. Borowiec, and U. Marmol, "Study of the Precise Determination of Pipeline Geometries Using UAV Scanning Compared to Terrestrial Scanning, Aerial Scanning and UAV Photogrammetry," *Sensors*, vol. 23, no. 19, Oct. 2023, doi: 10.3390/s23198257.

[13] N. Duc *et al.*, "Mind the Domain Gap: Measuring the Domain Gap Between Real-World and Synthetic Point Clouds for Automated Driving Development," May 23, 2025, *arXiv*: arXiv:2505.17959. doi: 10.48550/arXiv.2505.17959.

[14] T. Xia, J. Yang, and L. Chen, "Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning," *Autom. Constr.*, vol. 133, p. 103992, Jan. 2022, doi: 10.1016/j.autcon.2021.103992.

[15] X. Yang, E. Del Rey Castillo, Y. Zou, and L. Wotherspoon, "Semantic segmentation of bridge point clouds with a synthetic data augmentation strategy and graph-structured deep metric learning," *Autom. Constr.*, vol. 150, p. 104838, Jun. 2023, doi: 10.1016/j.autcon.2023.104838.

[16] J. Xu *et al.*, "Domain-generalizable point cloud instance segmentation of bridge components using class-balanced dynamic thresholding," *Autom. Constr.*, vol. 181, p. 106631, Jan. 2026, doi: 10.1016/j.autcon.2025.106631.

[17] X. Yang, Y. Fu, and J. Kim, "Synthetic-to-Real Domain Adaptation with Virtual Laser Scanning and Self-Training-Based Category-Aware Cuboid Mixing for Semantic Segmentation of Bridge Point Clouds," 2024, *SSRN*. doi: 10.2139/ssrn.4902362.

[18] C. Lin, S. Abe, S. Zheng, X. Li, and P. Chun, "A structure-oriented loss function for automated semantic segmentation of bridge point clouds," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 40, no. 6, pp. 801–816, Feb. 2025, doi: 10.1111/mice.13422.

[19] Y. Chen, C. Lin, X. Li, S. Kubo, T. Yamane, and P. Chun, "Automated dimension estimation of bridge components using semantic segmentation and geometric fitting of point cloud data," *Eng. Struct.*, vol. 342, p. 120837, Nov. 2025, doi: 10.1016/j.engstruct.2025.120837.

[20] "Point Cloud Dataset of Reinforced Concrete Bridges Captured with Matterport Pro3 Scanner in Japan." figshare, Mar. 08, 2025. doi: 10.6084/m9.figshare.28091453.v2.

[21] "Detection of Structural Components in Point Clouds of Existing RC Bridges." Accessed: Jan. 30, 2026. [Online]. Available: https://zenodo.org/records/1240534

[22] J. Lu and J. Deng, "Relation3D: Enhancing relation modeling for point cloud instance segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2025, pp. 8889–8899..

A Graph-RAG Framework for Natural Language Interaction with Dynamic Construction Project Graphs

Mohammad Saeed Heidary[1], Xuesong Shen[1], Milad Mousavi[1], Khalegh Barati[1], and James Linke[2]

[1] School of Civil and Environmental Engineering, the University of New South Wales, Sydney, NSW 2052, Australia
[2] GeoAI, Sydney, NSW 2019, Australia.
m.heidary@unsw.edu.au, x.shen@unsw.edu.au, milad.mousavi@unsw.edu.au, khalegh.barati@unsw.edu.au, j.linke@llsurveys.com.au

Abstract –

Construction projects generate heterogeneous data from schedules, design models, and site observations. Graph-based data models have recently been adopted to integrate these diverse data sources, but interacting with such models typically requires specialized technical knowledge, which limits their usability in construction practice. This study proposes a schema-aware Graph Retrieval-Augmented Generation (Graph-RAG) framework that enables natural language interaction with construction project graph data models. The framework integrates a large language model with a Neo4j graph database to interpret user queries, retrieve project data through Cypher queries, and generate evidence-grounded responses. A case study integrating schedule data, an IFC design model, and computer-vision-derived site observations produced a project graph with 218 nodes and 377 relationships. Evaluation using six project queries showed that the proposed framework achieved 100% accuracy, outperforming a Text-to-Cypher approach (33.3% accuracy), while direct LLM interaction failed to answer any questions (0% accuracy). These results demonstrate the effectiveness of schema-aware Graph-RAG for reliable natural language access to construction project data.

Keywords –

Construction Projects; Knowledge Graph; Graph-RAG; LLMs; Construction Data Interaction; Natural Language Interaction

1 Introduction

Construction projects generate large volumes of heterogeneous data throughout their lifecycle, including schedules, design models, site observations, and as-built records [1, 2]. Integrating these diverse data sources in a meaningful and consistent manner remains a significant challenge for project management and decision-making [3]. In recent years, graph-based data models and knowledge graphs have emerged as a promising solution for construction data integration [4]. By representing project entities such as activities, infrastructure elements, resources, and observations as interconnected nodes and relationships, graph models enable heterogeneous data sources to be linked in a semantically meaningful way, supporting richer representation of project knowledge and improved data integration across different project systems [5, 6].

Despite their strong capability for integrating heterogeneous construction data, interacting with graph-based data models often requires specialized technical knowledge. Accessing and querying graph databases typically require graph query languages and database technologies that are more familiar to computer scientists than to construction practitioners [7]. As a result, many project stakeholders may find it difficult to directly interact with graph-based project data models, limiting their usability in everyday construction management. This creates a gap between the availability of integrated project data and the ability of practitioners to effectively access and interpret that information.

Recent advances in large language models (LLMs) offer a promising opportunity to bridge this gap by enabling natural language interaction with complex data systems [8]. Because graph data models organize project information through semantically connected entities and relationships [9], LLMs can potentially serve as an interface that translates user questions into meaningful interactions with the project graph. Motivated by this potential, this study proposes a framework that leverages LLMs to enable natural language interaction with construction project graph data models. By allowing users to query project information using natural language while grounding responses in graph-based data, the proposed approach aims to improve the usability of

graph-based construction information systems and make integrated project knowledge more accessible to construction stakeholders.

2 Related Work

This section reviews prior research on construction information modelling, LLMs, and retrieval-augmented generation, highlighting current practices and limitations in construction-phase applications.

2.1 Construction Information Models and Graph-based Representations

Construction projects have traditionally relied on Building Information Modelling (BIM) platforms and relational databases to manage project information [10]. BIM has been widely adopted to support design coordination, quantity take-off, and visualization[11], while relational databases have been used to store schedules, cost records, and inspection data [12, 13]. For example, Heidary et al. [14] used a 4D BIM environment to link 3D design models with construction schedules, enabling improved visualization of planned activities over time and supporting hazard identification and safety management during construction, whereas Palomar et al. [15] implemented a relational database–based platform using SQL Server to centrally store and synchronize structured project data across multiple disciplines. Although these approaches have improved information organization, they typically operate in isolated silos and rely on predefined schemas that limit flexibility when project conditions change [3]. As construction progresses and new data sources emerge, such as site observations and as-built records, maintaining consistency and traceability across systems remains challenging [1, 2].

To address these limitations, recent research has explored graph-based data models and knowledge graphs for construction information management [4]. Graph representations allow entities such as activities, resources, building elements, and observations to be explicitly linked, supporting richer semantics and more flexible querying than traditional tabular models [5]. In this context, Sun et al. [16] reviewed text-based automatic knowledge graph construction for road infrastructure operations, demonstrating how document-centric sources can be transformed into structured knowledge graphs to support querying and reasoning across lifecycle activities. Knowledge graphs have been applied to integrate BIM data, schedules, safety information, and regulatory knowledge, enabling reasoning over relationships rather than isolated attributes [6, 17]. These approaches have demonstrated advantages in representing complex dependencies and enabling inference across heterogeneous data sources [18, 19].

Despite these advances, interaction with graph-based construction data models typically relies on predefined dashboards or fixed query templates [7]. Such interfaces are designed around anticipated information needs and require backend development when new questions arise [9]. As a result, responding to emerging or unexpected information demands during the construction phase remains difficult, limiting the practical usability of graph-based systems for day-to-day decision-making by project teams.

2.2 Large Language Models in Construction Projects

LLMs are generative AI systems trained to predict and produce natural language, enabling tasks such as question answering, summarization, and drafting based on learned linguistic patterns [8]. Recent evaluations show that ChatGPT-style models can perform competitively on domain problem-solving tasks, motivating their exploration as decision-support tools in construction management [20]. In the construction context, LLMs are increasingly positioned as conversational interfaces that can translate complex technical content into more accessible, actionable narratives for practitioners [21].

In construction applications, Prieto et al. [22] investigated ChatGPT for construction scheduling, examining how well it can generate schedule-related outputs and support planners through natural-language interaction. For project documentation, Xiao et al. [23] combined computer vision with ChatGPT to convert construction videos into automated daily reports, aiming to reduce manual reporting effort while preserving site progress details. For compliance, Yang and Zhang [24] proposed a prompt-based pipeline to transform building code text into structured representations usable for automated checking.

However, the direct use of LLMs also introduces important limitations. LLMs generate responses based on statistical patterns learned from training data rather than verified, project-specific information [24], which can lead to hallucinations, incorrect assumptions, or missing contextual constraints [25]. In construction projects, where decisions are often time-sensitive and safety-critical, such behavior poses significant risks. These limitations highlight the need for complementary mechanisms that can ground LLM outputs in authoritative, traceable, and project-specific information sources.

2.3 Retrieval-augmented Generation in Construction

Retrieval-augmented generation (RAG) has recently been adopted in construction research to improve the reliability of large language model (LLM) outputs by grounding responses in external knowledge sources. Lee et al. [26] compared a RAG-based approach with a fine-tuned LLM for construction safety knowledge retrieval by integrating GPT-4 with a safety knowledge graph derived from guidelines, demonstrating improved consistency in answering safety queries. Similarly, Baek et al. [27] developed a RAG-enhanced framework that retrieves activity- and equipment-specific cases from a construction accident repository to generate tailored safety risk management guidance, showing that domain-adapted embeddings improve retrieval quality. Beyond safety applications, Zheng et al. [28] integrated LLMs with a nested contract knowledge graph to improve the accuracy and interpretability of automated contract risk identification.

Recent studies have further extended RAG and LLM-grounded approaches to different construction tasks and data modalities. Guo et al. [29] proposed a multimodal framework combining a visual language model with a two-stage RAG pipeline to support automated safety compliance monitoring from site images. Zhou et al. [30] integrated an external construction engineering management knowledge base with general-purpose LLMs to enable question answering in construction management contexts when domain training data are limited. Related work has explored alternative grounding strategies, such as long-context LLMs for building code interpretation [31], RAG-based systems for generating personalized feedback in construction management training [32], and multimodal multi-agent frameworks for geotechnical design tasks involving both text and images [33]. In addition, recent work by Pan et al. [34] proposed an LLM-enabled multi-agent framework that enables natural language interaction with graph-based digital twins of existing buildings, where an agent pipeline interprets user queries, generates graph database commands, and retrieves information from the digital twin graph to produce grounded responses for facility management scenarios.

Nevertheless, existing applications of RAG in construction largely focus on static knowledge sources, such as codes, regulations, historical documents, or predefined training materials. The use of RAG to interact with live, evolving project data during the construction phase has received limited attention. In particular, the potential of Graph-RAG to enable natural language access to project knowledge graphs that represent the current state of construction activities and site conditions remains underexplored. Addressing this gap requires frameworks that combine graph-based project data with retrieval-grounded language models to support reliable natural language interaction with construction-phase project information.

3 Research Methodology

This study proposes a Graph Retrieval-Augmented Generation (Graph-RAG) framework to enable natural language interaction with a construction project graph data model. The system integrates a large language model (LLM) with a graph database so that users can query project information using natural language while responses remain grounded in structured graph data. When a user submits a question, the framework first retrieves the schema and ontology from the project graph to provide structural context. Using this information, the LLM decomposes the user query into simpler sub-questions aligned with the graph structure. For each sub-question, the LLM generates a Cypher query to retrieve relevant information from the graph database. The retrieved graph data are then used to generate grounded answers for each sub-question, which are finally synthesized by the LLM to produce a unified response to the original user query. The overall workflow of the proposed Graph-RAG framework is illustrated in Figure 1.

The LLM in the proposed framework performs three key tasks: (1) question decomposition, (2) graph query generation for retrieval, and (3) response generation and synthesis. These components enable the system to bridge natural language user queries and structured project graph data, allowing complex project information to be accessed through intuitive natural language interaction.

3.1 Schema-aware Question Decomposition

When a user submits a natural language query, the system first retrieves the schema and ontology information from the project graph database. This information describes the available node types, relationships, properties, and semantic structures within the project data model. The schema is provided as contextual input to the LLM to ensure that the query interpretation is consistent with the structure of the graph.

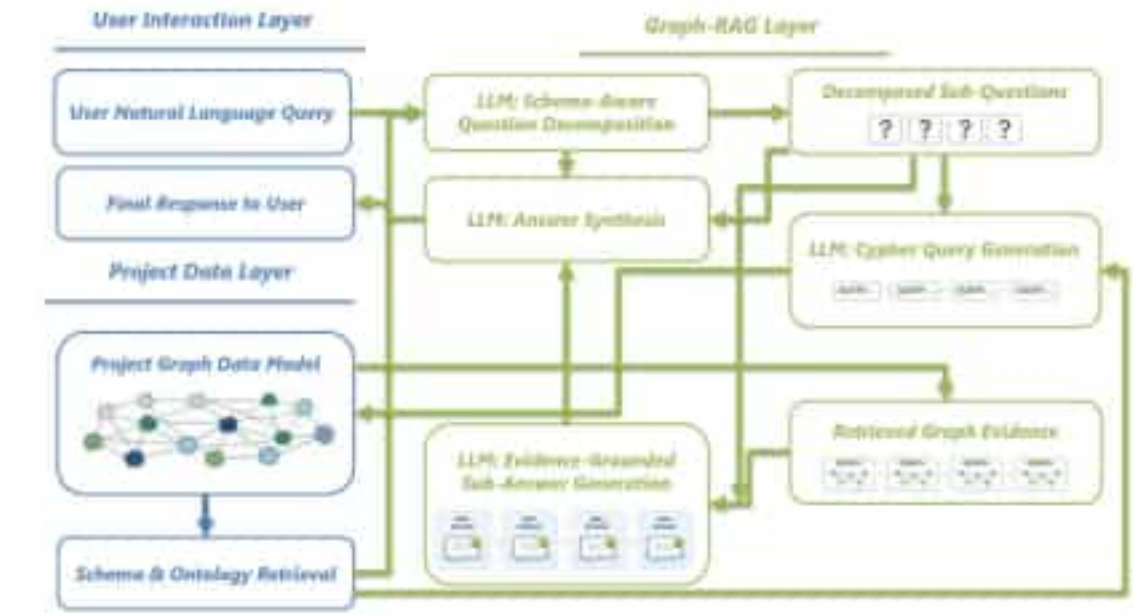


Figure 1. Overview of the proposed Graph-RAG framework for natural language interaction with construction project graph data.

Using this schema-aware context, the LLM refines the user query and decomposes it into a set of simpler sub-questions. Each sub-question represents a specific information retrieval task that can be mapped to graph entities and relationships. This decomposition step helps transform complex user queries into structured information retrieval tasks that align with the project graph schema.

3.2 Graph Query Generation and Retrieval

For each decomposed sub-question, the LLM generates a Cypher query that retrieves the relevant information from the graph database. The generated queries follow the structure of the project graph schema and are designed to retrieve entities, relationships, and contextual attributes required to answer the sub-question.

The Cypher queries are executed against the Neo4j graph database, and the retrieved results are returned as structured graph evidence. This retrieval step ensures that the responses generated by the LLM are grounded in actual project data rather than relying solely on language model inference.

3.3 Evidence-Grounded Response Generation and Synthesis

Once the relevant graph data are retrieved, the LLM generates a response for each sub-question using the retrieved evidence. The model is instructed to base its reasoning only on the retrieved graph information to ensure factual consistency and transparency.

Finally, the LLM synthesizes the responses to the individual sub-questions together with the original user query to generate a unified answer. This synthesis step enables the system to provide coherent explanations that integrate multiple pieces of graph-retrieved information into a final response for the user.

4 Case Study

To evaluate the proposed framework in a real-world scenario, a case study was conducted on an ongoing road intersection upgrade project in New South Wales, Australia. Construction of the project commenced in late 2023 and is scheduled for completion in 2027. The project involves multiple construction activities, resources, and observational data collected during the construction phase. These heterogeneous data sources provide a suitable environment for examining how a graph-based data model combined with large language models can support natural language interaction with project information.

4.1 Case Study Project Graph Data Model Preparation

In this case study, three types of project data were

integrated to construct the project graph data model: the project schedule (structured data), an IFC design model (semi-structured data), and construction site images (unstructured data). These heterogeneous data sources were mapped into a graph structure using ten top-level entities representing key concepts of the construction domain. Through predefined semantic rules, information from the schedule, design model, and site observations was linked to form a unified project graph.

Construction site images were processed using the YOLOv8 computer vision model to detect relevant construction elements. The detected elements were then matched with corresponding elements in the IFC design model based on spatial location and linked to the related construction activities defined in the project schedule. By connecting activities, design elements, and observed site conditions, the different data sources were semantically integrated into a single project graph data model. After processing one project schedule, one IFC model, and 32 construction site images (two images per observation date across 16 different dates), the resulting graph contained 218 nodes and 377 relationships representing the project state (Figure 2).



Figure 2. Snapshot of the dynamic construction project graph in Neo4j on 5 September 2025.

4.2 Developing the Graph-RAG Pipeline

The Graph-RAG pipeline was developed and implemented in Python by integrating the Neo4j graph database with the OpenAI GPT-4.1 model through the OpenAI API. The developed pipeline first retrieves the schema of the project graph and converts it into a structured text summary that can be provided to the LLM as contextual input. Based on this schema-aware context, the framework processes the user question through several LLM-driven stages. First, the LLM decomposes the user question into simpler sub-questions while considering the schema and ontology of the project graph. Second, for each sub-question, the LLM generates a Cypher query to retrieve relevant data from the Neo4j database. Finally, the retrieved graph evidence is used by the LLM to generate grounded responses for each sub-question, and these are then synthesized into the final answer returned to the user. Prompt engineering was used throughout these stages to control the LLM behavior and keep the outputs aligned with the graph structure and project semantics.

In the implemented prompts, each stage was guided by a specific instruction logic. For question decomposition, the LLM was instructed to act as an ontology-aware planning assistant that refines the user question conservatively, decides whether decomposition is needed, and creates schema-consistent sub-questions without introducing unsupported graph elements. For Cypher generation, the LLM was instructed to produce read-only Cypher queries, follow project-specific semantic rules, use only valid retrieval patterns, and avoid any write or administrative operations. For evidence-grounded answering, the LLM was instructed to answer each sub-question only from the retrieved graph evidence, explicitly mention uncertainty when evidence is incomplete, and avoid inventing unsupported facts. In the final synthesis stage, the LLM was instructed to answer the original user question directly by combining the sub-question answers while clearly preserving the distinction between graph-supported findings and uncertain information.

4.3 Testing Setup

To evaluate the effectiveness of the proposed framework, six questions were designed to test the system's ability to retrieve project information under different reasoning scenarios. These questions include temporal status queries, progress tracking queries, resource matching queries, and delay detection queries.

- Q1 — What was the construction status of the activity "8 Columns A – FW" on 2 June 2025?
- Q2 — What was the construction status of the activity "8 Columns A – FW" on 21 May 2025?
- Q3 — What was the construction status of the activity "8 Piles A" on 23 April 2025?
- Q4 — What was the construction status of the activity "8 Piles A" on 7 May 2025?
- Q5 — Which observed machinery types match the machinery planned for the associated construction activities?
- Q6 — Based on the latest available observations, which construction activities are currently identified as delayed?

Four different question-answering methods were evaluated:

- Direct LLM – only the user question is given to the LLM.
- LLM + Full Graph Data – the LLM receives the question and the entire project graph data model as input.
- Text-to-Cypher Retrieval – the LLM generates a Cypher query directly from the question and retrieves graph data in a single step.
- Proposed Graph-RAG Method – the schema-aware decomposition and retrieval framework proposed in this study.

A Python-based testing interface was developed to allow a user to submit a question and automatically generate responses from all four evaluation methods simultaneously for comparison, as illustrated in Figure 3.



Figure 3. Screenshot of the Python testing environment showing simultaneous outputs from the four evaluation methods.

4.4 Results

The experimental results demonstrate clear differences between the evaluated methods. When only the question was provided to the LLM, the model was unable to produce correct answers for any of the questions because it had no knowledge of the specific project data. This outcome is expected since general-purpose LLMs are trained on broad datasets and do not contain information about a particular construction project.

When the entire project graph data model (218 nodes and 377 relationships) was provided as input to the LLM, the API request exceeded the tokens-per-minute (TPM) rate limit of GPT-4.1, producing the following error:

Error: request exceeds the allowed token limit (429 rate limit error).

This indicates that processing the entire project graph as input exceeded the TPM rate limit of the GPT-4.1 API, making this approach impractical for real-world applications due to computational and processing constraints.

The remaining two approaches were able to produce responses. The one-step Text-to-Cypher method correctly answered only the last two questions (2/6), while it failed to correctly interpret the temporal context required for the first four questions. In contrast, the proposed Graph-RAG framework successfully answered all six questions while also providing evidence retrieved from the project graph. A summary of the performance of the four evaluated methods is presented in Table 1, demonstrating the improved accuracy and reliability of the proposed Graph-RAG framework in retrieving project-specific information.

Table 1. Accuracy comparison of the evaluated question-answering methods.

Method	Accuracy
Direct LLM	0/6 (0 %)
LLM + Full Graph Data	Not executable under standard API rate limits
Text-to-Cypher	2/6 (33.3 %)
Proposed Graph-RAG	6/6 (100 %)

5 Discussion and Conclusions

This study proposed a Graph Retrieval-Augmented Generation (Graph-RAG) framework to enable natural language interaction with construction project graph data models. By integrating a large language model with a graph database, the framework allows users to query project information in natural language while ensuring that responses are grounded in structured project data. The case study shows that the approach can effectively retrieve project-specific information from heterogeneous data sources integrated into a unified project graph. Compared with alternative approaches such as direct LLM interaction and one-step Text-to-Cypher retrieval, the proposed framework achieved higher accuracy and produced responses supported by explicit graph evidence.

Nevertheless, this study has some limitations. The evaluation was conducted on a single case study with a relatively small project graph, which may not fully represent the scale and complexity of larger construction projects. Future work will extend the framework to support multi-query conversations, causal and predictive

reasoning, and broader evaluation across different project types. Further research will also examine the performance, scalability, and user adoption of the proposed framework in larger construction project environments.

Acknowledgments

This research is supported by funding from the Australian Research Council (ARC) Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), and the ARC Research Hub for Human-Robot Teaming for Sustainable and Resilient Construction (IH240100016). We also extend our sincere appreciation to GeoAI, in particular Mr. Alastair Linke, Chief Executive Officer, Dr. Youheng Guo, and Mr. Barber Cameron, for providing access to the case study data and for offering valuable technical feedback throughout the project.

References

[1] W. Teclaw, J. O'Donnel, V. Kukkonen, P. Pauwels, N. Labonnote, and E. Hjelseth, "Federating cross-domain BIM-based knowledge graph," *Advanced Engineering Informatics,* vol. 62, 2024, doi: 10.1016/j.aei.2024.102770.

[2] C. Du, Z. Deng, S. Nousias, and A. Borrmann, "Predictive modeling: BIM command recommendation based on large-scale usage logs," *Advanced Engineering Informatics,* vol. 68, 2025, doi: 10.1016/j.aei.2025.103574.

[3] D. F. Bucher, J. J. Hunhevicz, R. K. Soman, P. Pauwels, and D. M. Hall, "From BIM to Web3: A critical interpretive synthesis of present and emerging data management approaches in construction informatics," *Advanced Engineering Informatics,* vol. 62, 2024, doi: 10.1016/j.aei.2024.102884.

[4] M. H. Rasmussen, M. Lefrançois, P. Pauwels, C. A. Hviid, and J. Karlshøj, "Managing interrelated project information in AEC Knowledge Graphs," *Automation in Construction,* vol. 108, 2019, doi: 10.1016/j.autcon.2019.102956.

[5] X. Wang and N. El-Gohary, "Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements," *Automation in Construction,* vol. 147, 2023, doi: 10.1016/j.autcon.2022.104696.

[6] H. Wang, S. Xu, D. Cui, H. Xu, and H. Luo, "Information Integration of Regulation Texts and Tables for Automated Construction Safety Knowledge Mapping," *Journal of Construction Engineering and Management,* vol. 150, no. 5, 2024, doi: 10.1061/jcemd4.Coeng-14436.

[7] R. Sacks, Z. Wang, B. Ouyang, D. Utkucu, and S. Chen, "Toward artificially intelligent cloud-based building information modelling for collaborative multidisciplinary design," *Advanced Engineering Informatics,* vol. 53, 2022, doi: 10.1016/j.aei.2022.101711.

[8] M. B. Jelodar, "Generative AI, Large Language Models, and ChatGPT in Construction Education, Training, and Practice," *Buildings,* vol. 15, no. 6, 2025, doi: 10.3390/buildings15060933.

[9] J. Schlenger, K. Pluta, A. Mathew, T. Yeung, R. Sacks, and A. Borrmann, "Reference architecture and ontology framework for digital twin construction," *Automation in Construction,* vol. 174, 2025, doi: 10.1016/j.autcon.2025.106111.

[10] L. Soibelman and H. Kim, "Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases," *Journal of Computing in Civil Engineering,* vol. 16, no. 1, pp. 39-48, 2002, doi: 10.1061/(asce)0887-3801(2002)16:1(39).

[11] M. Mousavi, A. TohidiFar, and A. Alvanchi, "BIM and machine learning in seismic damage prediction for non-structural exterior infill walls," *Automation in Construction,* vol. 139, 2022, doi: 10.1016/j.autcon.2022.104288.

[12] T. Hegazy and T. Ersahin, "Simplified Spreadsheet Solutions. II: Overall Schedule Optimization," *Journal of Construction Engineering and Management,* vol. 127, no. 6, pp. 469-475, 2001, doi: 10.1061/(asce)0733-9364(2001)127:6(469).

[13] T. Valenko and U. Klanšek, "An integration of spreadsheet and project management software for cost optimal time scheduling in construction," *Organization, Technology and Management in Construction: an International Journal,* vol. 9, no. 1, pp. 1627-1637, 2017, doi: 10.1515/otmcj-2016-0028.

[14] M. S. Heidary, M. Mousavi, A. Alvanchi, K. Barati, and H. Karimi, "Semi-automatic Construction Hazard Identification Method Using 4D BIM," presented at the Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC), 2021, doi: 10.22260/ISARC2021/0080

[15] I. J. Palomar, J. L. García Valldecabres, P. Tzortzopoulos, and E. Pellicer, "An online platform to unify and synchronise heritage architecture information," *Automation in*

Construction, vol. 110, 2020, doi: 10.1016/j.autcon.2019.103008.

[16] Y. Sun, X. Shen, S. Zlatanova, K. Barati, and J. Linke, "Text-based automatic knowledge graph construction for road infrastructure operations management," *Automation in Construction,* vol. 182, 2026, doi: 10.1016/j.autcon.2025.106733.

[17] R. Zhang and N. El-Gohary, "Transformer-based approach for automated context-aware IFC-regulation semantic information alignment," *Automation in Construction,* vol. 145, 2023, doi: 10.1016/j.autcon.2022.104540.

[18] S. Kosse, V. Betker, P. Hagedorn, M. König, and T. Schmidt, "A Semantic Digital Twin for the Dynamic Scheduling of Industry 4.0-based Production of Precast Concrete Elements," *Advanced Engineering Informatics,* vol. 62, 2024, doi: 10.1016/j.aei.2024.102677.

[19] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft, "Universal Digital Twin - A Dynamic Knowledge Graph," *Data-Centric Engineering,* vol. 2, 2021, doi: 10.1017/dce.2021.10.

[20] R. Nyqvist, A. Peltokorpi, and O. Seppänen, "Can ChatGPT exceed humans in construction project risk management?," *Engineering, Construction and Architectural Management,* vol. 31, no. 13, pp. 223-243, 2024, doi: 10.1108/ecam-08-2023-0819.

[21] L. Zhang and Q. Ge, "Offsite construction risk management agent – a GPTs based analysis model," *Engineering, Construction and Architectural Management,* 2025, doi: 10.1108/ecam-05-2024-0624.

[22] S. A. Prieto, E. T. Mengiste, and B. García de Soto, "Investigating the Use of ChatGPT for the Scheduling of Construction Projects," *Buildings,* vol. 13, no. 4, 2023, doi: 10.3390/buildings13040857.

[23] B. Xiao, Y. Wang, Y. Zhang, C. Chen, and A. Darko, "Automated daily report generation from construction videos using ChatGPT and computer vision," *Automation in Construction,* vol. 168, 2024, doi: 10.1016/j.autcon.2024.105874.

[24] F. Yang and J. Zhang, "Prompt-based automation of building code information transformation for compliance checking," *Automation in Construction,* vol. 168, 2024, doi: 10.1016/j.autcon.2024.105817.

[25] M. Oral, Ö. Alboga, S. Aydınlı, and E. Erdis, "Usability of large language models for building construction safety risk assessment," *Engineering, Construction and Architectural Management,* 2025, doi: 10.1108/ecam-08-2024-1143.

[26] J. Lee, S. Ahn, D. Kim, and D. Kim, "Performance comparison of retrieval-augmented generation and fine-tuned large language models for construction safety management knowledge retrieval," *Automation in Construction,* vol. 168, 2024, doi: 10.1016/j.autcon.2024.105846.

[27] S. Baek, C. Y. Park, and W. Jung, "Automated safety risk management guidance enhanced by retrieval-augmented large language model," *Automation in Construction,* vol. 176, 2025, doi: 10.1016/j.autcon.2025.106255.

[28] C. Zheng, S. Wong, X. Su, Y. Tang, A. Nawaz, and M. Kassem, "Automating construction contract review using knowledge graph-enhanced large language models," *Automation in Construction,* vol. 175, 2025, doi: 10.1016/j.autcon.2025.106179.

[29] K. X. Guo, P. K.-Y. Wong, J. C. P. Cheng, C.-F. Chan, P.-H. Leung, and X. Tao, "Enhancing visual-LLM for construction site safety compliance via prompt engineering and Bi-stage retrieval-augmented generation," *Automation in Construction,* vol. 179, 2025, doi: 10.1016/j.autcon.2025.106490.

[30] S. Zhou, X. Liu, D. Li, T. Gu, K. Liu, Y. Yang, and M. O. Wong, "Integrating domain-specific knowledge and fine-tuned general-purpose large language models for question-answering in construction engineering management," *Automation in Construction,* vol. 175, 2025, doi: 10.1016/j.autcon.2025.106206.

[31] J. Lee and G. Lee, "Long context window-based zero-shot legal interpretation of building codes and regulations," *Automation in Construction,* vol. 179, 2025, doi: 10.1016/j.autcon.2025.106450.

[32] X. Hu, Y. M. Goh, and J. Tay, "Mixed-methods evaluation of automated personalised feedback in construction management training using RAG and LLMs," *Automation in Construction,* vol. 182, 2026, doi: 10.1016/j.autcon.2025.106745.

[33] H.-R. Xu, N. Zhang, Z.-Y. Yin, and P. G. A. Njock, "Multimodal framework integrating multiple large language model agents for intelligent geotechnical design," *Automation in Construction,* vol. 176, 2025, doi: 10.1016/j.autcon.2025.106257.

[34] Y. Pan, M. Wang, L. Lu, R. Lamsal, E. Pärn, S. Zlatanova, and I. Brilakis, "LLM-enabled multi-agent framework for natural language interaction with graph-based digital twins," *Automation in Construction,* vol. 183, 2026, doi: 10.1016/j.autcon.2026.106791.

Safe and Robust Imitation Learning for Drone Navigation in Cluttered Construction Environments

Yun Seok Gwon[1] **and Heung Jin Oh**[1]

[1]Bowen School of Construction, Purdue University, USA

ygwon@purdue.edu, oh367@purdue.edu

Abstract -

Autonomous navigation in cluttered construction environments is challenging due to unstructured obstacles, GPS-denied conditions, and operational constraints on onboard computation. Optimization-based planners such as Covariant Hamiltonian Optimization for Motion Planning (CHOMP) offer obstacle-avoidance capabilities but are often unsuitable for high-frequency control due to their computational cost and sensitivity to local minima. This study investigates a safety-regularized imitation learning framework that integrates residual policy learning, a learned safety critic providing probabilistic risk estimates, and Conditional Value-at-Risk (CVaR) regularization. Through systematic evaluation in a procedurally generated PyBullet construction environment featuring dense scaffolding and debris, we show that safety-oriented loss shaping alone is insufficient to mitigate covariate shift in behavioral cloning, resulting in low task success despite conservative trajectories. In contrast, incorporating Dataset Aggregation (DAgger) enables the learned policy to acquire effective recovery behaviors, improving navigation success to 80% while maintaining safety margins comparable to the expert planner. The primary contribution of this work is an empirical demonstration that dataset aggregation, rather than safety regularization alone, is critical for robust imitation learning in cluttered construction environments. The proposed framework provides a foundation for distilling optimization-based planners into lightweight, reactive policies suitable for simulated aerial navigation in construction-like environments, while real-world deployment remains future work.

Keywords -

Imitation Learning; Drone Navigation; Control Barrier Functions; Robust Control; PyBullet; Robotics

1 Introduction

Construction sites are dynamic and hazardous environments for robotic deployment, characterized by narrow corridors, hanging scaffolding, and scattered debris such as pallets, pipes, and steel beams. These conditions limit the effectiveness of standard GPS-based waypoint following or map-heavy global planners for aerial robots tasked with site inspection, progress monitoring, or surveillance [1]. Therefore, navigation in such environments often requires a reactive controller that supports real-time and online obstacle avoidance control.

While optimization-based motion planners such as CHOMP [2] can generate feasible trajectories in cluttered environments, their computation costs can limit use in real-time control. Running such optimizers online at the control frequency required for quadrotors (often > 50 Hz) is difficult on standard embedded hardware [3]. Furthermore, these planners may generate trajectories with high jerk due to gradient-based updates on non-convex cost landscapes, potentially imposing mechanical loads on the airframe.

Given the computational load of optimization based motion planners, Imitation Learning (IL) has been explored as a practical alternative method for distilling the capabilities of a heavy planner into a lightweight neural policy. In this work, IL is used to learn a reactive policy from expert demonstrations [4]. However, common IL approaches, such as standard Behavioral Cloning (BC), are known to suffer from two flaws. First, it is susceptible to covariate shift, where small errors accumulate, which can drive the robot into unvisited states where the policy is undefined [5]. Second, BC mimics the expert directly; if the expert minimizes time by grazing obstacles (as CHOMP often does), the student policy inherits this behavior.

To address these limitations, this project implemented a safety-regularized imitation learning for cluttered construction environments. The approach combines residual policy learning, a learned approximation to a control barrier function, and risk-sensitive loss terms. The implementation focused on the following components:

1. A meta-residual policy learning architecture that combines a classical nominal controller with a learned neural correction to improve tracking performance and obstacle avoidance in cluttered environments.
2. A latent safety critic inspired by control barrier functions that estimates proximity to unsafe regions directly from raw LiDAR observations.
3. A risk-sensitive training objective incorporating Conditional Value-at-Risk (CVaR) to emphasize rare but

critical failure cases arising from complex obstacle geometries.

4. A systematic evaluation in a procedurally generated PyBullet construction environment to empirically demonstrate dataset aggregation, rather than safety regularization alone, is critical for robust imitation learning in highly cluttered spaces.

2 Research Background

2.1 Autonomous Navigation in Cluttered Construction Environments

Construction sites are challenging for autonomous navigation due to dense structural elements, narrow passages, unstructured debris, and frequent layout changes. These environments can be GPS-denied or unreliable, complicating GPS-dependent navigation [6]. UAVs are increasingly used for construction tasks such as site inspection, image- or video-based 3D mapping, and performance monitoring, motivating robust navigation in complex settings [7].

UAV navigation is commonly categorized into map-based and map-less (reactive) strategies. Map-based methods (e.g., SLAM and occupancy mapping) build an environment representation to support planning, whereas map-less methods react directly to sensor observations [8]. Prior work has extensively studied sampling-based and optimization-driven planners for collision-free trajectory generation, but these methods often assume stable environment representations and can be computationally demanding for onboard use on resource-constrained platforms [9].

In cluttered, dynamic settings, maintaining accurate and up-to-date maps is difficult due to layout changes, sensing artifacts, and moving obstacles, which can reduce the effectiveness of map-heavy global planning. This motivates reactive control approaches that map observations to actions without expensive online optimization. Learning-based navigation, including imitation learning, provides one such approach by distilling expert behavior into reactive policies [10]. However, robustness and safety remain challenging under distribution shift and rare failure cases, motivating imitation learning methods that incorporate safety constraints and distribution-alignment mechanisms [11].

2.2 Imitation Learning and Meta-Learning

IL has evolved beyond simple BC. While BC [12] treats control as a supervised regression problem, it lacks long-horizon stability due to covariate shift. DAgger [5] addresses this by querying the expert on student-induced states, a strategy employed in this project for data aggregation. Recent work has explored meta-learning formulations (e.g., MAML [13]) to improve adaptation across navigation tasks. While these methods enable rapid parameter updates, they primarily adjust the behavior of an underlying planner rather than altering the optimization landscape.

In highly cluttered environments, local minima often affect planner performance. Residual Policy Learning (RPL) [14] provides a complementary approach by enabling the learned component to modify the effective vector field guiding the trajectory. Generative Adversarial Imitation Learning (GAIL) [15] seeks to align occupancy measures between expert and learner but may produce unsafe behavior when expert demonstrations approach obstacle boundaries. Diffusion-based policies [16] have demonstrated multimodal action modeling; however, their iterative denoising procedure incurs computational overhead, presenting challenges for high-frequency aerial control loops.

2.3 Trajectory Optimization: CHOMP

The expert demonstrator used in this study is based on CHOMP (Covariant Hamiltonian Optimization for Motion Planning) [2]. While Model Predictive Control (MPC) [17] and sampling-based methods like RRT* are standard in local planning, MPC requires solving convex optimization problems at every timestep, which can be computationally demanding. RRT* produces probabilistically complete but often non-differentiable paths.

CHOMP treats trajectory generation as a functional gradient descent problem. It handles cost maps (ESDFs) and produces smooth trajectories, though it is prone to local minima and requires gradient computations. Our approach uses CHOMP as a *nominal* prior to guide the student policy training.

2.4 Safe Control and Barrier Functions

Control Barrier Functions (CBFs) [18] provide a method to ensure forward invariance of a safe set. If a valid barrier function $h(x)$ exists, safety can be enforced by constraining the control input u:

$$\dot{h}(x) + \alpha(h(x)) \geq 0 \tag{1}$$

Defining an analytical $h(x)$ for a raw point cloud is difficult in unstructured environments. Recent work [19] learns CBFs from data. This project implements a latent safety critic inspired by CBFs to estimate the safety margin directly from LiDAR features.

3 Preliminaries

This section defines the robot dynamics, environmental representation, and safety formulation used in the implementation.

3.1 Quadrotor Dynamics

The quadrotor is modeled as a kinematic system controlled by velocity commands. Let state $s_t \in \mathcal{S} \subset \mathbb{R}^n$ include the robot's position $p_t \in \mathbb{R}^3$, velocity $v_t \in \mathbb{R}^3$, and sensor observations o_t (LiDAR scans). The control input $u_t \in \mathcal{A} \subset \mathbb{R}^3$ specifies the target velocity vector.

$$p_{t+1} = p_t + u_t \Delta t \tag{2}$$

This kinematic model is used for high-level trajectory planning, assuming a lower-level flight controller handles the rotor dynamics.

3.2 Euclidean Signed Distance Fields (ESDF)

The expert planner relies on an implicit surface representation of the environment. An ESDF, denoted as $\phi(x)$, maps a point $x \in \mathbb{R}^3$ to the distance to the nearest obstacle boundary ∂O.

$$\phi(x) = \begin{cases} d(x, \partial O) & \text{if } x \notin O \\ -d(x, \partial O) & \text{if } x \in O \end{cases} \tag{3}$$

This representation allows for gradient computation $\nabla\phi(x)$, which is utilized by the CHOMP optimizer. In the simulation, this is computed from the LiDAR point cloud.

3.3 Control Barrier Functions (CBF)

A Control Barrier Function $h(s)$ is a scalar function defined over the state space such that the set $C = \{s \in \mathcal{S} : h(s) \geq 0\}$ represents the safe region. The boundary $\partial C = \{s : h(s) = 0\}$ represents the limit of safety. For a control system $\dot{s} = f(s, u)$, the safety condition is maintained if the control input u satisfies:

$$\frac{\partial h}{\partial s} f(s, u) \geq -\alpha h(s) \tag{4}$$

where $\alpha > 0$ determines how aggressively the system is allowed to approach the boundary. In this study, this condition is incorporated as a penalty term in the loss function rather than being solved online via QP.

3.4 Conditional Value-at-Risk (CVaR)

To address rare collision scenarios, CVaR is employed. For a loss distribution Z and confidence level $\alpha \in (0, 1)$, CVaR is defined as the expected loss in the worst α-percentile of cases:

$$\text{CVaR}_\alpha(Z) = \mathbb{E}[Z \mid Z \geq \text{VaR}_\alpha(Z)] \tag{5}$$

where VaR_α is the Value-at-Risk threshold. This term is intended to robustify the policy against outlier scenarios.

4 Methodology

4.1 System Architecture

The system consists of three main components: a global planner (providing the goal), a local expert planner (CHOMP), and the learning-based student policy. The student policy processes raw LiDAR data and state information to generate velocity commands.

4.2 Residual Policy Learning

We employed Residual Policy Learning (RPL) rather than mapping observations directly to absolute actions. The final control command u_t is given by:

$$u_t = v_{curr} + \pi_\theta(f(s_t)) + u^z_{PID} \tag{6}$$

where v_{curr} is the current velocity, π_θ predicts a delta-velocity correction (Δv) in the xy-plane, and u^z_{PID} is a fixed PID controller for altitude maintenance.

4.2.1 Feature Extraction

Raw LiDAR scans are converted into a local occupancy grid and processed into an Euclidean Signed Distance Field (ESDF) patch (size 7×7) centered on the drone. The policy network input is a concatenation of:

- The flattened ESDF patch (49 dimensions).
- Current velocity vector (3 dimensions).
- Goal delta vector (3 dimensions).

4.2.2 Network Architecture

The policy network π_θ is a Multi-Layer Perceptron (MLP) with two hidden layers of 256 units each, using LeakyReLU activations (slope 0.2).

4.3 Learned Safety Critic

A neural network $r_\phi(s)$ (Safety Critic) was trained to approximate the risk of a state. The ground truth risk R_{gt} is computed during expert data collection using an exponential decay of the minimum LiDAR clearance d_{min}:

$$R_{gt}(s) = \exp(-\gamma \cdot d_{min}) \tag{7}$$

where $\gamma = 2.0$. This signal is bounded in $[0, 1]$. The critic network mirrors the policy architecture but outputs a single scalar value through a Sigmoid activation, trained via Mean Squared Error (MSE):

$$\mathcal{L}_{critic} = \|r_\phi(s) - R_{gt}(s)\|^2 \tag{8}$$

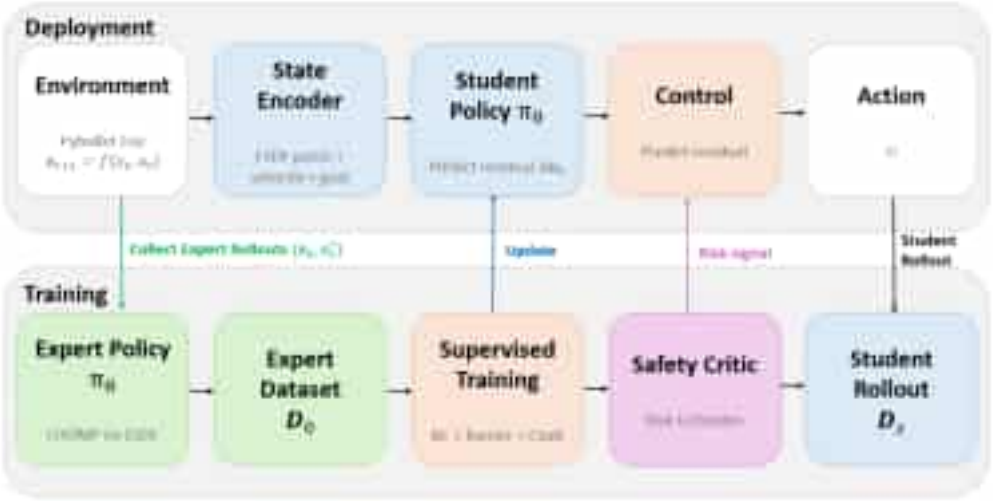


Figure 1. System architecture. The student policy learns a residual correction to the current velocity, guided by the expert's demonstrations and regulated by the Safety Critic.

4.4 Loss Function

The training process utilized a Safety-Regularized loss function, combining behavioral cloning with barrier-based regularization and risk sensitivity.

$$\begin{aligned}\mathcal{L}_{total} = & \underbrace{\|\pi_\theta(s) - \Delta u^*_{target}\|^2}_{\text{BC}} \\ & + \lambda_1 \underbrace{\mathbb{E}[B(s,u)]}_{\text{Barrier Mean}} \\ & + \lambda_2 \underbrace{\text{CVaR}_\alpha(B(s,u))}_{\text{Robust Safety}}\end{aligned} \tag{9}$$

where the Barrier Cost $B(s,u)$ is defined as:

$$B(s,u) = r_\phi(s) \cdot \|u_{final}\|^2 \tag{10}$$

This term penalizes high kinetic energy ($\|u\|^2$) when the estimated risk $r_\phi(s)$ is high. The $\text{CVaR}_{0.1}$ term calculates the mean of the top 10% of barrier costs in the batch. Parameters were set to $\lambda_1 = 0.01$ and $\lambda_2 = 0.05$ for the experiments.

4.5 Training Pipeline: DAgger

The DAgger algorithm was utilized to address distribution shift. The pipeline proceeds as follows:

1. **Expert Collection:** Collect initial trajectories using the CHOMP teacher.
2. **Pre-training:** Train π_θ and r_ϕ on expert data.
3. **Iterative Rollout:** Execute the student policy π_θ in the environment to collect new observations $s_{student}$.
4. **Expert Labeling:** Query the teacher for the optimal action u^* at these new states.
5. **Aggregation & Retraining:** Add the new (s, u^*) pairs to the dataset and retrain.

5 Experimental Setup

5.1 Simulation Environment

A custom simulation environment was developed using the PyBullet physics engine. The environment, "Construction Site," spans a $22m \times 22m$ area.

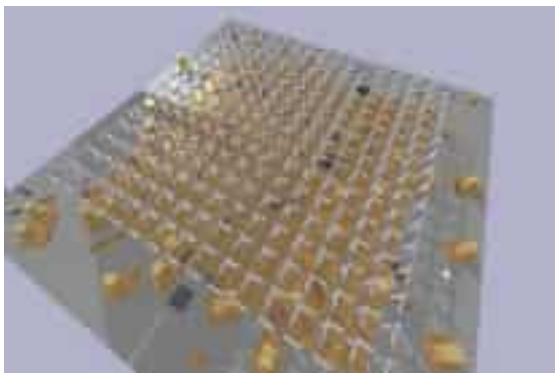


Figure 2. The PyBullet simulation environment featuring procedurally generated scaffolding and debris.

The scene generation is procedural to explicitly model the structural variability found across different construction scenarios. It consists of:

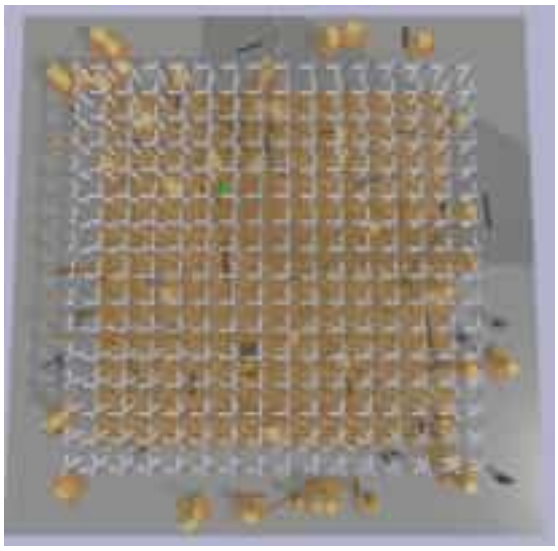


Figure 3. Top-down view of a generated episode. Red indicates the goal, blue indicates the start. Note the density of the scaffolding grid.

- **Scaffolding Grid:** A 15×15 grid of bays ($1m \times 1m$) constructed from vertical uprights and horizontal ledger pipes (radius $0.06m$).
- **Debris Scattering:** 40 wooden pallets ($0.6m \times 0.5m$) and 80 random debris items (steel beams, cylinders, pipes) scattered across the floor.
- **Spawn Logic:** The drone and goal are often spawned inside the scaffolding structure or atop debris piles.

5.2 Sensors and Robot Model

The robot is modeled as a quadrotor with a $0.18m$ radius, equipped with a simulated LiDAR sensor:

- **Range:** 10.0 meters.
- **Resolution:** 72 azimuth rays $\times$ 5 elevation layers (Total 360 rays).
- **Update Rate:** 60 Hz.
- **Feature Space:** Raw point cloud rasterized into a 7×7 ESDF patch.

Camera inputs were not used to reduce computational overhead, relying on geometric depth data. These procedurally generated scenes are intended to represent a type of cluttered construction-like layouts rather than all construction scenarios, and the present evaluation is limited to short-horizon navigation in dense obstacle fields.

5.3 Expert Demonstrator

The expert policy is based on CHOMP, utilizing the instantaneous ESDF to compute a velocity gradient minimizing:

$$\mathcal{U}(traj) = w_{obs}\mathcal{U}_{obs} + w_{smooth}\mathcal{U}_{smooth} + w_{att}\mathcal{U}_{att} \tag{11}$$

Parameters were tuned to $k_{att} = 1.5$ (goal attraction) and $k_{rep} = 0.8$ (obstacle repulsion).

6 Results

6.1 Quantitative Comparison

Five comparative baselines were evaluated to systematically analyze the performance of Imitation Learning variants against the optimization-based expert. The baselines are:

1. **Teacher (CHOMP):** The optimization-based expert.
2. **Vanilla BC:** Behavioral Cloning on expert data.
3. **BC + CVaR:** BC with risk-sensitive loss.
4. **BC + CVaR + CBF:** Full loss function on static expert data.
5. **BC + CVaR + CBF + DAgger:** The full method with iterative dataset aggregation.

Five comparative baselines were evaluated to systematically analyze the performance of Imitation Learning variants against the optimization-based expert.

Table 1 summarizes the results averaged over 15 evaluation episodes.

6.2 Observation of "Drift" Failure Mode (Methods 2-4)

Our initial hypothesis was that a robust loss function (CVaR + CBF) might enable training a safe policy from expert data alone. However, methods trained purely on static expert datasets (Methods 2, 3, and 4) yielded success rates below 15%. While these methods produced smooth trajectories (Jerk $< 8m/s^3$) and maintained high safety margins (Clearance $> 0.3m$), they consistently failed to reach the goal. Flight logs indicate these failures were due to covariate shift. The expert trajectories lie on a specific manifold; when the student drifted due to approximation error, it entered unvisited states. Lacking data in these regions, the policy output defaulted to passive behaviors, causing the drone to freeze or drift until timeout. This observation suggests that loss shaping alone was insufficient to resolve the distribution shift problem in this environment.

6.3 Impact of DAgger (Method 5)

Implementing DAgger (Dataset Aggregation) improved the success rate to 80%. By iteratively collecting data where the student operates and the expert corrects, the policy learned recovery behaviors. Notably, the jerk metric increased significantly (from 7.8 to $30.2m/s^3$). This

Table 1. Failure Analysis and Final Success (Averaged over 15 Episodes)

Method	Success	Jerk (m/s^3)	Min Clearance
Teacher (CHOMP)	100%	42.4	0.06 m
Vanilla BC	13.3%	4.3	0.55 m
BC + CVaR	6.7%	4.8	0.69 m
BC + CVaR + CBF	6.7%	7.8	0.34 m
Ours (with DAgger)	**80.0%**	**30.2**	**0.07 m**

suggests that the "smoothness" observed in the static-data policies was a result of passivity. The DAgger-trained policy exhibited more aggressive control behavior (closer to the Teacher's 42.4 jerk) to actively correct errors and navigate the clutter.

6.4 Flight Smoothness

The Teacher policy exhibited high jerk ($\approx 42.1 m/s^3$) due to the discrete nature of the ESDF gradient updates and the repulsion forces in the potential field. The policy with DAgger reduced jerk to $\approx 30.2 m/s^3$. This reduction indicates that the neural network effectively filtered some of the high-frequency optimization noise from the expert while maintaining the trajectory intent.

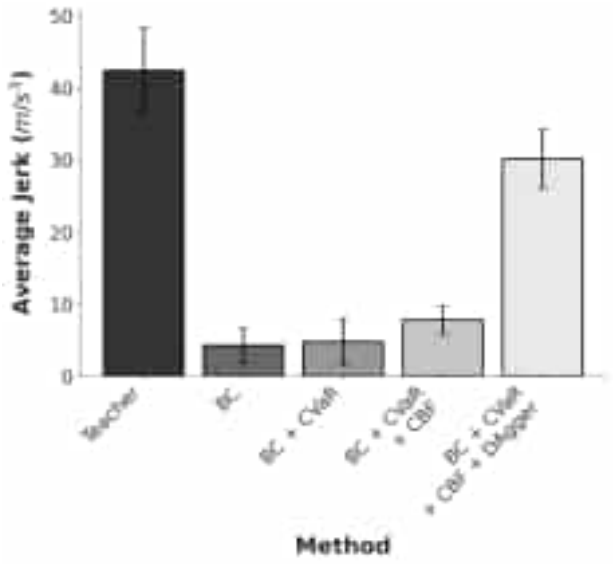


Figure 4. Comparison of average jerk. Lower values indicate smoother flight.

6.5 Safety Margins

The minimum clearance distance to obstacles was tracked during flight. The Teacher policy occasionally violated the safety margin, potentially due to local minima. The policy maintained a comparable clearance ($0.07m$ vs $0.06m$). The inclusion of the CVaR term was intended to penalize proximity to obstacles, though in the DAgger-trained model, the clearance remained tight due to the complexity of the environment.

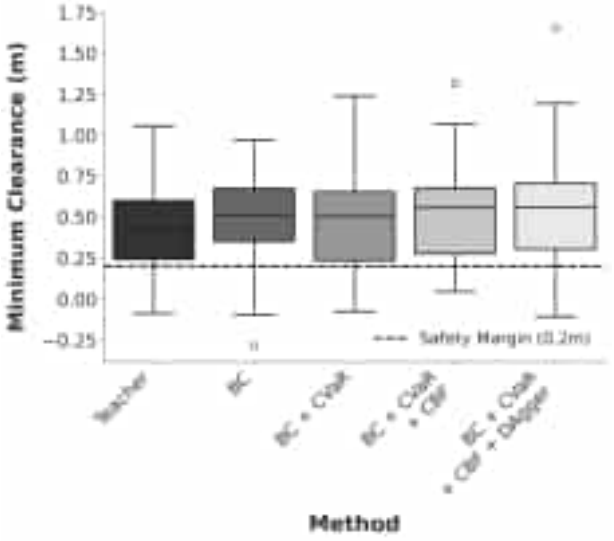


Figure 5. Distribution of minimum clearance distances.

7 Discussion

The quantitative evaluation highlights the challenges inherent in learning from static expert demonstrations. Despite incorporating safety-oriented regularization, methods trained without dataset aggregation consistently failed to achieve reliable goal-reaching performance. These failures are attributable to covariate shift: the expert trajectories occupy a narrow region of the state space, whereas deviations during execution place the learner in states for which no corresponding supervisory labels exist. Under these conditions, the policy exhibited unreliable corrective behavior.

Incorporating DAgger mitigated this issue by explicitly collecting supervisory signals in learner-induced states. This procedure enabled the policy to acquire effective recovery strategies and resulted in an improvement in success rate. The observed increase in jerk relative to static-training methods reflects a shift from conservative, low-

mobility behavior toward active correction, aligning more closely with the expert's operational characteristics. This suggests a trade-off where robustness requires active correction, which inherently increases control effort. The residual-learning formulation integrated the nominal controller with the learned correction. This structure allowed the system to default to the nominal controller when the residual was near zero, though the drift issues in static training suggest the residual component is critical for navigation in complex clutter.

7.1 Scalability

The results have implications for deployment on embedded hardware. The constant-time inference $O(1)$ of the neural policy contrasts with the scaling complexity of CHOMP. This efficiency is relevant for MAVs with limited onboard compute. Additionally, the Latent CBF operates on raw depth features rather than visual textures, which visual-domain dependence in future sim-to-real transfer to physical hardware using depth cameras.

7.2 Limitations

The current method relies on a static world assumption. The ESDF is computed instantaneously, ignoring the velocity of dynamic obstacles. Extending the method to dynamic environments would require a temporal component. Additionally, the LCBF provides a probabilistic estimate rather than a formal control-theoretic guarantee of safety. The present study is limited to simulation in PyBullet and does not include real-world UAV experiments, so the results should be interpreted as controlled empirical evidence rather than deployment-ready validation.

8 Conclusion

This paper presented a safety-regularized imitation learning framework for autonomous navigation in cluttered construction environments. By combining a residual policy architecture with a learned approximation of a control barrier function and a risk-sensitive objective, the method integrates safety considerations directly into policy learning for reactive aerial navigation. Experimental results demonstrate that, although safety-aware loss terms improve clearance behavior, they are insufficient to resolve covariate shift when trained solely on static expert demonstrations. Incorporating dataset aggregation through DAgger significantly improved task success by enabling the policy to learn recovery behaviors in learner-induced states. These findings highlight the importance of distribution-alignment mechanisms in imitation learning for risk-aware robotic navigation in cluttered environments.

Beyond the simulated evaluation presented in this study, several directions are particularly relevant for construction-oriented deployment. Future work will extend the environment complexity to include a broader range of construction clutter types, such as partially assembled structural elements, temporary formwork, and moving equipment. Additional sensing degradations common to construction sites, including dust, occlusion, sparse geometry, and sensor noise, will be incorporated to better reflect real-world operating conditions. Integrating prior information from Building Information Models (BIM) as structural or semantic priors may further improve navigation robustness and enable tighter coupling with construction workflows such as progress monitoring and safety inspection. Together, these extensions aim to bridge the gap between learning-based navigation research and practical deployment of aerial robots in active construction environments.

References

[1] Aiyu Zhu, Tianhong Dai, Gangyan Xu, Pieter Pauwels, Bauke De Vries, and Meng Fang. Deep reinforcement learning for real-time assembly planning in robot-based prefabricated construction. *IEEE Transactions on Automation Science and Engineering*, 20(3):1515–1526, 2023. doi:10.1109/TASE.2023.3236805.

[2] Nathan Ratliff, Matt Zucker, J. Andrew Bagnell, and Siddhartha Srinivasa. Chomp: Gradient optimization techniques for efficient motion planning. In *2009 IEEE International Conference on Robotics and Automation*, pages 489–494, 2009. doi:10.1109/ROBOT.2009.5152817.

[3] Daniel Mellinger, Nathan Michael, and Vijay Kumar. *Trajectory Generation and Control for Precise Aggressive Maneuvers with Quadrotors*, pages 361–373. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-642-28572-1. doi:10.1007/978-3-642-28572-1_25. URL https://doi.org/10.1007/978-3-642-28572-1_25.

[4] Dean A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In D. Touretzky, editor, *Advances in Neural Information Processing Systems*, volume 1. Morgan-Kaufmann, 1988. URL https://proceedings.neurips.cc/paper_files/paper/1988/file/812b4ba287f5ee0bc9d43bbf5bbe87fb-Paper.pdf.

[5] Stéphane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. No-regret reductions for imitation learn-

ing and structured prediction. *CoRR*, abs/1011.0686, 2010. URL http://arxiv.org/abs/1011.0686.

[6] Yingxiu Chang, Yongqiang Cheng, Umar Manzoor, and John Murray. A review of uav autonomous navigation in gps-denied environments. *Robotics and Autonomous Systems*, 170:104533, 2023. ISSN 0921-8890. doi:https://doi.org/10.1016/j.robot.2023.104533. URL https://www.sciencedirect.com/science/article/pii/S0921889023001720.

[7] Mohammad Nahangi, Adam Heins, Brenda McCabe, and Angela Schoellig. Automated localization of uavs in gps-denied indoor construction environments using fiducial markers. In Jochen Teizer, editor, *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 88–94, Taipei, Taiwan, July 2018. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-3-00-060855-1. doi:10.22260/ISARC2018/0012.

[8] Dipraj Debnath, Fernando Vanegas, Juan Sandino, Ahmad Faizul Hawary, and Felipe Gonzalez. A review of uav path-planning algorithms and obstacle avoidance methods for remote sensing applications. *Remote Sensing*, 16(21), 2024. ISSN 2072-4292. URL https://www.mdpi.com/2072-4292/16/21/4019.

[9] Gopi Gugan and Anwar Haque. Path planning for autonomous drones: Challenges and future directions. *Drones*, 7(3), 2023. ISSN 2504-446X. doi:10.3390/drones7030169. URL https://www.mdpi.com/2504-446X/7/3/169.

[10] Caleb Harris, Youngjun Choi, and Dimitri Mavris. Imitation learning for uas navigation in cluttered environments. 01 2021. doi:10.2514/6.2021-0452.

[11] Chenxin Yang. Review of ai-based uav navigation in gps-denied environments. In *Proceedings of the 2025 2nd International Conference on Electrical Engineering and Intelligent Control (EEIC 2025)*, pages 583–596. Atlantis Press, 2025. ISBN 978-94-6463-864-6. doi:10.2991/978-94-6463-864-6_51. URL https://doi.org/10.2991/978-94-6463-864-6_51.

[12] Faraz Torabi, Garrett Warnell, and Peter Stone. Behavioral cloning from observation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, page 4950–4957. AAAI Press, 2018. ISBN 9780999241127.

[13] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks, 2017. URL https://arxiv.org/abs/1703.03400.

[14] Tom Silver, Kelsey R. Allen, Joshua B. Tenenbaum, and Leslie Pack Kaelbling. Residual policy learning. *ArXiv*, abs/1812.06298, 2018. URL https://api.semanticscholar.org/CorpusID:56327855.

[15] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 4572–4580, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

[16] Kostas Bekris, Kris Hauser, Sylvia Herbert, Jingjin Yu, Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *Int. J. Rob. Res.*, 44(10–11): 1684–1704, September 2025. ISSN 0278-3649. doi:10.1177/02783649241273668. URL https://doi.org/10.1177/02783649241273668.

[17] Carlos E. García, David M. Prett, and Manfred Morari. Model predictive control: Theory and practice—a survey. *Automatica*, 25(3):335–348, 1989. ISSN 0005-1098. doi:https://doi.org/10.1016/0005-1098(89)90002-2. URL https://www.sciencedirect.com/science/article/pii/0005109889900022.

[18] Aaron D. Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. In *2019 18th European Control Conference (ECC)*, pages 3420–3431, 2019. doi:10.23919/ECC.2019.8796030.

[19] Alexander Robey, Haimin Hu, Lars Lindemann, Hanwen Zhang, Dimos V. Dimarogonas, Stephen Tu, and Nikolai Matni. Learning control barrier functions from expert demonstrations. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3717–3724, 2020. doi:10.1109/CDC42340.2020.9303785.

PFAB: A Generic Software Framework for Predictive Fabrication in Construction

Luca Bettermann[1,2], Sebastian Esser[1,2], Martin Slepicka[1,2] and André Borrmann[1,2]

[1]Chair of Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2]Georg Nemetschek Institute, Technical University of Munich, Germany

luca.bettermann@tum.de, sebastian.esser@tum.de, martin.slepicka@tum.de, andre.borrmann@tum.de

Abstract -

As digital fabrication systems in construction become more sensorized and controllable, they expose growing opportunities for data-driven process understanding that remain largely underutilized. Consequently, additive manufacturing workflows continue to rely on expert intuition and trial-and-error, which limits their industrial adoption in dynamic environments, such as construction sites. This paper presents PFAB, an open-source Python-based framework that enables predictive fabrication workflows by structuring data, coordinating learning tasks, and guiding parameter selection. PFAB formalizes fabrication as a data-driven process that generates predictive models through active learning methods and supports performance-based optimization of process parameters. Its modular, schema-first design separates framework-level orchestration from fabrication-specific logic, allowing the same predictive workflow to be applied across fabrication systems. The framework is demonstrated through Learning–by–Printing, an exemplary implementation for extrusion-based additive manufacturing, and is complemented by an executable reference implementation illustrating end-to-end usage.

Keywords -

Predictive fabrication; Data-driven additive manufacturing; Active learning; Learning by printing

1 Introduction

1.1 Motivation

Additive Manufacturing (AM) for construction is widely regarded as a promising technology to address long-standing challenges in productivity and sustainability. By connecting parametric design with automated fabrication and machine control, AM offers unprecedented flexibility and enables the efficient fabrication of bespoke components for reduced material use [1]. Yet despite rapid growth in academic activity and high-visibility demonstrators, AM has not realized this potential in practice [2]. Most systems continue to depend on expert intuition and iterative trial-and-error, resulting in a lack of first-time-right manufacturing and limiting broader industrial adoption.

A central difficulty emerges from the interaction of multiple interdependent factors, including design parameters, material behavior, environmental conditions, and machine settings. Small variations in any of these aspects can produce disproportionately different results [3]. This coupling and variability make AM processes difficult to model explicitly and challenging to control using traditional methods. As a result, there is a growing need for approaches that can leverage the digital nature of AM not only for automation, but also for acquiring, structuring, and interpreting process data to enable more predictable and robust fabrication [4].

1.2 A layered view on digital fabrication

Digital fabrication workflows can be understood as a sequence of tightly connected layers. At the design layer, geometric intent, structural considerations, and material selection are defined using computational and parametric tools. The fabrication layer translates this intent into a task plan governed by process parameters and converts it into machine-executable instructions for physical execution. These two layers form the basis of current practice and are well supported by existing digital toolchains.

Modern digital fabrication systems increasingly operate as cyber-physical systems (CPS): sensors capture the *as-fabricated* reality, data feedback is used to synchronize fabrication tasks, monitor quality, and maintain stable operation [5]. However, the use of this data remains largely reactive. Data is collected, inspected, or used for post-process corrections, but rarely interpreted in a way that reveals the underlying relationships between the design, material, process, and environmental factors that shape fabrication outcomes. As a result, current workflows still lack a systematic mechanism for learning from data, and the potential of cyber-physical systems remains underutilized.

To fully exploit the data feedback provided by modern CPS architectures, a predictive layer is proposed (Fig. 1). The Predictive Fabrication (PFAB) framework introduces this layer through a modular and generalizable architecture that orchestrates the entire data-processing and learning pipeline. PFAB models the general behavior of a fabri-

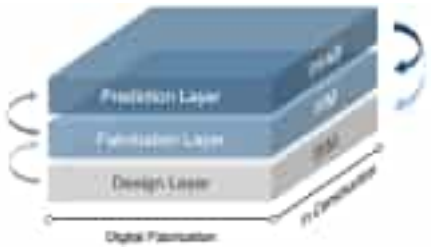


Figure 1. Layered view of digital fabrication in construction.

cation system to enable performance-driven optimization of process parameters. The resulting prediction model can be integrated into the fabrication layer for experiment-specific refinement and into the design layer for outcome-informed decision-making. This perspective provides the conceptual foundation for the generic software framework presented in this work.

2 Related Works

Recent work in digital fabrication highlights the strong coupling between design geometry, material behavior and process parameters, underscoring the need for structured process descriptions [6]. Frameworks such as Fabrication Information Modeling (FIM) formalize the digital-physical data flows required for cyber-physical operation [5]. Building on this foundation, data-driven approaches in AM have focused on feature extraction and learning methods to interpret sensing data and enable predictive modeling [4].

2.1 Data-driven additive manufacturing

Additive manufacturing processes build components layer by layer through controlled material deposition, giving rise to sensor measurements that are inherently spatially and temporally structured. Feature extraction forms a foundational component of data-driven workflows, as the performance of downstream learning processes is ultimately bounded by the quality and relevance of the extracted features. Wolfs et al. synthesized key quality indicators across scales and emphasized the importance of in-line and on-line measurement for achieving first-time-right manufacturing [7]. Versteege et al. demonstrated how knowledge-driven feature engineering can transform high-dimensional sensor data into compact, physically meaningful representations suitable for monitoring and learning [8, 9].

Complementary efforts use machine learning across a range of AM technologies to predict specific outcomes, such as buildability constraints [10] or microcrack investigations for the optimization of material performance [11]. Physics-informed and hybrid modeling strategies demonstrate how physical principles can be embedded into learning models, as shown in co-driven approaches that incorporate physical equations to improve rheological prediction and parameter selection [12]. Zhao et al. highlight the need for real-time monitoring to capture process deviations in-situ, as demonstrated by transformer-based detectors for in-process defect identification and analysis [13].

2.1.1 ML-oriented frameworks in related domains

Other fields have recognized the need for modular learning workflows that structure data processing, model training and deployment. In computational design, AIXD provides a low-code environment that integrates surrogate modeling and inverse design within an orchestrated pipeline [14]. General AutoML frameworks such as TPOT and Auto-Sklearn 2.0 [15, 16] further exemplify this principle by automating feature preprocessing, model selection, and validation within standardized pipelines. These frameworks demonstrate how modular architectures coordinate data, models and decisions, lowering the barrier to deploying data-driven techniques in complex learning workflows.

2.2 Summary of related works

Existing research shows that robust feature extraction remains fundamental, as predictive accuracy depends critically on the information derived from process measurements. However, current sensing, prediction, and hybrid modeling efforts address specific components of the fabrication process, leaving substantial potential in unifying these contributions within a holistic framework. More recently, studies have pointed to the need for real-time monitoring and adaptation during fabrication, where deviations must be addressed as they emerge to enable first-time-right manufacturing. At the same time, most approaches remain tightly coupled to specific materials, machines, or sensing setups, limiting their applicability across the diverse and variable characteristics of digital fabrication methods.

3 PFAB Framework

PFAB is a generic software framework for predictive fabrication that unifies data structuring, learning methods, and optimization routines within digital fabrication workflows. This transforms fabrication experiments into a systematic learning process, supporting parameter calibration and establishing the architecture for predictive closed-loop control. The PFAB framework is released as open-source software under the MIT license and cited as a versioned research artifact [17].

3.1 Overview

General-purpose machine learning frameworks assume the availability of static datasets and do not address a defining characteristic of fabrication systems: the ability to actively generate high-quality data through controlled physical experiments. PFAB is built around this capability, emphasizing targeted experimentation to maximize information acquisition while limiting resource expenditure.

The `PfabAgent` initiates this process by orchestrating the fabrication system through a structured data acquisition and evaluation workflow across an arbitrary number of sensing modalities. The resulting dataset enables the training of predictive models that support holistic, data-driven optimization of process parameters for unseen experimental configurations. The framework is built upon three architectural pillars:

- **Modular system architecture:** PFAB distinguishes between a core layer that defines the data object structure, a logic layer hosting fabrication-specific and internal processing, and an orchestration layer that integrates these components into a unified workflow.
- **Schema-first design:** To manage heterogeneous data, PFAB enforces a strict schema contract that guarantees consistency and enables automated validation across all downstream learning tasks.
- **Stateless agent:** The logic and orchestration are encapsulated in a stateless agent that operates as a pure function of the dataset, ensuring reproducibility and allowing robust recovery from interruptions.

Building on this architecture, PFAB structures parameter selection across three calibration modes: a *baseline mode* that proposes initial experiments through maximin spacing, an *exploration mode* focused on systematic information gain, and an *inference mode* that leverages trained models for performance-driven parameter optimization.

3.2 System requirements

PFAB assumes a digitally controlled fabrication system in which design intent, process parameters, and machine instructions are explicitly parameterized and programmable. Parameter configurations proposed by the predictive layer must be executable without manual intervention. The system must further provide sensor-based feedback capturing *as-fabricated* process information. This may include visual sensing, force or torque measurements, energy consumption, acoustic signals, or laser-based measurements. No specific sensing modality is required, as PFAB operates on derived features rather than raw signals.

3.3 Predictive layer workflow

The PFAB workflow consists of a sequence of stages that connect data structuring, user-defined logic, model training, and parameter calibration. While these steps can vary depending on the fabrication setup, their core structure follows a consistent pattern that integrates fabrication logic with PFAB's learning and calibration mechanisms. The key stages are outlined below.

3.3.1 Problem formalization

PFAB follows a schema-first architecture inspired by AIXD [14], in which all dimensions, parameters, features and performance attributes are defined explicitly before any data is collected. These data objects form a hierarchical structure: `Dimensions` describe the component structure, `Parameters` describe all inputs of the fabrication system, `Features` represent processed sensory information, and `PerformanceAttributes` score the extracted features to evaluate experiment outcomes. `DataBlocks` group these objects into coherent categories, and the resulting `Schema` serves as a strict contract that governs validation, storage and all downstream learning tasks. Once the schema is defined, a `Dataset` instance acts as a lightweight container that stores structured `ExpData` objects, one per fabrication experiment, and automatically manages their storage across local and external data sources (Listing 1).

```
# Definition of Dimensional Domain
dim1 = Dimension("layers", iterator="l_idx")
dim2 = Dimension("segments", iterator="s_idx")
d1   = Domain("spatial", [dim1, dim2])

# Definition of Data Objects
p1 = Parameter.real("speed", max_val=80.0)
f1 = Feature.array("filament_width", domain=d1)
a1 = PerformanceAttribute.score("print_quality")

# Create Data Blocks
params  = Parameters([p1])
domains = Domains([d1])
feats   = Features([f1])
attrs   = PerformanceAttributes([a1])

# Define Schema and initialize Dataset objects
schema = Schema("name", params, domains, feats, attrs)
dataset = Dataset(schema, local_folder="local/")
```

Listing 1. Initialization of schema and dataset.

Fabrication-specific logic is introduced by defining how raw sensor measurements are transformed into structured features and how these features are scored against task-specific target values to evaluate fabrication outcomes. Since the appropriate learning algorithm depends on the structure of the data, PFAB allows users to configure an appropriate prediction model. Once these components are specified and validated against the schema, the `PfabAgent` initializes the systems that orchestrate all subsequent evaluation, learning, and parameter calibration steps (Listing 2).

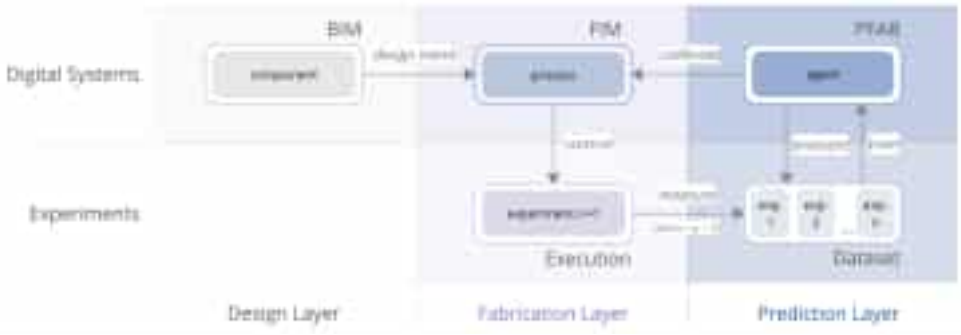


Figure 2. Illustration of the fabrication–learning loop with respect to the design, fabrication and prediction layers.

```python
# Initialize PfabAgent
agent = PfabAgent()

# Import and register user-defined models
import FeatModel, EvalModel, PredModel
agent.register_model(FeatModel)
agent.register_model(EvalModel)
agent.register_model(PredModel)

# Initialize agentic systems and validate against schema
agent.initialize_systems(schema)
```

Listing 2. Agent initialization and user-defined model integration.

3.3.2 Calibration modes

PFAB structures parameter proposal across three named calibration modes that differ only in their objective function, all executed through a single optimization engine, the `CalibrationSystem`. This allows the framework to guide fabrication from the first experiment until production-ready deployment within a unified interface.

Mode	Prediction model	Objective
Baseline	-	Maximin spacing
Exploration	Surrogate	Information-driven
Inference	Trained	Performance-driven

Table 1. Modes of the `CalibrationSystem`.

Baseline mode. Before active model development can start, PFAB generates an initial dataset using the `PfabAgent.baseline_step`, which proposes experiment configurations by maximizing spacing within the defined parameter bounds, requiring no predictive model. The fabricated experiments form an initial schema-based `Dataset` that enables subsequent calibration steps to operate on empirical observations rather than priors.

Exploration mode. The structured dataset is interfaced with the `DataModule`, which handles normalization, batching, and dataset splitting according to the schema and user inputs, ensuring consistent preprocessing across all learning steps. In the exploration mode, the user invokes `PfabAgent.exploration_step` on the current datamodule, and PFAB proposes the most informative next parameter configuration within the defined bounds. The proposed experiment is fabricated and added to the dataset, incrementally improving data coverage and predictive fidelity. This closes the fabrication–learning loop illustrated in Fig. 2, yielding a progressively enriched dataset promoting predictive generalization. The prediction model is periodically trained and validated on the growing dataset using the datamodule via `PfabAgent.train_step`. Once sufficient data coverage and validation accuracy are achieved, the predictive layer is considered production-ready.

Inference mode. The trained prediction model is queried through `PfabAgent.inference_step` to map a given design intent, expressed through its parametric representation, to an optimized set of process parameters for fabrication. Unlike exploration, inference is guided by a performance-driven optimization objective rather than information acquisition. The objective is defined through weighted performance attributes, which encode task-specific trade-offs and are passed directly to the inference step, preserving the stateless nature of the agent while allowing objectives to be adjusted without corrupting the workflow.

3.4 Executable Reference Implementation

An executable reference repository is provided as the primary entry point to the framework. It instantiates the required fabrication-specific interfaces and executes the full agentic workflow using synthetic data, enabling end-to-end execution without hardware or sensing dependencies. The repository is publicly available at `https://github.com/luca-bettermann/pred-fab-mock/tree/release/isarc-2026`.

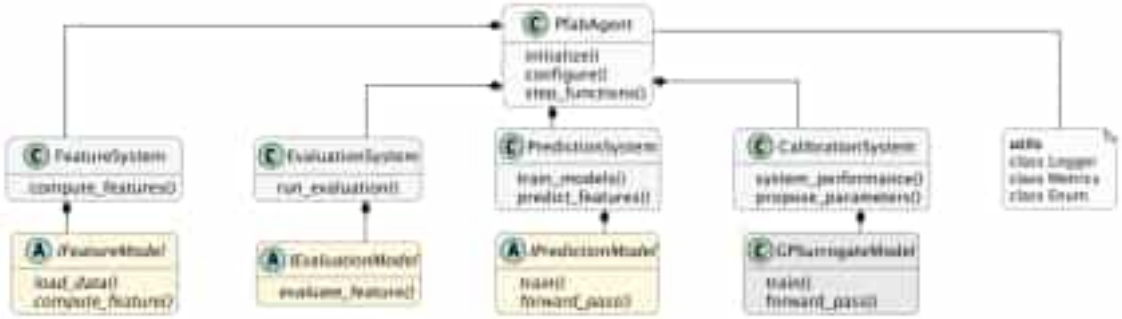


Figure 3. Schematic illustration of the modular `PfabAgent` system architecture.

3.5 Cross-layer integration

The predictive layer described in Sec 3.3 is self-contained and constitutes the methodological scope of this paper. However, as indicated in Fig. 1, the PFAB architecture is explicitly designed to leverage the predictive capabilities in preceding layers of the digital fabrication pipeline. These integrations reuse the foundational data structures and agent abstractions described in this section, and are briefly discussed to clarify the architectural intent of PFAB.

Fabrication-time adaptation. In principle, predictive models trained within the PFAB workflow can be fine-tuned and queried incrementally during fabrication as new *as-fabricated* data becomes available. By operating on explicit dimensions (e.g., printing layers), such integration would enable localized parameter recalibration and predictive closed-loop control. While architecturally supported, the methodology of fabrication-time adaptation is not detailed in this paper.

Design-time prediction. Similarly, trained prediction and evaluation models can be integrated in the design layer to estimate expected features and performance without physical execution. This enables outcome-informed design decisions, but remains outside the scope of this work.

4 Methods

The PFAB architecture is separated into framework-level orchestration and user-defined fabrication logic, as illustrated by Fig. 3. The `PfabAgent` serves as the single entry point and coordinates exploration, training and inference by delegating execution to specialized orchestration systems. These systems automatically manage any number of feature, evaluation and prediction components while keeping the core workflow unchanged.
Fabrication-specific logic is integrated through the highlighted abstract interfaces. Feature extraction is implemented via the `IFeatureModel`, which transforms raw sensor data into structured features. The `IEvaluationModel` maps these features to normalized performance attributes. Predictive models defined through the `IPredictionModel` contract are trained on the resulting dataset and queried during calibration. Together, these interfaces define the low-level extension points of PFAB, enabling fabrication-specific implementations while preserving a fixed orchestration flow.

4.1 Dimensional Domains

In digital fabrication, scalar process parameters are defined per experiment, whereas sensor-derived features are inherently dimensional, measured at specific positions within the fabrication structure (e.g., per layer, per path segment, or per structural node). PFAB formalizes this through the concept of a *Domain*: a named, ordered iteration space composed of *Dimension* objects, each representing one structural axis of the fabrication system (Listing 1). Features declare the domain they belong to, clearly defining their dimensional structure. The schema enforces that all user-defined models contain a coherent domain declaration, validated at system initialization (Listing 2). This explicit structure makes the agent's systems directly tailored to the structural element being fabricated, and forms the architectural basis for fabrication-time adaptation and predictive closed-loop control.

4.2 Calibration strategy

PFAB treats parameter selection as a constrained numerical optimization problem executed by a single solver orchestrated by the `CalibrationSystem`. In both the exploration and inference modes, the system calibrates process parameters by solving the same optimization problem, differing only in the objective function being optimized.
Calibration objectives are expressed through the scalar system performance $S(\cdot)$ that aggregates multiple task-specific performance attributes into a single normalized score. Each attribute is assumed to be normalized to $[0, 1]$

and assigned a user-defined weight w_i reflecting its relative importance. The system performance is computed as

$$S(\mathbf{a}) = \frac{\sum_i w_i \, a_i}{\sum_i w_i}, \tag{1}$$

where $\mathbf{a}$ denotes the vector of performance attributes.

4.2.1 Active learning objective

During dataset generation, calibration is driven by an active learning approach that reduces epistemic uncertainty while maintaining alignment with the user-defined system performance. PFAB employs a lightweight Gaussian Process-based surrogate model `GPSurrogateModel` that maps input parameters directly to expected performance attributes, while providing the required uncertainty estimates. The acquisition function follows an Upper Confidence Bound (UCB) formulation [18]:

$$\alpha(\mathbf{x}) = (1 - \kappa)\, S\big(\boldsymbol{\mu}(\mathbf{x})\big) + \kappa\, S\big(\boldsymbol{\sigma}(\mathbf{x})\big), \tag{2}$$

where $\boldsymbol{\mu}(\mathbf{x})$ and $\boldsymbol{\sigma}(\mathbf{x})$ denote the predicted mean and uncertainty, respectively, and $\kappa \in [0, 1]$ controls the exploration–exploitation balance *within the exploration mode*. The aggregation function $S(\cdot)$ is applied consistently to both quantities to prioritize information gain in regions relevant to the task objective.

4.2.2 Performance-driven objective

In the inference mode, calibration switches from the surrogate model to the full prediction pipeline to prioritize accuracy over uncertainty. The `PredictionSystem` maps input parameters $\mathbf{x}$ to predicted physical features $\mathbf{y}$, which are scored by the `EvaluationSystem` and aggregated into a scalar system performance:

$$J(\mathbf{x}) = S\big(E\big(f_\theta(\mathbf{x})\big)\big)\,, \tag{3}$$

where f_θ denotes the prediction model and $E(\cdot)$ the evaluation mapping.

5 Case Study: Learning–by–Printing

While PFAB is designed as a fabrication-agnostic predictive framework, this paper grounds its concepts through an instantiation in extrusion-based additive manufacturing. In this context, predictive fabrication manifests as Learning–by–Printing (LbP), in which fabrication outcomes are evaluated and used to calibrate process parameters such as nozzle speed and layer height. PFAB provides the structured data representation, learning pipeline, and calibration logic required to implement the fabrication–learning loop (Fig. 2).

5.1 Proof of Concept

The feasibility of Learning–by–Printing was demonstrated in a prior study on data-driven parameter calibration for extrusion-based 3D printing in construction [19]. That work aimed to validate whether a physical printing system can autonomously improve fabrication quality with respect to a defined objective using sensor data and predictive models.

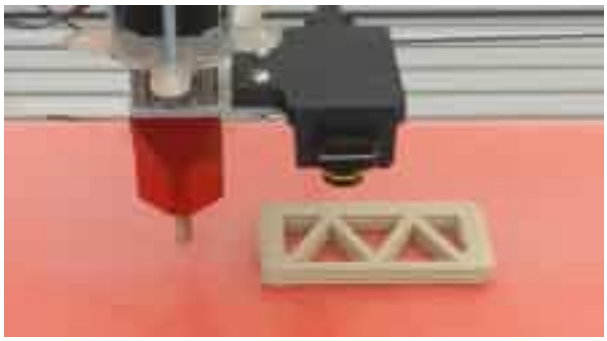


Figure 4. Extrusion-based clay 3D printing setup used in the Learning–by–Printing proof of concept [19].

The study executed the fabrication–learning loop in inference mode. A prediction model was trained from structured fabrication data and subsequently used to calibrate process parameters for future prints. Repeated fabrication trials demonstrated measurable improvements in geometric accuracy and structural consistency, validating the practical effectiveness of embedding predictive models directly into the fabrication workflow.

The Learning–by–Printing study is implemented using PFAB's schema-first formulation. Table 2 summarizes the parameters, features, and performance attributes used in the case study.

Table 2. Schema instantiation in the Learning–by–Printing case study [19].

Parameters	Features	Performance Attrs.
layer time	node overlap	p_{node}
layer height	filament width	$p_{filament}$
path offset	print time	$p_{process}$

Fabrication logic was implemented by extracting geometric features from computer vision-based models and evaluating them against task-specific target values derived from the intended print geometry. These steps directly realize PFAB's `IFeatureModel` and `IEvaluationModel` interfaces, while the `IPredictionModel` was instantiated using Gaussian Process Regression.

The calibration objective is formulated as the weighted

system performance,

$$S(\mathbf{a}) = \frac{w_1 * p_{node} + w_2 * p_{filament} + w_3 * p_{process}}{w_1 + w_2 + w_3}, \tag{4}$$

with weights $w_1 = 3.0$, $w_2 = 2.0$, and $w_3 = 1.25$ reflecting the relative importance of the performance terms. Optimizing this objective enabled automatic parameter calibration toward improved fabrication quality without any manual intervention.

6 Discussion and Conclusion

Prior research in data-driven digital fabrication has established sensing, feature extraction, prediction, and monitoring as key enablers for improving process reliability. However, these contributions typically address isolated aspects of the fabrication pipeline and remain tightly coupled to specific machines, materials, or experimental setups. This fragmentation limits their applicability to first-time-right manufacturing beyond the predefined setting of individual studies. Robust digital fabrication therefore requires jointly addressing process complexity and sensitivity to local, context-dependent disturbances in a generalizable manner.

PFAB contributes a unifying software framework that formalizes predictive fabrication as a structured, end-to-end process linking data acquisition, learning, and parameter calibration. By enforcing consistent data structures and modular interfaces, PFAB enables the development of predictive models that capture general fabrication behavior across experiments. The presented proof of concept in extrusion-based additive manufacturing demonstrates that such models, when embedded directly into the fabrication workflow, can guide parameter calibration and measurably improve fabrication outcomes without manual intervention.

At a conceptual level, PFAB addresses fabrication complexity by modeling the interactions between design, material, process, and environment in a data-driven manner. While this paper focuses on the predictive layer and its calibration workflow, the framework's architecture is explicitly designed to support fabrication-time adaptation as well as design-time prediction. Predictive closed-loop control targets local, non-generalizable sensitivities during execution, while an outcome-informed design stage promotes desirable fabrication outputs. In this way, PFAB provides a foundation for systematically addressing both complexity and sensitivity across the layers of digital fabrication systems.

The framework is presented as a conceptual and architectural contribution referencing an initial proof of concept; extensive empirical validation will be reported in upcoming publications. In practice, predictive performance remains bounded by the quality of feature extraction and evaluation criteria, which cannot be fully standardized and must remain customized to the fabrication context. Extending the predictive layer to fabrication-time adaptation introduces constraints on sensing latency and computational cost, which scale with the complexity of the prediction model's forward pass. These challenges are intrinsic to digital fabrication and reinforce the need for a system-level framework that makes such dependencies explicit and manageable. By standardizing how data, models, and fabrication systems interact, PFAB promotes modularity, transferability, and reproducibility across fabrication contexts. We argue that such data-driven frameworks form a necessary foundation for advancing digital fabrication toward reliable, first-time-right manufacturing in construction.

Acknowledgement

The authors gratefully acknowledge the funding received from EU Pathfinder Project AM2PM (Grant Agreement No. 101162318) as well as from German Research Foundation (DFG) in the frame of Transregio 277/2 project (project number 414265976).

References

[1] Alexander Paolini, Stefan Kollmannsberger, and Ernst Rank. Additive manufacturing in construction: A review on processes, applications, and digital planning methods. *Additive manufacturing*, 30:100894, 2019.

[2] R. Wolfs. The status quo of 3d concrete printing: are we there yet? *RILEM Technical Letters*, 8:182–189, 2023.

[3] Richard A Buswell, WR Leal De Silva, Scott Z Jones, and Justin Dirrenberger. 3d printing using concrete extrusion: A roadmap for research. *Cement and concrete research*, 112:37–49, 2018.

[4] X. Zhang, D. Chu, X. Zhao, C. Gao, L. Lu, Y. He, and W. Bai. Machine learning-driven 3d printing: A review. *Applied Materials Today*, 39:102306, 2024.

[5] M. Slepicka and A. Borrmann. Fabrication information modeling for closed-loop design and quality improvement in additive manufacturing for construction. *Automation in Construction*, 168:105792, 2024.

[6] R. A. Buswell, W. R. Leal da Silva, F. P. Bos, H. R. Schipper, D. Lowke, N. Hack, H. Kloft, V. Mechtcherine, T. Wangler, and N. Roussel. A process classification framework for defining and describing digital fabrication with concrete. *Cement and Concrete Research*, 134:106068, 2020.

[7] R. Wolfs, D. Bos, J.-F. Caron, M. Gerke, R. Mesnil, R. Buswell, N. Ducoulombier, N. Hack, E. Keita, P. Kinnell, et al. On-line and in-line quality assessment across all scale levels of 3d concrete printing. *Cement and Concrete Research*, 185:107646, 2024.

[8] J. Versteege, R. J. M. Wolfs, and T. A. M. Salet. Data-driven additive manufacturing with concrete: Enhancing in-line sensory data with domain knowledge, part i: Geometry. *Automation in Construction*, 172:106020, 2025.

[9] J. Versteege, R. J. M. Wolfs, and T. A. M. Salet. Data-driven additive manufacturing with concrete: Enhancing in-line sensory data with domain knowledge, part ii: Moisture and heat. *Automation in Construction*, 177:106327, 2025.

[10] S. Senthilnathan and B. Raphael. Predicting buildability using the surface texture of 3d printed concrete elements. *Journal of Architectural Engineering*, 31(2):04025010, 2025.

[11] H. Zhao, X. Wang, Z. Chen, X. Liu, Y. Wang, J. Wang, and J. Sun. Microcrack investigations of 3d printing concrete using multiple transformer networks. *Automation in Construction*, 172:106017, 2025.

[12] S.-Y. Geng, B.-Y. Cheng, W.-J. Long, Q.-L. Luo, B.-Q. Dong, and F. Xing. Co-driven physics and machine learning for intelligent control in high-precision 3d concrete printing. *Automation in Construction*, 176:106294, 2025.

[13] H. Zhao, J. Sun, X. Wang, Y. Wang, Y. Su, J. Wang, and L. Wang. Real-time and high-accuracy defect monitoring for 3d concrete printing using transformer networks. *Automation in Construction*, 170: 105925, 2025.

[14] A. Maissen, A. A. Apolinarska, S. V. Kuhn, L. Salamanca, M. A. Kraus, K. Tatsis, G. Casas, R. Bischof, R. Rust, W. Kaufmann, et al. Aixd: Ai-extended design toolbox for data-driven and inverse design. *Computer-Aided Design*, page 103945, 2025.

[15] R. S. Olson and J. H. Moore. *TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning*, pages 151–160. Springer International Publishing, Cham, 2019. ISBN 978-3-030-05318-5. doi:10.1007/978-3-030-05318-5_8. URL `https://doi.org/10.1007/978-3-030-05318-5_8`.

[16] M. Feurer, K. Eggensperger, S. Falkner, M. Lindauer, and F. Hutter. Auto-sklearn 2.0: Hands-free automl via meta-learning. *Journal of Machine Learning Research*, 23(261):1–61, 2022.

[17] L. Bettermann. PFAB: Predictive Fabrication Framework (v0.1.0), December 2025. URL `https://github.com/luca-bettermann/pred-fab`.

[18] N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 1015–1022, Madison, WI, USA, 2010. Omnipress.

[19] L. Bettermann, M. Slepicka, S. Esser, and A. Borrmann. Data-driven parameter calibration in additive manufacturing for construction. In *Proc. of the 32nd Int. Conference on Intelligent Computing in Engineering (EG-ICE)*, 2025.

Text-enriched graph embeddings for HVAC-related maintenance triage

Sajith Wettewa[a,b], Lei Hou[a,b*], Kevin Zhang[a,b]

[a] School of Engineering, RMIT University, Melbourne, Victoria, Australia
[b] Centre for Future Construction, RMIT University, Melbourne, Victoria, Australia
[*]Corresponding author

sajith.wettewa@rmit.edu.au, lei.hou@rmit.edu.au, kevin.zhang@rmit.edu.au

Abstract-
Facilities management teams encounter complex HVAC maintenance scenarios as faults propagate across interconnected subsystems. This study presents HVACONNECT, a text-enriched heterogeneous graph learning framework for HVAC-related maintenance triage. Using six years of CMMS work orders and BAS signals from a 19-building campus, a heterogeneous graph capturing building hierarchies, HVAC systems, and work orders was constructed. A large language model (LLM) pipeline was used to extract semantic data from raw work order text. A relation-aware heterogeneous GATv2 was then used to generate node embeddings via masked feature reconstruction. Feature ablation showed that removing text features reduced clustering quality by 44%, highlighting the critical value of semantic context. Downstream priority prediction achieved an operationally useful 83% precision despite severe class imbalance, and the learned component embeddings enabled meaningful maintenance-zone clustering.

Keywords-
Graph Deep Learning, GNNs, Priority Prediction, Maintenance Triage

1 Introduction

Heating, ventilation, and air conditioning (HVAC) systems dominate commercial building operational loads, requiring proactive maintenance to prevent comfort complaints and energy waste [1]. Traditionally, Fault Detection and Diagnosis (FDD) relies on Building Automation System (BAS) sensor streams. However, scaling BAS analytics across diverse portfolios is hindered by cross-building variability, inconsistent instrumentation, and high operational overhead [2].

Computerised Maintenance Management Systems (CMMS) offer a highly accessible alternative. Work orders routinely capture symptoms, actions, and priority signals without needing extra sensors. While recent Natural Language Processing (NLP) based methods can mine this text for triage, they typically treat records independently, ignoring the interconnected nature of building subsystems [3,4]. Graph learning naturally models these interdependencies. Graph Neural Networks (GNNs) couple equipment topology, operational context, and text-based maintenance events into a unified formulation. Heterogeneous graph architectures, in particular, allow the integration of mixed node and edge types while preserving narrative semantics [5,6].

To bridge this gap, this study introduces HVACONNECT (Heating Ventilation and Air Conditioning Operations Network for Node Embedding, Clustering, and Triage). By unifying BAS signals and CMMS work orders into a heterogeneous graph, our framework learns robust node representations. We demonstrate its operational utility through a three-level priority prediction task and maintenance zone clustering. Therefore, Section 2 reviews related work, followed by the proposed methodology in Section 3. Section 4 presents experimental results, and Section 5 concludes the paper.

2 Related work

BAS-driven fault detection and diagnosis has been widely studied using both domain-informed modelling and data-driven learning. However, persistent organisational and technical barriers exist, including inconsistent data quality, portfolio

variability, and weak alignment between algorithm outputs and actionable maintenance workflows, which are explored below.

2.1 CMMS work order analytics

Maintenance logs were repeatedly shown to contain predictive signals when analysed using text mining and machine learning. Work order narratives support component-level fault frequency analysis, while portfolio benchmarking indicates that these could be used to distinguish maintenance scenarios by building user populations [3,7] More recent work that applied NLP to automate labour assignment used neural networks for priority prediction, while clustering and taxonomy generation supported structured knowledge discovery from unstructured records [4,8]. However, these approaches rarely modelled explicit relationships between assets, symptoms, and interventions as a unified relational system.

2.2 Graph learning for maintenance representation

GNNs were progressively used to encode relational structure for diagnostics and prognostics [5]. In HVAC-specific settings, semi-supervised Graph Convolutional Networks (GCNs) improved fault diagnosis under limited labels, which indicated that topology and relational neighbourhood signals were beneficial when data were scarce [6]. Broader graph learning methods, including neighbourhood aggregation, attention-based message passing, relational convolutions, and heterogeneous graph transformers, enabled scalable modelling of multi-typed nodes and edges [9]. Text-based graph learning further demonstrate the capabilities of integrating text narratives into embeddings [10]. Together, these methods motivate our approach of treating maintenance planning as learning over a heterogeneous, text-bearing building graph rather than as isolated prediction over sensors or tickets.

3 Methodology

3.1 Data set

The data set covered BAS variables with CMMS work orders from a North American university complex as summarised in Table 1.Work orders were available in text format as exported tabular records. Each record contained a text description field alongside structured metadata fields such as dates, component identifiers, and cost values, which were directly ingested into the processing pipeline.

The scope focused on centralised HVAC assets aligned with Uniformat II[11] and Whitestone specifications [12]. Work orders were organised by system, subsystem, and component using unified identifiers for cross-building consistency, and each work order was originally linked to a HVAC component. Only reactive maintenance events were analysed.

Table 1.Work order dataset characteristics

Dataset attribute	Description
Data collection timeline	2014-2019 (6 years)
Institutional distribution	19 Buildings
Dataset volume	2999 distinct work order records
HVAC classification architecture	System, Subsystem, Component
Performance metrics	Labour hours, Material costs, Total maintenance costs
BAS Data	Temperature, Humidity, Atmospheric pressure, Wind Conditions

3.2 Analysis of unstructured work order descriptions

CMMS descriptions used inconsistent terminology and documentation styles. A data extraction step was applied to recover problem and action narratives, floor designations, and room markers, as illustrated in Figure 1. A 300-row gold standard was constructed by two domain experts with over five years of facilities management experience, who independently extracted problem and action labels from raw CMMS descriptions. Disagreements were resolved by a third expert. Four extraction sources were scored against this gold standard. LLM variants comprised three Gemini 2.5 Pro [13] prompt configurations (zero-shot, one-shot, and few-shot, with an increasing number of examples) and an open-source model Qwen3-30B-A3B-Instruct [14]. Each row was scored as correct when all fields matched the gold standard. The few-shot Gemini prompt achieved 99.7% accuracy, with similar performance from Qwen. The few-shot Gemini was retained as the production configuration. Across the full dataset, only 9 rows were flagged for manual checks, confirming that the pipeline operates reliably at scale with minimal human intervention.

Two candidate embedding models, OpenAI text-embedding-3-large [15] and Gemini embedding-001 [15], were evaluated across five metrics to be used as the text embedding model. While Gemini achieved higher thematic cohesion (0.730 vs 0.586), OpenAI demonstrated stronger inter-theme separation (0.646 vs 0.840, lower is better) and higher negation contrast (0.309 vs 0.198). OpenAI was selected as separation and negation sensitivity are more critical for distinguishing opposite system

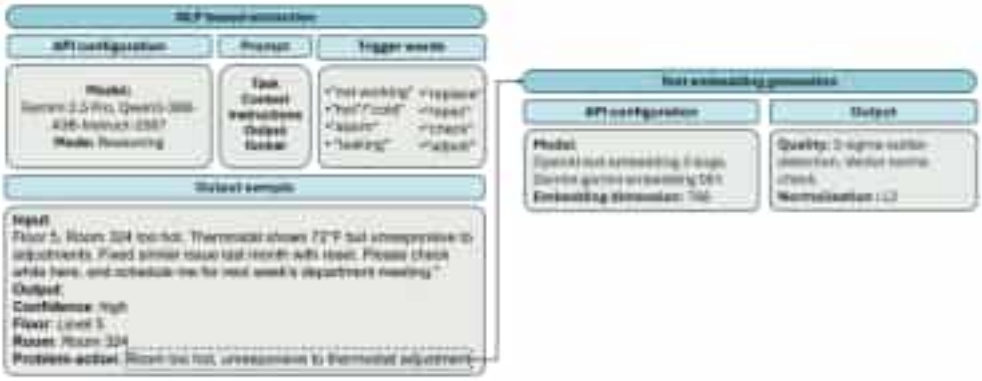


Figure 1. NLP extraction and text embedding generation workflow with other leading models used for comparison of extraction and embedding quality

states, such as “no airflow” versus “adequate airflow” in HVAC maintenance records. 768 embedding dimension is chosen as a standard widely used sentence transformer dimension [16]

3.3 Heterogeneous graph and fault propagation modelling

A heterogeneous HVAC operations graph was constructed to represent potential fault propagation through spatial, functional, and temporal relationships. The design followed a tripartite framing that aligned the representation with standards for system classification, maintenance documentation, and control mapping. The physical infrastructure layer encoded six containment levels, linking buildings to floors, floors to rooms, rooms to HVAC systems, systems to components, and components to work orders. Beyond containment, system-to-system edges captured control interactions and functional flow, and work order-to-work order edges captured temporal adjacency. Edge types and semantics are listed in Table 2.

Table 2. Graph edges and relationships

Source node	Target node	Relationship type
Building	Floor	contains floor
Floor	Room	contains room
Room	System	has system
System	Component	has component
Component	Work Order	has workorder
System	System	controls system
System	System	flows to system (flow type)
Work Order	Work Order	temporal (time difference)

An HVAC-specific flow topology was encoded using system flow edges to reflect service delivery pathways. Cooling generation supplied distribution networks that served downstream terminals, and heat generation followed a similar structure, while control edges represented feedback relationships. Temporal edges linked work orders within a seven-day window to reflect operational interdependence. The median inter-event time between consecutive maintenance events on related components is 4.2 days, with 62.4% of events falling within seven days. Temporal edge weights follow a linear decay from 1.0 at same-day to near-zero at the window boundary. A sensitivity analysis across 3, 7, 14, and 30-day windows confirmed stable results, with chain counts ranging from 225 to 357 and the functional similarity threshold having negligible effect, confirming that the seven-day window is a balanced operating point rather than an arbitrary choice.

3.4 Heterogeneous graph feature engineering

After graph construction, feature sets were assigned to each node type using BAS and CMMS data, as summarised in Table 3. Work order nodes combined text embeddings with numeric attributes representing problem diversity, thermal context, cost signals, and priority information. Features were normalised to stabilise message passing using z-score standardisation for scalar features and L2 normalisation for text embeddings.

Table 3. Summary of feature engineering

Node type	Feature groups	Representation
Building	Size, age, condition and cost indicators	Normalised numerics
Floor	Floor index, room count, subsystem counts	Normalised counts
Room	Floor index, subsystem presence	Indices, binary flags

System	System category, component count	One hot, normalised counts
Component	System family, component class	One hot, categorical codes
Work order	Text semantics, environment, priority, cost, problem diversity	Embeddings, engineered numerics, frequency features

3.5 Graph representation learning and downstream evaluation

Standard graph attention on homogeneous graphs does not preserve relation-specific semantics, therefore a customised heterogeneous GATv2 was implemented to learn embeddings under edge-type attention and multimodal feature integration. The building mask (BM) module given in Eq. (1), (2) reduces cross-building information leakage for within-building analyses. The overall architecture is shown in Figure 2.Figure 2. HVACONNECT model architecture

3.6 Heterogeneous graph attention architecture

Each message passing layer applied relation-specific GATv2 operators and aggregated relation-wise messages using a mean operator to avoid domination by high-degree relations. Four edge categories, building hierarchy, HVAC system structure, HVAC flow, and temporal linkage, were handled by separate operators. Each operator used four attention heads with 32 dimensions per head, producing 128-dimensional hidden representations after concatenation. Work order nodes used a two-stage projection with increasing dropout, while other node types used a single projection to 128 dimensions.

3.7 Domain-informed learning objective

The model was trained transductively using masked feature reconstruction. During training, subsets of input features were masked and the model reconstructed only masked positions, which encouraged reliance on neighbourhood information and reduced identity mapping. Masking rates were 30% for 768-dimensional language features and 10% for engineered scalar features. Reconstruction loss in Eq. (3) used mean squared error on masked indices only. A diversity regulariser Eq. (4), adapted from redundancy reduction principles in Barlow Twins [17] and VICReg [18], maximised mean pairwise Euclidean distance across all component embeddings to counteract global embedding collapse. A separation loss Eq. (6), grounded in supervised contrastive learning [19], operated on building-masked embeddings using cosine similarity to increase within-building alignment while reducing across-building similarity, preserving building-level identity distinct from the global geometric spread enforced by the diversity term. Total loss is given in Eq. (7).

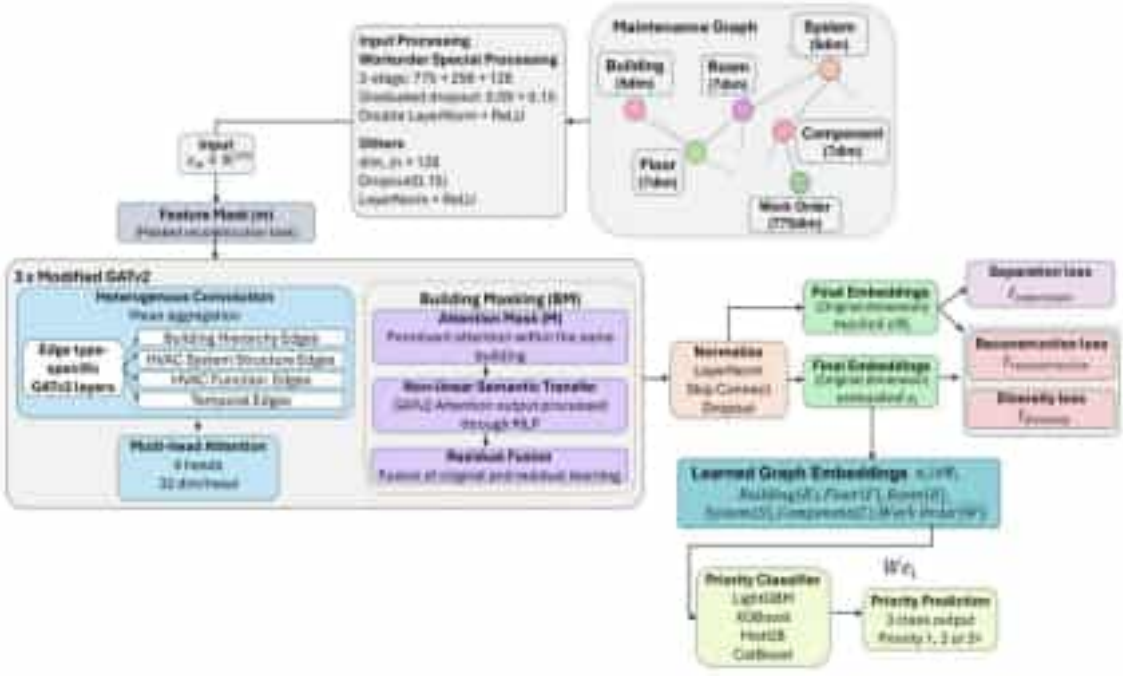


Figure 2. HVACONNECT model architecture

$$M_{ij} = \begin{cases} 0, & b(i) = b(j) \\ -\infty, & b(i) \neq b(j) \end{cases} \quad (1)$$

$$\alpha_{ij}^{BM} = \text{softmax}_j(e_{ij} + M_{ij}) \quad (2)$$

$$\ell_{\text{reconstruction}} = \frac{1}{|m|} \sum_{i \in m} \| D_t(e_i) - x_i^{in} \|^2 \quad (3)$$

$$\ell_{diversity} = -\frac{1}{N(N-1)} \sum_{i \neq j} \| e_i - e_j \|_2 \quad (4)$$

$$\tilde{e}_i^M = \frac{e_i^M}{\|e_i^M\|_2}, S_{ij} = (\tilde{e}_i^M)^\top \quad (5)$$

$$\ell_{\text{separation}} = -\text{mean}(S_{ij} \mid b(i) = b(j), i \neq j) + \text{mean}(S_{ij} \mid b(i) \neq b(j)) \quad (6)$$

$$\ell_{total} = \sum_t w_t \, \ell_{reconstruction}^{(t)} + \lambda + \gamma \, \ell_{separation} \quad (7)$$

Here, $b(i)$denotes the building membership of node i, $e_{ij}$the raw GATv2 attention score, and $\alpha_{ij}^{BM}$the attention weight after building masking. The set $m$contains the masked feature positions, $x_i^{in}$is the original input feature of node i, $e_i$and $e_i^M$are the node embeddings before and after masking, $D_t$is the decoder associated with node type t, $N$is the number of nodes, $S_{ij}$is the cosine similarity between L2-normalised masked embeddings, and $w_t$is the reconstruction weight for node type t. In total loss, $w_t$is the reconstruction weight for node type t, and λand γare the weights for the diversity and separation losses, respectively.

4 Experiments and results

4.1 Priority triage

The applicability of GNN-derived embeddings was first analysed using a priority prediction task using downstream classifiers. Raw CMMS priority values were grouped into three classes to reduce sparsity while retaining operational distinctions of urgent, high, and all remaining priorities.

Models were evaluated using two splits as given in Table 4. Priority labels were severely imbalanced, with 1,186 urgent work orders, 1,731 high-priority work orders, and 82 work orders in the remaining group. Under the stratified split, gradient boosting models achieved strong discrimination. With XGBoost, macro precision reached 0.832 and macro AUROC reached 0.837, indicating that the embedding representation separated classes despite macro averaging penalising minority performance. Under the temporally ordered split, performance decreased modestly. A text-only baseline trained on problem action embeddings underperformed the full HVACONNECT representation, which indicated that triage benefited from combining text semantics with relational and contextual signals learned through heterogeneous message passing.

4.2 Embedding validation and ablation evidence

Ablation analysis (Figure 3) reveals that removing problem and action text features causes the largest degradation in clustering quality, followed by thermal context feature Figure 3. Feature ablation. All changes are sign-normalised (negative:degradation). Baseline: Silhouette:0.896, DB:0.352, IBC:0.221, IBD:0.189. Overall is the unweighted mean of all four metrics. Removing priority deviation has minimal impact, confirming the model does not rely on it as an implicit shortcut. Clustering metrics, namely Intra-Building Cohesion (IBC), Inter-Building Distinction (IBD), Silhouette, and Davies-Bouldin (DB), confirm the embeddings successfully capture building context while maintaining coherent, well-separated structure.

Table 4. Priority prediction metrics

Model	Precision (weighted)	Recall (weighted)	F1 (weighted)	Precision (macro)	Recall (macro)	F1 (macro)	AUROC (macro)
80/20 Split							
LightGBM	0.7652	0.7650	0.7622	0.7958	0.6454	0.6925	0.8349
XGBoost	0.7615	0.7583	0.7563	0.8316	0.6424	0.6961	0.8374
HistGB	0.7612	0.7583	0.7558	0.8325	0.6415	0.6959	0.8340
Temporal Split							
CatBoost	0.7189	0.7200	0.7192	0.5622	0.5381	0.5473	0.7779
XGBoost	0.7349	0.7480	0.7409	0.4958	0.5002	0.4976	0.7950
LightGBM	0.7314	0.7427	0.7362	0.4935	0.4957	0.4940	0.7694
Text embedding only (CatBoost)	0.6252	0.6275	0.6239	0.4412	0.4399	0.4384	0.7050

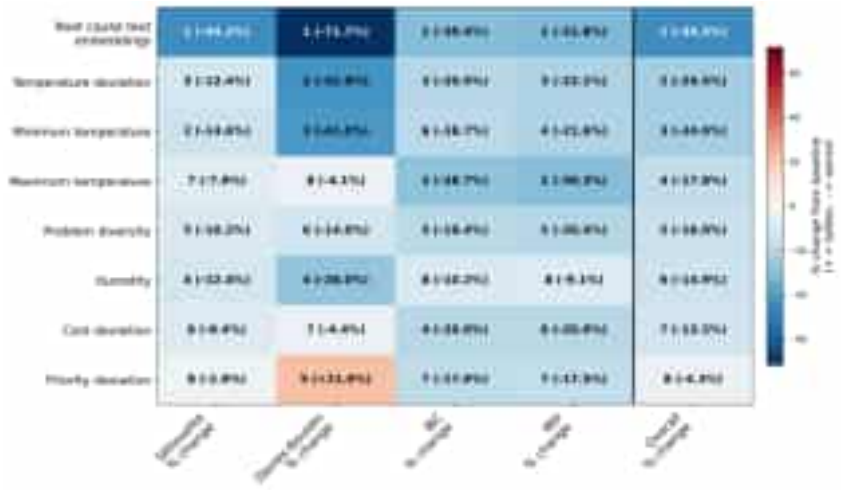


Figure 3. Feature ablation . All changes are sign-normalised (negative = degradation). Baseline: Silhouette:0.896, DB:0.352, IBC:0.221, IBD:0.189. Overall is the unweighted mean of all four metrics

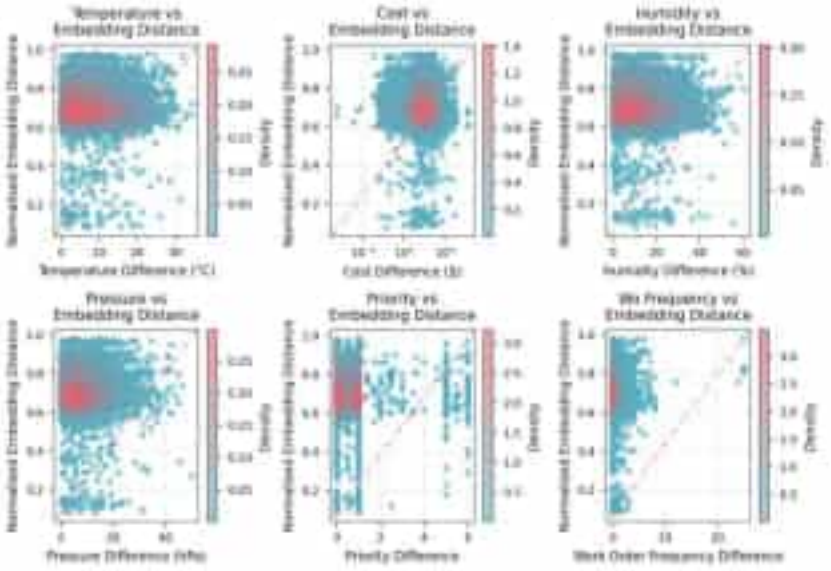


Figure 4. Embedding distance visualisation for feature reconstruction using HVACONNECT

Figure 4 maps pairwise operational attribute differences against normalised embedding distances. Most points sit above the hypothetical y = x line (where a single attribute would explain separation), proving HVACONNECT integrates multiple features simultaneously. Dense clusters with near-zero attribute differences but high embedding distances confirm that contextual factors, such as spatial location, component type, and message-passing dependencies, heavily drive the representations. Additionally, the log-scaled cost subplot shows components with drastically different maintenance costs can remain close in embedding space. This demonstrates that the attention mechanism effectively balances environmental, operational, and economic attributes without allowing any single feature to dominate.

4.3 Maintenance zone identification

A further application of the learned HVACONNECT component embeddings focused on optimising maintenance zones. Traditional zone-based strategies often relied on static, generalised boundaries that were not sensitive to operational conditions or functional interdependencies, which could create inefficiencies and overlapping responsibilities [20]. HVACONNECT addressed this gap by deriving data-driven maintenance zones directly from learned embeddings.

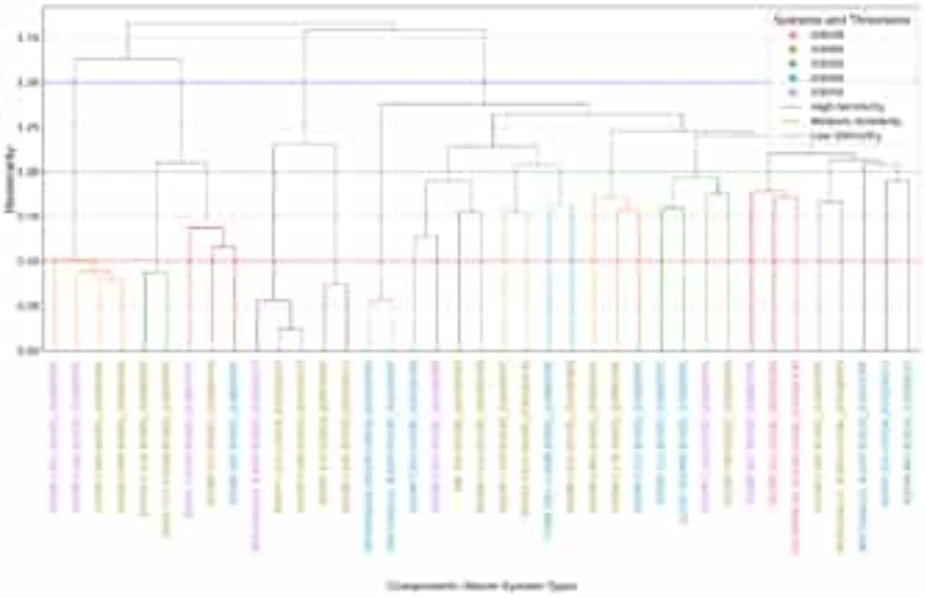


Figure 5. Hierarchical component relationships in building F280

4.4 Portfolio-scale zoning

Portfolio-scale zones were formed by applying DBSCAN to L2-normalised component embeddings using cosine-based dissimilarity. A semantic cut-off of $\theta \geq 0.80$ was used, corresponding to a neighbourhood radius of $\varepsilon = 1 - \theta$, which yielded $\varepsilon = 0.20$ in cosine distance. This configuration produced five dense and well-separated zones that broadly aligned with HVAC subsystem groupings, including heating, cooling, distribution, terminal units, and controls. Location attributes and work order history were attached after clustering to characterise each zone for scheduling and resourcing, while ε remained adjustable to match the operational setting.

The derived zones offered a compact view of resource demand and historical patterns. For example, a controls-centric zone captured a large share of thermostats and panels and exhibited substantial historical labour and cost, while a critical cooling zone isolated a small number of chillers and cooling towers that required specialist coverage despite a comparatively low cost profile.

4.5 Granular building-level relationships

To support building-specific planning, hierarchical clustering was applied to building-masked component embeddings using Ward's method to generate per-building dendrograms. This analysis provided a more granular view of how components clustered based on maintenance histories, and it supported three operational tiers defined by dendrogram distance thresholds (Figure 5). Tier A captured near-identical relationships for $0 \leq d \leq 0.50$, Tier B captured moderate coupling for $0.50 < d \leq 1.00$, and Tier C captured weak coupling for $d > 1.00$. The dendrogram visualisation was reported per building in a Room-System-Component format, which supported rapid inspection of within-system structure and cross-system relationships.

4.6 Limitations and future work

The framework's reliability depends on the quality of historical CMMS data, as inconsistent documentation can degrade the extracted semantics and embedding quality. Validation is limited to a single campus. While operational deployment requires pilot testing with facility management teams, the framework's outputs support several potential workflows. For reactive maintenance, priority predictions and nearest-neighbour searches could enable dispatcher triage and ticket bundling for co-located or similar components. In terms of planned maintenance, aggregated priority predictions by zone could inform labour and inventory forecasting. Additionally, for strategic maintenance, annual summaries of zone workloads could support threshold reviews and long-term replacement planning. Future work should evaluate model transferability across diverse portfolios, integrate real-time sensor attributes, and apply temporal graph learning to capture evolving operational states.

5 Conclusion

This study presented HVACONNECT, a text-enriched heterogeneous graph representation of

HVAC maintenance that unified physical hierarchy, operational relations, and work order history. The learned work order embeddings supported triage through priority prediction under severe class imbalance, while ablation evidence indicated that problem-action text contributed strongly to meaningful embedding structure. Beyond triage, the same embedding space supported data-driven maintenance zoning at portfolio scale and more granular building-level relationship inspection through dendrograms. Together, these outputs suggested a single representation learning backbone that could support reactive routing, planned scheduling, and strategic review of maintenance activities when deployed with human-in-the-loop processes.

6 References

[1] J.H. Lai, Maintenance performance: Examination of the computer-aided maintenance data of a large commercial building, Journal of Performance of Constructed Facilities 29 (2015) 04014118.

[2] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Automation in Construction 50 (2015) 81–90.

[3] H.B. Gunay, W. Shen, C. Yang, Text-mining building maintenance work orders for component fault frequency, Building Research & Information 47 (2019) 518–533.

[4] Y. Li, Y. Liu, J. Zhang, L. Cao, Q. Wang, Automated analysis and assignment of maintenance work orders using natural language processing, Automation in Construction 165 (2024) 105501.

[5] T. Li, Z. Zhou, S. Li, C. Sun, R. Yan, X. Chen, The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study, Mechanical Systems and Signal Processing 168 (2022) 108653.

[6] C. Fan, Y. Lin, M.S. Piscitelli, R. Chiosa, H. Wang, A. Capozzoli, Y. Ma, Leveraging graph convolutional networks for semi-supervised fault diagnosis of HVAC systems in data-scarce contexts, Build. Simul. 16 (2023) 1499–1517.

[7] P. Nojedehi, W. O'brien, H.B. Gunay, Benchmarking and visualization of building portfolios by applying text analytics to maintenance work order logs, Science and Technology for the Built Environment (2021).

[8] M. Ensafi, W. Thabet, X. Gao, Work order prioritization using neural networks to improve building operation, Journal of Information Technology in Construction (ITcon) 29 (2024) 324–346.

[9] S. Wettewa, L. Hou, G. Zhang, Graph Neural Networks for building and civil infrastructure operation and maintenance enhancement, Advanced Engineering Informatics 62 (2024) 102868.

[10] Z. Yu, D. Jin, Z. Liu, D. He, X. Wang, H. Tong, J. Han, Embedding text-rich graph neural networks with sequence and topical semantic structures, Knowl Inf Syst 65 (2023) 613–640.

[11] R.P. Charette, H.E. Marshall, UNIFORMAT II elemental classification for building specifications, cost estimating, and cost analysis, US Department of Commerce, Technology Administration, National Institute of …, 1999.

[12] D. Abate, M. Towers, R. Dotz, L. Romani, The Whitestone facility maintenance and repair cost reference 2009-2010, Whitestone Research, California (2009).

[13] Gemini Pro - Google DeepMind, On-line:https://deepmind.google/technologies/gemini/pro.

[14] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, Qwen3 technical report, arXiv Preprint arXiv:2505.09388 (2025).

[15] New embedding models and API updates, On-line:https://openai.com/index/new-embedding-models-and-api-updates.

[16] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019: pp. 3982–3992.

[17] J. Zbontar, L. Jing, I. Misra, Y. LeCun, S. Deny, Barlow twins: Self-supervised learning via redundancy reduction, in: International Conference on Machine Learning, PMLR, 2021: pp. 12310–12320.

[18] A. Bardes, J. Ponce, Y. LeCun, Vicreg: Variance-invariance-covariance regularization for self-supervised learning, arXiv Preprint arXiv:2105.04906 (2021).

[19] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised Contrastive Learning, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020: pp. 18661–18673.

[20] B. Atkin, A. Brooks, Total facility management, (2021).

Ensembled Machine Learning-Based Cost Prediction Model for Modular and Offsite Construction Projects in Canada

Qudrati Al Wasiew[1], Amirhossein Mehdipoor[2], Aryan Hojjati[2] and SangHyeok Han[1,3]

[1]Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada
[2]Construction Research Centre, National Research Council Canada, Ottawa, ON, Canada
[3]Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Gina Cody School of Engineering and Computer Science, Concordia University, Montréal, QC H3G 1M8, Canada
qudratial.wasiew@mail.concordia.ca, amirhossein.mehdipoor@nrc-cnrc.gc.ca, aryan.hojjati@nrc-cnrc.gc.ca, sanghyeok.han@concordia.ca

Abstract –

Recent studies have highlighted the necessity for a conceptual cost prediction model for Modular and Offsite Construction (MOC) projects. Barriers such as high costs and the absence of scientific methods for justifying cost comparisons between MOC and conventional construction methods impede the adoption of MOC. This study presents an ensemble machine learning (ML) technique that assesses the performance of three distinct algorithms: linear regression, support vector machine regression, and random forest. It illustrates the effectiveness of this approach for preliminary cost predictions in the conceptual stage. The analysis of predicted versus actual costs demonstrated strong correlation, with R^2 approximately 0.78, root mean square error (RMSE) around 13340 (0.053 when normalized), and mean absolute error (MAE) approximately 10331 (0.041 when normalized). Additionally, the implementation of the synthetic minority oversampling technique (SMOTE) eliminated data-related stochasticity, while the timeline-based data collection addressed time-related stochasticity. This ensemble machine learning technique serves as a foundational database for subsequent learning-based models in MOC projects.

Keywords –

Modular and Offsite Construction; Ensembled Machine Learning; Conceptual Cost Prediction

1 Introduction

Modular and Offsite Construction (MOC) is acknowledged as a superior approach for improving sustainability and productivity compared to conventional on-site construction. MOC is a procedure wherein diverse building elements, including walls, floors, roofs, and staircases, are manufactured in a controlled environment and subsequently transported to the construction site for assembly. In this study, MOC or modular refers to factory-produced housing units manufactured in controlled environments and transported to site for final assembly, typically consisting volumetric modules and panelized structural components. However, the execution of MOC has been obstructed by uncertainty concerning the costs and advantages of employing these novel approaches in the construction industry. MOC approaches, including prefabrication, panelization, and modularization, utilize lean manufacturing principles to tackle conventional in-situ building issues, such as productivity, logistics, safety, emissions, waste, quality, and the trade-offs between environmental limitations and prices [1]. Notwithstanding these advantages, there have been increasing obstacles to the adoption of MOC as an alternative in nations such as Canada, Hong Kong, Singapore, the United Kingdom, and the United States. The recent threats to Canada's construction industry resulting from post-pandemic inflation highlight the necessity of adopting MOC as a potential solution to mitigate the negative impacts of inflation in this vital sector. However, progress is impeded by the absence of a cost database that can effectively illustrate the feasibility of MOC compared to conventional methods [2]. The primary obstacle to MOC in Canada is its inadequate comprehension. In the United Kingdom, prefabricated houses have not achieved commercial success due to obstacles such as elevated costs and restricted opportunities for clients and designers to customize [3]. Researchers have observed that ambiguity surrounding initial expenses, capital expenditures, payback timeframes, and supply chain challenges has impeded the implementation of MOC in China [4]. The principal challenges related to insufficient information for adopting MOC include: (1) the conceptual ambiguity surrounding MOC costs, (2) the perception that offsite solutions incur higher costs than traditional alternatives, (3) the absence of cost data and information regarding MOC, and (4) the uncertainty regarding methods to

reduce construction costs while enhancing efficiency.

To tackle with the cost related problems, researchers have followed traditional approaches, statistical methods, machine learning (ML) techniques and hybrid approaches. Traditional cost estimating methods employ detailed quantity take-offs, whereas comparative cost estimates depend on parameters including type, size, capacity, time, material, and other related factors. Traditional cost estimating relies on blueprints and specification breakdowns, which are time-consuming. Additionally, comparative cost estimates, such as case-based reasoning (CBR), assume a linear relationship between the final cost and fundamental design variables of the project. The assumption of a linear relationship is problematic and frequently leads to inaccurate cost estimates [5, 6]. Recent studies have increasingly focused on machine learning and deep learning algorithms with multiple layers, which are effective for comprehending the non-linear patterns inherent in the data [7, 8]. Stochastic parameters include temporal factors such as inflation rates and discount rates, as well as elements like missing data and outliers, which contribute to uncertainty in learning models. Researchers have employed methods such as Monte Carlo Simulation and Fuzzy Logic Theory to address these types of problems by completing the missing data [9, 10].

To address all these problems together, this study employs an ensemble machine learning-based cost prediction methodology, utilizing data from the US Census Bureau survey and RSMeans Square Foot Costs Book, to fulfill the following objectives: (1) to create a conceptual stage MOC project cost prediction model, (2) to establish a model for cost comparison between MOC and traditional alternatives, (3) to compile an initial cost database for MOC projects in Canada, and (4) to analyze the cost savings realized through the application of lean manufacturing techniques in previously completed projects across the U.S. The cross-country data were used because no complete dataset from the Canadian Modular Industry was available, and both the US Census Bureau and RSMeans have been widely adopted as international reference benchmarks in construction cost research. RSMeans provides normalized cost indexes that are transferable across regions and have been directly adopted in this study using two cross-country datasets. RSMeans Square Foot Cost data for Canada are already internally normalized using Canadian labor rates, material pricing, and provincial location factors. The objective of this study is not to establish absolute cost parity between the U.S. and Canada, but to examine whether the learned cost relationships from modular construction data remain valid when applied to region-specific Canadian baseline costs. Section 2 delineates the study approach employed to attain these aims. Section 3 delineates the findings and discourse of this investigation.

2 Methodology

As illustrated in Figure 1, the proposed methodology consists of following five steps: (1) data collection for modular buildings in U.S., data collection for conventional buildings in U.S., and data collection for conventional buildings in Canada; (2) data pre-processing stage combining preparation of data, exploratory data analysis, principal component analysis and mandatory pre-processors for a machine learning model such as one hot encoding and normalization; (3) developing the machine learning model using an ensembled approach with three algorithms; (4) data augmentation using the synthetic minority oversampling technique (SMOTE) and (5) model evaluation and validation. The following sections describe each step of the overview of the research framework.

2.1 Data Collection

The data collection step was mainly conducted through two major data sources which are the manufactured houses survey from US Census Bureau [11] and RSMeans Square Foot Costs Book [12]. 528 data points were collected from the survey results of U.S. Census Bureau ranging from houses 874 square feet gross area to 1879 square feet gross area. The survey was conducted across 44 states of U.S. from 2018 to 2023. Majority of the buildings were constructed using vinyl siding wood frame with concrete blocks and pressure treated wood as the foundation. Average buildings had 2 to 3 bedrooms as per the published information available. The price collected was the average sales price up to transportation to the client site excluding taxes and land acquisition fees. Depending on the available data from the survey results, required data were collected from the RSMeans Square Foot Costs Book for conventional houses in U.S. and conventional houses in Canada. 4483 data points were collected for conventional houses in U.S. ranging from 874 square feet gross area to 1879 square feet gross area to match the available manufactured housing data pattern, from 2018 to 2023. The data was collected for three different material types; wood siding wood frame, brick veneer wood frame, and stucco on wood frame. The data included three different number of floor representatives as in 1 floor, 1.5 floor and 2 floor as per RSMeans data set. The type selected was the single-family residential houses of average grade with continuous RCC footing and average mechanical and

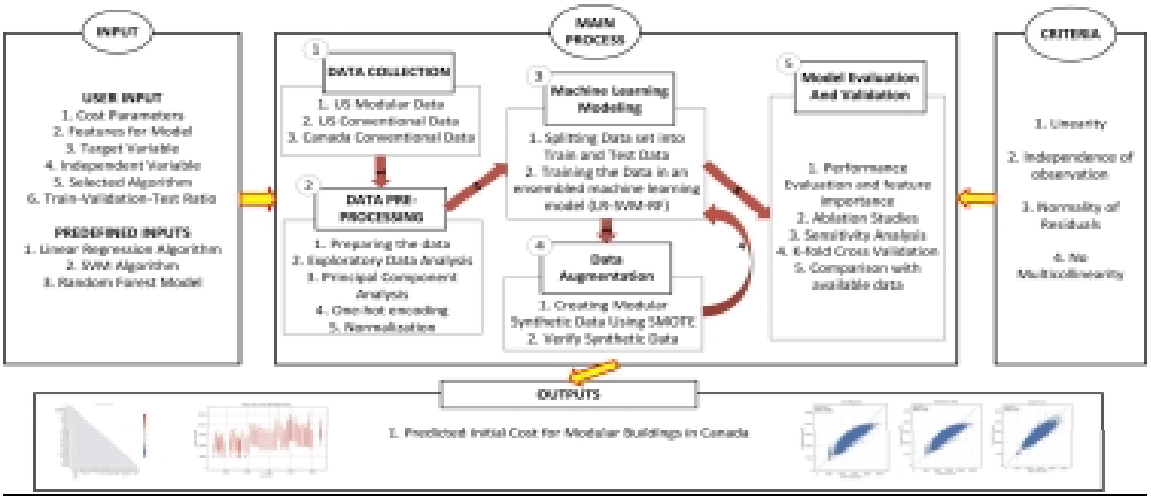


Figure 1: Overview of the research framework.

electrical equipment. The price collected was a sales price with general contractor overhead and profit without tax, which resembles a similar pattern from the conveyed survey results from U.S. Census Bureau. Similarly, conventional house price was collected for the 10 Provinces in Canada as well with a similar pattern of parameters such as area, floors, material and average building details. The total dataset was divided into 80:20 training and testing set ratio. The training dataset was used for model learning and hyperparameter tuning, while remaining 20% was used for independent model evaluation. Table 1 shows the statistical distribution of the collected data across different parameters.

Table 1: Summary of Dataset Distribution and Characteristics

Description	Details
Total Data Points	8738
Location	44 States/10 Provinces
Country	U.S. and Canada
Year	2018-2023
Material	4 Types
Number of Floors	1, 1.5, 2.0
Average Building	Single Family Single Section
Dataset Distribution	
Category	Number of Samples
Modular (U.S.)	528
Conventional (U.S.)	4483
Conventional (Canada)	3727
Total	8738
Model Training Configuration	
Parameter	Value
Train-Test Split	80:20

The Canadian dataset for conventional buildings collected using RSMeans was not used to train the final machine learning model, rather it was used up to data augmentation as a baseline cost for conventional buildings with Canadian parameter. This helped the dataset to learn cost pattern differences among U.S. and Canada regional barrier and the U.S. modular and conventional data helped to learn cost patterns between modular and conventional buildings.

2.2 Data Pre-processing

The second step included preparation of data according to the requirement for the model, exploratory data analysis to understand the existing relationship and patterns, principal component analysis to find out influential parameters connected to the target variable and finally one hot encoding along with normalization before feeding the data into the machine learning model.

2.3 Machine Learning Modeling

A three-algorithm ensemble machine learning model was selected to execute the core methodology. The three algorithms utilized were linear regression, support vector machine, and random forest. The objective of choosing these three algorithms was to furnish a list with diverse capabilities for capturing both linear and nonlinear patterns in the dataset. Linear regression forecasts cost by representing the linear increase or decrease of the target variable as a function of the independent variables. A support vector machine can be optimized by the selection of the kernel and hyperparameters to accommodate variations in the dataset's pattern. Furthermore, random forest is a tree-based algorithm that has recently demonstrated efficacy in identifying high-dimensional

nonlinear patterns inside datasets [13]. Figure 2 illustrates the architecture of the generated machine learning model.



Figure 2: ML model architecture.

2.4 Data Augmentation

The data augmentation process was executed to address some gaps in the U.S. Census Bureau survey data. Data for gross areas ranging from 1202 to 1452 square feet and from 1576 to 1768 square feet were absent. The initial model exhibited subpar performance due to an inadequate variation in the distribution of the chosen independent variable. As some ranges of modular housing sizes were missing in the U.S. Census Bureau data, to ensure a balanced dataset and avoid bias toward certain building sizes, the Synthetic Minority Oversampling Technique (SMOTE) was used to generate additional samples for the underrepresented ranges. At first more data were collected from RSMeans for the underrepresented ranges for conventional U.S. and conventional Canada buildings. Finally, data augmentation was applied to generate synthetic data for the missing ranges in the U.S. modular data as well as some data for Canada modular so that the model is not overfitting to any independent variables. Various data augmentation approaches were examined, including the SMOTE, random oversampling technique, adaptive synthetic oversampling technique, TOMEK under-sampling technique, and k-means and k-medians imputation. Given the availability of real-world datasets pertaining to conventional buildings in RSMeans as a primary data source, the SMOTE was selected for the data augmentation phase. This method generates synthetic data by establishing neighbors within the existing minority data, while a custom relationship logic can elucidate the observed relationships between the majority and minority data throughout the entire dataset. The generated synthetic data was assessed by distribution skewness, correlation preservation tests (such as Pearson and Spearman), and the Kolmogorov-Smirnov (KS) test of the cumulative distribution function [14]. Table 2 shows the data type ratio before and after augmentation process.

Table 2: Data type ratio for collected and augmented dataset

Dataset	Data Type	Ratio
Collected Data (8738)	Real Modular: Real Conventional	6:94
First Augmented Data (36042)	Real Modular: Synthetic Modular: Real Conventional Including Missing Area Ranges	1.5:48.5:50
Final Model Dataset (18153)	Real Modular: Synthetic Modular	2.9:97.1

2.5 Model Evaluation and Validation

The model evaluation and validation phase commenced with an assessment of the created models' performance regarding correctness and errors, with the resultant feature importance in relation to the original principal component analysis. Upon attaining the requisite performance level, multiple ablation tests were undertaken, including the elimination of existing outliers in the dataset, to ascertain whether performance could be enhanced further or if the outliers were merely noise. The building type was concurrently utilized as a variable across all accessible data, including conventional data, to evaluate potential enhancements in model performance and to ascertain the presence of overfitting. A sensitivity analysis was conducted to assess variations in the target variable and to identify any excessively sensitive elements within the model. K-fold cross-validation was ultimately conducted to confirm that the final model was neither overfitting nor underfitting. To make sure the synthetic data behaved like real data, statistical tests such as Pearson, Spearman, and Kolmogorov-Smirnov were used to compare distributions. The results showed no significant difference between real and synthetic samples, meaning the added data followed the same patterns as the original dataset.

3 Results and Conclusions

To assess the proposed machine learning-based cost prediction model, various evaluation methodologies were employed at this stage such as the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) to yield a thorough comprehension of the model and identify the optimal algorithm as shown in Figure 3. Subsequently, feature importance, ablation experiments, sensitivity analysis, and k-fold cross-validation were conducted.

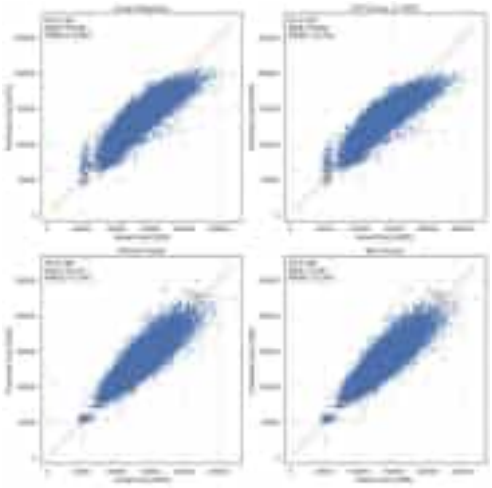


Figure 3: Performance of the machine learning model.

3.1 Performance Evaluation and Feature Importance

As per the ensembled model architecture, random forest is chosen as the optimized algorithm for the dataset. Random forest feature importance was conducted to analyse the influential features that influence the final model. Table 3 shows the performance comparison against existing conceptual cost studies in conventional construction as there is no comparable study in modular construction till date.

Table 3: Performance Comparison

Article	Algorithm	R^2	Dataset
[15]	LR and ANN	0.69 – 0.98	Historical Residential Buildings from Germany
[16]	LR	0.95	Historical Care Retirement Facility
[17]	ANN	0.89	Water Rehabilitation Projects
[8]	GB	0.99	RSMeans Structural Assemblies
Proposed Study	RF	0.78	Combined US Census Bureau and RSMeans

The ensemble architecture was initially developed to evaluate three candidate algorithms' (LR, SVM and RF) performance on the final dataset. As shown in Figure 3, RF was the best performer as per developed ensembled architecture, so final feature importance was conducted for RF. Depending on internal pattern of the dataset, and hyperparameter tuning, the best performer can change in case of any other dataset or this dataset when additional data from Canada modular industry is available. Testing the model's performance on any other dataset is considered outside the scope of this study. Figure 4 shows the output for the random forest feature importance test. Area and year appear as dominant predictors because construction cost estimation is inherently dependent on building size and temporal cost escalation. However, other variables such as location, material type, and number of floors also contributed to model predictions. The scope of this study is within the prediction for a conceptual stage cost for modular buildings, and with the availability of more granular data regarding phases and components future work can investigate more into the dominant predictors such as specific modular building parameters like module sizes, panel sizes and joints.

3.2 Ablation Studies

Two varieties of ablation investigations were performed subsequent to the selection of the final machine learning model. The final model dataset had 18,153 data points, encompassing both actual and synthetic modular building data. The dataset had outliers across various model features, including area, location, material, and year. Subsequent to the final model selection, these outliers were eliminated from the final dataset to ascertain if the model's performance would

improve. The outliers were identified using interquartile range (IQR) method, where observations outside 1.5xIQR from the first and third quartiles were considered potential outliers. The R^2 value decreased from 0.78 to 0.75, suggesting that the remaining outliers in the sample represented significant distributional variance rather than mere noise, and their removal led to underfitting. A dataset with typical RSMeans data included 36,042 data points. The inclusion of "building type" as an additional training feature was examined, yielding an R^2 value of 0.95 and the maximum importance for this feature suggesting overfitting, which resulted in the model's exclusion. The inclusion of "building type" resulted in a significant increase of R^2 value because the model captured dataset-specific correlations rather than generalizable patterns. This is expected since RSMeans data encodes strong cost differences directly through building type (e.g., economy, average etc.). A comparison between initial principal component analysis and this model's random forest feature importance revealed that "building type" did not consistently appear as dominant contributor on principal component analysis explaining overall data variance, however in this model it emerged as an overly dominant predictor, suggesting potential bias. Therefore, instead of applying explicit regularization, a feature exclusion strategy was adopted to improve model generalizability.

3.3 Sensitivity Analysis

A sensitivity study was conducted to evaluate the model's performance across all chosen features. The primary objective of this test was to identify any oversensitive characteristics within the designed model and to ascertain whether the model remained susceptible to uncertain parameters, such as absent outliers, incomplete data, temporal variables like inflation, and other deficiencies.

Figure 5 illustrates the outcomes of the sensitivity analysis. The final model demonstrated a balanced sensitivity across all selected features. The previously observed unclear behavior, stemming from ambiguous parameters, was largely mitigated in the final model.

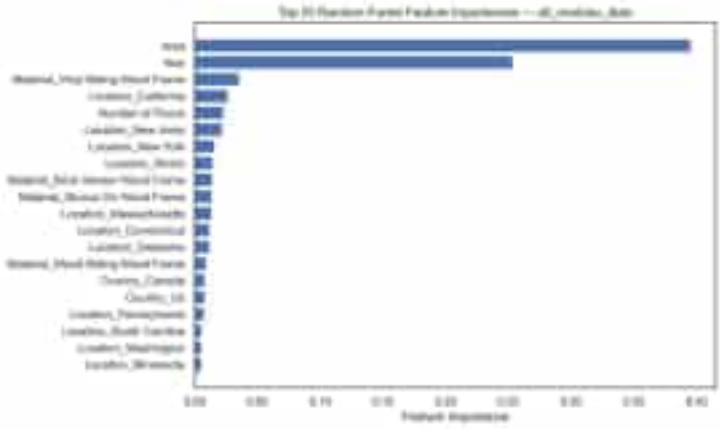


Figure 4: Random Forest feature importance of the final tested model.

3.4 K-fold Cross Validation and Comparison with Available Data

The predictions from the machine learning model were ultimately confirmed by 5-fold cross-validation and by comparing them with the cost data available from the U.S. Census Bureau survey results and the RSMeans database. Figure 6 illustrates that there were no significant alterations in the R^2 value during the 5-fold cross-validation, demonstrating that the model was not overfitted to any subset of the dataset and performs consistently on unseen data. The model achieved a mean R^2 value of 0.779 with a standard deviation of 0.008 and standard error of 0.0036. The 95% confidence interval of the mean R^2 was estimated as [0.77, 0.79], indicating low variability and stable performance across different data partitions.

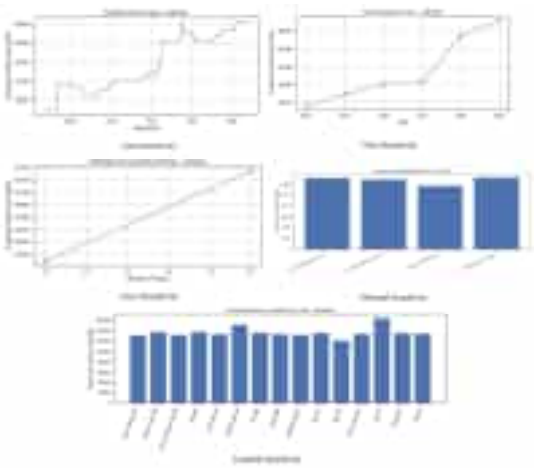


Figure 5: Sensitivity analysis results from the final model.

Subsequently, the projected modular construction cost data was compared with the actual data. The comparison involved analyzing the cost data disparity between modular and conventional buildings from the U.S. Census Bureau survey and contrasting it with the conventional costs from RSMeans in Canada and the projected modular costs derived from the model.

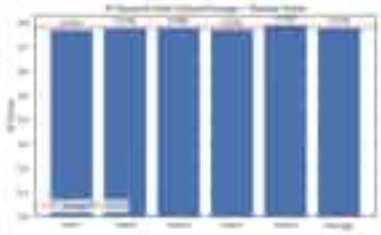


Figure 6: 5-fold cross validation results.

The comparison of analogous features among the selected independent variables in the U.S. and Canada revealed a mere 0.9% discrepancy between the two. This further demonstrates that the model has discerned the relational patterns present in the empirical data. The objective of the comparison in Table 4 is not statistical inference but illustrative comparison of cost patterns between modular and conventional construction. A detailed statistical hypothesis testing framework is considered outside the scope of this conceptual stage study and will be explored in future research. The purpose of this cross-country comparison is also not to establish exact cost parity but to evaluate whether learned cost relationships remain consistent across datasets. Full economic normalization between U.S. and Canadian markets is considered outside the scope of this conceptual model.

Table 4: Comparison with real world data for similar features

Building Type and Country	Area (Sq. ft.)	Cost (US $)	Average % Difference
Conventional, U.S.	1777	239138	-
Modular, U.S.	1777	141200	40%
Conventional, Canada	1777	299350	-
Predicted Modular, Canada	1777	176892	40.9%

3.5 Conclusion

This paper presents an ensemble machine learning strategy to tackle current cost-related issues in conceptual cost prediction for MOC. The proposed framework establishes a thorough methodology to create a verified learning-based cost prediction tool for the MOC sector in Canada.

This study's contributions are as follows. A complete strategy for creating a conceptual-stage cost estimation tool for MOC projects is provided. This framework tackles the dual aims of forecasting modular construction expenses and establishing a preliminary database for MOC projects, which may be utilized for subsequent cost prediction models. This study determines the most significant cost parameters for a conceptual cost prediction model by the integration of comprehensive feature selection and feature importance analysis. This model does not differentiate between specific cost components rather, it illustrates the cost reductions of a MOC project relative to a conventional project. Nevertheless, it can acquire and convey information and expertise from one regional research to another by employing conceptual cost factors. This may prove efficient for future research in sectors lacking data, but notable advancements have occurred in other regional areas. Knowledge and information produced in one region can be effectively assimilated and disseminated using a comparable machine learning model to enhance the designated industry.

Moreover, the established machine learning model can be calibrated to function as a cost-comparison tool between MOC and traditional projects. The ambiguity regarding the advantages of MOC projects can be alleviated. Despite the initial overhead expenses associated with establishing factories and machines potentially surpassing those of traditional projects, the elimination of waste, enhancement of quality, and acceleration of timelines can markedly decrease costs, hence boosting sustainable construction practices.

Applying cross-country normalization is considered outside the scope of this conceptual-stage model. Future

work will investigate explicit cross-country normalization once sufficient real modular cost data become available in Canada. Moreover, although synthetic data were used to fill missing ranges, the model was carefully validated to ensure these data did not distort results. Future work will include more real modular housing data from Canadian projects to further verify model robustness. As the objective of this study is conceptual-stage cost prediction rather than detailed cost breakdown modeling, component-level decomposition such as production, transportation and overhead cost will be addressed in future work when more granular modular factory data become available.

This research facilitates the creation of a viable machine learning-based cost predictor by the implementation of methodologies including timeline data collection, data augmentation, and synthetic data synthesis, which mitigate uncertainties stemming from inflation rates, absent data, and outliers. While the model is proficient in making certain predictions, enhancements can be implemented to evaluate its capacity to address uncertainties. The model could be expanded to function as a more detailed cost breakdown structure for MOC projects when analogous data is accessible from factories. Ultimately, the model may be employed to formulate a multi-objective optimization problem whereby cost serves as either a constraint or the objective function of the favored methodology.

Acknowledgement

The authors gratefully acknowledge the National Research Council Canada (NRC) for their collaboration and funding for this project (Grant ID: CSDP-004-01).

References

[1] C. Mao, et al. Cost analysis for sustainable off-site construction based on a multiple-case study in China. *Habitat International*, 57:215-222, 2016.

[2] A. Mehdipoor, I. Iordanova and M. Al-Hussein. Off-Site Construction Resilience: A Strategic Response to Inflation Challenges in Construction in Postpandemic Canada. *Journal of architectural engineering*, 31(1):04024045, 2025.

[3] E. Vakaj, et al. An ontology-based cost estimation for offsite construction. *Journal of information technology in construction (ITcon)*, 28:220-245, 2023.

[4] C. Mao, et al. Major Barriers to Off-Site Construction: The Developer's Perspective in China. *Journal of Management in Engineering*, 31(3):04014043, 2015.

[5] S. Z. Doğan, D. Arditi and H. M. Günaydın. Determining attribute weights in a CBR model for early cost prediction of structural systems. *Journal of Construction Engineering and Management*, 132(10):1092-1098, 2006.

[6] R. Sonmez. Range estimation of construction costs using neural networks with bootstrap prediction intervals. *Expert Systems with Applications*, 38(8):9913-9917, 2011.

[7] M.-Y. Cheng, H.-C. Tsai and E. Sudjono. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Systems with Applications*, 37(6):4224-4231, 2010.

[8] D. Chakraborty, et al. A novel construction cost prediction model using hybrid natural and light gradient boosting. *Advanced Engineering Informatics*, 46:101201, 2020.

[9] M. Ammar, T. Zayed and O. Moselhi. Fuzzy-based life-cycle cost model for decision making under subjectivity. *Journal of Construction Engineering and Management*, 139(5):556-563, 2013.

[10] O. S. Alshamrani and A. Alshibani. Automated decision support system for selecting the envelope and structural systems for educational facilities. *Building and Environment*, 181:106993, 2020.

[11] U. S. C. Bureau. Manufactured Housing Survey Annual Data On-line: https://www.census.gov/data/tables/time-series/econ/mhs/annual-data.html, Accessed: 11/07/2025.

[12] GORDIAN. Square Foot Costs with RSMeans Data. On-line: https://www.rsmeans.com/products/books/2025-cost-data-books/2025-square-foot-costs-book?, Accessed: 11/07/2025.

[13] H. A. Salman, A. Kalakech and A. Steiti. Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024:69-79, 2024.

[14] V. Buggineni, C. Chen and J. Camelio. Enhancing manufacturing operations with synthetic data: a systematic framework for data generation, accuracy, and utility. *Frontiers in Manufacturing Technology*, 4:1320166, 2024.

[15] O. Dursun and C. Stoy. Conceptual estimation of construction costs using the multistep ahead approach. *Journal of Construction Engineering and Management*, 142(9):04016038, 2016.

[16] R. Sonmez. Conceptual cost estimation of building projects with regression analysis and neural networks. *Canadian Journal of Civil Engineering*, 31(4):677-683, 2004.

[17] T. Shehab and M. Farooq. Neural network cost estimating model for utility rehabilitation projects. Engineering, *Construction and Architectural Management*, 20(2):118-126, 2013.

A Synergized LLM-KG Framework for Domain Knowledge Cognition

Ruiyan Zheng[1] and Jinying Xu[2]

[1, 2]Department of the Built Environment, National University of Singapore, Singapore
zhengruiyan@u.nus.edu, jinying.xu@nus.edu.sg

Abstract –

Large language models (LLMs) evolve rapidly, yet their probabilistic generation is inherently non-traceable and leaves a persistent bottleneck in domain knowledge cognition. Retrieval augmentation via texts or knowledge-graphs (KGs) improves coverage and factual accuracy, but existing studies still depend largely on static knowledge structures and manually governed updating, so the obstacle to sustained and reliable domain knowledge cognition remains. This study therefore proposes a synergized LLM–KG framework for domain knowledge, centered on ontology-governed KG automatic generation (AutoGen) and update, and hybrid vector–graph retrieval for clause-grounded reasoning. Specifically, guided by a manually developed ontology, the KG is automatically constructed, incrementally enriched, and jointly used with vector retrieval to ground each query against both textual evidence and graph structure. A case study is conducted in construction contract management. Results demonstrate the robust KG AutoGen performance with average F1 scores of 0.904 for entity extraction and 0.856 for triple extraction. Across three query stages, the framework further improves multi-hop reasoning and clause-grounded accuracy, with its F1 advantage over GraphRAG widening from 0.00 to 0.37 and that over LLM-only prompting increasing from 0.25 to 0.45. This study provides a traceable route to deeper, more reliable, and continually refined domain knowledge cognition.

Keywords –

LLM; Knowledge graph; Domain knowledge; Knowledge cognition

1 Introduction

Recent advances in large language models (LLMs) have remarkably improved machines' natural language processing performance even in general but complex contexts. However, when application scenarios shift from general-purpose settings to domain-specific environments, the performance is no longer determined solely by interpretative and generative ability [1]. Rather, it fundamentally depends on domain knowledge cognition, defined as the processes by which input information is transformed, processed, stored, retrieved, and used [2]. Despite their strong parametric memory, LLMs remain susceptible to hallucinations due to their probabilistic generative nature, so factual reliability remains a concern in specialized domain knowledge. More importantly, LLMs do not naturally convert unstructured evidence into reusable, explicit knowledge representations, and dependable domain deployment remains difficult [3].

As structured representations of entities and their relationships, knowledge graphs (KGs) are widely regarded as a promising approach to domain knowledge cognition due to their support for logical reasoning and cross-task reuse [4]. Nevertheless, constructing and maintaining high-quality KGs remains costly, and practical deployments often suffer from limited coverage and delayed updates, and high maintenance effort. To address these limitations, increasing attention has been devoted to integrating LLM and KG. Existing studies largely fall into two categories: (i) LLM for KG, where LLMs are leveraged to accelerate entity alignment and relation extraction for automated KG construction; and (ii) KG for LLM, where KGs serve as external knowledge sources that enhance LLM reasoning through approaches such as Graph-based Retrieval-Augmented Generation (GraphRAG) and Think-on-Graph (ToG) [5]. Despite progress along both directions, most existing work focuses on isolated stages of the cognitive process. LLMs are typically used for one-off triple extraction and KG construction, whereas KGs primarily serve as static or manually updated stores for retrieval. As domain evidence is often ambiguous and context-dependent, there is an increasing need for a synergized LLM–KG framework that leverages their complementarity to update domain cognition through query-driven interaction, in a manner akin to human experts [6].

To address this gap, this study proposes a synergized LLM–KG framework tailored for domain knowledge cognition, which integrates ontology-governed KG

AutoGen and update with hybrid retrieval-augmented reasoning. Departing from conventional functionally isolated designs, this framework aims to construct a continuously improving cognitive system that synergizes KG and LLM. Specifically, the domain ontology is first manually established as the governing schema and representational boundary for domain knowledge. Building on this, an LLM-driven and ontology-governed KG AutoGen pipeline is developed to support domain KG initialization and incremental update. At the query stage, the framework can acquire relevant external resources online, update the vector database (DB), and incorporate domain knowledge via KG AutoGen into KG DB. More importantly, it introduces a hybrid reasoning mechanism, in which the ontology schema, vector DB, and KG DB are jointly leveraged to ground evidence and strengthen reasoning capability.

A construction contract management case study is conducted to evaluate the proposed framework. Using the Public Sector Standard Conditions of Contract for Construction Works 2020 (PSSCOC CW 2020) in Singapore as the source corpus, this study excerpts 3 subparts for the comparative experiment. To ensure a fair comparison with baseline methods, online search is excluded. Instead, the first subpart is used for initialization, while the remaining two are introduced sequentially in two query-driven update stages. The evaluation focuses on two aspects: the accuracy and stability of KG AutoGen, assessed against manually annotated entity- and triple-level gold standards for each subpart; and reasoning performance, evaluated through a 15-question hierarchical benchmark with manually defined gold-standard references, in comparison with LLM-only prompting and GraphRAG under the same information setting at each stage. After systematic examination, the results demonstrate both the reliability of ontology-governed KG construction and update, and the advantage of hybrid retrieval in improving multi-hop reasoning and clause-grounded answering during query stage.

The remainder of this paper is organized as follows. Section 2 reviews related work on LLM-KG integration for domain knowledge cognition. Section 3 presents the proposed methodology. Section 4 describes the construction contract management case study and the corresponding experimental evaluation. Section 5 concludes the paper and discusses future work.

2 Related work on domain knowledge augmentation for LLM

Among existing domain knowledge augmentation research for LLM, domain knowledge cognition can be treated as an operational capability that supports (i) perception of domain evidence, (ii) memory organization and reliable recall, (iii) reasoning over constraints and relational structure, and (iv) decision-oriented use of the derived knowledge [7]. Building on this, existing approaches can be positioned into four representative paradigms: prompting engineering, Naive RAG, Graph RAG, and synergized LLM–KG, as shown in Figure 1.



Figure 1. Comparison of Prompt engineering, Naive RAG, Graph RAG and Synergized LLM-KG.

(1) Prompt engineering

Prompt-based methods treat cognition primarily as an in-model process, with manual prompt construction based on scaffolds, including few-shot exemplars and structured reasoning instructions, to elicit task behavior. Their main advantage is low cost and rapid adaptation, since no external memory layer is required. However, domain knowledge remains inadequate, largely implicit and non-auditable, its reliability depends heavily on manual prompt design and the model itself, and systematic provenance tracking is typically unavailable.

(2) Naive RAG

Naive RAG grounds cognition in external text source materials, i.e., vector database (Vector DB), utilizing retrieval mechanisms as a proxy for memory recall. While this approach effectively mitigates hallucinations and supports efficient retrieval over large document collections, it relies on unstructured, fragmented text chunks. Consequently, the system often lacks the connectivity required for cross-document consistency, struggling to reconstruct the complex causal chains necessary for robust knowledge reasoning [8].

(3) Graph RAG

Exemplified by approaches like GraphRAG and ToG, this paradigm introduces a structured memory in the form of a KG to act as a semantic index. By retrieving entities, navigating relational paths, and aggregating neighborhood contexts alongside text, these methods significantly enhance precision and support for compositional queries. While this improves explainability and reasoning depth, conventional pipelines often treat the KGs as static DBs with high maintenance costs, lacking a mechanism for dynamic, user-driven updates to the knowledge state.

(4) Synergized LLM–KG

This paradigm integrates LLMs and KGs as mutually complementary components in domain knowledge cognition. The KG serves as structured domain memory that preserves entities, relations, and constraints, while the LLM supports extraction, retrieval orchestration, and reasoning over the grounded evidence. Although prior studies have demonstrated the promise of combining LLMs with structured knowledge, most focus on either KG-assisted reasoning or LLM-assisted KG construction in isolation. Much less attention has been given to ontology-governed KG automatic generation and incremental update under newly introduced evidence [6]. Consequently, existing synergy remains largely static at the retrieval stage, rather than extending to controlled knowledge expansion.

Motivated by these observations, the unresolved issue is not simply how to retrieve more evidence, but how to support a continuous domain cognition process in which knowledge can be structured, queried, verified, and selectively updated. This study accordingly tries to address this through two coupled capabilities: ontology-governed KG AutoGen and update, and hybrid vector-graph retrieval for clause-grounded reasoning.

3 Methodology

To operationalize the proposed paradigm, this study develops a synergized LLM-KG framework to support domain knowledge cognition, as illustrated in Figure 2. The framework comprises three parts: document parsing, ontology-governed KG AutoGen and update, and QA with hybrid reasoning. Operationally, the framework starts from a parsed corpus and a manually prepared ontology. The domain ontology is manually developed to provide a consistent schema of concepts, relations, and constraints. Document parsing then converts raw documents into a cleaned corpus and a vector database for retrieval. Governed by the ontology, KG AutoGen transforms extracted evidence into ontology-aligned entities and triples, which are consolidated into the KG as structured domain memory and incrementally updated when new evidence is introduced. At the query stage, the QA module jointly exploits the ontology schema, vector DB, and KG DB to support clause-grounded reasoning, with agentic external search activated when internal evidence is insufficient.

3.1 Document parsing

The process begins with document parsing, where domain documents are uploaded as raw sources. The system first applies structural segmentation to preserve the original information organization, such as section boundaries and figure-related content, so later retrieval can return evidence in coherent units rather than fragmented text. After segmentation, text cleaning and normalization will be conducted to reduce noise introduced by formatting and heterogeneous layouts. The cleaned text is then tokenized and stored in a vector DB, which serves as the retrieval substrate for later evidence access. In short, document parsing turns unstructured materials into a normalized corpus and a vector index, which later supports both ontology construction and evidence grounding in QA.

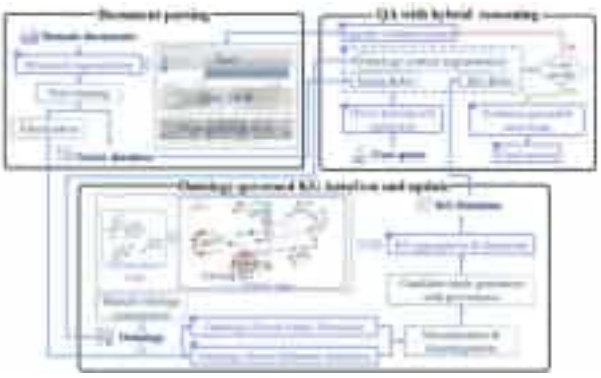


Figure 2. Research methodology.

3.2 Domain ontology manual development

Domain ontology is developed to provide a stable semantic boundary for KG AutoGen and update by defining the admissible classes, relation types, and constraints required for the target domain. The development process follows four steps: identifying core domain concepts from existing domain ontology (such as IfcOWL for BIM/IFC semantics; BOT for building structure; Brick for building systems) or authoritative domain sources (IFC for BIM data structure; bSDD for the built environment; OSHA 29 CFR 1926 for construction safety; FIDIC/NEC for construction contract); defining relation types and semantic constraints to capture domain logic; normalizing synonymous or ambiguous expressions into controlled labels and stable identifiers; and formalizing the resulting schema in RDF/OWL as a machine-readable governance layer. The ontology is then evaluated through competency questions to verify its adequacy for entity typing, relation extraction, and reasoning support. Once established, it serves as the schema boundary for KG extraction and validation, while the KG is incrementally constructed and updated under its governance.

3.3 KG AutoGen

The ontology specifies the normative schema and constraints, while the KG serves as the instance-level memory populated and governed under this schema for subsequent retrieval and reasoning. Given the ontology,

KG AutoGen therefore provides a systematic pathway to transform dispersed textual evidence into a structured and auditable KG. The ontology is maintained in RDF/OWL to provide a formally specified schema of concepts, relations, and constraints, and it functions as a constraint layer that steers extraction toward schema-consistent structures. Specifically, entities are extracted in an ontology-driven manner so that mentions are typed by predefined professional concepts rather than ad hoc labels, and relations are derived under the ontology's relation schema to reduce structural drift and improve constraint consistency. To ensure global consistency across sources and iterations, normalization and disambiguation are performed so that variant surface forms resolve to stable identifiers.

During implementation, the operational KG is stored in Neo4j as a labeled property graph (LPG), enabling efficient subgraph operations and low-latency incremental updates through Cypher queries [9]. Candidate triples are generated with provenance, including source documents and evidence spans, which supports traceable verification before consolidation. To enforce ontology governance over an LPG-based KG, KG AutoGen applies a lightweight schema mapping and validation layer that projects ontology classes and properties to Neo4j labels, relationship types, and admissible property constraints, and uses these constraints to gate triple insertion. Importantly, the generation and consolidation behavior evolve between iterations. In the first run, no existing KG is available, and the newly generated triples are connected and organized to form an initial KG structure. In other runs, the system operates in an incremental mode: newly generated triples are aggregated and aligned against the existing KG, where entity matching, relation alignment, and duplicate resolution enable integration, refinement, and expansion of the graph. Finally, evaluation and validation filter out inconsistent or weakly supported triples before accepted updates are committed to the KG database. Through this iterative construction-and-alignment process, KG AutoGen establishes a validated memory substrate that the QA module can reliably retrieve and reason over, while also enabling controlled assimilation of domain knowledge over time.

3.4 QA system with hybrid reasoning

When user makes a query, the framework runs a QA pipeline that tightly couples synergized retrieval and structured reasoning. A user query is first parsed and expanded into retrieval-ready representations, including key entities, constraints, and decomposed sub-questions. Based on this structured interpretation, the system launches two complementary retrieval streams. The first stream follows a vector-led Naive RAG workflow that retrieves semantically relevant text segments from the vector database. The second stream follows GraphRAG on the KG DB, where Cypher queries retrieve query-conditioned communities and construct a task-specific subgraph that captures relational structure and constraint dependencies that plain text retrieval does not expose. This study further augments the LLM with ontology text, which supplies schema and constraint information to enforce schema-consistent interpretation and reduce structurally invalid inferences. During the process, the LLM acts as an orchestrator that aligns text evidence with graph evidence, checks consistency against ontology constraints, and resolves mismatches between retrieved passages and KG assertions.

When internal evidence is insufficient or outdated, external evidence retrieval is activated. Agentic web search acquires supplementary sources, which are subsequently processed through document parsing and ontology-governed KG AutoGen to produce query-relevant structured evidence. This evidence is then cross-checked against the existing subgraph and ontology constraints before entering the reasoning pipeline. On this basis, the final answer is generated through evidence-grounded synthesis, with explicit reconciliation between text-derived evidence and graph-derived structure.

Reasoning is not treated as a terminal step when externally retrieved evidence is verified. Instead, the framework performs a controlled update, under which only qualified statements are written back to the KG DB. Specifically, a candidate statement is committed to the KG only when (i) its supporting evidence is traceable to concrete sources, (ii) it does not contradict existing KG facts or the evidence base, and (iii) it complies with the ontology-defined schema and constraints. Through this update process, each interaction not only produces an evidence-grounded response but also incrementally improves the knowledge base, increasing coverage and precision over repeated use while preventing unverified errors from being written back and amplified.

4 Case study

To evaluate the proposed framework, a construction contract management case study is conducted using the Public Sector Standard Conditions of Contract for Construction Works 2020 in Singapore (PSSCOC CW 2020). This corpus is well suited to the study because contract reasoning is clause-intensive, procedurally constrained, and inherently multi-hop: practical questions often require navigation across notices, causal events, entitlement conditions, payment procedures, and dispute routes.

4.1 Experimental design

Three subparts are excerpted from PSSCOC CW 2020. Subpart A covers time, claims, loss and expense,

and payment workflow; Subpart B adds instructions, contract documents, ambiguity, and delay causation; and Subpart C further adds variations, valuation, dispute routes, and service of notices. To ensure fair comparison, online search is excluded. The system is initialized with Subpart A only, then updated with Subpart B in the first growth stage and with Subpart C in the second growth stage. This design makes knowledge expansion explicit and allows the effects of KG update on reasoning depth to be examined stage by stage.

4.2 Experimental setting

The evaluation follows two tracks. First, KG AutoGen is assessed against manually annotated gold standards at the entity and triple levels for each subpart. The gold annotations cover 20 entities and 18 triples in Subpart A, 15 entities and 12 triples in Subpart B, and 15 entities and 12 triples in Subpart C. Second, reasoning performance is evaluated using a 15-question hierarchical benchmark with manually defined gold-standard clause references. The benchmark contains five Stage-1 questions over Subpart A only, five Stage-2 questions over Subparts A+B, and five Stage-3 questions over Subparts A+B+C. Their reasoning depth increases from single-chain 3-4 hop, to cross-subpart 5-6 hop, to multi-subpart 7-9 hop, while the number of gold-standard sections rises from 8-10 in Stage 1, to 12-14 in Stage 2, and to 18-24 in Stage 3. LLM-only prompting and GraphRAG are adopted as baselines, and all models receive the same information at each stage. Precision, recall, and F1 are reported for both KG AutoGen and reasoning experiments. Figure 4 illustrates the implemented interface for clause-grounded answering and graph inspection.

4.3 Results and discussion

For the construction contract management case, the domain ontology was manually developed as a contract-variant-specific governance layer for KG AutoGen and update. Rather than aiming to cover construction contracts in general, the ontology was bounded to the conceptual and procedural scope of the target corpus, namely the Public Sector Standard Conditions of Contract for Construction Works 2020 in Singapore (PSSCOC CW 2020), with emphasis on the three selected subparts used in the staged experiment: time extension, claims and payment workflow; instructions, document hierarchy, ambiguity, and delay causation; and variations, valuation, disputes, and notice service. Within this boundary, it specifies the main entity classes, relation types, and semantic constraints needed for clause-grounded reasoning, including actors such as Contractor and Superintending Officer, instruments such as notices and instructions, events such as delay and ambiguity, and procedural objects such as claims, payment certificates, variations, valuation steps, and dispute routes. The ontology was formalized in RDF/OWL and validated through competency questions derived from the benchmark, such as whether the schema could represent the notice and substantiation chain for time and loss-and-expense claims, the contractual effectiveness of oral instructions, the resolution path for ambiguity-linked delay and cost, the variation-to-valuation-to-payment route, and the effect of service-of-notice rules on later dispute procedures.

Building on the ontology, KG AutoGen constructs the knowledge graph through staged initialization and incremental update. In the initialization stage, ontology-typed entities and schema-consistent relations are extracted from the initial contract corpus, transformed into provenance-backed triples, and consolidated into an initial KG. In later stages, when additional contract subparts are introduced, KG AutoGen performs incremental consolidation by aligning newly generated triples with the existing graph through entity matching, relation alignment, and duplicate resolution. In this way, the KG grows progressively with the expanding corpus while remaining governed by the same ontology-defined schema and constraints. Figure 3(a) visualizes the resulting graph at the final stage, and Figure 3(b) shows a representative clause-centered subgraph.



Figure 3. Visualization of (a) the 3rd update stage KG and (b) the detailed subgraph.

The framework is implemented as an end-to-end operational system, where LLM services via the DeepSeek API are used to orchestrate query interpretation, evidence alignment, and verification, while the KG is stored in Neo4j to support efficient subgraph retrieval and incremental updates. As illustrated in Figure 4, users interact with the system through a QA interface. For each query, the system retrieves both passage-level evidence from the vector database and structure-level evidence from the KG, constructs a query-conditioned subgraph, and generates an evidence-grounded answer with clause-linked support. In parallel, the interface exposes the retrieved related

knowledge as an interpretable graph view and triple list, enabling users to inspect the reasoning substrate rather than receiving a black-box response. For more details, please check the demo exhibition video from the website: https://www.youtube.com/watch?v=IDWH5f-LsZE.

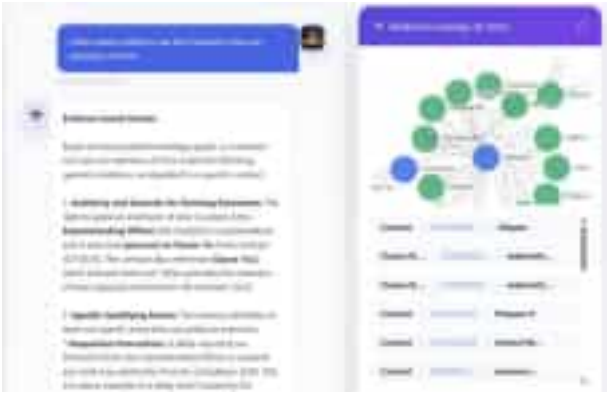


Figure 4. The synergized LLM-KG cognition system.

To further assess hybrid reasoning performance, the proposed framework is compared with two baselines: DeepSeek (LLM-only) and Neo4j GraphRAG (GraphRAG). To ensure a fair comparison, the LLM-only systems operate without access to the internal KG or any online search; the same base LLM and embedding model are used across GraphRAG and the proposed framework, while the difference lies in whether ontology-governed KG AutoGen, KG incremental update, and ontology-augmented retrieval are enabled. All methods are evaluated on the same set of contract management questions under the same staged information setting.

Tables 1 and 2 summarize KG AutoGen performance across the three subparts. Entity extraction remains strong across all stages, with F1 scores of 0.927, 0.903, and 0.882 for Subparts A, B, and C, respectively, yielding an average F1 of 0.904. Triple extraction is more demanding but still robust, with F1 scores of 0.889, 0.800, and 0.880 and an overall average of 0.856. These results show that ontology governance provides a stable extraction boundary not only during initialization but also during incremental update, even when later subparts introduce more cross-clause and procedural relations.

Table 1. Performance of machine annotated entity across subparts.

Subpart	Precision	Recall	F1
A	0.905	0.950	0.927
B	0.875	0.933	0.903
C	0.833	0.938	0.882

Table 2. Performance of machine annotated triple across subparts.

Subpart	Precision	Recall	F1
A	0.889	0.889	0.889
B	0.769	0.833	0.800
C	0.846	0.917	0.880

Table 3 shows that the reasoning advantage of the proposed framework becomes more evident as the knowledge state grows and the questions become more multi-hop. At initialization, the proposed method matches GraphRAG at F1 = 0.76 and already exceeds LLM-only prompting by 0.25. After the first update, its F1 rises to 0.79, outperforming GraphRAG (0.55) and LLM-only prompting (0.43). After the second update, the advantage widens further to 0.82 versus 0.45 for GraphRAG and 0.37 for LLM-only prompting. This trend indicates that the main benefit does not lie merely in storing more information, but in updating the KG under ontology governance and then retrieving evidence jointly from vector space and graph structure.

Table 3. Results of evaluation metrics for compared models.

Stage	LLM-only			GraphRAG			Ours		
	P	R	F1	P	R	F1	P	R	F1
Startup	0.42	0.66	0.51	0.72	0.81	0.76	0.72	0.82	0.76
Update 1	0.36	0.53	0.43	0.49	0.62	0.55	0.75	0.84	0.79
Update 2	0.31	0.46	0.37	0.39	0.52	0.45	0.78	0.86	0.82

In qualitative terms, LLM-only can produce fluent but weakly grounded answers, whereas GraphRAG improves structural retrieval but remains limited when the graph is incomplete and processing capability is limited at one time. By contrast, the proposed framework returns clause-linked answers together with an inspectable subgraph, making the reasoning path auditable. This is particularly important for contract management, where whether a claim succeeds often depends on the exact interaction among notice clauses, evidentiary requirements, and procedural escalation.

5 Conclusion

This study proposes a synergized LLM-KG framework for domain knowledge cognition, centered on ontology-governed KG automatic generation and update, and hybrid vector-graph retrieval for clause-grounded reasoning. Unlike conventional approaches that treat KG construction, updating, and querying as isolated functions, the proposed framework connects these stages within a unified workflow, in which structured knowledge can be automatically constructed from domain evidence, incrementally expanded under

ontology governance, and jointly exploited with vector retrieval during reasoning. In this way, the framework provides not only evidence-grounded answers, but also a traceable and continuously improvable knowledge base.

To ensure reliability, the framework uses a manually developed ontology as the schema boundary for entities and triples extraction and validation, thus KG is constructed and updated through provenance-backed and schema-consistent triples. At the query stage, textual evidence and graph structure are jointly retrieved to support clause-grounded reasoning. The construction contract management case study demonstrates that the proposed framework achieves robust KG AutoGen performance across initialization and update stages, while also improving multi-hop reasoning and clause-grounded answering over both LLM-only prompting and GraphRAG baselines. These results indicate that the main advantage lies not merely in storing more domain knowledge, but in governing KG growth through ontology constraints and integrating vector-based and graph-based evidence during reasoning.

Several limitations also remain. Future work will further validate the framework compared to broader advanced LLMs and retrieval frameworks with real-world dynamic document settings. Other important directions include more systematic handling of conflicting evidence, more principled criteria for ontology refinement, and broader evaluation of long-term knowledge growth under repeated updates. In addition, future research may explore closer integration between external evidence acquisition and internal KG consolidation, so that domain knowledge can be expanded in a more timely and trustworthy manner.

6 Acknowledgement

This research is funded by NUS Start-up Grant Project: “Digitally Empowered Project Cognition for Sustainability”.

References

[1] Y. Wan, Y. Liu, C. Chen, M. Packianather, Empowering LLMs by hybrid retrieval-augmented generation for domain-centric Q&A in smart manufacturing, Advanced Engineering Informatics 65 (2025) 103212. https://doi.org/10.1016/j.aei.2025.103212.

[2] U. Neisser, Cognitive Psychology: Classic Edition, Psychology Press, New York, 2014. https://doi.org/10.4324/9781315736174.

[3] M. Li, Z. Wang, BuildingGPT: Query building semantic data using large language models and vector-graph retrieval-augmented generation, Building and Environment 287 (2026) 113855. https://doi.org/10.1016/j.buildenv.2025.113855.

[4] B. Mo, K. Yu, J. Kazdan, J. Cabezas, P. Mpala, L. Yu, C. Cundy, C. Kanatsoulis, S. Koyejo, KGGen: Extracting Knowledge Graphs from Plain Text with Language Models, (2025). https://doi.org/10.48550/arXiv.2502.09956.

[5] C. Zheng, S. Wong, X. Su, Y. Tang, A. Nawaz, M. Kassem, Automating construction contract review using knowledge graph-enhanced large language models, Automation in Construction 175 (2025) 106179. https://doi.org/10.1016/j.autcon.2025.106179.

[6] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, X. Wu, Unifying Large Language Models and Knowledge Graphs: A Roadmap, IEEE Trans. Knowl. Data Eng. 36 (2024) 3580–3599. https://doi.org/10.1109/TKDE.2024.3352100.

[7] G. Ciatto, A. Agiollo, M. Magnini, A. Omicini, Large language models as oracles for instantiating ontologies with domain-specific knowledge, Knowledge-Based Systems 310 (2025) 112940. https://doi.org/10.1016/j.knosys.2024.112940.

[8] Q. Chen, X. Yin, B. Yuan, Q. Chen, Personalized safety training for construction workers: A large language model-driven multi-agent framework integrated with knowledge graph reasoning, Computers in Industry 174 (2026) 104399. https://doi.org/10.1016/j.compind.2025.104399.

[9] D. Zhang, G. Ma, T. Qu, X. Wang, W. Zhou, X. Wang, A knowledge graph-enhanced large language model for question answering of hydraulic structure safety management, Advanced Engineering Informatics 66 (2025) 103468. https://doi.org/10.1016/j.aei.2025.103468.

Robot assisted selective disassembly for buildings

Kara Williams[1], Anupam Satumane[1], Shayan Shayesteh[1] and Benjamin Sanchez[1]

[1]Department of Sustainable Technology and the Built Environment, Appalachian State University, USA

williamskd9@appstate.edu, satumanead@appstate.edu, shayestehs@appstate.edu, sanchezandradeb@appstate.edu

Abstract –

This study explores the integration of Building Information Modeling (BIM)-based selective disassembly planning with robotic automation to facilitate the efficient and autonomous retrieval of building components, addressing critical challenges in the construction industry's transition toward a Circular Economy (CE). The BIM-based disassembly planning system developed in previous studies forms the core of this framework, generating optimized sequences for disassembly while considering environmental impacts, cost efficiency, and structural dependencies. This information is processed and utilized by robotic operating systems to execute precise disassembly tasks, as demonstrated through virtual simulations and pilot tests. The results indicate the technical feasibility of robotic selective disassembly for simple building archetypes, highlighting the potential for scaling this approach to more complex scenarios. The proposed method offers substantial decarbonization benefits by enabling material recovery and reuse, reducing reliance on virgin resources, and minimizing waste generation. These advancements directly contribute to reducing greenhouse gas emissions associated with material extraction, processing, and disposal. Additionally, this approach aligns with CE principles by extending the lifecycle of building components and promoting sustainable construction practices. Future work will focus on validating the framework in real-world applications, addressing challenges such as variability in building systems, structural interdependencies, and environmental constraints. The integration of advanced robotics, including AI-driven perception and adaptive control, will further enhance scalability and robustness. Moreover, potential applications in extreme environments, such as extra-terrestrial construction, underscore the versatility of this approach.

Keywords –

Disassembly planning; Robot assisted disassembly; Building Information Modeling; Cradle-to-Cradle; Circular Economy

1 Introduction

Selective disassembly sequence planning (SDSP) is emerging as a transformative technology in the disassembly of buildings, offering several advantages. It facilitates the efficient retrieval of reusable building components and materials [7,9], supports adaptive reuse [15], renovations [16], and maintenance [16], and contributes to the development of superior life-cycle strategies like Design for Disassembly (DfD) [1,5,10,12] and the concept of buildings as future material banks [1,14]. Simultaneously, the field of robotic automation for assembly and disassembly in construction is gaining momentum [8]. The deployment of robotics in this context is proving to increase the efficiency and quality of work, demonstrate both technical and economic feasibility, and address labor shortages in response to the substantial current and future demand for buildings worldwide [4].

The need for implementing SDSP in a systemic manner is critical in today's context for a transition towards a Circular Economy (CE) in the built environment. The construction industry is currently responsible for the depletion of 40% of global natural resources and generates approximately 40% of waste ending up in landfills [1]. These issues are largely due to the fact that the existing built environment was not designed with disassembly, reuse, adaptation, or transformation in mind [12,16]. Additionally, practical SDSP methods and tools have been lacking [14]. However, recent advances have seen the implementation of SDSP [15] and robotic automation [18] using Building Information Modeling (BIM). This convergence has created a clear pathway for leveraging the combined benefits of both technologies to achieve optimized, systemic disassembly of building structures, thereby supporting the decarbonization of the built environment.

The primary aim of this study is to develop a framework for the integration of BIM-based SDSP into robotic automation processes. The framework focuses on the information processing using a previously developed BIM-based SDSP system for enriching the intelligence of robots for the autonomous disassembly of targeted

structure components. The information processing will be elucidated for the technical implementation using different types of robotic solutions (e.g. robotic arms and wheeled robots) and virtual simulations (e.g. robotic arm simulator). A functional demonstration has been created to showcase the preparation, processing, and integration of information necessary for the BIM-based SDSP system and the operating system of a robotic solution.

2 Background

2.1 Selective disassembly for buildings

Selective disassembly sequence planning (SDSP) is the process of recognizing the required consecutive steps for dismantling targeted building components, defining deconstruction activities, and ordering them logically. While SDSP for manufacturing products has been developed over the past five decades, SDSP for buildings has gained more attention only in the last decade due to the growing need for a CE in the built environment [15]. In 2018, the BIM-based Sequential Disassembly Planning for Buildings (SDPB) method was developed for the selective disassembly of building components. The SDPB method analyses the disassembly capacity of building assemblies by generating optimized sequences for retrieving targeted components, considering environmental impact, building cost, and rule-based analysis [14]. Mahmoudi and Hosseini Nourzad [9] developed the Hybrid Disassembly Sequence Planning (HDSP) method and tested it with the same building archetypes presented in the SDPB model. Denis et al. [5] developed the Disassembly Network Analysis (DNA) method for selective disassembly planning by applying network analysis. In a subsequent study, Vandervaeren et al. [16] developed the Material Flow Analysis for Demountable buildings (MFAD) method to evaluate the effect of the disassembly capacity on the building material flow. In 2024, Allam and Nik-Bakht [2] proposed a deconstruction planning method using Knowledge Graph (KG) technology to facilitate the evaluation of disassembly sequences for targeted components. In their work, Abu-Ghaida, et al. [1] developed a disassembly network-based approach for accounting for product recovery potential in building life cycle assessments. The primary distinction between the KG technology [2] and the closed-loop lifecycle information flow [18] lies in their respective structures and flexibility. The KG framework is centered on data integration for lifecycle modeling within a structured and organized arrangement, whereas the closed-loop lifecycle information flow illustrates the circulation of data within a dynamic, circular system. Collectively, these data structuring models are complementary, as KGs can facilitate the utilization of closed-loop lifecycle information flows for continuous enhancement and incorporation into BIM. In 2021, Beurskens [3] presented an assessment method to evaluate the reuse potential of buildings and building products. The author proposed the use of Design Structure Matrices (DSM) for establishing the relationships among building components, with the purpose of identifying the disassembly sequence of the components using Steward's algorithm.

2.2 Robotics for disassembly of building structures

The field of construction robotics has predominantly focused on assembly processes, with numerous successful implementations such as the DFAB House project at ETH Zurich demonstrating robotic fabrication of concrete walls, the SAM100 (Semi-Automated Mason) robot for bricklaying, and the Japanese construction giant Obayashi Corporation's automated systems for steel beam assembly. In contrast, relatively limited research has been directed towards robotic disassembly, particularly in the context of selective disassembly operations [13]. Unlike total disassembly, which involves the complete deconstruction of a structure, selective disassembly represents a more sophisticated approach requiring precise targeting of specific components while preserving the integrity of surrounding elements [14,15]. This strategic methodology, when combined with robotic automation, holds significant potential for enabling adaptive reuse, reconfigurability, and transformability of buildings. Recent advances in robotic technology, particularly when integrated with BIM-based SDSP systems, are beginning to address the complexities inherent in selective disassembly operations, offering promising solutions for more sustainable and circular approaches to building modification and material recovery. However, significant challenges remain in developing robust systems capable of handling the variability and complexity of building disassembly tasks, particularly in terms of component recognition, manipulation strategies, and real-time adaptation to changing site conditions.

Little of the existing literature regarding the integration of robotics with on-site construction activities has focused on disassembly or similar topics. Meta-analyses of robotics in construction (RiC) focused on revealing trends in the technology's implementation, but did not identify disassembly as a significant topic of recent research [6,17]. In addition, human-robot collaboration (HRC) literature sees a much heavier importance placed on construction than deconstruction [11]. Opportunities for the integration of robotics in total disassembly and selective disassembly are relevant. The use of BIM in building design creates a network of spatial data regarding building materials and assemblies that can

be utilized by robotic operating systems for robot assisted selective disassembly. The primary contribution of this paper is the structuring of the SDPB output as an information package compliant with the Industry Foundation Classes (IFC) schema for robotic operating systems, alongside a pilot demonstration of this pipeline simulation.

3 A framework for robot assisted selective disassembly for buildings

The framework is centred on information processing using a BIM-based SDSP system to enhance the intelligence of robots for the autonomous disassembly of targeted components, leveraging a system developed in earlier studies (see Figure 1). This system provides crucial information for autonomous selective disassembly, including optimized sequences for disassembling building parts (components and connections), geometrical and topological information for building parts, three-dimensional coordinates for each part (x, y, z), hosted components, hosted connections, connection type, physical interface characterization, structural composition relationship, graph data structure, extraction directions for each building part ($+x$, $-x$, $+y$, $-y$, $+z$, $-z$), and physical constraining parts in each extraction direction. All this information is used and managed by the robot operating system for developing the disassembly tasks.

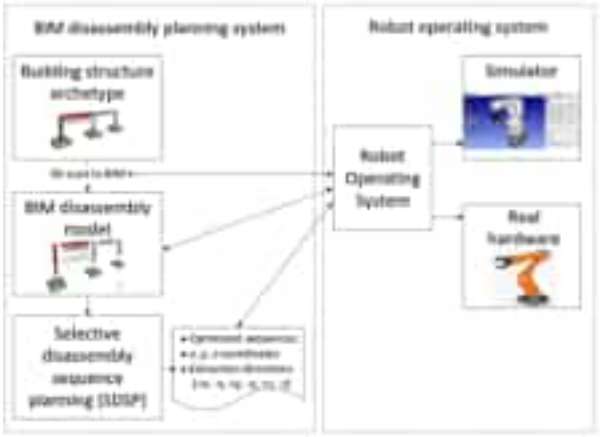


Figure 1. Framework for the integration of BIM-based SDSP for robotic automation

3.1 BIM-based disassembly planning

The BIM-based SDSP system selected for this research is the Sequential Disassembly Planning for Buildings (SDPB) method developed in previous studies. The SDPB method is based on an approach developed for product manufacturing, namely the Disassembly Sequence Structure Graph (DSSG) method. The SDPB method is used to generate optimized disassembly sequences for retrieving targeted components from BIM-based models by considering environmental impact, building cost, and rule-based analysis. The information generated by the system is as follows:

- Optimized sequences for disassembling building components.
- Geometrical and topological information for building parts is the information related to the geometrical characteristics of the disassembly parts. This information is fundamental to the creation of an analytical disassembly model that is consistent with the real conditions of the objects in the space as well as for determining the physical constraints among objects.
- Three-dimensional coordinates for each part (x, y, z) define the origin of a 3D Cartesian coordinate system associated with all the parts of the disassembly model. This characteristic is mainly associated with selective disassembly methods that evaluate extraction directions in a free 3D space, such as matrix-based methods and petri net methods.
- Hosted components and hosted connections register the parts that structurally depend on a host component. This characteristic is exclusive to buildings' disassembly models, in which the static equilibrium plays a critical factor to be able of estimating realistic disassembly patterns.
- Connection type indicates the condition of a fastener for being removed in one direction (e.g. $+x$) or multiple directions (e.g. $+x$, and $-x$).
- Physical interface characterization defines the geometrical and mechanical specifications associated with different types of connections (e.g., bolts, screws, rivets, washers, welding). The physical interface characterization plays a critical role in the accurate study of the disassembly properties of a connection type and the technical feasibility for disassembly.
- Structural composition relationships are an important characteristic for disassembly models of buildings. This characteristic indicates the structural connectivity among parts inside a disassembly model. For manufacturing products, any fastener can be removed from the disassembly model if the fastener is physically accessible. In comparison, in a building composition, some fasteners could be physically accessible but without them, the structural stability of the assembly would be compromised.
- The graph data structure is necessary for establishing the mathematical configuration of the disassembly model. A graph data structure is the

abstract representation of a disassembly model through nodes (vertices) and liaisons (edges).

- Extraction directions are the possible paths for removing any disassembly part. As a generality for SDSP methods, the extraction directions are four in a 2D environment (+*x*, -*x*, +*y*, -*y*) and six in a 3D environment (+*x*, -*x*, +*y*, -*y*, +*z*, -*z*). However, the extraction directions can be increased by adding as many extraction directions as necessary and by adding compounded extraction paths.
- Physical constraining parts are the physical restrictions of a disassembly part in any extraction direction (contact constraints, motion constraints, and projection constraints).

In this study, we tested the BIM-based SDSP system in a building archetype. The BIM-based disassembly model was developed, and the information package was structured according to the IFC. IFC is the most important international standard for open BIM data exchange. IFC is a standardized and continuously expanded data model that facilitates the exchange of BIM data in the AEC domain. Figure 2 shows the building archetype under study, as well as the information package generated by the system.

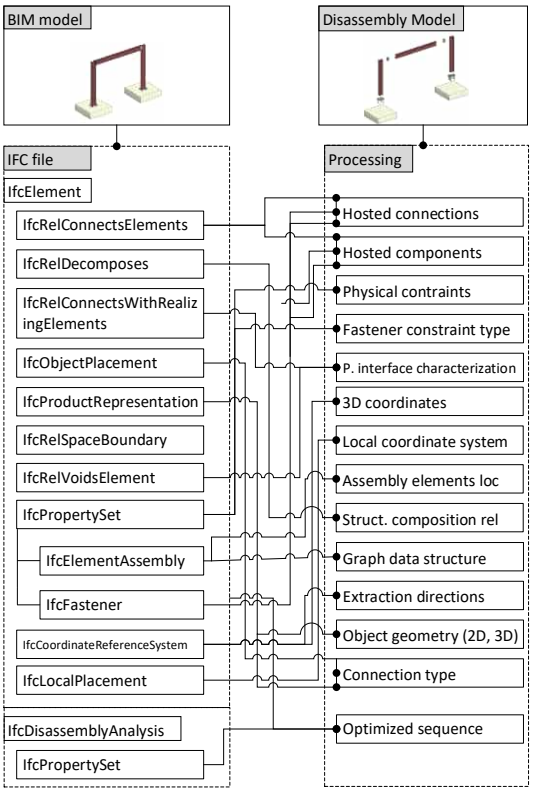


Figure 2. Disassembly sequence IFC information package

3.2 Robotic control architecture for selective disassembly

At the current stage of this research, the BIM-based SDSP framework has been implemented to automatically generate a structured disassembly information package for targeted building components. The robotic simulation and control architecture, however, is proposed as the next stage of development and is therefore presented here as an implementation pathway. The purpose of this phase is to transform the SDSP-generated information into robot-readable commands and environment representations. In this framework, the SDSP output serves as the high-level task planner, while the robotic platform is responsible for motion planning, collision detection, execution logic, and state updates during the disassembly process.

A central motivation for the proposed method is that while IFC preserves rich building information, such as component identities, geometry, spatial relationships, and semantic properties, it is not designed for robot control. Specifically, direct IFC export lacks robot-oriented parameters required for task execution, leaving the challenge of autonomous disassembly unresolved. The proposed SDSP framework addresses this limitation by enriching and reorganizing the BIM-derived data into a structured disassembly information package. In this regard, the SDSP layer acts as an intermediate semantic translation step between the building information model and the robotic execution environment.

The robotic system proposed for this phase consists of a fixed-base articulated manipulator operating within a physics-based simulation. A medium-reach, six-degree-of-freedom (6-DoF) robotic arm is considered appropriate for the initial implementation, as the pilot study involves localized disassembly actions on a simple frame and requires controlled positioning, directional extraction, and manipulation of structural and connection elements. The software stack integrates Gazebo as the simulation environment, ROS 2 as the middleware layer, and MoveIt 2 as the motion planning and manipulation framework. Within this architecture, Gazebo maintains the world model and physics-based actuation, ROS 2 manages inter-node communication for task planning and control, and MoveIt 2 handles collision-aware trajectory planning, kinematic feasibility, and scene updates.

To support this integration, the SDSP information package is translated into a robot-oriented task model. For each component or connection to be removed, the package should provide, at minimum, a unique identifier, sequence index, pose data, end-effector orientation, approach and extraction vectors, dependency list, clearance requirements, and obstacle data. It should also maintain status variables indicating whether an element is intact, detached, grasped, or removed. For connections, attributes such as type, disengagement mode, and any required motion constraints are included. Although some

of these fields may be partially inferred from the BIM model, the key contribution of the SDSP framework is that it assembles them into a task-centered representation that is aligned with robotic disassembly planning.

The proposed control architecture follows a hierarchical structure. At the top level, the SDSP module provides the ordered sequence of disassembly actions and inter-component dependencies. At the middle level, a task-planning node in ROS 2 interprets each action and decomposes it into motion primitives such as approach, align, engage, detach, extract, retreat, and place. At the lower level, MoveIt 2 generates feasible trajectories subject to workspace constraints, while the controller executes joint-space or Cartesian commands in Gazebo. After each successful disassembly action, the planning scene and simulation state are updated to reflect the removal of the corresponding component, allowing the next step in the sequence to be planned against the modified environment.

While a full perception layer is outside the scope of the current study, the framework is designed to ultimately incorporate vision and force sensing for closed-loop execution. Vision will confirm component identity and pose consistency relative to the digital model, while force-torque sensing will monitor contact conditions during extraction. In early simulation-based development, these functions will be approximated through perfect-state feedback; later stages can replace these assumptions with simulated depth cameras, joint feedback, and force-sensitive control. Therefore, the proposed system will function as an integrated architecture that bridges BIM-derived disassembly logic with robot-level planning and execution.

3.3 Proposed validation approach for robot assisted selective disassembly

For functional demonstration, a pilot study was developed for robot-assisted selective disassembly of a simple frame (see Figure 3). The BIM model was structured according to the specified needs for a robot operating system. The information package was created automatically for the selective disassembly of the component c_1 (structural column). The optimized disassembly path is: f_1-f_2-c_1. The information package was exported to the robot operating system for the disassembly task. To maintain scope, the current study focused on the BIM-to-simulation information flow, treating robotic execution as a conceptual demonstration rather than a fully implemented control pipeline.

Consequently, the next development stage involves validating the BIM-to-robot workflow in a simulated environment. In this phase, the disassembly environment will be represented in Gazebo, the robot manipulator will be integrated through ROS 2 and MoveIt 2, and the SDSP information package will be used as the task-level input for robotic execution. Under this workflow, the robotic arm executes the prescribed disassembly sequence step-by-step, updating the simulation and planning scene after each completed action. For the frame example, the robot will first perform the task associated with f_1, then update the environment to reflect its removal, then proceed to f_2, and finally execute the extraction of c_1 once the prerequisite steps have been completed. This will allow the proposed framework to test whether the SDSP-generated information can be correctly interpreted and executed within a robotic simulation.

The initial simulation-based validation targets three technical objectives. The first is data interoperability, namely whether the SDSP package can be translated into simulation and motion-planning inputs. The second is kinematic feasibility, namely whether the robot can generate valid, collision-free motions for each step in the prescribed sequence. The third is state consistency, namely whether the simulation environment can be updated correctly after each removal to reflect the evolving disassembly state. Assessment metrics will include sequence completion status, motion-planning success rate, collision occurrences, and any violations of the prescribed extraction or dependency constraints. Additional qualitative observations will be recorded regarding failure modes, such as unreachable approach poses, geometric interference, or inconsistencies between the SDSP representation and the simulation environment.

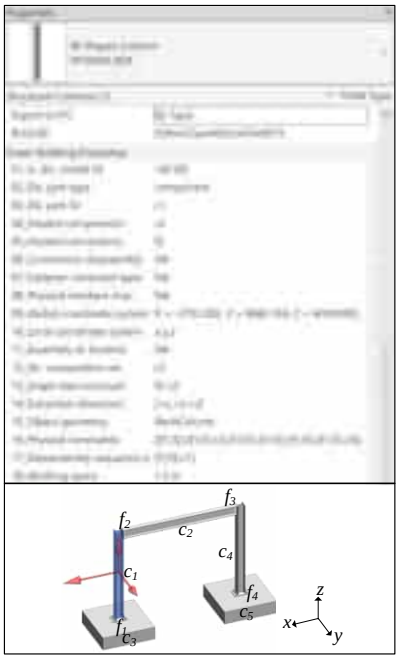


Figure 3. BIM-based Sequential Disassembly Planning for Buildings (SDPB)

The pilot study highlights the framework's ability to

integrate BIM-based SDSP data with robotic control systems, enabling the autonomous execution of selective disassembly. These findings underscore the importance of robust data structuring and compatibility with IFC standards to ensure seamless data exchange between BIM models and robotic systems. Future validation efforts will extend the proposed framework to more complex building structures with diverse materials, connection types, and architectural configurations. Additionally, subsequent physical implementations with real-world robotic solutions will address challenges such as unstructured environments, sensor integration, and scalability. These advancements will refine the SDSP framework and support the application in practical construction and deconstruction scenarios.

4 Conclusions and Future work

The study's findings indicate that the BIM-based SDSP system is highly effective in enriching a robot's ability to process information, sense, plan, and execute the retrieval of specific components from a building structure. In this study, we propose the information package that can be automatically produced by a BIM-based SDPB system for performing the disassembly works in an autonomous way by a robotic solution. The results demonstrate the technical feasibility of implementing SDSP with robotic solutions for simpler building structures. This approach offers significant benefits for decarbonization in the construction industry by enabling the efficient recovery and reuse of building materials, thus reducing the demand for virgin resources and minimizing waste. By integrating robotics with BIM-based selective disassembly, the method supports sustainable construction practices and aligns with Circular Economy principles, directly contributing to the reduction of greenhouse gas emissions associated with material extraction, processing, and disposal.

Looking forward, the next steps involve applying this approach to more complex building archetypes with diverse subsystems in real-world scenarios. This entails addressing challenges such as increased variability in component types, structural interdependencies, and environmental constraints. The integration of advanced robotic capabilities, including AI-based perception and adaptive control systems, will be crucial in overcoming these challenges and enhancing the system's scalability and robustness. Additionally, this methodology has potential applications in futuristic scenarios, such as the disassembly of building structures designed for extra-terrestrial construction (e.g., space infrastructure and settlements). Such applications would benefit from the autonomous and precise nature of the proposed approach, specifically through the expansion of the system's future operational capacity to function in extreme conditions.

Future research will also focus on improving the integration between BIM-based SDSP systems and robotic solutions, particularly in terms of real-time data exchange and dynamic task optimization. Exploring human-robot collaboration for selective disassembly tasks could further enhance safety and efficiency. Lastly, validating the proposed framework through experimental testing with physical robotic systems and expanding its application to larger-scale projects will be critical to demonstrating its practical viability and impact on advancing Circular Economy principles in the built environment.

References

[1] [1] H. Abu-Ghaida, M. Ritzen, A. Hollberg, S. Theissen, S. Attia, S. Lizin, Accounting for product recovery potential in building life cycle assessments: a disassembly network-based approach, The International Journal of Life Cycle Assessment (2024) 1-26.

[2] [2] A.S. Allam, M. Nik-Bakht, Knowledge Graph-based Deconstruction Planning of Building's Products, ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, Vol. 41, IAARC Publications, 2024, pp. 1018-1024.

[3] [3] P.R. Beurskens, The development of an evaluation method to support circular building design, (2021).

[4] [4] M. Daneshmand, F. Noroozi, C. Corneanu, F. Mafakheri, P. Fiorini, Industry 4.0 and prospects of circular economy: a survey of robotic assembly and disassembly, The International Journal of Advanced Manufacturing Technology 124 (9) (2023) 2973-3000.

[5] [5] F. Denis, C. Vandervaeren, N. De Temmerman, Using network analysis and BIM to quantify the impact of design for disassembly, Buildings 8 (8) (2018) 113.

[6] [6] M. Gharbia, A. Chang-Richards, Y. Lu, R.Y. Zhong, H. Li, Robotic technologies for on-site building construction: A systematic review, Journal of Building Engineering 32 (2020) 101584.

[7] [7] S. Kim, S.-A. Kim, A design support tool based on building information modeling for design for deconstruction: A graph-based deconstructability assessment approach, Journal of cleaner production 383 (2023) 135343.

[8] [8] A. Kunic, R. Naboni, A. Kramberger, C. Schlette, Design and assembly automation of the Robotic Reversible Timber Beam, Automation in Construction 123 (2021) 103531.

[9] [9] M. Mahmoudi Motahar, S.H. Hosseini Nourzad, A hybrid method for optimizing selective

disassembly sequence planning in adaptive reuse of buildings, Engineering, Construction and Architectural Management 29 (1) (2022) 307-332.
[10] [10] M. Mahmoudi Motahar, S.H. Hosseini Nourzad, F. Rahimi, Integrating complete disassembly planning with deconstructability assessment to facilitate designing deconstructable buildings, Architectural Engineering and Design Management 20 (1) (2024) 150-167.
[11] [11] C.C. Ohueri, M.A.N. Masrom, M. Noguchi, Human-robot collaboration for building deconstruction in the context of construction 5.0, Automation in Construction 167 (2024) 105723.
[12] [12] L.-M. Ottenhaus, Z. Yan, R. Brandner, P. Leardini, G. Fink, R. Jockwer, Design for adaptability, disassembly and reuse–A review of reversible timber connection systems, Construction and Building Materials 400 (2023) 132823.
[13] [13] H. Poschmann, H. Brueggemann, D. Goldmann, Disassembly 4.0: A review on using robotics in disassembly tasks as a way of automation, Chemie Ingenieur Technik 92 (4) (2020) 341-359.
[14] [14] B. Sanchez, M. Honic, F. Leite, P. Herthogs, R. Stouffs, Augmenting materials passports to support disassembly planning based on building information modelling standards, Journal of Building Engineering 90 (2024) 109083.
[15] [15] B. Sanchez, C. Rausch, C. Haas, R. Saari, A selective disassembly multi-objective optimization approach for adaptive reuse of building components, Resources, Conservation and Recycling 154 (2020) 104605.
[16] [16] C. Vandervaeren, W. Galle, A. Stephan, N. De Temmerman, More than the sum of its parts: Considering interdependencies in the life cycle material flow and environmental assessment of demountable buildings, Resources, Conservation and Recycling 177 (2022) 106001.
[17] [17] B. Xiao, C. Chen, X. Yin, Recent advancements of robotics in construction, Automation in Construction 144 (2022) 104591.
[18] [18] J. Zhang, H. Luo, J. Xu, Towards fully BIM-enabled building automation and robotics: A perspective of lifecycle information flow, Computers in Industry 135 (2022) 103570.

A Particle System Approach to Simulating Construction Sprays: Shotcrete, Painting, and Abrasive Blasting

Mohammad Reza Yazdi Samadi[1] **, Ralf Waspe**[1] **, Ali Muhammad**[1] **, Christian Schlette**[1]

[1]SDU Center for Large Structure Production, Maersk Mc-Kinney Moller Institute,
University of Southern Denmark, Odense, Denmark

mry@mmmi.sdu.dk,

Abstract -

Spraying processes, such as shotcrete, painting, and abrasive blasting, are critical to construction but remain labor-intensive, hazardous, and challenging to automate. Realistic and computationally efficient simulation models are crucial for supporting robotic process planning, training, and control. However, existing approaches are typically process-specific and lack generality across different material systems. This study presents a unified, physics-informed particle system-based simulation that can represent diverse spraying processes within a single computational model. The formulation integrates stochastic adhesion and rebound mechanisms with process-dependent material parameters, enabling the reproduction of characteristic behaviors—from cohesion and detachment in shotcrete to smooth deposition in painting and particle scattering in abrasive blasting. The framework operates in near real-time and has been quantitatively validated for the shotcrete process using published experimental datasets, while additional spraying processes such as painting and abrasive blasting are demonstrated qualitatively to illustrate the generality of the modeling approach. The results demonstrate that only a limited set of parameter adjustments is required to emulate distinct process dynamics, establishing a scalable foundation for robotic spraying and trajectory optimization.

Keywords -

Physics-Based Simulation; Construction Sprays; Particle System; Robotic Spraying

1 Introduction

A wide range of spraying processes is used in construction, including surface protection and coatings, sprayed concrete (shotcrete), insulation and sealant foams, spray painting, adhesive sprays, and abrasive blasting. All of these processes are highly dynamic: material is ejected from the nozzle and projected onto a target surface, requiring the operator to both control deposition and continuously adjust based on visual feedback. Despite their importance, most of these processes are inherently hazardous—exposing workers to dust, rebound, chemical agents, or physical strain. Combined with a severe labor shortage and a declining skilled workforce in the construction sector [1], the need for robotics and automation to support, augment, or partially replace human operators becomes urgent.

To achieve human-level expertise autonomously, a robotic system must first acquire a reliable understanding of the underlying spraying process. Developing such understanding requires robust simulation models. Shotcrete provides a compelling example: it is a complex and skill-intensive process that requires operators to undergo rigorous training and certification. However, physical shotcrete tests are costly, wasteful, and unsustainable, whereas simulation-based training (e.g., [2]) offers a safer and more scalable alternative. Beyond training, simulation models enable technical analysis, waste reduction, and process optimization, provided they are sufficiently reliable and accurate.

The main contribution of this study is the extension of our previously developed shotcrete simulation model [3] to a broader class of construction spraying processes. The underlying particle-based framework has been quantitatively validated for shotcrete deposition against experimental data [4, 5]. Building on this foundation, the present work investigates whether the same physics-based simulation architecture can be generalized to other spraying applications, including painting and abrasive blasting.

Unlike prior work that focused exclusively on concrete spraying, we show that these processes can be represented within a unified particle-based framework by adjusting a limited set of material and process parameters governing adhesion, cohesion, and rebound behavior. This enables the reproduction of distinct spraying regimes while preserving a consistent simulation architecture. The additional processes considered in this study are, therefore, intended to demonstrate the generality of the framework through parameter adaptation, rather than to provide independent quantitative validation.

The remainder of this article is organized as follows. Section 2 reviews the state of the art in simulation technologies and robotic advancements relevant to each process. A theoretical overview of particle methods is provided in

Section 3. Section 4 details our simulation model, including the parameter settings and ranges for each process. The corresponding results are presented in Section 5, followed by a discussion and concluding remarks in Section 6.

2 Background

Spraying processes have long been the subject of engineering analysis and robotics research. From concrete placement to protective coatings and abrasive blasting, each domain has developed its own simulation methods and automation strategies. However, most of the existing work has remained domain-specific, with few attempts at unifying frameworks that can handle multiple processes under a standard methodology. This section reviews the state of the art in simulation in three representative applications: shotcrete, spray painting, and abrasive blasting.

2.1 Sprayed concrete

Shotcrete simulation has attracted considerable attention due to the complex interactions between aggregates, mortar, and substrate surfaces. Early work applied the Discrete Element Method (DEM) to explicitly represent aggregate and mortar phases, modeling collision, rebound, and packing effects in sprayed layers [6]. DEM-based formulations have also been extended to capture rheological behavior and yield-stress effects of fresh concrete mixtures by combining particle kinematics with macroscopic rheology models [7, 8].

Coupled Computational Fluid Dynamics–Discrete Element Method (CFD–DEM) approaches later emerged to approximate both the granular and fluid aspects of fresh concrete [9]. Nevertheless, DEM and CFD–DEM remain computationally intensive and struggle to meet real-time constraints. Velez et al. [10] traded physical realism for speed (via ray-casting adhesion) and proposed a real-time training simulator. More recently, particle system-based models have been proposed to strike a balance between realism and efficiency [4], introducing a physics-informed particle system capable of reproducing spray dynamics and cohesive failure phenomena. A voxel-based model [11], capable of modeling material evolution in real time, further underscores the growing interest in bridging the divide between computational efficiency and fidelity. These developments highlight a spectrum of approaches, from high-fidelity but slow solvers to simplified, high-performance models.

2.2 Spray painting

Spray painting is among the most extensively studied spraying processes, primarily due to its widespread use in the automotive and manufacturing industries [12]. Simulation models have been developed to analyze droplet transport, deposition, and film formation, typically by resolving the airflow field using CFD [13]. However, accurately predicting paint thickness on complex geometries remains a challenge and requires the development of advanced coupled CFD models.

Coupling CFD with a Lagrangian particle-tracking approach enables the prediction of droplet trajectories and the final film thickness. Earlier, Böttner and Sommerfeld [14] developed a comprehensive CFD–DEM model for electrostatic powder painting, combining a finite-volume approach for the airflow field with discrete particle tracking and the solution of the Laplace equation to incorporate electrostatic effects. Later, Fogliati et al. [15] employed the time-averaged Navier–Stokes equations, comprising the k–ϵ turbulence model, to predict deposition outcomes across various geometries and operating conditions.

In robotics, automated spray-painting systems often integrate simulation or predictive modeling into offline path-planning and optimization pipelines. These systems aim to achieve uniform coating thickness, minimize material waste, and improve process efficiency through model-based trajectory optimization [16, 17].

2.3 Abrasive blasting

Abrasive blasting (sandblasting) is a jet-cleaning process used for surface preparation, roughening, or coating removal. It involves high-velocity particles impacting a target surface, typically modeled as a purely particulate flow with negligible adhesion. Classical simulations of abrasive blasting have predominantly relied on the DEM. To achieve greater physical realism, Longmore et al. [18] employed DEM and represented each granule as a tetrahedral lattice of four particles, implicitly capturing static friction through the interlocking of neighboring granules. Alternative approaches have utilized CFD to analyze the influence of parameters such as particle size, blasting pressure, and nozzle distance [19].

Across all three domains, notable progress has been achieved in either high-fidelity process simulation or task-specific robotic automation. However, the literature remains fragmented: most simulation tools are tailored to a single material system, and their formulations are not readily transferable across different spraying processes. This limitation motivates the present work, which seeks to unify shotcrete application, spray painting, and abrasive blasting within a single, particle-based simulation framework—distinguished only by their process parameters.

3 Particle methods

Particle methods are a broad family of mesh-free numerical techniques in which a continuum or system is modeled through discrete entities—"particles"—whose motion and

interactions collectively approximate the behavior of the underlying process. In contrast to grid-based (Eulerian) schemes, particle methods adopt a Lagrangian viewpoint that follows material motion, thereby reducing numerical diffusion during advection, simplifying the treatment of free surfaces and interfaces, and naturally handling complex boundaries and heterogeneous materials without the need for complex mesh generation. These properties make particle methods well-suited to highly dynamic processes, such as sprays, splashes, fragmentation, and other multiphase phenomena. These advantages, however, come at a cost: particle methods generally demand greater computational resources, careful implementation and tuning, and extensive validation to achieve predictive reliability.

Following the formalization proposed by [20], particle methods can be described in terms of five foundational concepts:

1. **Representation:** the basic unit or object that is being manipulated (particle, point, droplet).
2. **Interaction:** the ability of particles to affect one another (e.g., via collisions, constraints, or forces).
3. **Neighborhood:** the criterion that determines when and with which neighbors interactions occur.
4. **Evolution:** changes to a particle's state governed by external physics, independent of neighbors.
5. **Iteration:** the system's progression through discrete timesteps.

They further distinguished three levels of description for particle methods:

- A `general algorithmic framework`, including the data structures and functions common across particle-based approaches.
- A `particle method instance`, which specifies the problem or application context to which the method is applied.
- A `state transition function`, which governs how an instance evolves from its initial state to later states through repeated application of the algorithm.

This layered definition separates the abstract algorithmic design from application-specific implementations, providing a general foundation for particle-based simulation in a wide range of domains.

4 Simulation model

The proposed simulation model is implemented as a modular, physics-informed particle system in C++. Designed for generality, the same core framework can simulate a wide range of spraying processes—shotcrete, painting, sandblasting, and others—by modifying only material parameters and process-specific plugins. Figure 1 presents a Unified Modeling Language (UML) class diagram that organizes the framework into a particle emitter, particle data structure, surfaces, obstacles, and a physics engine for integration and collision handling.

At the core lies the `Swarm`, which manages all active and inactive particles. Each particle is represented by a lightweight structure containing its position, velocity, and lifetime, which are updated at discrete timesteps (default: $\Delta t = 0.01\,\mathrm{s}$). The swarm handles activation, deactivation, and efficient culling of particles outside the domain, enabling scalability to tens of thousands of particles and supporting real-time simulation.

Particle generation is handled by the `Emitter`, derived from an abstract interface. Emitters are parameterized by particle spawn rate, velocity range, shape, aperture, and angular spread, allowing them to replicate the behavior of different nozzle types. For example, paint spraying can be represented by high emission rates with a point emitter (near-zero aperture) for air spray guns, a rectangular emitter for a fan-shaped airless spray guns, or an annulus for rotary bell atomizers. Sandblasting, by contrast, is modeled with a circular emitter, coarser particles than spray paint, and higher initial velocities. Each emitter initializes particles with randomized attributes sampled from empirical distributions.

Within the modular framework, particles emitted by the `Emitter` are managed by the `Swarm` and updated under external forces. Each particle is represented as a geometric, massless point with optional user-defined attributes (e.g., shape, size, color) that affect only visualization. Motion evolves under external forces such as gravity, aerodynamic drag, or viscoelastic effects; in the simplest configuration, trajectories are purely ballistic. The `PhysicsEngine` advances motion explicitly using forward Euler integration at discrete timesteps, ensuring deterministic, computationally efficient updates suitable for real-time use. For collision handling, trajectories are approximated as straight-line segments per timestep, and intersections are tested against geometric obstacle primitives (planes, cylinders, boxes, spheres) using algebraic methods [21]. Impacts are then resolved through specular reflection with attenuation by a coefficient of restitution, rather than by DEM-style contact solvers. This abstraction deliberately omits particle–particle interactions to maintain scalability and real-time performance while retaining sufficient fidelity for spray-process simulation.

Whether a particle adheres or rebounds upon impact is a stochastic process. In shotcrete, a rebound probability $p_r(\theta)$ is computed from the particle's impact angle θ rel-

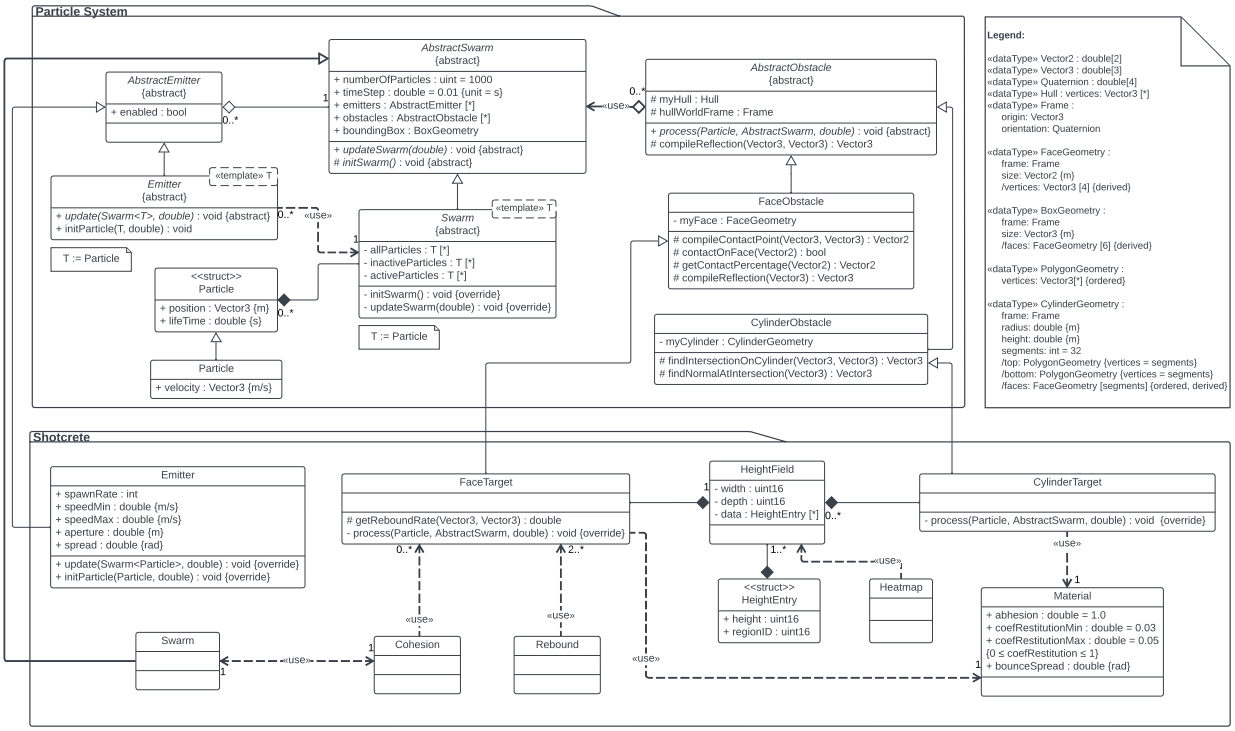


Figure 1. Unified Modeling Language (UML) class diagram of the particle system library, together with an example shotcrete extension. The model specifies abstract and concrete classes for particle emission, swarm management, and obstacle handling, conforming to the UML 2.5.1 specification.

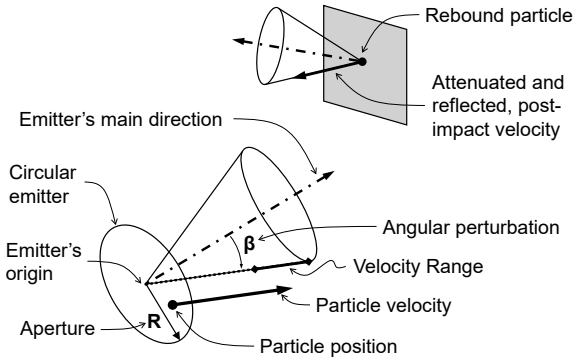


Figure 2. Circular particle emitter model. Particles are stochastically initialized within the emitter aperture and assigned velocity vectors perturbed from the primary emission axis. Perturbations are constrained by angular and magnitude bounds, thereby controlling the spread of the emission. Post-impact dynamics are represented by a reflected direction about the local surface normal, with the speed attenuated to account for energy dissipation.

ative to the surface normal using an empirical polynomial fit to Melbye's curves [22]. A uniform variate $u \sim U[0,1]$ is drawn and scaled by a material-specific *abhesion* (i.e., anti-adhesion) factor $\alpha \in [0,1]$; the outcome is obtained by comparing this draw with $p_r(\theta)$. For shotcrete, we use $\alpha = 1$, so the angle–dependent probability governs adhesion/rebound; for paint, we set $\alpha = 0$, suppressing rebound so that all impacts adhere. For abrasive blasting, we do *not* increase α; instead, we bypass the angle curve by setting $p_r(\theta) \equiv 1$, which forces every impact to rebound regardless of angle. Rebounding particles leave with the specularly reflected direction and an attenuated speed given by the coefficient of restitution, with a small random scatter added to match experimental spread.

Surfaces and obstacles are implemented through an abstraction hierarchy. The `AbstractObstacle` interface provides collision detection and reflection utilities, with concrete implementations such as `CylinderObstacle` and `FaceObstacle`. Targets like `FaceTarget` extend this functionality by recording deposition into a `HeightField`, a grid structure that tracks layer thickness and identifiers for cells lacking adequate support. This enables both visualization and quantitative analysis (e.g., thickness uniformity, rebound ratio), supported by a

`Heatmap` utility for graphical overlays.

Cohesion—a key sprayability parameter for achieving satisfactory in-situ properties of shotcrete mixes [23]—quantifies mesoscale structural stability and is modeled through post-processing of the heightfield. At user-defined intervals, cells that exceed local support thresholds are flagged and clustered into connected regions, identifying potential cohesive failure zones. This approach provides a computationally efficient surrogate for more detailed rheological modeling, enabling real-time analysis of failure risk while omitting costly particle–particle or fluid–structure interactions. In painting, analogous effects such as dripping or sagging can be represented but are generally negligible, whereas abrasive blasting does not exhibit cohesive failure at all. A summary of the default parameters for these three processes is provided in Table 1.

The recommended heightfield resolution depends on the target size; nevertheless, higher values such as 1024×1024 are generally advised. Since the simulation model decouples resolution from runtime performance, increasing resolution improves the accuracy of adhesion (i.e., material accumulation) and cohesion analysis without incurring significant computational penalties [4].

The simulation timestep Δt was selected in the range of 10^{-4}–10^{-2} s to balance accuracy and throughput. To avoid missed collisions, Δt must be small enough that the maximum particle displacement per step is less than the smallest obstacle. A default timestep of 10^{-3} s is recommended, with adjustments possible depending on performance and accuracy requirements.

The model achieved near real-time performance without employing common acceleration strategies, despite the availability of effective and well-established techniques such as spatial hashing [25] and data-parallel computation [26]; in typical configurations involving on the order of 10^4–10^5 active particles, real-time execution was achieved on standard computing hardware, consistent with previously reported benchmarks [3, 4]. Incorporating such methods in future work could further extend scalability to simulations involving millions of particles.

5 Results

As described in Section 4, emission directions are sampled from probability distributions constrained by the spray cone angle. Figure 3 compares two sampling strategies: a uniform distribution and a truncated normal distribution. The uniform distribution produces a stronger concentration of particles near the spray cone axis, whereas the normal distribution allows the spread to be adjusted via its mean (μ) and standard deviation (σ). In earlier work [4], we reproduced on-target accumulation profiles of wet-mix shotcrete reported by Ginouse and Jolin [27] by setting $\mu = 0.255$ and $\sigma = 0.102$. The simulation achieved errors of approximately **5%**, **3%**, and **2.8%** for maximum absolute, root-mean-square, and mean absolute error, respectively, relative to empirical data.

Figure 4 demonstrates four representative runs of the same robotic spraying task under different process models and obstacle configurations, highlighting the simulator's flexibility and the qualitative effects of process physics and geometry. Panel (a) illustrates shotcrete with reinforcements, showcasing pronounced shadowing and underdeposition behind reinforcement bars. Rebounded particles are visible in flight and are deposited on the ground heightfield, while cohesive buildup and detachment are possible. Panel (b)—shotcrete without reinforcements—exhibits a largely radially symmetric accumulation pattern with far fewer localized voids, demonstrating how obstacles strongly affect coverage even for identical nozzle kinematics. Panel (c) depicts spray painting, displaying near-complete immediate adhesion and a uniformly covered surface, consistent with high-adhesion, no rebound, and negligible sagging. Abrasive blasting is illustrated in Panel (d), which contrasts sharply: adhesion is suppressed, nearly all particles rebound with significant scattering, and no material accumulates on the target; the heatmap therefore represents hit coverage and suggests erosion intensity.

All simulations were executed using process parameters derived from Table 1; specifically, cohesion and the angle-dependent rebound model were enabled for the shotcrete cases, rebound was suppressed for the painting, and forced

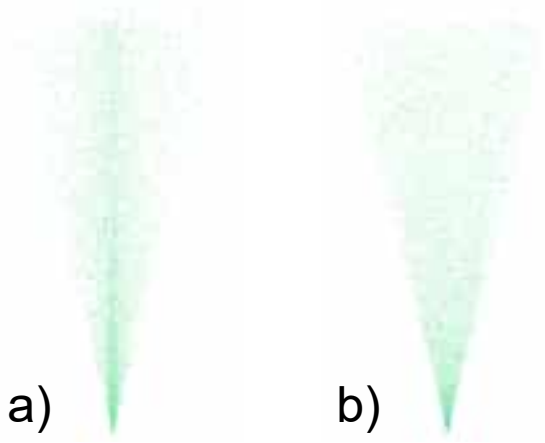


Figure 3. Comparison of particle emission patterns when directions are sampled from **a)** a uniform distribution and **b)** a truncated normal distribution ($\mu = 0.255$, $\sigma = 0.102$). The normal distribution offers tunable control over spread, enabling closer alignment with empirical spray patterns. All other parameters were held constant.

Table 1. Default simulation parameters for three spraying processes.

Parameter (units)	**Shotcrete**	**Spray painting**	**Abrasive blasting**
Emitter type	Circular	Point / rectangular / annulus	Circular
Particle spawn rate (particles/s)	10^4	10^4–10^5	10^3–10^5
Particle size*	3.0 (aggregate)	0.05 (droplet)	0.5 (grain)
Particle velocity (m/s)†	20–30 (wet-mix)	20–60	100–200
Emitter aperture (mm)	32–76	0.5–2.0 (point emitter)	3.2–19.1
Spray cone angle (°)	9–18	15–120+	15
Abhesion α	1.0 (angle-based)	0.0 (rebound suppressed)	N/A (bypassed, forced rebound)
Rebound model	Angle-dependent polynomial	N/A	forced rebound ($p_r = 1$)
Coefficient of restitution	0.03–0.05	N/A	0.6–0.9
Cohesion enabled?	Yes	No (optional drip)	No
Scattering spread	Small angular noise	Negligible	Moderate–large
Notes	High cohesion, angle rebound	Near-instant adhesion	Elastic impacts, erosion

* Particle size refers to the rasterized diameter (in pixels) of points drawn with `GL_POINTS`, stored as a floating-point value.
† Particle velocity is highly affected by input airflow, air pressure, and aggregate size [24].

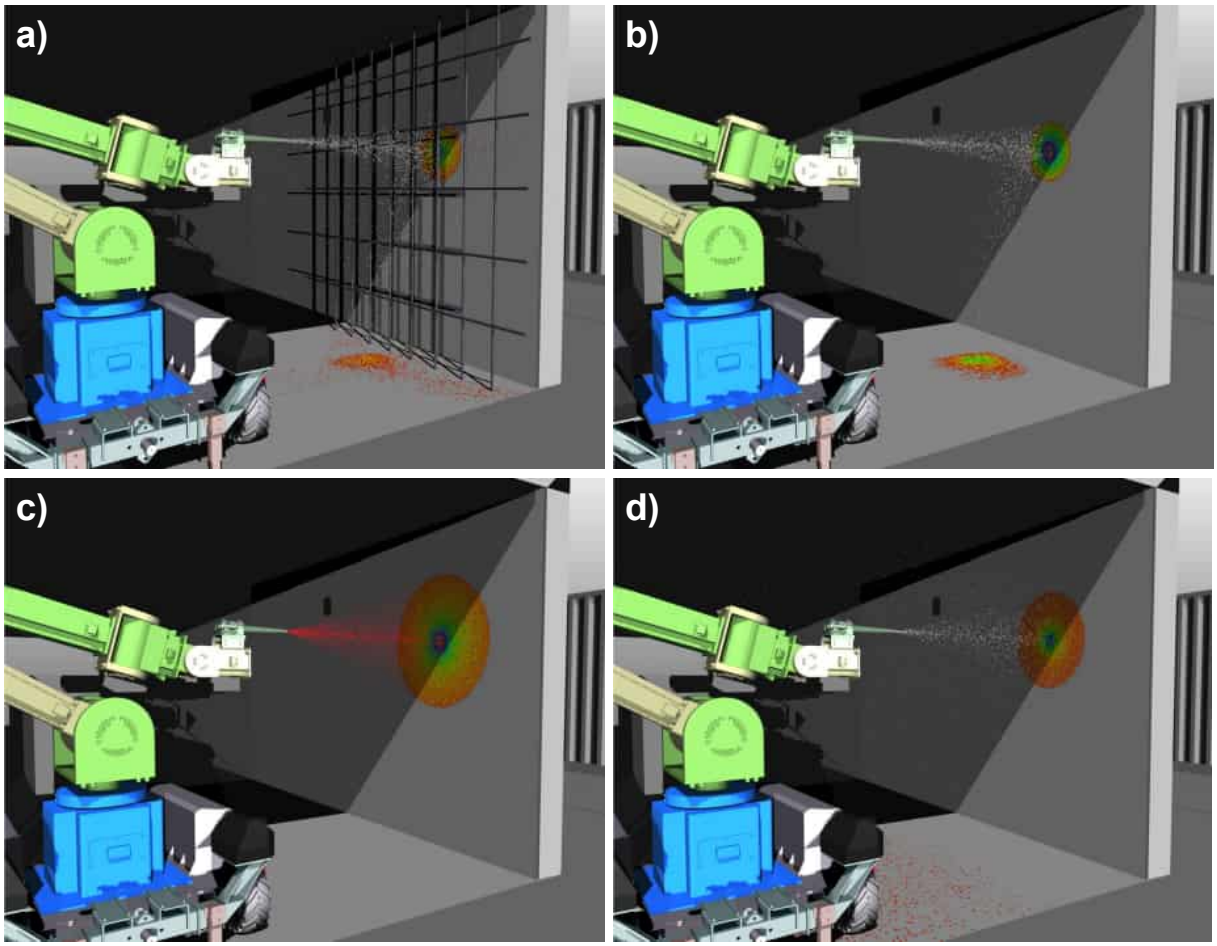


Figure 4. Comparison of robotic spraying simulations across four processes: **a)** shotcrete with reinforcement elements, **b)** shotcrete deposition without reinforcement, **c)** spray painting, and **d)** abrasive blasting, where adhesion is disabled, and all particles rebound with high scattering. Parameters for each scenario were selected according to Table 1.

rebound (high restitution and scatter) was used for blasting. These visual comparisons demonstrate that changing only plugin parameters (while keeping the same emitter/nozzle kinematics) produces distinct, physically interpretable outcomes.

6 Discussion and conclusion

The results confirm that the unified framework can simulate characteristic behaviors in different spraying processes. For instance, in the shotcrete configuration, the model successfully captures cohesive failure and material detachment phenomena, which are absent in the painting mode. Conversely, in painting mode, high adhesion and low rebound parameters produce a smooth surface finish; however, sagging may still occur. In abrasive blasting, the system exhibits purely particulate scattering without accumulation. These qualitative distinctions are consistent with the expected physical behavior of each process, supporting the generality of the proposed model. In this context, variations in physical sprayer characteristics such as flow rate, nozzle geometry, and supply pressure are implicitly captured through emitter parameters (particle spawn rate, velocity distribution, and angular spread), whose influence on deposition behavior has been previously analyzed for the shotcrete process [4], while further process-specific validation under controlled operating conditions remains a direction for future work.

A key advantage lies in the modular software architecture. The same core simulation engine operates across all modes, with behavior determined solely by parameter sets that define material and process properties. This demonstrates the feasibility of our central objective: to enable a single, physics-informed particle system to emulate diverse spraying processes by varying only their governing parameters. Such an approach can be directly embedded into robotic programming environments or Digital Twins, enabling engineers to switch between simulation-based design and control processes seamlessly.

The automation potential of this unified approach is substantial. A robotic cell could employ simulation-in-the-loop optimization to refine spray patterns until predicted deposition or coverage meets desired criteria. Similarly, robotic abrasive blasting systems could test and evaluate different nozzle trajectories in simulation to maximize cleaning efficiency while minimizing material loss. Because the simulation executes in near real time, it also provides a practical foundation for machine learning–based process control or operator training in virtual environments.

The current formulation also excludes several classes of physical phenomena, including particle–particle interactions in dense flow regimes, complex fluid behaviors such as splashing and droplet coalescence, and post-deposition volumetric evolution such as lateral flow or expansion (e.g., in foam-based materials). As a result, the model is most appropriate for particle-dominated spraying processes where deposition is primarily governed by particle–surface interactions, including adhesion, rebound, and accumulation, and where a heightfield representation provides a sufficient approximation of material buildup. This limitation also prevents simulating processes characterized by significant post-deposition evolution, such as expanding foam. Transitioning to a three-dimensional cellular automaton (3D CA) would substantially enhance geometric expressiveness and enable simulation of time-dependent material behavior, including lateral motion and post-impact spreading. Overall, while the framework is robust and extensible, advancing its fidelity will require volumetric modeling, physically grounded parameter calibration, and broader experimental validation. These developments will further strengthen its applicability as a tool for robotic construction, process optimization, and virtual skill assessment across diverse spraying applications.

References

[1] C. Klein and J. Smith. Addressing labour and skills shortages in a fast-changing economy. *Documents de travail du Département des Affaires économiques de l'OCDE*, 2024. doi:10.1787/757311cb-en.

[2] Edvirt — training for the mining and tunneling industry. online: `https://www.edvirt.com/`, Accessed: 25/9/2025.

[3] M. R. Yazdi Samadi, R. Waspe, and C. Schlette. Toward shotcrete process simulation to support robotic operation. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 707–713, Lille, France, 2024. doi:10.22260/ISARC2024/0092.

[4] M. R. Yazdi Samadi, R. Waspe, and C. Schlette. Physics-based particle system modeling of shotcrete process for robotic placement. *Construction Robotics*, 9(30), 2025. doi:10.1007/s41693-025-00172-y.

[5] M. R. Yazdi Samadi, R. Wu, S. Gholami, R. Waspe, A. Muhammad, A. Billard, and C. Schlette. From human to height-field: Predictive shotcrete simulation with a physics-informed particle system. In *Proceedings of the 22nd International Conference on Advanced Robotics (ICAR)*, pages 826–832, San Juan, Argentina, 2025. doi:10.1109/ICAR65334.2025.11338646.

[6] U. C. Puri and T. Uomoto. Characterization of distinct element modeling parameters for fresh concrete

and its application in shotcrete simulations. *Journal of materials in civil engineering*, 14:137–144, 2002. doi:10.1061/ ASCE!0899-1561 2002!14:2 137!

[7] K. Anjneya and A. Deb. A review of the discrete element method for modelling the rheology of fresh concrete. In *Recent Advances in Computational and Experimental Mechanics, Vol—I*, pages 81–95, Singapore, 2022. doi:10.1007/978-981-16-6738-1_7.

[8] V. Mechtcherine, A. Gram, K. Krenzer, J.-H. Schwabe, S. Shyshko, and N. Roussel. Simulation of fresh concrete flow using Discrete Element Method (DEM): theory and applications. *Materials and structures*, 47:615–630, 2014. doi:10.1617/s11527-013-0084-7.

[9] G. Ma, H. Ma, and Z. Sun. Simulation of two-phase flow of shotcrete in a bent pipe based on a CFD–DEM coupling model. *Applied Sciences*, 7:3530, 2022. doi:10.3390/app12073530.

[10] G. Velez, L. Matey, A. Amundarain, Á. Suescun, J. Marín, and C. de Dios. Modeling of shotcrete application for use in a real-time training simulator. *Computer-Aided Civil and Infrastructure Engineering*, 28(6):465–480, 2013. doi:10.1111/j.1467-8667.2012.00788.x.

[11] S. Yamakawa, K. Vazquez-Santiago, Y. Xia, H. Ogura, S. Yamamoto, R. Nakanishi, and K. Shimada. Concrete spray 3D printing simulator for nozzle trajectory planning. *IEEE Robotics and Automation Letters*, 2025. doi:10.1109/lra.2025.3615038.

[12] M. R. Pendar, F. Rodrigues, J. C. Páscoa, and R. Lima. Review of coating and curing processes: Evaluation in automotive industry. *Physics of Fluids*, 34(10):101301, 2022. doi:10.1063/5.0109376.

[13] Y. Chen, W. Chen, B. Li, G. Zhang, and W. Zhang. Paint thickness simulation for painting robot trajectory planning: a review. *Industrial robot*, 44(5): 629–638, 2017. doi:10.1108/IR-07-2016-0205.

[14] C. U. Böttner and M. Sommerfeld. Numerical calculation of electrostatic powder painting using the Euler/Lagrange approach. *Powder technology*, 125(2):206–216, 2002. doi:10.1016/S0032-5910(01)00508-3.

[15] M. Fogliati, D. Fontana, M. Garbero, M. Vanni, and G. Baldi. CFD simulation of paint deposition in an air spray process. *Journal of Coatings Technology Research*, 3(2):117–125, 2006. doi:10.1007/s11998-006-0014-5.

[16] W. Chen, D. Zhao, and B. Zhang. Path planning for spray painting robot of workpiece surfaces. *Mathematical problems in engineering*, 2013:1–6, 2013. doi:10.1155/2013/659457.

[17] P. J. From and J. T. Gravdahl. A real-time algorithm for determining the optimal paint gun orientation in spray paint applications. *IEEE transactions on automation science and engineering*, 7(4):803–816, 2010. doi:10.1109/TASE.2009.2033567.

[18] J. P. Longmore, P. Marais, and M. M. Kuttel. Towards realistic and interactive sand simulation: A GPU-based framework. *Powder technology*, 235:983–1000, 2013. doi:10.1016/j.powtec.2012.10.056.

[19] X. Yu, X. Wei, X. Huang, and L. Zhang. Study on the modeling method and influencing parameters of sandblasting process for blade grinding. *Advances in Materials Science and Engineering*, 2022:1–19, 2022. doi:10.1155/2022/7905927.

[20] J. Pahlke and I. F. Sbalzarini. A unifying mathematical definition of particle methods. *IEEE Open Journal of the Computer Society*, 4:97–108, 2023. doi:10.1109/OJCS.2023.3254466.

[21] D. H. Eberly. *3D Game Engine Design: A Practical Approach to Real-Time Computer Graphics*. Elsevier, Amsterdam, 2007. doi:10.1201/b18212-12.

[22] T. Melbye. Sprayed concrete for rock support. MBT International, 2001.

[23] N. Trussell and S. Jacobsen. Review of sprayability of wet sprayed concrete. *Nordic Concrete Research*, 63(2):21–41, 2020. doi:10.2478/ncr-2020-0016.

[24] N. Ginouse and M. Jolin. Experimental and numerical investigation of particle kinematics in shotcrete. *Journal of Materials in Civil Engineering*, 26: 06014023, 2014. doi:10.1061/(ASCE)MT.1943-5533.0001092.

[25] S. Lefebvre and H. Hoppe. Perfect spatial hashing. *ACM transactions on graphics*, 25(3):579–588, 2006. doi:10.1145/1141911.1141926.

[26] K. Sims. Particle animation and rendering using data parallel computation. *SIGGRAPH Computer Graphics*, 24:405–413, 1990. doi:10.1145/97880.97923.

[27] N. Ginouse and M. Jolin. Mechanisms of placement in sprayed concrete. *Tunnelling and Underground Space Technology*, 58:177–185, 2016. doi:10.1016/j.tust.2016.05.005.

Structural Semantic Segmentation and Fine-Grained Surface Feature Extraction in Heritage Buildings

Jingxuan Li [1, 2], Jian Bi [3], Vincent J.L. Gan [1, 2], Michael Yit Lin Chew [1]

[1] Department of the Built Environment, National University of Singapore, Singapore
[2] Centre of Digital Building Technology, National University of Singapore, Singapore
[3] College of Mechanical and Vehicle Engineering, Hunan University, China

Abstract

Heritage buildings exhibit complex and irregular legacy environments that pose significant challenges for conservation. This study introduces a framework grounded on high-precision terrestrial laser scan, which serves as input for both semantic segmentation of structural components and fine-grained surface feature extraction. Leveraging the TLS scans, the proposed approach supports 3D scene understanding of heritage buildings, including structural segmentation of walls, floors and ceilings, as well as extracting surface-level details (e.g., cracks) that are layered onto the main structural components. The framework (1) integrates 3D point cloud-based geometric representation of heritage buildings and (2) displays the primary building 3D geometry with updated surface attributes (crack spatial pattern) to uncover new insights. This method improves spatial interpretation in complex heritage environments, offering a platform for analysing potential heritage conservation issues.

Keywords –
Heritage Building; Scene Understanding; Semantic Segmentation; Fine-Grained Surface Feature

1 Introduction

Heritage buildings serve as vital touchstones of cultural identity; yet their conservation faces challenges such as weathering, material aging, and urban development pressures that accelerate deterioration. The shophouse is a predominant heritage building typology in Southeast Asia and Southern China [1, 2]. Singapore is one of the most significant centres for shophouse development [2]. Figure 1 shows a typical shophouse (i.e., NUS Baba House), featuring a ground floor designated for retail or commercial use and upper floors reserved for residence. Compared to modern commercial complexes, shophouse architecture is a less spatially efficient typology for commercial operation. Furthermore, it is characterised by fragile building materials that present unique vulnerabilities to age-related degradation. Such surface conditions may occur independently on specific floors or manifest in combination within particular areas, making the intensive monitoring of shophouse conditions increasingly important to preserve and sustain their cultural and historical significance.



Figure 1. Sample point clouds of shophouse obtained from terrestrial laser scanning

In Singapore, there are over 6,000 conserved shophouses required on-site monitoring and surveying which is a tremendous work [3]. Nowadays, reality capture technologies such as LiDAR or photogrammetry have been adopted to provide a digital archive of the shophouse. But these methods require intensive human labour and time to accurately scan and document the building information and geometry in digital format [4], followed by evaluating the surface conditions and recording relevant conservation issues.

To solve this problem, many studies have attempted to integrate scanning to support as-built data capture as well as automated structural segmentation and fine-grained surface feature extraction (material change,

surface pattern etc.). Applying 3D point clouds for building surface feature extraction and assessment (such as cracks) has been attempted by many studies [5-7]. The studies could achieve defect identification accuracy from 42.5% to 99.7%. To achieve this level of accuracy, close-range scanning of the defect area and high-resolution scanned data are simultaneously required [5, 6]. However, digital documentation at such level-of-detail is impracticable for thousands of heritage buildings. For example, a typical 3-storey shophouse has 8 to 12 rooms, which require 16 to 36 scans to generate comprehensive point clouds. Manual labelling of point clouds in this level-of-detail across different structural vulnerabilities like cracks, spalling, moisture infiltration could be tedious and time consuming. More importantly, methods that directly use 3D point cloud data to undertake semantic segmentation or other related tasks are restricted by inherent quality limitations, including the inability to capture local structural details, sensitivity to non-uniform sampling densities, and a lack of translation invariance, which limits their generalisability to complex, irregular geometries often found in heritage sites [8, 9].

As such, more studies nowadays intend to leverage image-based sensing to address the semantic segmentation in heritage buildings. Image segmentation offers several advantages. First, computational efficiency is improved as they require less processing power and memory compared to direct processing of point clouds, making them more efficient and light-weighted for practical applications. Secondly, image-based methods are more flexible as the collection and labelling of images is inherently more straightforward and cost-effective than processing 3D point clouds [10-13]. Additionally, the abundance of open-source dataset, pretrained models and well-established networks provides a robust foundation for transfer learning specific to heritage buildings.

Other studies explored image semantic segmentation with foundational models introducing robust capabilities for segmenting objects in images and videos [14, 15]. Further advancements have focused on open-vocabulary segmentation [11, 12, 16]. Since heritage buildings present more complicated architectural and ornamental elements than modern architectures, adopting open vocabulary models allows the network to understand the various unseen or arbitrary elements present in heritage building environments.

While image-based methods significantly improve efficiency for rapid data acquisition and are easier to deploy at scale, they still face challenges in fully capturing the heterogeneous materials and diverse surface conditions present in historical architecture. Compared with LiDAR-based scanning, image-based methods often exhibit lower geometric fidelity and weaker spatial representation, making it difficult to integrate fine-grained surface details into a precise spatial context, particularly for depth-related features. This limitation reduces their ability to quantify defect dimensions or reveal underlying spatial patterns and their potential impacts for heritage buildings.

This study introduces a framework that utilises high-precision terrestrial laser scanning (TLS) as the primary input for both structural semantic segmentation and fine-grained surface feature extraction. By leveraging TLS, the approach enables comprehensive building 3D scene understanding, encompassing semantic segmentation of structural components like walls, floors, and ceilings alongside fine-grained surface feature extraction that layers onto the structural components. The framework achieves this by (1) integrating point cloud-based 3D geometric representation of heritage buildings and (2) displaying primary building geometry with fine-grained surface attributes. The 3D integration enhances interpretability and enables spatial-level structural analysis by providing a quantified crack dimensions and localization within the building's coordinate system.

2 Methodology

The methodology of this work is shown in Figure 2. For each individual scan, the TLS generates a panoramic colormap for colour point clouds. The colormap contains abundant information about the scene and can be used for 3D scene understanding.

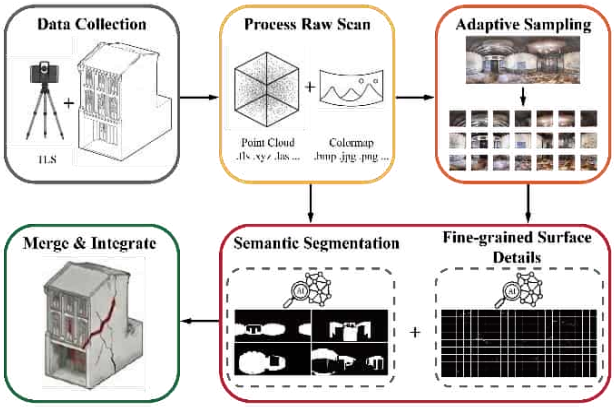


Figure 2. Overview of proposed method

This study uses two pipelines dedicated to processing the colormap to segment structural walls, floors, and ceilings alongside fine-grained surface detail extraction that is layered onto the main structural segmentation results. The colormap overlay embeds surface-level labels, allowing each point to carry both the structural component type and its surface condition through updated surface attributes. The resulting masks from both pipelines are integrated to provide a unified point cloud-based 3D geometric representation of the heritage

building, updated with surface attributes on the segmented structural components such as cracks on walls etc. to reveal their spatial patterns. This method can be extended to other surface features such as openings, material transitions and surface patterns, thereby enhancing 3D scene understanding tasks.

2.1 Dataset Collection and Processing

To enable automatic crack identification and semantic segmentation of structural components on panoramic colormap, experiments are conducted on a shophouse to collect its point cloud using TLS. The dataset consists of the following: point clouds, panorama colormaps that are extracted from point clouds, and ground truth colormaps. They are processed to align the requirement of machine learning pipelines.

2.1.1 On-site Scanning

The point cloud dataset is acquired by scanning a heritage shophouse, the NUS Baba House. This building was built around 1895 and undergone several renovations [17]. To capture a comprehensive dataset, the shophouse is scanned from 24 positions (shown in Figure 3) to cover all interior surfaces and façade elevation.

A FARO® TLS was used to acquire the point cloud data. Scanner parameters were set at indoor 10 meters for interior scenes and outdoor 20 meters for exterior and semi-exterior scenes. The acquired raw point clouds are processed and the colormap of each scene is exported. Figure 4 shows the registered point cloud and on-site scanning operation photo.

2.1.2 TLS Panoramic Colormap

The machine learning pipeline for fine-grained surface feature extraction exhibits varying constraints regarding input resolution; while some architectures support native resolution processing via internal slicing, others require fixed dimensions or down-sampling. However, relying on down-sampling is problematic because pixel density is a critical determinant of prediction accuracy; reducing colormap quality inevitably leads to information loss and compromises feature extraction results. Furthermore, preserving the exact original resolution is essential for the subsequent accurate integration of 2D predictions back into the 3D point cloud. Consequently, to bypass the resolution constraints of specific networks and standardize testing across diverse pipelines without sacrificing fidelity, a systematic slicing and tiling approach was implemented. This process partitions the original panoramic colormap into a grid of unified patches. A significant challenge in standard tiling is the "boundary effect," where semantic information is corrupted because object instances (e.g., cracks) traversing tile borders are severed, leading to incomplete feature extraction.

To mitigate this issue, an overlap sampling mechanism is applied. This principle introduces a defined overlapping area between adjacent tiles, ensuring that objects located at boundaries are captured fully within the receptive field of at least one tile, thereby increasing model sensitivity. The tiling logic is mathematically defined as follows: for a colormap of dimensions $w \times h$, with tile size T, overlap distance d, and step size $s = T - d$, each tile is indexed by its column x and row y. Given these parameters, the total number of tiles required to cover this colormap is computed as $Nx = \lceil (w - d) / (T - d) \rceil$ along the horizontal direction and $N_y = \lceil (h - d) / (T - d) \rceil s$ along the vertical direction.

2.1.3 Test Scenes

Indoor scans 07 (1F atrium), 13 (2F front hall), 16 (2F corridor), 19 (3F front hall), 21 (3F staircase), 22 (3F corridor), 23 (3F atrium), 24 (3F rear hall) shown in Figure 4 are selected to test the defect feature extraction methods' performance. These scenes are selected considering the complexity of background context. The simplest scene 19 features one significant crack on an unobstructed wall. Scene 23 contains multiple cracks on multiple walls, while the background context is increased but the defect area is relatively clean. Scene 07 includes many defect types and is captured in higher resolution, bringing challenge to the computing. Scene 24 features insignificant crack and spalling, the background context is extremely complex, introducing many interferences to the scene.

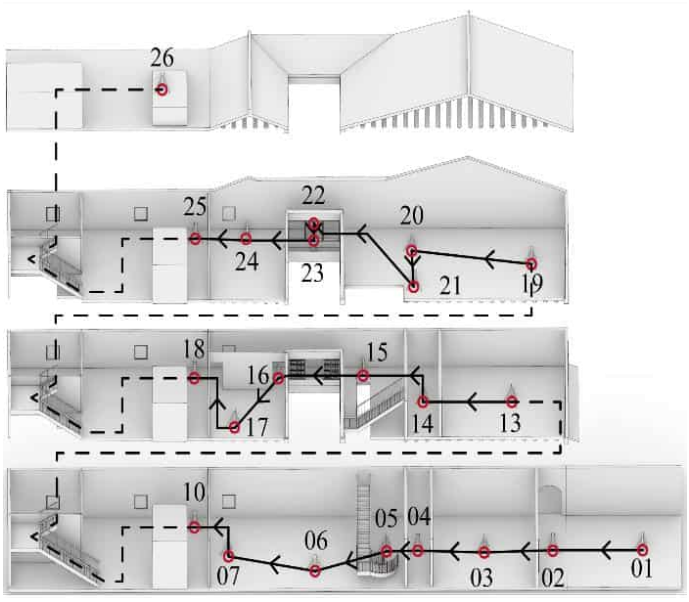


Figure 3. On-site scanning in NUS Baba House

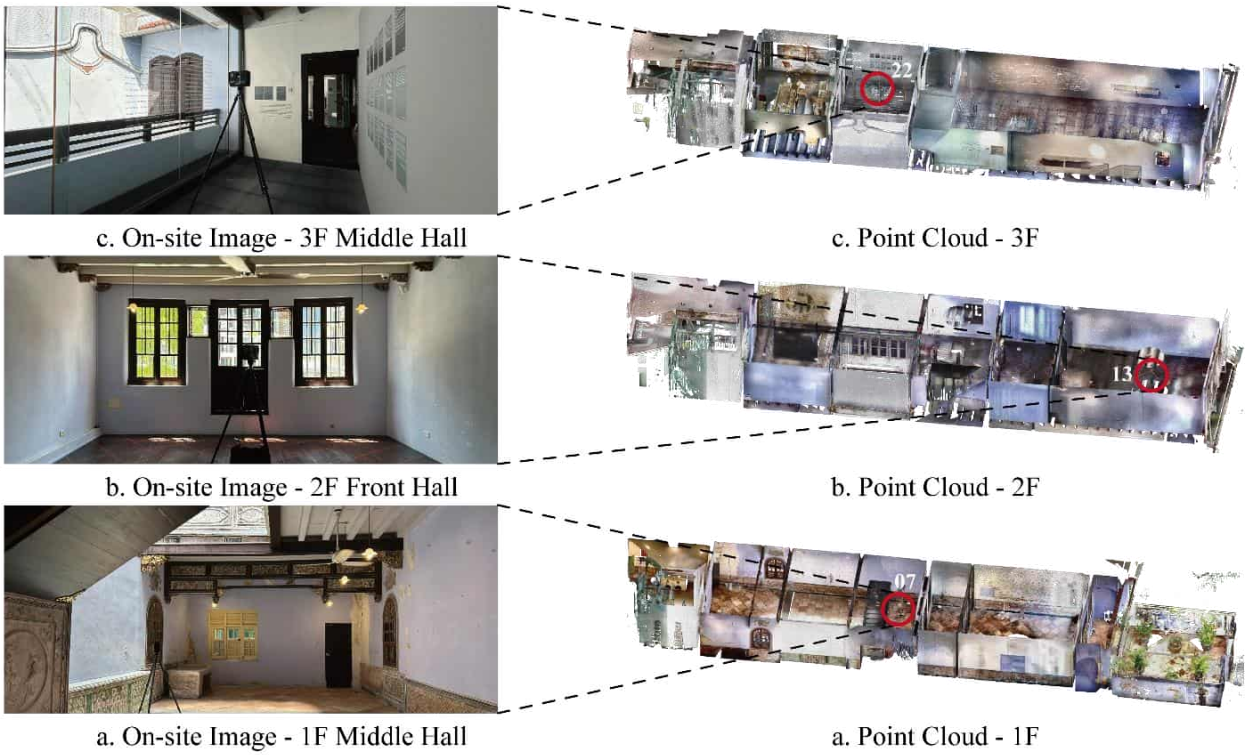


Figure 4. On-site scanning photo and corresponding registered point cloud

2.2 Fine-Grained Surface Feature Extraction

Identification of fine-grained surface features has achieved remarkable performance in structural health monitoring domains [18-22]. Advanced architectures like Vision Transformers (ViTs) significantly outperform traditional CNNs in capturing fine-grained crack details. While widely applied for precise segmentation, these methods remain restricted by application limitations, as existing models often exhibit limited generalisability, struggling to maintain accuracy when facing domain shifts caused by diverse lighting conditions, complex surface textures, or unseen structural materials.

In this study, crack identification is undertaken using multiple pipelines, each applied to tiled colormap. The pipelines generate complementary crack masks that highlight varying crack widths, orientations, and intensities across different surface conditions. The results are analysed using Precision (Pr), Recall (Re), F1-score (F1) and IoU to compare and evaluate all tested pipelines under all scenes and test settings, the metrics are defined in Equations (1)-(4).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{4}$$

2.3 Segmentation of Structural Components

When applying crack identification methods to the panoramic colormaps together with 3D point clouds, numerous false positive predictions are generated due to the complexity and scale of the scanned scenes. Successfully identifying the true positive predictions from the false alarms is critical for increasing the accuracy of crack identification. Incorporating the segmentation of structural components-such as walls and columns-facilitates the removal of false predictions detected in the non-structural area. In this study, the 3D scene is semantically interpreted using a semantic segmentation masking network, whereby structural components are extracted as combined masks to exclude false predictions that are detected on irrelevant elements.

Structural component segmentation is achieved using semantic masking generated by state-of-the-art segmentation networks [16, 23]. These masks restrict the crack identification process to structural regions, effectively suppressing noise and reducing false positives. This approach enables more accurate crack identification across complex indoor environments characterized by

occlusions and elements geometrically similar to cracks.

2.4 Crack Identification in Point Clouds

Mapping crack extraction results in 3D point cloud representation provides an intuitive visualisation of the surface conditions, revealing the crack's spatial distribution patterns and their potential structural implications. To achieve this, detected defects are labelled on the panoramic colormap, which replaces the original colormap. Since the colormap is inherently pixel-to-point aligned with the XYZ coordinates, this replacement automatically colourizes the corresponding 3D points to highlight defects. These colour labelled scenes are subsequently registered into a holistic 3D model of the building. Following registration, redundant low-resolution overlapping regions are removed to preserve only the high-resolution in the final point cloud.

3 Results and Discussion

3.1 Crack Identification

Tables 1 and 2 show the Recall and IoU of crack identification [19-22]. Figure 5 shows the results which reveal insight into the proposed tile and merge method. Initial testing of the crack identification revealed a dichotomy in performance. While high accuracy and specificity were observed in 75% of the raw colormap, the models struggled significantly in scenes where cracks pixels are minute relative to total pixels. Despite the successful identification of certain cracks, the overall quantitative metrics for Precision, Recall, and IoU are severely limited by scene complexity. Across all unmasked test scenes, the baseline maximum precision peaked at only 25.44%, with a maximum IoU of 20.75%. These limited results are primarily attributed to an excessive amount of false positive pixels generated from background noise and non-structural elements. Therefore, a robust mechanism is required to filter the noise and retain only valid predictions on relevant building surfaces for subsequent 3D integration. While overlap sampling is applied to preserve the cracks on tile boundaries, its contribution to the correct identification of true defects was limited.

3.2 Structural Wall Segmentation

To address the limitations identified in the baseline crack extraction, a two-stage optimisation strategy involving semantic masking and overlap sampling is implemented. Semantic segmentation networks [16, 23] are utilised to generate semantic masks, as illustrated in Figure 6, based on specific structural prompts, including "wall," "brick wall," "plaster wall," "column," and

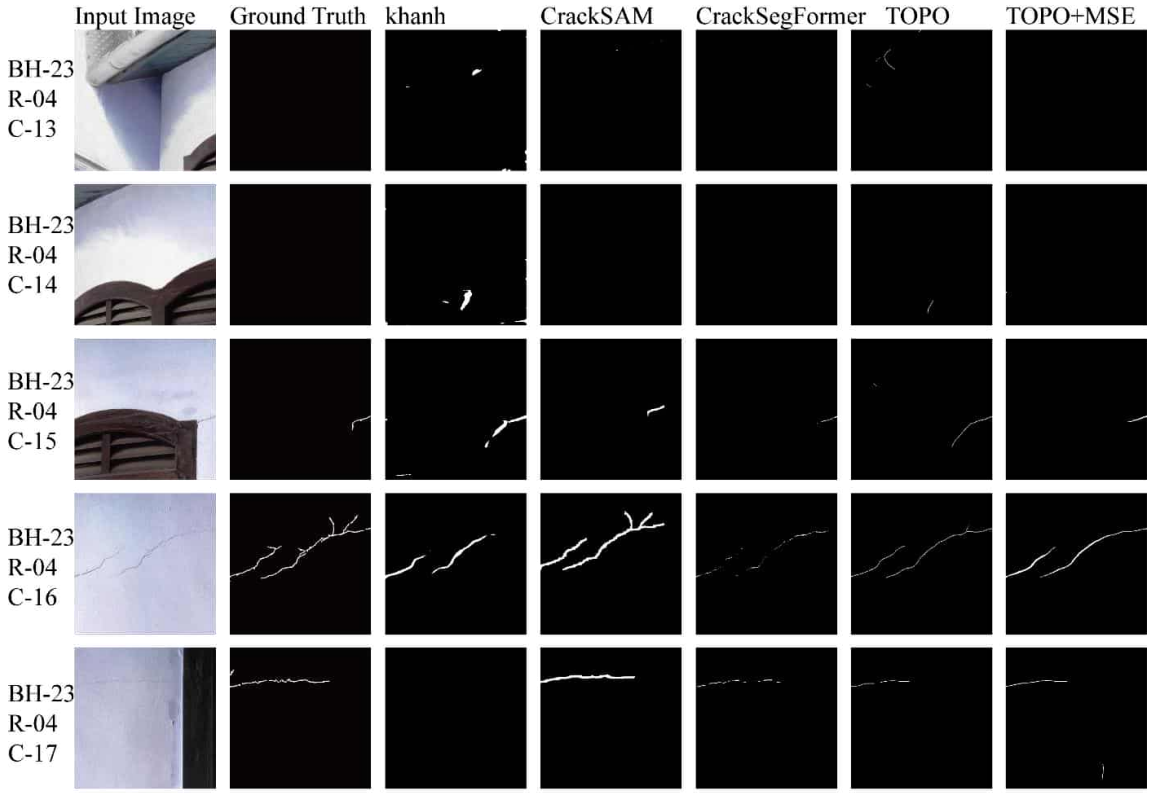


Figure 5 Examples for crack extraction using overlap sampling

"building surface." Among the evaluated networks, masks generated by GroundedSAM2 [16] were selected for its superior stability and prediction accuracy.

Experiment results show that semantic masking and overlap sampling combined approach enhances robustness in prediction. The introduction of semantic masking layer could significantly improve the overall mean precision across all methods from 4.55% to 21.18% and the mean IoU from 3.48% to 10.88%.

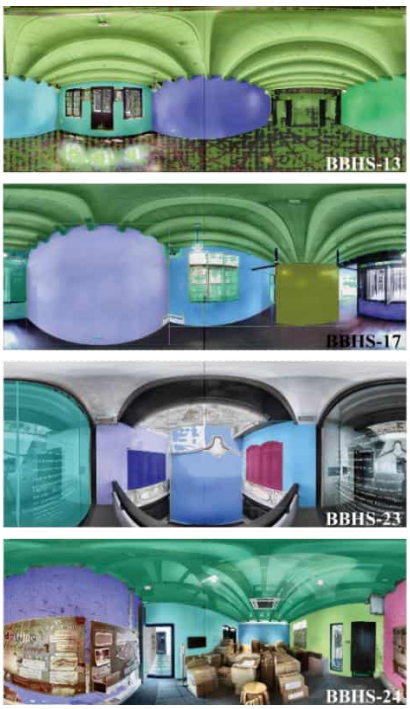


Figure 6. Semantic masks of structural components

The efficiency of this optimisation approach is most evident in complex scenes where baseline prediction generates more false positives. For instance, in test scene bbhs-22 using the CrackSegFormer method, the application of semantic masking combined with overlap sampling raised precision from 8.25% to 63.68%, marking a substantial recovery from the noise-dominated baseline. Similarly, CrackSAM in scene bbhs-21 increased IoU by 37.20% from a negligible 0.33% to 37.53%. These results confirm that introducing semantic masks effectively eliminates false positives on non-structural elements, fundamentally improved the precision and efficacy of the crack extraction pipeline. However, the improvement in Precision and IoU involves a trade-off in Recall. Because the semantic masks are isolating the predicted structural elements, some true defect pixels located at the boundaries of these surfaces could be excluded. While this leads to reductions in Recall depending on the architectural complexity of the scene, the trade-off is considered justifiable for structural health monitoring. Rather than pursuing pixel level accurate detection, the introduction of semantic masks significantly reduces background noise to improve the global detection reliability.

3.3 Surface Attributes in Point Clouds

Figure 7 displays representative results from the pipeline, illustrating fine-grained crack extraction within the segmented structural walls. Beyond visual inspection, the integration of these detections into a 3D point cloud environment enables several critical downstream tasks.

First, it allows quantification of defect length and area, providing a basis for severity evaluation. Second, visualizing surface defects in 3D reveals their spatial

Table 1 Defect extraction metrics - Recall

Detection Method	Test Scene							
	bbhs-07	bbhs-13	bbhs-16	bbhs-19	bbhs-21	bbhs-22	bbhs-23	bbhs-24
No Semantic Segmentation								
CrackSAM	84.18%	45.25%	26.66%	88.51%	79.82%	85.05%	75.42%	33.75%
CrackSegFormer	19.40%	4.23%	0.00%	3.89%	0.00%	26.41%	7.38%	8.37%
khan	48.11%	29.44%	2.42%	86.03%	21.69%	60.16%	15.46%	28.72%
TOPO	22.15%	10.86%	1.18%	24.96%	31.93%	23.23%	16.00%	13.23%
TOPO+MES	29.42%	10.50%	4.89%	50.25%	36.14%	34.00%	20.01%	17.20%
Semantic Segmentation								
CrackSAM	82.49%	44.83%	25.76%	88.51%	74.40%	84.33%	58.92%	25.58%
CrackSegFormer	19.16%	4.23%	0.00%	3.89%	0.00%	26.29%	6.64%	5.15%
khan	47.93%	29.44%	2.25%	86.03%	21.69%	60.03%	14.27%	18.21%
TOPO	22.14%	10.70%	1.18%	24.96%	31.93%	23.15%	14.65%	8.28%
TOPO+MES	29.42%	10.50%	3.32%	50.25%	36.14%	33.96%	19.01%	11.07%
Semantic Segmentation + Overlap Sampling								
CrackSAM	81.24%	47.65%	11.19%	99.83%	79.52%	89.45%	57.85%	27.48%
CrackSegFormer	21.24%	5.82%	0.51%	4.23%	0.00%	28.17%	10.16%	3.90%
khan	44.69%	31.39%	4.22%	86.31%	18.98%	59.15%	14.21%	18.04%
TOPO	23.37%	11.31%	1.01%	36.79%	31.93%	25.39%	15.76%	8.38%
TOPO+MES	30.49%	10.50%	8.10%	54.37%	36.14%	36.07%	20.24%	11.16%

Table 2 Defect extraction metrics - IoU

Detection Method	Test Scene							
	bbhs-07	bbhs-13	bbhs-16	bbhs-19	bbhs-21	bbhs-22	bbhs-23	bbhs-24
No Semantic Segmentation								
CrackSAM	0.80%	1.95%	0.36%	1.62%	0.33%	20.75%	13.44%	3.58%
CrackSegFormer	0.30%	0.26%	0.00%	0.29%	0.00%	6.95%	3.67%	2.59%
khan	0.21%	0.24%	0.01%	0.75%	0.02%	4.47%	0.48%	1.93%
TOPO	1.71%	0.34%	0.04%	2.42%	0.70%	10.71%	7.57%	4.89%
TOPO+MES	2.13%	0.65%	0.14%	4.70%	0.74%	17.03%	10.55%	6.90%
Semantic Segmentation								
CrackSAM	4.35%	22.04%	1.07%	9.18%	37.54%	24.75%	23.79%	13.04%
CrackSegFormer	5.53%	3.60%	0.00%	3.26%	0.00%	22.08%	6.27%	4.06%
khan	4.07%	10.89%	0.04%	16.76%	1.94%	9.24%	12.12%	4.69%
TOPO	12.35%	9.04%	0.18%	17.19%	9.49%	14.70%	12.89%	6.72%
TOPO+MES	11.05%	7.68%	0.38%	29.18%	12.86%	22.41%	16.28%	8.11%
Semantic Segmentation + Overlap Sampling								
CrackSAM	4.66%	24.62%	0.41%	14.01%	31.02%	25.92%	24.70%	12.89%
CrackSegFormer	4.41%	4.57%	0.05%	2.94%	0.00%	24.27%	9.53%	3.34%
khan	3.54%	10.72%	0.07%	14.97%	1.54%	8.02%	11.58%	4.41%
TOPO	12.30%	9.44%	0.14%	23.90%	8.85%	15.53%	13.75%	6.75%
TOPO+MES	10.52%	7.79%	0.77%	28.33%	11.93%	23.06%	16.82%	7.60%

distribution across the building, extending analysis from isolated scenarios to a comprehensive assessment. Ultimately, the 3D point cloud serves as foundation for BIM and digital-twin platforms, supporting efficient management and long-term building condition updating.



Figure 7. Point clouds of scan19 with cracks

4 Conclusions

This study presents a structural semantic segmentation and fine-grained surface feature extraction for heritage buildings. The proposed method addresses the inherent challenges of conducting structural health assessments on point cloud data captured in complex heritage environments. By incorporating semantic understanding of the scanned scenes, the method isolates structural components, refining crack identification on 3D point cloud data. The research was validated through experiments on a shophouse, benchmarking automated predictions against human visual inspections. Experimental results highlight the feasibility of TLS-derived point clouds for structural health assessment, confirming improvements in defect identification metrics in real-world heritage environment.

The proposed method has certain limitations for improvement. As the prediction was deployed using default parameters and generic datasets, fine-tuning these models on a dedicated heritage defect dataset could significantly bolster identification performance and robustness against domain-specific noise. Furthermore, current state-of-the-art networks predominantly focus on single-defect extraction (i.e., cracks), often yielding suboptimal results for multi-defect scenarios. Integrating diverse defect datasets with open-vocabulary semantic understanding models offers a promising pathway toward a holistic, multi-objective structural health assessment framework. Finally, while the high-resolution data provided by TLS is essential for detecting minute cracks, it introduces a trade-off in computational efficiency, the inherent operational constraints of TLS also limit its scalability in rapid field deployments. To bridge the gap between precision and operational practicality, future research will adapt this framework for data derived from mobile SLAM. Successfully adapting this pipeline to mobile data would not only mitigate the operational limitations of data acquisition, but also ready the framework for efficient structural health monitoring.

Acknowledgements

The authors would like to thank NUS Museum, Baba House, and the team for granting access and supporting the 3D scanning work. The authors also thank Centre of Digital Building Technology for supporting the experiment. This research is supported by Singapore MOE AcRF Tier-1 Grant (Ref: A-8004446-00-00). Any opinions, findings and conclusions or recommendations expressed in the material are those of the author(s) and do not reflect the views of the grantors.

References

[1] W. Han and J. Beisi. Urban morphology of commercial port cities and shophouses in Southeast Asia. Procedia Engineering, 142:190–197, 2016.

[2] M. Tohiguchi and H. S. Chong. Shophouse: Asian urban composite housing. In Planning for a Better Urban Living Environment in Asia, pages 212–230, Routledge, London, 2000.

[3] Urban Redevelopment Authority. Master Plan 2019 SDCP Conserved Building layer (2024). Data.gov.sg. On-line: https://data.gov.sg/datasets/d_3d73d61e5f5fe3ce52ac5d86b395d0b2/view, Accessed: 04/12/2025.

[4] M. Buldo, L. Agustín-Hernández, and C. Verdoscia. Semantic enrichment of architectural heritage point clouds using artificial intelligence: The Palacio de Sástago in Zaragoza, Spain. Heritage, 7(12):6938–6965, 2024.

[5] M. E. Mohammadi, R. L. Wood, and C. E. Wittich. Non-temporal point cloud analysis for surface damage in civil structures. ISPRS International Journal of Geo-Information, 8(12):527, 2019.

[6] Y. Jing, J.-X. Zhong, B. Sheil, and S. Acikgoz. Anomaly detection of cracks in synthetic masonry arch bridge point clouds using fast point feature histograms and PatchCore. Automation in Construction, 168:105766, 2024.

[7] Y. Jing, W. Lin, B. Sheil, and S. Acikgoz. 3D multimodal feature for infrastructure anomaly detection. Automation in Construction, 178:106388, 2025.

[8] C. R. Qi, Y. Li, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413, 2017.

[9] R. Pierdicca, M. Paolanti, F. Matrone, M. Martini, C. Morbidoni, E. S. Malinverni, E. Frontoni, and A. M. Lingua. Point cloud semantic segmentation using a deep learning framework for cultural heritage. Remote Sensing, 12(6):1005, 2020.

[10] R. Xu, X. Wang, T. Wang, Y. Chen, J. Pang, and D. Lin. PointLLM: Empowering large language models to understand point clouds. arXiv preprint arXiv:2308.16911, 2024.

[11] P. Nguyen, T. D. Ngo, E. Kalogerakis, C. Gan, A. Tran, C. Pham, and K. Nguyen. Open3DIS: Open-vocabulary 3D instance segmentation with 2D mask guidance. Open3DIS: Open-vocabulary 3D instance segmentation with 2D mask guidance. arXiv preprint arXiv:2312.10671, 2024.

[12] S. Peng, K. Genova, C. Jiang, A. Tagliasacchi, M. Pollefeys, and T. Funkhouser. OpenScene: 3D scene understanding with open vocabularies. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 815–824, Vancouver, Canada, 2023.

[13] Y. Yang, X. Wu, T. He, H. Zhao, and X. Liu. SAM3D: Segment anything in 3D scenes. arXiv preprint arXiv:2306.03908, 2023.

[14] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020, 2021.

[15] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. K. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. B. Girshick, P. Dollár, and C. Feichtenhofer. SAM 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714, 2024.

[16] T. Ren, S. Liu, A. Zeng, J. Lin, K. Li, H. Cao, J. Chen, X. Huang, Y. Chen, F. Yan, Z. Zeng, H. Zhang, F. Li, J. Yang, H. Li, Q. Jiang, and L. Zhang. Grounded SAM: Assembling open-world models for diverse visual tasks. arXiv preprint arXiv:2401.14159, 2024.

[17] NUS Museum. About NUS Baba House. On-line: https://babahouse.nus.edu.sg/explore/about-baba-house/, Accessed: 04/12/2025.

[18] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing, 338:139–153, 2019.

[19] C. Benz, P. Debus, H. K. Ha, and V. Rodehorst. Crack segmentation on UAS-based imagery using transfer learning. In Proceedings of the 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), pages 1–6, Dunedin, New Zealand, 2019.

[20] K. Ge, C. Wang, Y. T. Guo, Y. S. Tang, Z. Z. Hu, and H. B. Chen. Fine-tuning vision foundation model for crack segmentation in civil infrastructures. Construction and Building Materials, 431:136573, 2024.

[21] B. G. Pantoja-Rosero, D. Oner, M. Kozinski, R. Achanta, P. Fua, F. Perez-Cruz, and K. Beyer. TOPO-Loss for continuity-preserving crack detection using deep learning. Construction and Building Materials, 344:128264, 2022.

[22] H. Li, H. Zhang, H. Zhu, K. Gao, H. Liang, and J. Yang. Automatic crack detection on concrete and asphalt surfaces using semantic segmentation network with hierarchical Transformer. Engineering Structures, 307:117903, 2024.

[23] L. Medeiros. Language Segment-Anything. GitHub Repository. On-line: https://github.com/luca-medeiros/lang-segment-anything, Accessed: 04/12/2025

A scalable reinforcement learning framework for task allocation in flexible construction scenarios

Yifan Wang[1], Bin Yang[1]

[1] College of Civil Engineering, Tongji University, China
2410444@tongji.edu.cn, yangbin@tongji.edu.cn

Abstract

The construction process requires close collaboration among multiple crews, which relies fundamentally on efficient task allocation. However, the performance and efficiency of existing scheduling approaches remain limited and are insufficient to support fine-grained decision-making. This paper proposes a scalable Reinforcement Learning (RL) framework for Multi-Crew Task Allocation (MCTA) in flexible construction scenarios. First, a disjunctive graph is developed to model construction tasks and their internal constraints, upon which the MCTA problem is formulated as a Markov Decision Process. Then, a four-module RL network architecture is designed to address the MCTA problem, and the overall policy is trained using Proximal Policy Optimization (PPO). Case study shows that the proposed framework can effectively allocate tasks, achieving a 8.63% reduction in project duration and a 15.88% decrease in crew idle time compared with existing methods.

Keywords –

Construction scheduling; Task allocation; Reinforcement learning; Disjunctive graph; Proximal policy optimization

1 Introduction

The construction process involves multiple work crews that can independently execute specific tasks. The completion of these tasks collectively leads to the final delivery of the entire project. The organization of the construction workflow plays a crucial role in ensuring project efficiency[1,2]. Construction Schedule Optimization (CSO) aims to enhance this organization by systematically planning and coordinating the timing and sequence of construction activities. In this process, the overall project is decomposed into several manageable and analyzable task units, commonly referred to as the Work Breakdown Structure (WBS). Based on this, the sequence of task execution is determined under the constraints of process precedence and resource availability, and each task is assigned to corresponding work crews for execution.

The problem of determining the optimal start times together with executor assignment under resource and precedence constraints is defined as the Flexible Resource-Constrained Project Scheduling Problem (FRCPSP)[3,4], which is known to be NP-hard[5]. In existing studies, such problems are primarily addressed using mathematical methods[6,7], heuristic methods[8,9] and metaheuristic approaches[10,11]. With the increasing complexity and expanding scale of modern construction projects, the construction process has evolved into a dynamically changing complex system characterized by intricate spatiotemporal constraints[9,12] and expansion of project scale. However, these methods rely on static rules or complex computational mechanisms, resulting in insufficient performance and generalization in flexible construction scenarios with strong dynamics of tasks and resources[13,14].

Reinforcement Learning (RL) is an effective approach for solving sequential decision-making problems in dynamic environments. Several studies have applied RL to determine construction project schedules[15–17]. However, these studies exhibit two key limitations: (1) they suffer from limited scalability, which leads to the curse of dimensionality in large-scale problems and an inability to handle variable problem sizes caused by dynamic changes or unexpected events; (2) they often fail to explicitly model the internal spatiotemporal constraints among construction tasks, forcing the neural network to implicitly learn project-specific decision rules in a black-box manner. These limitations hinder the practical applicability of RL in complex and flexible construction scenarios and restrict its ability to generate high-quality, fine-grained schedules.

To address these limitations, this study proposes an RL framework for modeling and solving fine-grained construction scheduling problem Multi-Crew Task Allocation (MCTA) problem from the perspective of Multi-Agent Simulation (MAS)[18]. The main contributions of this study are summarized as follows: (1)

A disjunctive graph is developed to model MCTA by explicitly representing spatiotemporal constraints among tasks. (2) A scalable RL framework is proposed to learn scheduling policies over the disjunctive graph, integrating Relational Graph Convolutional Network (RGCN) [19] and self-attention-based aggregation.

2 Related Work

2.1 Mathematical methods for construction scheduling

Early studies addressed the CSO problem using mathematical methods, among which the most common and influential is the Critical Path Method (CPM)[20]. CPM is a project scheduling technique that predicts the total project duration and identifies critical activities by locating the longest path of dependent tasks within a project. However, CPM has proven insufficient to handle the multi-objective, multi-constraint nature of modern scheduling problems, particularly in balancing resource continuity and cost trade-offs.

Consequently, Integer Programming (IP) and Linear Programming (LP) have been introduced to solve CSO problems. For example, Mattila et al.[21] proposed an integer linear programming-based linear scheduling model for resource leveling. Gomar et al.[22] developed a linear programming model to optimize the allocation of multi-skilled labor resources, minimizing labor costs while satisfying project demands. Ipsilands et al.[23] formulated a multi-objective linear programming model for scheduling linear and repetitive projects. To address the dynamic nature of the construction process, Robinson[24] applied a Dynamic Programming (DP) approach to determine project schedules that minimize overall duration, enabling the optimal allocation among activities with arbitrary cost – time functions.

2.2 Heuristic and metaheuristic methods for construction scheduling

Large-scale CSO problems are typically NP-hard, making it difficult for exact algorithms to obtain optimal solutions within a reasonable computation time. Heuristic methods, on the other hand, can efficiently produce near-optimal and satisfactory solutions by designing problem-specific decision rules. For example, Elazouni[25] proposed a heuristic algorithm that enumerates and ranks parallel project schedules under credit constraints, aiming to minimize project duration while maintaining real-time cash flow matching. Gajpal and Elazouni[26] introduced an enhanced heuristic algorithm based on a polynomial-shifting strategy, which replaces exhaustive search and reduces computation time from exponential to polynomial order without compromising solution accuracy. Similarly, Al-Shihabi et al.[27] developed a heuristic method that balances project duration and financial progress through a dual-priority rule enabling rapid generation of feasible scheduling solutions.

Metaheuristic methods improve candidate solutions through iterative computation based on predefined criteria, without requiring strong assumptions about the underlying CSO problem. Chan et al.[28] proposed a Genetic Algorithm (GA)-based construction resource scheduling method, which learns the structural features of project networks through selection and recombination mechanisms. Elazouni et al.[29] employed GA, Simulated Annealing (SA), and the Shuffled Frog-Leaping Algorithm (SFLA) to solve project scheduling problems under financial constraints, achieving the shortest total project duration with SA in large-scale networks. Ye et al.[11] further developed a two-stage task allocation framework for multiple heterogeneous construction robots based on an improved genetic algorithm, effectively enhancing collaboration efficiency in multi-robot systems.

2.3 RL for construction scheduling

The reliance of heuristic and metaheuristic approaches on predefined rules and static optimization limits their adaptability to the dynamic construction processes. RL is an effective approach for solving dynamic decision-making problems[30], and has been applied to address CSO challenges. Soman et al.[16] proposed a hybrid method that integrates RL with Linked-Data to automatically generate look-ahead schedules for construction projects. Kedir et al.[17] combined RL with agent-based modeling to obtain practical solutions that support decision-making in construction activity sequencing. Yuan et al.[15] considered both process precedence and resource constraints in bridge construction and employed RL to generate construction schedules. In addition, Agrawal et al.[31] developed a novel approach integrating Proximal Policy Optimization (PPO) with Monte Carlo sampling, which enables automatic generation of optimized assembly sequences and schedules for prefabricated buildings. These studies did not decompose construction projects into fine-grained task units and thus overlooked the precedence and spatial overlap constraints among tasks. Liu et al.[32] propose a Construction Markov Decision Process framework that leverages multi-agent RL to evaluate goals and plan paths, reducing project duration and labor idle time. However, they did not provide a systematic formulation of the scheduling problem, and therefore the methods still exhibit limitations in both performance and practical applicability.

3 Methodology

3.1 Problem formulation

This study focuses on a flexible construction process involving multiple crews, each possessing distinct skill sets, and models each crew as an agent. The entire project is decomposed into multiple tasks using WBS[33], and all agents collaboratively complete the construction through interleaved task execution, following the workflow illustrated in Figure 1. In this research, the MCTA problem incorporates three types of constraints, which are designed to explicitly capture the interactions among the three key elements in MCTA: tasks, labor resources, and workspace. Specifically, capability constraints reflect the matching relationship between tasks and crew skills, precedence constraints model the logical dependencies among tasks, and spatial overlap constraints characterize the conflicts arising from limited workspace.

(1) Capability constraint: Each crew can only execute tasks corresponding to its specific trade (e.g., a rebar-tying crew cannot perform formwork installation tasks).

(2) Precedence constraint: A task cannot be started until all its adjacent predecessor tasks are completed (e.g., formwork installation must precede concrete pouring).

(3) Spatial overlap constraint: If the workspace of a target task overlaps with that of a neighboring ongoing task, the target task cannot be initiated.

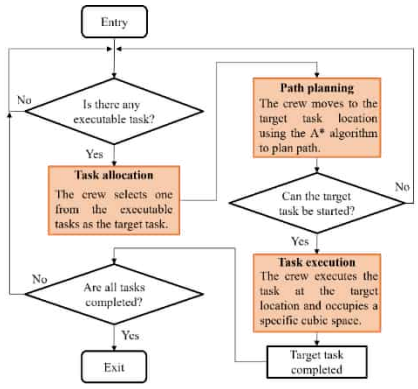


Figure 1 Workflow of crew agents

3.2 RL mathematical Model for MCTA

To address the MCTA problem, we adopt a RL approach, which requires formulating the scheduling process as a Markov Decision Process (MDP). The MDP can be formally described by a tuple $\langle \mathbf{S}, \mathbf{A}, \mathrm{P}, \mathrm{R}, \gamma \rangle$. $\mathbf{S}$ and $\mathbf{A}$ denote the state and action spaces, respectively, while P is the state transition probability function. R represents the reward function used to evaluate task allocation performance, and γ is the discount factor.

3.2.1 State

The state of the MCTA is defined by tasks attributes, agent attributes, and relationships among tasks. Attributes of tasks and agents are categorized into static and dynamic attributes, as illustrated in Table 1. $\mathbf{S}_{\mathrm{loc}}^{\mathrm{task}}$ and $\mathbf{S}_{\mathrm{loc}}^{\mathrm{agent}}$ are represented by two-dimensional coordinates on the plane. $\mathbf{S}_{\mathrm{type}}^{\mathrm{task}}$ and $\mathbf{S}_{\mathrm{type}}^{\mathrm{task}}$ are encoded as one-hot vectors, with the element corresponding to its type set to 1 and all others set to 0. $\mathbf{S}_{\mathrm{dur}} = [d]$ is determined based on the task quantity and the corresponding crew productivity rate. Four task pools—$task_{\mathrm{wait}}$, $task_{\mathrm{queue}}$, $task_{\mathrm{on}}$, and $task_{\mathrm{end}}$—are used to represent the execution status of tasks[9]. $\mathbf{S}_{\mathrm{po}}$ is represented as a one-hot encoded vector, where the element corresponding to the current pool of a task is set to 1 and all others are set to 0.

Table 1 Summary of state space

Temporality	Attribute	Symbol	Representation
Static	Task location	$\mathbf{S}_{\mathrm{loc}}^{\mathrm{task}}$	2D vector
	Task type	$\mathbf{S}_{\mathrm{type}}^{\mathrm{task}}$	One-hot vector
	Execution time	$\mathbf{S}_{\mathrm{dur}}$	1D scalar
	Agent type	$\mathbf{S}_{\mathrm{type}}^{\mathrm{agent}}$	One-hot vector
Dynamic	Task pool	$\mathbf{S}_{\mathrm{po}}$	One-hot vector
	Agent location	$\mathbf{S}_{\mathrm{loc}}^{\mathrm{agent}}$	2D vector

The relationships between tasks can be represented with a disjunctive graph[34] $G = (T, P, S)$, as shown in Figure 2. Here, T denotes the set of all tasks (nodes in black). P is a set of directed edges representing the precedence constraints between consecutive tasks, directed from each predecessor task to its successor (illustrated in orange). S is a set of undirected edges, in which each edge connects a pair of tasks that have a spatial overlap relationship (illustrated in blue).

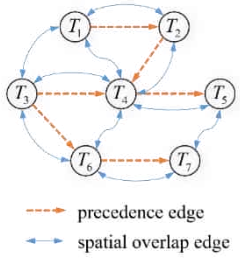


Figure 2 Disjunctive graph for MCTA

3.2.2 Action

Each crew agent outputs one discrete action a_t at a decision step t, where each optional action corresponds to a task. The dimensionality of the action space equals the total number of tasks. Due to the three types of constraints mentioned in Section 3.1, a large portion of the action space consists of invalid actions, which may lead to inefficient exploration during the agent's learning process. To address the problem, this study introduces action mask technique to handle a variable action space. Specifically, the logits of invalid actions are set to negative infinity at the output layer of the policy network, ensuring that their probabilities become zero after the softmax operation. Consequently, action sampling is performed only among valid actions, effectively improving exploration efficiency and training stability.

3.2.3 State transition

The state transition P of MCTA corresponds to the update of the overall construction progress. When an agent reaches an available task location, the task is moved from $task_{\text{queue}}$ to $task_{\text{on}}$. Upon task completion, the task transitions from $task_{\text{on}}$ to $task_{\text{end}}$, while its adjacent successor tasks are simultaneously moved from $task_{\text{wait}}$ to $task_{\text{queue}}$, thereby becoming eligible for execution.

3.2.4 Reward

The objective of MCTA is to minimize the total project duration, with all crew agents sharing a common reward defined as $\mathrm{R} = \mathrm{R}_{\text{finish}} + \mathrm{R}_{\text{time}}$. Here, $\mathrm{R}_{\text{finish}}$ is a terminal reward granted to all agents upon the completion of all construction tasks, while R_{time} is a stepwise penalty applied at each simulation timestep to encourage faster overall project completion.

3.3 Scalable RL network architecture for MCTA

This study proposes a scalable network architecture comprising four modules: a feature extraction module, a global representation module, a task scoring module, and a decoder model as illustrated in Figure 3.

3.3.1 Feature extraction module

The disjunctive graph represents precedence and spatial overlap constraints through different types of edges, thereby transforming spatiotemporal dependencies into connection features. To process such a heterogeneous graph with multiple edge types, a RGCN (parameterized by θ_{R}) is employed to extract task-level relational features.

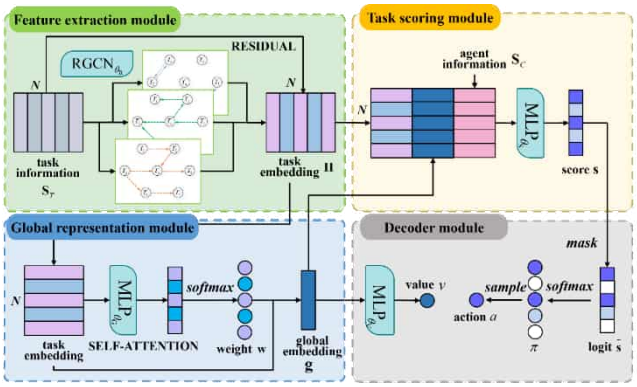


Figure 3 RL network architecture for MCTA For a given task T_i, its raw features can be defined as:

$$\mathbf{S}_{T_i} = (\mathbf{S}_{\text{loc}}^{\text{task}} \,\|\, \mathbf{S}_{\text{type}}^{\text{task}} \,\|\, \mathbf{S}_{\text{dur}} \,\|\, \mathbf{S}_{\text{po}}) \tag{1}$$

where $\|$ is the vector concatenation operator. To mitigate over-smoothing problem during multi-layer information aggregation, the original features are concatenated with the RGCN output (Equation (2)) to form the d-dimensional embedding of n tasks $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_n)^{\mathrm{T}} \in \mathbb{R}^{n \times d}$, completing the feature extraction process.

$$\mathbf{h}_i = (\mathbf{S}_{T_i} \,\|\, \mathrm{RGCN}_{\theta_{\mathrm{R}}}(\mathbf{S}_{T_i}; \theta_{\mathrm{R}})) \tag{2}$$

3.3.2 Global representation module

The amount of information increases linearly with the number of tasks, so task-level information must be compressed into a graph-level representation to prevent dimensionality explosion. This study proposes a self-attention–based global representation method to identify critical tasks and reduce information loss A multi-layer perceptron (MLP) parameterized by θ_{G} is used to compute the attention weight $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_n)^{\mathrm{T}}$ of each task based on its task embedding (Equation (3)(4)), after which a weighted average of all task embeddings is performed to obtain a global embedding $\mathbf{g}$ with the same dimensionality as each individual task embedding (Equation (5)).

$$w_i = \mathrm{MLP}_{\theta_{\mathrm{G}}}(\mathbf{h}_i; \theta_{\mathrm{G}}) \tag{3}$$

$$\alpha_i = \frac{\exp(w_i)}{\sum_{j=1}^{n} \exp(w_j)} \tag{4}$$

$$\mathbf{g} = \sum_{i=1}^{n} \alpha_i \mathbf{h}_i \in \mathbb{R}^d \tag{5}$$

3.3.3 Task scoring module

We designed a scalable task-shared scorer, which shares feature-mapping parameters across all tasks. For

each task T_i, its task embedding $\mathbf{h}_i$ is concatenated with the global embedding $\mathbf{g}$ obtained from the global representation module and information $\mathbf{S}_c = (\mathbf{S}_{\text{type}}^{\text{agent}} \,\|\, \mathbf{S}_{\text{loc}}^{\text{agent}})$ of the agent, and then passed through an MLP scorer parameterized by θ_S to compute its score s_i. (Equation (6)).

$$s_i = \text{MLP}_{\theta_S}(\mathbf{h}_i \,\|\, \mathbf{g} \,\|\, \mathbf{S}_c; \theta_S) \tag{6}$$

Task scoring module not only preserves the independent comparability of individual nodes but also allows the network to learn the relative importance among tasks through backpropagation.

3.3.4 Decoder module

The decoder module utilizes the encoded information from upstream modules to generate both the value and action outputs. The value decoder maps the global information contained in the global embedding $\mathbf{g}$ into a scalar value v using MLP parameterized by θ_V:

$$v = \text{MLP}_{\theta_V}(\mathbf{g}; \theta_V) \in \mathbb{R} \tag{7}$$

The action decoder receives the task scores $\mathbf{s} = (s_1, s_2, ..., s_n)^T$ from the shared scorer. To handle the variable action space, invalid tasks are masked by setting their corresponding logits to negative infinity. Then softmax function is then applied to obtain the action probabilities. Finally, the selected action a_t is sampled from this probability distribution.

3.3.5 PPO for MCTA

PPO[35] is a widely used RL algorithm that improves training stability and efficiency by introducing a clipped objective function to limit abrupt policy updates. It alternates between sampling trajectories through agent–environment interaction and performing several epochs of gradient-based optimization, making it simple to implement yet highly effective for scheduling problems like MCTA. This study employs PPO to train agent strategy.

4 Case study

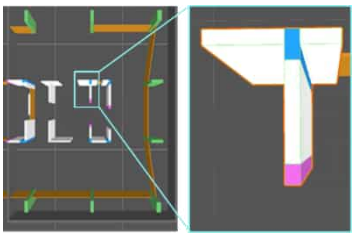
4.1 Case description and modeling process

To validate the effectiveness of the proposed method, we adopts a case study based on vertical component construction for a prefabricated structure. The case involves 7 construction crews (N=7) across 5 trade types, and a total of 86 construction tasks (n=86) associated with 36 structure components(Table 2). The 7 crews consist of one Joint cutting and Cleaning Crew (JCC), one Hoisting and Installation Crew (HIC), one Grouting Crew(GC), two Reinforcing Crews (RC1 and RC2), and two Form-work Crews (FC1 and FC2). Under capability constraints, each crew can only perform tasks corresponding to its trade. The simulation parameters are configured in strict accordance with reference[32]. We developed an RL environment based on Unity 3D.

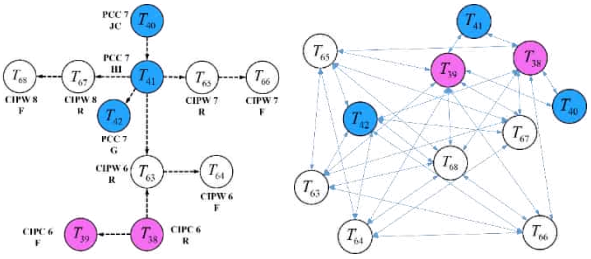
Table 2 Basic attributes of components

Component type	ID	Required tasks
Prefabricated columns	PCC 1-6	JC,HI,G
Cast-in-place columns	CIPC 1-5	R,F
Prefabricated walls	PCW 1-8	JC,HI,G
Cast-in-place walls	CIPW 1-17	R,F

A local structure segment (shown in Figure 4(a)) is selected to demonstrate the disjunctive graph modeling process. Within this segment, 11 tasks are involved, connected by 10 precedence relations, which are directed from predecessor tasks to successor tasks to represent precedence constraints, as shown in Figure 4(b). These precedence relationships are derived not only from general construction knowledge but also from the connection relationships among components and the specific technical requirements of the project. The 34 spatial overlap edges are illustrated in Figure 4(c). Each edge indicates that the corresponding tasks have overlapping bounding volumes and therefore cannot be executed simultaneously under the spatial overlap constraint. In Figure 4(b) and (c), the nodes are arranged using different layouts to better visualize the distinct connection patterns of the two relations.



(a) Selected segment of the structure



(b) Precedence relations (c) Spatial overlap relations

Figure 4 Modeling the case with disjunctive graph

4.2 Configurations

The simulation experiments were conducted on a computer equipped with a 13th Gen Intel® Core™ i9-13900K CPU (3.00 GHz), 64 GB RAM, an NVIDIA GeForce RTX 4090 GPU, and a 64-bit Windows 11 Professional operating system. In the simulation environment, each time step corresponds to 2.5 seconds, and each spatial unit represents one meter in the real-world project. Key network parameters are summarized in Table 3.

Table 3 Network parameters

Parameter	Value
Task feature dimension	12
Agent feature dimension	7
Number of RGCN layers	3
Hidden layer dimension of RGCN	64
Output layer dimension of RGCN	64
Embedding dimension	64+12=76
Number of MLP_{θ_G} layer	2
Hidden layer dimension of MLP_{θ_G}	64
Number of MLP_{θ_S} layer	3
Hidden layer dimension of MLP_{θ_S}	512
Number of MLP_{θ_V} layer	3
Hidden layer dimension of MLP_{θ_V}	512
Activation function	ReLU
Learning rate	0.0001
Epsilon	0.2
Number of epoch	3
Max step	330000
Buffer size	1280
Batch size	128

4.3 Results

Through iterative interaction with the environment (Figure 5), the proposed method successfully generated a fine-grained construction schedule for the case project. All 86 tasks were finished without constraints violated. Reference [36] evaluated 12 rules for task sequencing, from which we select three well-performing rules that are suited to MCTA as baselines. (1) First In First Out (FIFO): Tasks that enter the task pool $task_{\mathrm{queue}}$ earlier are given higher priority, reflecting a fairness-oriented scheduling strategy based on task availability. (2) Shortest Operation Time (SOT): Tasks with the shortest execution time are prioritized to improve efficiency. (3) Shortest Distance (SD): Tasks that are geographically closest to the crew agent are prioritized, aiming to reduce travel time. In addition, we import the Priority Rule (PR)[9], that comprehensively considers multiple factors.

Among metaheuristic methods, we selected GA as a baseline, as it is widely applied in scheduling research. Following the approach in [37], each gene represents the priority of a task, and the total project duration is used as the fitness function. The population size is set to 100. The optimization process is terminated after 80 generations, as the best fitness value shows no improvement for 20 consecutive generations. Also, an RL-based method in construction scheduling, FC-PPO[32], is included as a baseline. For fair comparison, this baseline adopts the same number and dimension of hidden layers as the scorer and value networks proposed in this study, and is trained for the same number of steps.



Figure 5 Runtime visualization of the simulator

Three performance indicators are defined to quantitatively evaluate the efficiency and stability of the MCTA system (as shown in Table 4), including:(1) Project Duration (PD): the total duration of the construction process.(2) Average Idle Time (AIT): the average idle time of crews caused by the unavailability of executable tasks.(3) Crew Utilization Rate (CUR): the ratio of working time to total time for crews. The results are averaged over 30 simulation runs with different random seeds.

Among the heuristic methods, FIFO, SOT, and SD each focus on a single factor and lack adaptability to complex construction scenarios. For highly dynamic construction scenarios, the GA method based on pre-calculated task priorities struggles to achieve reliable performance within limited iterations, with an average PD of up to 19.83 h. Among all seven evaluated methods, the proposed approach achieves the shortest average PD of 14.40 h, representing a 8.63 % reduction compared with the best competing method (FC-PPO). Moreover, its standard deviation is only 0.35 h, demonstrating excellent stability in performance. AIT and CUR reflect the utilization of human resources in the MCTA problem. The proposed method significantly reduces AIT of crews, achieving a 15.88 % reduction compared with the best among the other seven methods (FC-PPO), while also attaining the highest CUR. These results strongly demonstrate the superior performance of the proposed approach.

Table 4 Performance comparison of different methods

Method	PD (h)	AIT (h)	CUR (%)
FIFO	19.98±0.45	6.85±0.36	50.63±1.21
SOT	19.57±0.37	6.48±0.29	50.54±0.89
SD	16.30±1.04	4.88±0.61	56.68±2.15
PR	16.64±0.63	4.99±0.37	54.93±1.18
GA	19.83±0.27	6.79±0.23	50.75±0.65
FC-PPO	15.76±0.75	4.22±0.34	56.39±1.49
Ours	**14.40**±0.35	**3.55**±0.27	**60.29**±0.88

To further investigate whether the learned representations effectively capture real-world construction conditions, an additional analysis is conducted from two perspectives. (1) An ablation study is performed to evaluate the contribution of the graph-based representation. Specifically, the RGCN is replaced with a task-wise MLP that directly processes raw features. The results show that the cumulative reward does not exhibit an increasing trend during training, indicating that features extracted from the disjunctive graph are indispensable for decision-making. (2) A correlation analysis is conducted to examine the relationship between extracted features and the task importance reflected by the self-attention mechanism. The attention weights, together with the task scores, provide an indication of which features are most relevant to the construction conditions. The results show that the extracted features achieve an average Spearman rank correlation coefficient ρ of 0.46, with 70.0 % of feature dimensions exhibiting moderate or higher correlation ($\rho >$ 0.4), and 17.5 % showing strong correlation ($\rho >$ 0.7). In contrast, the raw features yield an average ρ of only 0.05, indicating a very limited contribution to representing construction conditions.

5 Conclusion and future work

This study proposes a disjunctive graph–based representation for flexible construction processes, together with a scalable RL framework to address the MCTA problem. The core network architecture integrates a feature extraction module, a task scoring module, a global representation module, and a decoder module. Within this architecture, a RGCN is employed to capture multi-relational features from the disjunctive graph, while a self-attention mechanism aggregates global information in a scale-agnostic manner. These designs enable the learned policy to generate high-quality task allocation decisions that shorten overall project duration and significantly improve labor resource utilization.

Future research should focus on developing large, standardized datasets for various construction types to learn generalized scheduling policies with stronger transferability and robustness. Advancing these directions will further strengthen the proposed framework's capability to support practical, real-time decision-making in complex construction management environments.

References

[1] FA. Agrama. Multi-objective genetic optimization for scheduling a multi-storey building. Automation in Construction, 44:119-128, 2014.

[2] M. Tomczak and P. Jaśkowski. Harmonizing construction processes in repetitive construction projects with multiple buildings. Automation in Construction, 139:104266, 2022.

[3] SH. Dike. Project scheduling with resource constraints. IEEE Transactions on Engineering Management, EM-11(4):155-157, 1964.

[4] H. Ding, C Zhuang, and J. Liu. Extensions of the resource-constrained project scheduling problem. Automation in Construction, 153:104958, 2023.

[5] M. Taghaddos, A. Mousaei, H. Taghaddos, U. Hermann, Y. Mohamed, and S. AbouRizk. Optimized variable resource allocation framework for scheduling of fast-track industrial construction projects. Automation in Construction, 158:105208, 2024.

[6] A. Kallantzis and S. Lambropoulos. Critical path determination by incorporating minimum and maximum time and distance constraints into linear scheduling. Engineering, Construction and Architectural Management, 11(3):211-222, 2004.

[7] AM. Elazouni and AA. Gab-Allah. Finance-based scheduling of construction projects using integer programming. Journal of Construction Engineering and Management, 130(1):15-24, 2004.

[8] W. Liu, J. Zhang, and W. Liu. Heuristic Methods for Finance-Based and Resource-Constrained Project Scheduling Problem. Journal of Construction Engineering and Management, 147(11):04021141, 2021.

[9] B. Liu, B. Yang, M. Dong, S. Jiang, and J. Xiao. Spatio-temporal deduction of floor construction based on the agent modeling of construction actors. Automation in Construction, 142:104487, 2022.

[10] N. Wongwai and S. Malaikrisanachalee. Augmented heuristic algorithm for multi-skilled resource scheduling. Automation in Construction, 20(4):429-445, 2011.

[11] X. Ye, H. Guo, and Z. Luo. Two-stage task allocation for multiple construction robots using an improved genetic algorithm. Automation in Construction, 165:105583, 2024.

[12] TQ. Nguyen, DKN. Chua. Preemptive Constraint Analysis in Construction Schedules. Journal of

Computing in Civil Engineering, 29(5):04014062, 2015.
[13] Q. Amarkhil and E. Elwakil. Enhanced planning and scheduling in building construction projects: an innovative approach to overcome scheduling challenges. International Journal of Construction Management, 24(16):1719-1729, 2024.
[14] L. Xie, Y. Chen, and R. Chang. Scheduling Optimization of Prefabricated Construction Projects by Genetic Algorithm. Applied Sciences, 11(12):5531, 2021.
[15] Y. Yao Y, VWY. Tam, J. Wang, KN. Le, A. Butera. Automated construction scheduling using deep reinforcement learning with valid action sampling. Automation in Construction, 166:105622, 2024.
[16] RK, Soman, M. Molina-Solana. Automating look-ahead schedule generation for construction using linked-data based constraint checking and reinforcement learning. Automation in Construction, 134:104069, 2022.
[17] NS. Kedir, S. Somi, AR. Fayek, and PHD. Nguyen. Hybridization of reinforcement learning and agent-based modeling to optimize construction planning and scheduling. Automation in Construction, 142:104498, 2022.
[18] A. Khodabandelu and J. Park. Agent-based modeling and simulation in construction. Automation in Construction, 131:103882, 2021.
[19] M. Schlichtkrull, TN. Kipf, P. Bloem, R. Berg, I. Titov, and M. Welling. Modeling Relational Data with Graph Convolutional Networks. arXiv, 1703.06103, 2017.
[20] JE. Kelley and MR. Walker. Critical-path planning and scheduling. In Eastern Joint IRE-AIEE-ACM Computer Conference, pages 160-173, Boston, USA, 1959.
[21] Mattila KG, Abraham DM. Resource leveling of linear schedules using integer linear programming. Journal of Construction Engineering and Management, 124(3):232-244, 1998.
[22] JE. Gomar, CT. Haas, and DP. Morton. Assignment and allocation optimization of partially multiskilled workforce. Journal of Construction Engineering and Management, 128(2):103-109, 2002.
[23] PG. Ipsilandis. Multiobjective linear programming model for scheduling linear repetitive projects. Journal of Construction Engineering and Management, 133(6):417-424, 2007.
[24] D. Robinson. Dynamic-Programming Solution to Cost-Time Tradeoff for Cpm. Management Science, 22(2):158-166, 1975.
[25] W. Liu, J. Zhang, W. Liu. Heuristic Methods for Finance-Based and Resource-Constrained Project Scheduling Problem. Journal of Construction Engineering and Management, 147(11):04021141, 2021.
[26] Y. Gajpal, A. Elazouni. Enhanced heuristic for finance-based scheduling of construction projects. Construction Management and Economics, 33(7):531-553, 2015.
[27] S. Al‐Shihabi, MM. AlDurgam. The contractor time–cost–credit trade‐off problem: integer programming model, heuristic solution, and business insights. International Transactions in Operational Research, 27(6):2841-2877, 2020.
[28] WT. Chan, DKH. Chua, and G. Kannan. Construction resource scheduling with genetic algorithms. Journal of Construction Engineering and Management, 122(2):125-132, 1996.
[29] A. Elazouni, A. Alghazi, SZ Selim. Finance-based scheduling using meta-heuristics: discrete versus continuous optimization problems. Journal of Financial Management of Property and Construction, 20(1):85-104, 2015.
[30] RS. Sutton and AG. Barto. Reinforcement Learning: An Introduction. 2nd ed. MIT Press, Cambridge, Massachusetts and London, England, 2018.
[31] AK. Agrawal, Y. Zou, H. Jin, M. Abdelmegid, and V. Gonzalez. Automated assembly sequence planning and scheduling in precast building projects using reinforcement learning. In Proceedings of the 42nd ISARC, Montreal, Canada, 2025.
[32] B. Yang, B. Liu, Y. Han Y, X. Meng, Y. Wang, H. Yang, and J. Xiao. Multi-agent reinforcement learning-enhanced decision-making of crew agents in floor construction processes. Advanced Engineering Informatics, 69:103996, 2026.
[33] B. Yang, B. Liu, J. Xiao, B. Zhang, Z. Wang, and M. Dong. A novel construction scheduling framework for a mixed construction process of precast components and cast-in-place parts in prefabricated buildings. Journal of Building Engineering, 43:103181, 2021.
[34] S. Wang, X. Li, L. Gao, J. Li, A multi-disjunctive-graph model-based memetic algorithm for the distributed job shop scheduling problem, Advanced Engineering Informatics 60 (2024)
[35] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal Policy Optimization Algorithms. arXiv, 1707.06347, 2017.
[36] H.-H. Doh, J.-M. Yu, J.-S. Kim, D.-H. Lee, and S.-H. Nam. A priority scheduling approach for flexible job shops with multiple process plans. International Journal of Production Research, 51(12):3748–3764, 2013.
[37] M. Taghaddos, H. Taghaddos, U. Hermann, Y. Mohamed, and S. AbouRizk. Hybrid multi-mode simulation and optimization for subarea scheduling in heavy industrial construction. Automation in Construction, 125:103616, 2021.

A Framework for Multi-scale Quantitative Defect Inspection of Concrete Bridge Surface Using Multi-view RGB-D Data

Sungwoo Cho[a], Seunghun Im[a], Taegyu Kim[a], Duho Chung[a], Ilhyoung Shin[b], Hyeongu Ji[b] and Hyoungkwan Kim[a*]

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
[b]Dangjin Branch, Korea Expressway Corporation, Korea
E-mail paul0524@yonsei.ac.kr, imseunghun4006@yonsei.ac.kr, tacoskywalker@yonsei.ac.kr, jungduho1@yonsei.ac.kr, niceguy@ex.co.kr, jeehw@ex.co.kr and hyoungkwan@yonsei.ac.kr

Abstract -

Robotic bridge inspection systems are emerging as crucial technologies to improve efficiency and safety in infrastructure maintenance, particularly in accessibility-constrained environments. However, the challenge of processing multi-view data and integrating detection, quantification, and report generation into a unified workflow remains unresolved. To address this challenge, this paper proposes an integrated inspection framework using multi-view RGB-D image sequences for a robotic bridge inspection. The framework supports multi-scale inspection by enabling both wide-area inspection and local-area inspection, while employing a two-stage voting mechanism to mitigate duplicate detections and false positives in 3D space. Experiments at a Precast Segmental Method (PSM) concrete box-girder bridge demonstrated that the framework successfully detected nine defect types, generated 2D inspection maps, and produced millimetre-scale quantification results. This work demonstrates the feasibility of an integrated robotic bridge inspection system that encompasses multi-view data analysis and quantification for practical workflows.

Keywords –

Bridge inspection; Multi-view integration; Defect detection; Quantification; Multi-scale inspection

1 Introduction

Surface defect inspection of concrete bridge structures is critical for ensuring structural safety and supporting maintenance decision-making. In South Korea, the proportion of bridges in service for over 30 years has been increasing [1], with surfaces such as box-girder interiors and exterior abutments representing typical hard-to-access regions due to their locations and structural characteristics [2]. Practical bridge inspection requires multi-scale approaches combining wide-area inspection of the overall structure and local-area inspection of defect-concentrated regions [3]. Beyond detection, defect measurement and quantification are essential components of concrete bridge inspection [4]. Obtaining quantitative measurements in physical units is essential for condition assessment and maintenance decision-making in practical bridge inspection. However, existing processes rely heavily on visual observation by inspectors, causing persistent issues including subjectivity, lack of consistency, and labour-intensive procedures [5]. Accordingly, there is growing need for automated inspection technologies providing objective and consistent quantitative results even in environments with limited accessibility.

In response to these demands, recent research has actively employed deep learning techniques for bridge surface defect detection and quantification [6, 7]. Torok et al. [8] investigated methods to identify global defect location and distribution through Structure-from-Motion (SfM), while Deng et al. [9] explored image-based Simultaneous Localization and Mapping (SLAM) approaches. Research on surface defect detection using Light Detection and Ranging (LiDAR) sensors has also been presented. Feng et al. [10] quantified crack geometry and location in 3D space through sensor fusion SLAM combining LiDAR and camera. Lin et al. [11] proposed surface defect detection methodologies utilizing 2D projection of LiDAR-acquired point clouds, demonstrating image and point cloud integration in defect inspection. However, these prior efforts have been primarily limited to specific defect types such as cracks and have not addressed comprehensive quantification across multiple defect types or inspection report generation. Furthermore, multi-scale inspection strategies encompassing both wide-area and local-area inspection have been largely overlooked.

Meanwhile, research on robot-based data acquisition systems for accessibility-constrained environments has also been conducted [12], and quadruped robots have demonstrated stable mobility and continuous image sequences acquisition across various environments [13]. However, these robotic data

acquisition technologies and defect analysis algorithms have been developed independently, and methodologies to systematically address duplicate detections and false positives arising from multi-view image sequences remain insufficient.

This paper proposes an integrated inspection framework for multi-scale defect detection, quantification, and report generation using bridge surface image sequences data. The proposed framework addresses the detection, classification, and quantification of nine defect types specified in Korean regulations [14], while mitigating duplicate detections and false positives. To validate the proposed framework, experiments were conducted using data acquired by a quadruped robot at an exterior abutment of a box-girder bridge. Additionally, a multi-scale inspection strategy was validated through experiments using interior diaphragm data captured by a handheld camera. These experiments demonstrated the practical applicability of the framework in real inspection environments.

2 Methodology

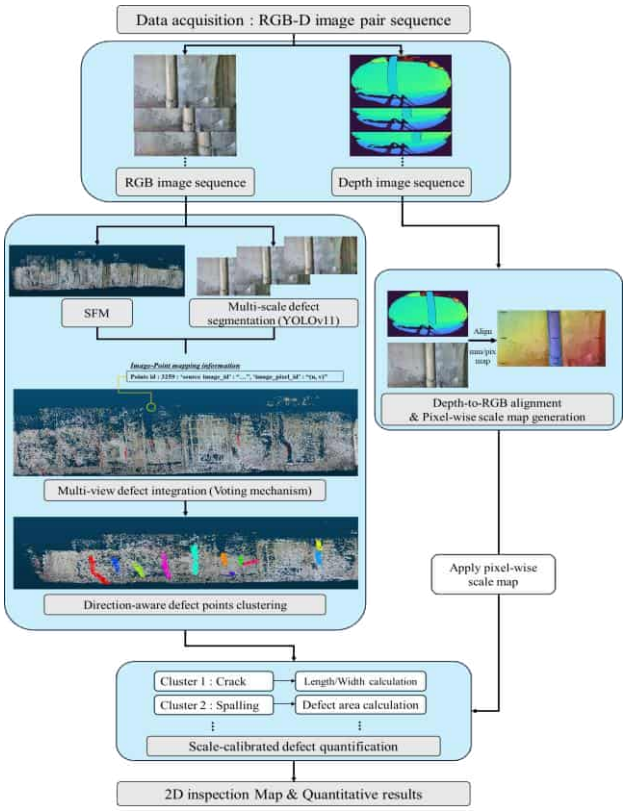


Figure 1. Overall Framework

Figure 1 shows the overall workflow of the proposed framework. The framework processes RGB-D image sequences, leveraging both 2D information from RGB images and 3D spatial information from depth data. Each processing module is described in detail in the following subsections.

2.1 SfM-based 3D reconstruction

SfM-based reconstruction is performed to integrate 2D RGB images into 3D space. The key advantage of SfM is that it establishes 2D-3D correspondence by identifying which pixel in which source image generated each point in the reconstructed point cloud. In overlapping regions between images, a single 3D point is mapped to multiple images, which forms the basis for multi-view integration. The proposed framework utilizes this 2D-3D correspondence to address duplicate detections and false positives arising from multi-view image sequences, as well as to support defect clustering and quantification. In addition, the reconstructed point cloud serves as the foundation for generating a 2D inspection map that visualizes the relative spatial distribution of defects within the inspected area.

2.2 Multi-scale defect detection

A YOLOv11x-seg model [15] is applied to each RGB image to detect and segment defects. The model was trained on nine types of surface defects in the dataset: crack, detachment, spalling, rebar exposure, material separation, damage, leak, lifting, and efflorescence. This paper used datasets from AI-hub [16]. All images are processed at the standard input resolution of 640 by 640 pixels. The model outputs a defect mask in polygon format for each detected defect, enabling the identification of defect type and boundary information within each RGB image.

To support both wide-area inspection and local-area inspection, the proposed framework provides two detection modes. Standard Mode uses the original image as input to the YOLOv11 model and detects defects directly. In contrast, Crop Mode divides the image into a grid of user-specified dimensions and performs YOLOv11 model for inference independently on each cropped image. To prevent defects from being split across crop boundaries, adjacent cropped images are overlapped by a specified number of pixels. For example, when using a 2 by 2 grid with 150-pixel overlap, defects detected in each cropped image can be geometrically merged. The continuous shape of defects that span multiple crops is then restored by computing the geometric union of overlapping defect polygons. While the grid size in Crop Mode can be selected flexibly without compromising detection performance, larger grids increase computational cost. Users can therefore choose the grid size based on site conditions and required detection precision.

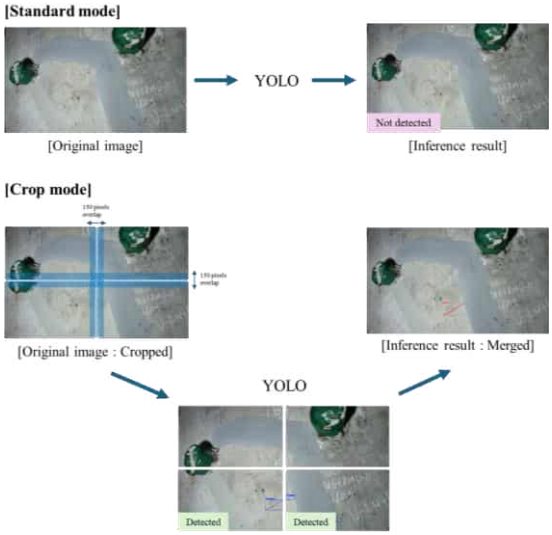


Figure 2. Standard Mode and Crop Mode for multi-scale defect detection (2×2 grid example)

2.3 Pixel-wise scale map generation

Since SfM-based reconstruction is inherently scale-free, depth information is required to convert pixel-based measurements into real-world physical units. To address the perspective-induced scale variation where the same physical dimension is represented by different numbers of pixels depending on the camera-to-surface distance, a pixel-wise scale map is generated for each image using depth information obtained from the RGB-D camera. This scale map computes the physical dimension per pixel (in millimetres per pixel) at each location in the image based on the depth value and camera focal length.

First, the depth image is aligned to the coordinate frame of the corresponding RGB image. This process is performed through coordinate-transformation and resolution matching using the extrinsic parameters between the RGB and depth cameras and the intrinsic parameters of each camera.

Each pixel (u_d, v_d) in the depth image and its corresponding depth value Z_d are back-projected to a 3D point $\mathbf{P}_d = (X_d, Y_d, Z_d)^\top$ in the depth camera coordinate frame using the depth camera intrinsic parameters. The depth camera intrinsic parameter matrix $\mathbf{K}_d$ is defined as:

$$\mathbf{K}_d = \begin{bmatrix} f_{x,d} & 0 & c_{x,d} \\ 0 & f_{y,d} & c_{y,d} \\ 0 & 0 & 1 \end{bmatrix}$$

where $f_{x,d}$, $f_{y,d}$ are the focal lengths of the depth camera and $c_{x,d}$, $c_{y,d}$ represent the principal point. The back-projection process is expressed as:

$$\begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix} = Z_d \cdot \mathbf{K}_d^{-1} \begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix}$$

The 3D point $\mathbf{P}_d$ expressed in the depth camera coordinate frame is then transformed to the RGB camera coordinate frame. This transformation is performed using the extrinsic parameters between the depth and RGB cameras, consisting of a rotation matrix $\mathbf{R}_{d\to r}$ and a translation vector $\mathbf{t}_{d\to r}$:

$$\mathbf{P}_r = \mathbf{R}_{d\to r}\,\mathbf{P}_d + \mathbf{t}_{d\to r}$$

where $\mathbf{R}_{d\to r} \in \mathbb{R}^{3\times 3}$ is the rotation matrix that rotates the depth camera coordinate frame to the RGB camera coordinate frame and $\mathbf{t}_{d\to r} \in \mathbb{R}^3$ is the translation vector representing the relative position difference between the two cameras.

The 3D point $\mathbf{P}_r = (X_r, Y_r, Z_r)^\top$ which has been transformed into the RGB camera coordinate frame, is projected onto the RGB image plane using the RGB camera intrinsic parameter matrix $\mathbf{K}_r$:

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \frac{1}{Z_r}\mathbf{K}_r \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}$$

Through this process, each pixel in the depth image is mapped to a corresponding pixel (u_r, v_r) in the RGB image coordinate frame, establishing pixel-level correspondence between the RGB and depth images.

From the aligned depth map, the physical scale $s_x(u_s, v_s)$, $s_y(u_s, v_s)$ at each pixel is computed using the depth value $d(u_d, v_d)$ (in millimetres) and the RGB camera focal lengths f_x , f_y :

$$s_{x\,or\,y}(u_s, v_s) = \frac{d(u_d, v_d)}{f_{x\,or\,y}}$$

The pixel-wise scale map generated in this manner is created for each RGB image and is subsequently used in the defect quantification stage to convert pixel-based measurements into physical units.

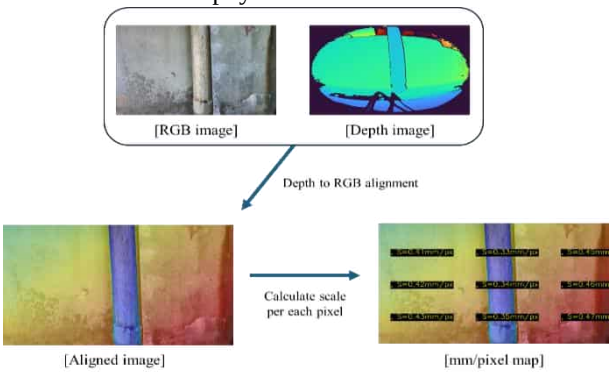


Figure 3. Scale map generation process

2.4 Multi-view defect integration in 3D space via two-stage voting mechanism

Defect polygons detected by the YOLOv11 model in each RGB image are integrated into 3D space through projection onto the point cloud generated by SfM. Using the 2D-3D correspondence established in the SfM database, each 3D point is back-projected to all RGB images in which it is visible. If the projected pixel coordinate falls inside a detected defect polygon, the point is designated as a defect point candidate. Through this process, each 3D point receives semantic information regarding defect presence and defect class from multiple RGB images. Since each 3D point uniquely represents a specific spatial location, the same defect can be represented as a single entity in 3D space even when observed in multiple images or captured in fragmented views.

During the projection process, false detections from the defect segmentation model must be considered. When acquiring numerous image sequences for inspection, false detections can occur frequently due to variations in lighting conditions and viewing angles. False detections can manifest in two forms: false positives where non-defect regions are incorrectly identified as defects, and misclassifications where defect classes are incorrectly predicted. To mitigate these issues, a two-stage voting mechanism is designed to leverage the multi-view nature of robotic inspection.

In Stage 1, the Detection Voting stage, the presence or absence of a defect at each point is determined. Each 3D point P_j has an associated set of images $\mathcal{I}_j$ in which it is observed. For each image $I_i \in \mathcal{I}_j$, the 3D point P_j is projected onto the image, and it is verified whether the projected pixel location falls inside any defect mask (including all defect classes). The defect presence voting ratio $v_d(P_j)$ for point P_j is defined as:

$$v_d(P_j) = \frac{1}{|\mathcal{I}_j|}\sum_{I_i \in \mathcal{I}_j} \mathbb{1}_{mask}(I_i, P_j)\ ,\ \tau_1 = 0.5$$

where $|\ \mathcal{I}_j\ |$ is the number of images in which point P_j is observed, and $\mathbb{1}_{mask}(I_i, P_j)$ is an indicator function that returns 1 if the projected location of point P_j in image I_i falls inside any defect mask, and 0 otherwise. In the proposed framework, a threshold τ_1 is applied, and only points satisfying $v_d(P_j) \geq \tau_1$ are selected as defect-present 3D points.

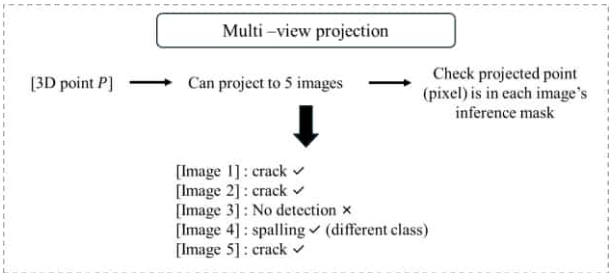
In Stage 2, the Class Voting stage, the final defect class is determined for each defect-present 3D point selected from Stage 1 by aggregating the observation frequency for each defect class. The voting ratio $v_c(P_j, c)$ for point P_j to be classified as class c is defined as:

$$v_c(P_j, c) = \frac{1}{|\mathcal{I}_j|}\sum_{I_i \in \mathcal{I}_j} \mathbb{1}_{mask}(I_i, P_j, c)$$

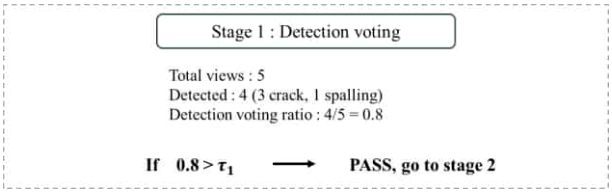
where $\mathbb{1}_{mask}(I_i, P_j, c)$ returns 1 if the projected location of point P_j in image I_i, falls inside a defect mask of class c, and 0 otherwise. The final defect class is determined through the following condition, which eliminates cases where defect classes are inconsistently predicted:

$$v_c(P_j, c) \geq \tau_2\ ,\ \tau_2 = 0.5$$

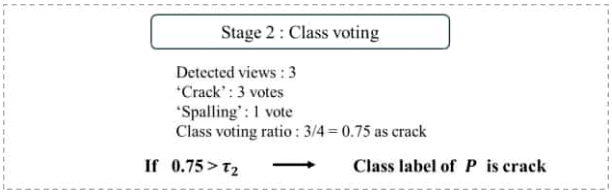
This two-stage voting mechanism mitigates false detections and misclassifications by labelling each point in the point cloud with a defect class based on the projection results. Through this process, multi-view defects are integrated into a unified 3D representation. Figure 4 illustrates this process.



(a)



(b)



(c)

Figure 4. Two-stage voting mechanism example; (a) multi-view projection, (b) detection voting, (c) class voting

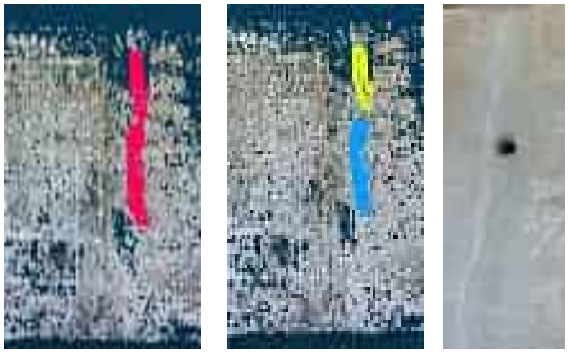
2.5 Direction-aware defect points clustering

The 3D defect points selected through voting stages must be clustered into individual defect instances. For defects such as cracks, conventional distance-based clustering using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) can erroneously merge intersecting cracks into a single cluster. To address

this issue, a direction-aware clustering strategy that leverages local orientation information is applied.

First, density-based primary clustering is performed using DBSCAN to remove noise and extract candidate defect regions. Second, the local orientation of points within each cluster is estimated using Principal Component Analysis (PCA) and point groups exhibiting large orientation differences are separated to split intersecting defects into independent instances. Finally, remerging is performed based on centroid distance and principal component orientation similarity to restore contiguous defects that may have been inadvertently separated. This direction-aware clustering approach overcomes the limitations of simple distance-based clustering, enabling more accurate clustering of complex defect geometries in multi-view environments. Figure 5 illustrates the differences between clustering methods.

Through this process, the final defect clusters can be visualized in 3D space, and this information is subsequently utilized to generate a 2D inspection map. This 2D inspection map provides the spatial locations of defects within the structure and can be used in practical inspection workflows.



(a) (b) (c)

Figure 5. Differences by clustering method; (a) DBSCAN clustering, (b) proposed clustering, (c) original crack RGB image.

2.6 Scale-calibrated defect quantification

Using the pixel-wise scale map established in Section 2.3, each clustered defect instance is quantified in millimetres. Since point cloud cluster data lacks clear boundary definition and exhibits sparsity, leading to potential measurement errors, defects are quantified using the RGB images associated with points in each cluster through the 2D-3D correspondence. As illustrated in Figure 6, each 3D point within a defect cluster can be projected to multiple RGB images through the 2D-3D correspondence. By aggregating these projection relationships, the image most frequently associated with the cluster points is selected as the measurement source. This image selection process is independently performed for each defect cluster, ensuring that every cluster is quantified from its own most representative image. The selected image's pixel-wise scale map and defect mask are then loaded, and different measurement strategies are applied depending on the defect type.

2.6.1 Crack measurement: length and width

For cracks, the segmentation mask is narrow, requiring measurement of length and width, but direct calculation from the mask is impractical. Therefore, the single-pixel-width crack centreline skeleton extraction and measurement algorithm proposed by [6] is adopted to measure length and width. However, [6] presented a measurement approach for single images. In the proposed framework, the image with the highest number of point observations from the 3D crack cluster is selected, and measurement is conducted on that image. Furthermore, while [6] used a fixed scale factor, the proposed approach applies the spatially varying scale map to convert pixel-based measurements into real-world units, thereby reducing measurement errors arising from depth changes and perspective effects.

2.6.2 Non-crack defect measurement: Area

For non-crack defects such as detachment, spalling, and rebar exposure, area is used as the primary quantification metric. In the selected image for the defect instance, the physical area of each pixel within the defect mask is computed by multiplying the scale values in the u and v directions from the pixel-wise scale map, and the total defect area is obtained by summing these individual pixel areas in square millimeters. Unlike cracks, non-crack defects do not require single-pixel-width precision, allowing the segmentation mask to be used directly.

The quantification results for each defect are stored in JSON format. Using the 2D inspection map and defect quantification results, the final inspection report is generated.

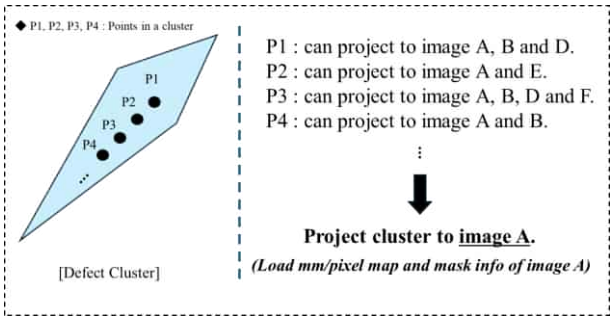


Figure 6. Scenario of image selection for defect cluster quantification

3 Experiments and Results

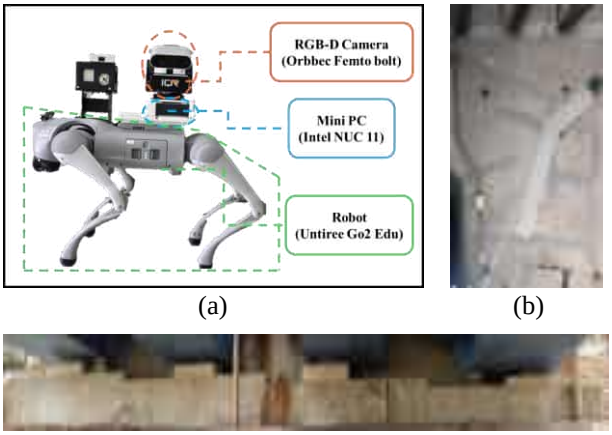
3.1 Data acquisition

For the experiments, data were collected using an RGB-D camera (Femto Bolt by Orbbec). In robot-based acquisition, the RGB-D camera was mounted on a quadruped robot (Go2 Edu by Unitree Robotics), with a mini-PC (Intel NUC 11) also equipped to store and manage the captured RGB-D image sequences (Figure 7(a)). In handheld acquisition, image sequences were acquired using the same camera model.

RGB and depth image sequences were captured as synchronized pairs. To ensure overlap for SfM-based 3D reconstruction, pairs were acquired at 0.5-second intervals. RGB-D images were captured at different resolutions per camera specifications (RGB: 3840 by 2160 pixels, depth: 512 by 512 pixels). Additionally, the RGB and depth lenses had different Fields of View (FOV). These characteristics and the extrinsic and intrinsic parameters were incorporated into the scale map generation described in Section 2.3.

3.2 Experiment Site

Images were acquired at an exterior abutment and an interior diaphragm of a Precast Segmental Method (PSM) concrete box-girder bridge (Figure 7 (b), (c)). Both sites represent different inspection scales within the multi-scale framework. At the abutment site, requiring wide-area inspection, the entire wall surface was captured using a quadruped robot, acquiring 698 RGB-D image pairs. At each position, images were taken at three camera orientations (upward, horizontal, downward tilt) from 50 to 70 centimetres standoff distance, advancing less than one metre between positions with five-second stationary periods to prevent motion blur. The diaphragm site represents a critical area where load-bearing prestressing tendons run through, requiring local-area inspection. A total of 212 RGB-D image pairs were acquired using a handheld camera. Since the diaphragm is inside the bridge structure, images were acquired with supplementary lighting due to limited natural lighting.



(a) (b)

(c)

Figure 7. (a) Hardware configuration on quadruped robot, (b) diaphragm site, (c) abutment site

3.3 Results

3.3.1 Abutment site

Figure 8 presents the wide-area inspection results obtained by applying the proposed framework to the exterior abutment. Figure 8(a) shows the 3D point cloud of the abutment reconstructed through SfM using the RGB-D image sequences acquired by the quadruped robot. This reconstruction enables verification of the overall geometry and inspection coverage of the target structure at a structure-wide scale. Figure 8(b) presents the integrated defect information obtained through multi-view integration and clustering of detections from multi-view image sequences. While defects may be observed repeatedly or captured in fragmented views across different viewpoints, the proposed framework consolidates them into single defect instances, mitigating duplicate detections. Different colours are assigned to each cluster for visualization purposes.

Figure 8(c) shows the 2D inspection map generated based on the identified 3D defect instances. The map visualizes defect locations and types in a structured format, enabling clear visualization of wide-area inspection results. Figure 8(d) presents quantification results for each defect type in tabular form.

To quantitatively evaluate the detection performance, 33 unique defects were manually identified across the abutment surface and used as Ground Truth (GT). YOLO alone produced 1,762 detection masks across 689 images, where duplicate detections from multi-view image sequences make it impossible to determine the actual number of unique defects without 3D projection. When detections were projected into 3D space and clustered without the voting mechanism, the framework achieved a Precision of 0.429, Recall of 0.758, and F1 Score of 0.556. With the multi-view voting mechanism applied, Precision improved to 0.889 with an

F1 Score of 0.781, demonstrating that the voting mechanism substantially reduces false positives while maintaining comparable Recall (0.697).

3.3.2 Diaphragm site

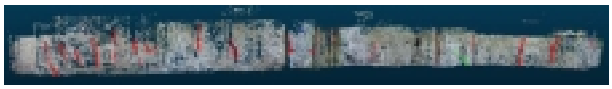
The diaphragm site represents a localized surface and requires finer inspection detail. Figure 9 demonstrates the effectiveness of the multi-scale inspection strategy at the interior diaphragm of the PSM box-girder bridge.

Figure 9 (a) shows the results obtained by applying Standard Mode to identify the overall defect distribution across the structure. In contrast, Figure 9 (b) presents the results of applying Crop Mode of the proposed framework to the region requiring local-area inspection. This comparison was conducted with all processing steps identical except for the detection scale. Based on YOLOv11 model inference on cropped images, defects that were not detected in Standard Mode were additionally detected.

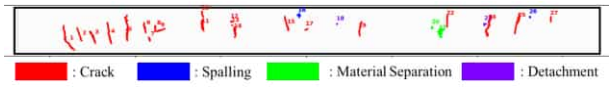
These results demonstrate that the proposed framework can accommodate both wide-area inspection and local-area inspection within a unified framework, as required in practical inspection environments.



(a)



(b)



(c)

No.	Type	Length (mm)	Width (mm)	Area (mm^2)
5	Crack	1145.6	0.39	-
15	Crack	650.0	0.44	-
23	Detachment	-	-	2515.3
26	Spalling	-	-	888.7

(d)

Figure 8. (a) point cloud from SfM algorithm, (b) all defect points visualization, (c) 2D inspection map, (d) example of defect quantification results



(a) (b)

Figure 9. Multi-scale inspection comparison at diaphragm site; (a) wide-area inspection using Standard Mode, (b) local-area inspection using Crop Mode

4 Discussion

The experimental results demonstrated that the proposed framework operates as a robotic inspection system for real concrete bridge surfaces. In inspection environments using multi-view images, issues such as duplicate detections, fragmented observations, and false positives frequently occur. The proposed framework mitigates these issues in 3D space and subsequently performs defect quantification. As shown in Section 3.3.1, the quantitative evaluation confirmed that the multi-view voting mechanism substantially reduces false positives while maintaining comparable Recall. However, direct comparison with the manual inspection map remains limited because the manual inspection was conducted at a different time than the experimental data acquisition.

Meanwhile, 3D reconstruction through SfM was advantageous in establishing explicit 2D-3D correspondence but is limited by the scarcity of extractable features in the acquired images. Sparse point clouds were generated in feature-poor regions, such as areas covered with repair tape on the interior diaphragm of the box girder, due to the lack of distinctive visual features or motion blur. To overcome these limitations, future work could explore RGB-D Visual SLAM or fusion with LiDAR sensors. Additionally, while the diaphragm site experiments demonstrated the effectiveness of multi-scale inspection, the use of handheld camera data can also be considered a limitation.

While the detection performance has been quantitatively evaluated in Section 3.3.1, further validation of the quantification process remains necessary. Ground truth measurements of defect dimensions should be obtained through manual inspection conducted concurrently with robotic data acquisition to enable direct comparison with the framework's quantification outputs.

5 Conclusion

This paper proposed an integrated framework that unifies defect detection, quantification, and report generation into a continuous workflow for accessibility-constrained concrete bridge surfaces using multi-view RGB-D image sequences. The framework addresses multi-scale inspection by supporting wide-area and local-area inspection while systematically mitigating duplicate detections and false positives in 3D space.

Experiments at the exterior abutment and interior diaphragm of a PSM box-girder bridge confirmed reliable operation in real environments, generating quantification results and 2D inspection maps for various defect types. This work demonstrates the feasibility of advancing robotic bridge inspection from simple data acquisition to an integrated system encompassing multi-view analysis and quantification.

Future research should address the limitations identified in Section 4. More robust reconstruction approaches must be developed to ensure stable 3D reconstruction. Rigorous validation of the quantification process is required to demonstrate that the proposed framework provides reliable results.

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. 2018R1A6A1A08025348) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00457308). ※ MSIT: Ministry of Science and ICT. This research was conducted with the support of i-thetto Co., Ltd. and Korea Expressway Corporation.

References

[1] CERIK(Construction & Economy Research Institute of Korea). Construction Trends Briefing No.941 for 2024. On-line : https://www.cerik.re.kr/report/briefing/2821, Accessed: 15/01/2026.

[2] Abdallah, A. M., Atadero, R. A., & Ozbek, M. E. (2022). A state-of-the-art review of bridge inspection planning: Current situation and future needs. *Journal of Bridge Engineering*, *27*(2), 03121001.

[3] Zhang, Y., & Yuen, K. V. (2022). Review of artificial intelligence-based bridge damage detection. *Advances in Mechanical Engineering*, *14*(9), 16878132221122770.

[4] Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y. J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning.

[5] Moore, Mark, et al. *Reliability of visual inspection for highway bridges, volume I*. No. FHWA-RD-01-105. Turner-Fairbank Highway Research Center, 2001.

[6] Yang, K., Bao, Y., Li, J., Fan, T., & Tang, C. (2024). Deep learning-based YOLO for crack segmentation and measurement in metro tunnels. *Automation in Construction*, *168*, 105818.

[7] Lin, C., Chen, Y., Itakura, K., Maharjan, S., & Chun, P. J. (2025). Bridge inspection using image–point cloud fusion with image filtering, damage detection and 3D registration. *Automation in Construction*, *180*, 106538.

[8] Torok, M. M., Golparvar-Fard, M., & Kochersberger, K. B. (2014). Image-based automated 3D crack detection for post-disaster building assessment. *Journal of Computing in Civil Engineering*, *28*(5), A4014004.

[9] Deng, L., Sun, T., Yang, L., & Cao, R. (2023). Binocular video-based 3D reconstruction and length quantification of cracks in concrete structures. Automation in Construction, 148, 104743.

[10] Feng, C. Q., Li, B. L., Liu, Y. F., Zhang, F., Yue, Y., & Fan, J. S. (2023). Crack assessment using multi-sensor fusion simultaneous localization and mapping (SLAM) and image super-resolution for bridge inspection. Automation in Construction, 155, 105047.

[11] Li, X., Liu, C., Yang, X., & Yan, H. (2025). 3D surface defect detection and measurement using point cloud unwrapping via dimension reduction. *Automation in Construction*, *177*, 106349.

[12] Lattanzi, D., & Miller, G. (2017). Review of robotic infrastructure inspection systems. *Journal of Infrastructure Systems*, *23*(3), 04017004.

[13] Halder, S., Afsari, K., Chiou, E., Patrick, R., & Hamed, K. A. (2023). Construction inspection & monitoring with quadruped robots in future human-robot teaming: A preliminary study. *Journal of Building Engineering*, *65*, 105814.

[14] *Enforcement Decree of the Special Act on the Safety Control and Maintenance of Establishments* (Presidential Decree No. 35948, Jan. 2, 2026, Korea).

[15] Yan, T. YOLOv11 by Ultralytics. On-line : https://github.com/yt7589/yolov11, Accessed: 15/01/2026

[16] AI-Hub. (2023). *The SOC facility crack-pattern image dataset.* On-line : https://www.aihub.or.kr, Accessed: 15/01/2026

A Framework for 3D Reconstruction of Concrete Box-Girder Interiors from Point Cloud Data

Taegyu Kim[a], Seunghun Im[a], Duho Chung[a], Sungwoo Cho[a], Ilhyoung Shin[b], Hyeongu Ji[b] and Hyoungkwan Kim[a*]

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
E-mail tacoskywalker@yonsei.ac.kr, imseunghun4006@yonsei.ac.kr, jungduho1@yonsei.ac.kr, paul0524@yonsei.ac.kr and hyoungkwan@yonsei.ac.kr
[b]Dangjin Branch, Korea Expressway Corporation, Korea
E-mail niceguy@ex.co.kr and jeehw@ex.co.kr

Abstract –

Inspection of concrete box-girders is essential for ensuring the structural performance of a PSM bridge throughout its service life. With advances in 3D sensing and point cloud processing, 3D data-based approaches have been increasingly applied to bridge inspection. This paper presents a 3D reconstruction pipeline for concrete box-girder interiors that generates CAD models directly from scan data. Point cloud data of girder interiors acquired by a quadruped robot equipped with a LiDAR sensor are preprocessed and then semantically segmented into interior components using a deep learning model. Segmented point cloud data is subsequently converted into a 3D CAD model through a sliced patch-based reconstruction algorithm. The proposed method was validated on real concrete box-girder interior data. Point cloud segmentation achieved a mIoU of 94.1%. The geometric accuracy of the generated CAD models was evaluated using mesh-to-point cloud distance, yielding mean distances of 0.025 m against point cloud data. These results demonstrate that the proposed pipeline can effectively construct as-built 3D models of concrete box-girder interiors, supporting efficient data-driven inspection.

Keywords –

Concrete Box-Girder; Point Cloud Data; Deep Learning; 3D Semantic Segmentation; 3D CAD Modeling; Robotic Data Acquisition.

1 Introduction

Bridge inspection is essential for ensuring structural performance throughout the entire service life. In particular, the internal inspection of bridges constructed using the Precast Segmental Method (PSM) has become an issue of growing concern. In 2016, corrosion and rupture of tendons were observed inside the concrete box-girder of Jeongneungcheon Overpass [1], and in 2020, tendon corrosion was identified at Cheongdam 1st Bridge [2]. Such cases of inadequate internal inspection have shown the critical importance of interior inspection for concrete box-girders.

Recent advances in 3D sensing technologies and point cloud processing have highlighted the potential of Scan-to-BIM and digital twin approaches for improving bridge inspection [3]. Jing et al. [4] proposed a method for semantic segmentation of masonry bridge components using a segmentation model trained solely on synthetic data, followed by instance segmentation using the DBSCAN algorithm. Kim et al. [5] applied SuperPoint Transformer to segment bridge point cloud data acquired by UAV-mounted LiDAR, training the model on both synthetic and real data. Yang et al. [6] presented a Scan-to-BIM framework for highway bridges, in which a deep-learning segmentation model trained on both synthetic and real datasets segments LiDAR-acquired point clouds into individual components, and the resulting segmented point clouds are then projected to extract component outlines that are subsequently parameterized and instantiated as BIM elements. Shu et al. [7] proposed an automated method to generate geometric digital twins of box-girder bridges, using a model trained on synthetic and TLS point clouds for component segmentation and cross-section key point extraction, and aligning internal structures generated from design drawings to the global coordinate system. Zhu et al. [8] proposed a method for cable-supported bridges with girders that uses TLS to scan the girder interior and processes interior and exterior point clouds to estimate the girder volume, including the internal void geometry.

Existing research have primarily focused on reconstructing external geometry or have relied on design drawings to model internal structures. Despite the importance of concrete box-girders, research on 3D data-driven inspection of box-girder interiors remains limited.

This highlights the need for methodologies to construct as-built 3D models specifically for concrete box-girder interiors.

This paper proposes a 3D reconstruction pipeline for concrete box-girder interiors based on point cloud semantic segmentation and a sliced patch-based 3D reconstruction algorithm. The proposed pipeline semantically segments structural components from interior scan data and reconstructs them into 3D models. Figure 1 illustrates an overview of the proposed pipeline. First, interior point cloud data are obtained from raw scans acquired by a LiDAR sensor mounted on a quadruped robot. Second, wall surfaces are extracted and reconstructed as parametric models to define the interior geometry. Third, the remaining interior components are semantically segmented using a deep learning-based segmentation model. Fourth, 3D CAD models of the segmented components are generated through a sliced patch-based algorithm.

2 Methodology

2.1 Point Cloud Data Acquisition and Preprocessing

As shown in Figure 1(a), interior data were acquired using a quadruped robot (Unitree Go2) equipped with a LiDAR sensor (Livox Avia) and a mini computer (Intel NUC). The acquired data were registered into a global map using FAST-LIO2 [9], and a loop-closure module was additionally integrated to correct accumulated drift and improve global consistency [10]. First, roll and pitch correction and axis alignment are performed to align the coordinate system such that the longitudinal direction of the girder is parallel to the X-axis. Then, the diaphragms at both ends of the girder were manually removed.

2.2 Wall Extraction and 3D CAD Model Generation

To extract wall surfaces from the preprocessed point cloud, this paper developed an alpha shape-based wall extraction algorithm. Figure 2(a) illustrates the stages of the proposed algorithm. First, the point cloud is projected

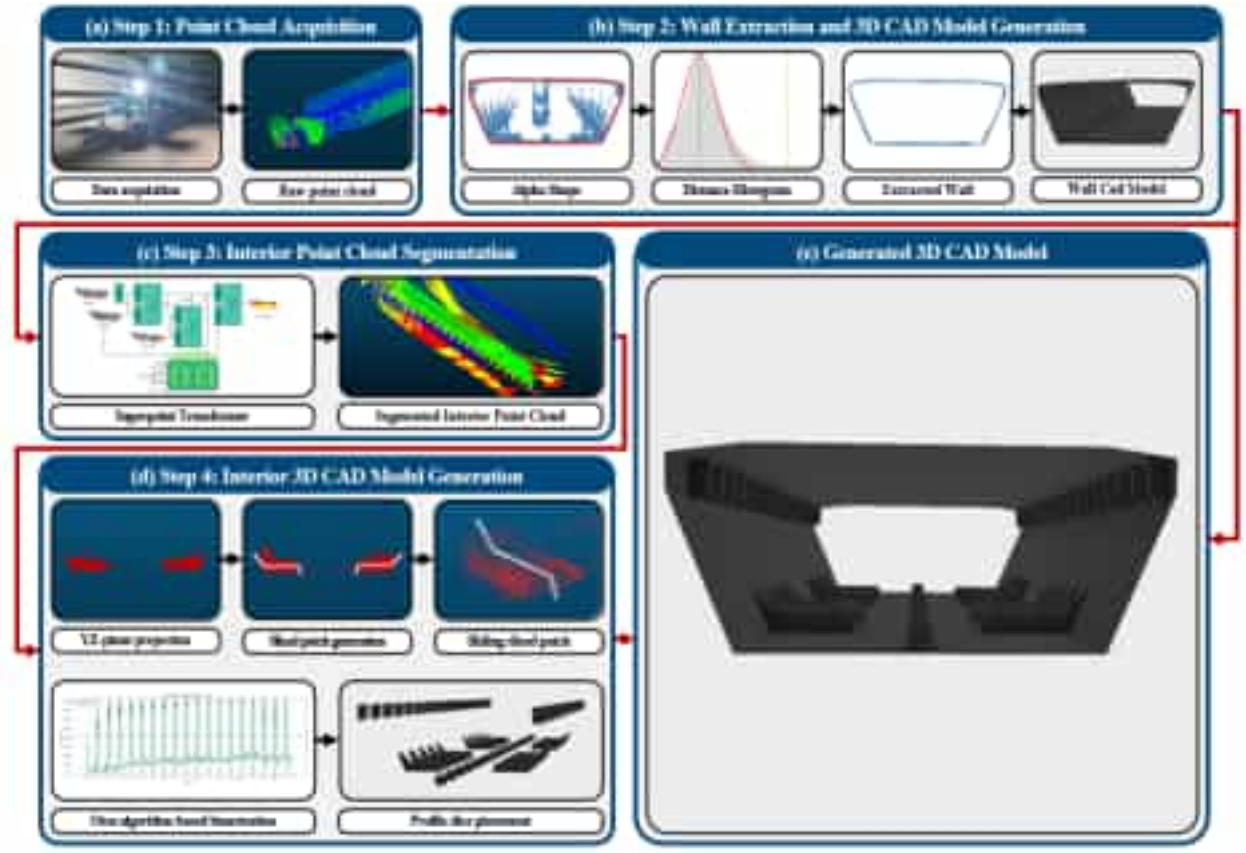


Figure 1. Overview of the proposed framework; (a) acquisition of box-girder interior point cloud data using a quadruped robot, (b) wall extraction and 3D CAD model generation, (c) point cloud segmentation using SuperPoint Transformer, (d) 3D CAD model generation of segmented components, (e) generated 3D CAD model.

onto the YZ plane, and the alpha shape algorithm is applied to estimate the outer contour of the girder cross-section. Statistical Outlier Removal (SOR) is applied prior to prevent errors caused by outliers. Second, the minimum Euclidean distance from each point in the full point cloud to the contour is computed. Third, the distribution of wall point distances is modeled as a Gaussian distribution. The peak of the distance distribution is defined as the mean μ. Since the portion of the histogram beyond the peak contains interior points, the Gaussian distribution is fitted only to the portion up to the peak. A threshold is then set at $\mu + n\sigma$ to separate wall points from interior points. In this study, a threshold of $\mu + 6\sigma$ was used, corresponding to a 99.99% inclusion probability. Finally, points within this threshold are classified as wall points, while the remaining points are classified as interior points. Figure 2(b) illustrates the generation of a 3D CAD model based on the extracted wall points. RANSAC-based line feature extraction [11] is first applied to the contour to extract parametric line segments representing the wall surfaces. The extracted line segments are then inwardly offset to μ to align them with the actual wall surface location. Finally, the cross-section is extruded along the longitudinal axis by the length of the girder to produce a parametric 3D model of the wall surfaces.

2.3 Point Cloud Segmentation

The preprocessed point cloud data is semantically segmented using SuperPoint Transformer, a deep learning-based 3D segmentation model [12]. SuperPoint Transformer (SPT) processes large-scale point clouds by first computing a multi-scale hierarchical partition of the scene into geometrically homogeneous superpoints. At each partition level, adjacent superpoints are connected to form a superpoint graph. The network then performs sparse self-attention on these graphs to propagate information at each scale, enabling large receptive fields and context aggregation while predicting semantic labels at the finest superpoint level. The interior components of the concrete box-girder are classified into five categories: deviator block, tendon, top temporary Post Tensioning (PT) bar anchor block, bottom temporary PT bar anchor block, and cable tray. A deviator block is a concrete block installed to redirect the alignment of external tendons, with embedded ducts through which tendons pass. The top and bottom temporary PT bar anchor blocks are used to apply compressive stress when joining segments during construction and are located at the top and bottom of the girder, respectively [13]. A tendon is a steel strand, and a cable tray is a metal structure that accommodates electrical and communication cables.

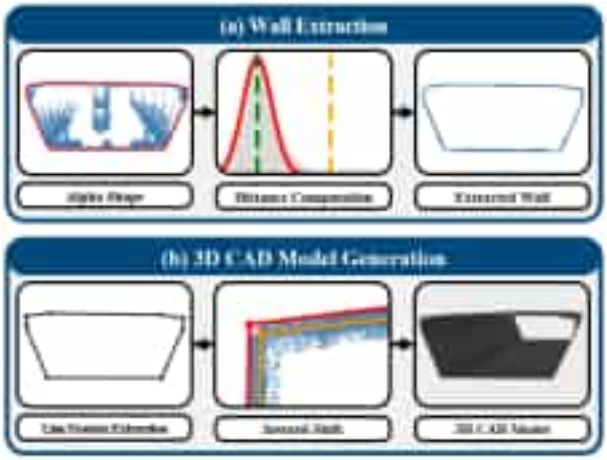


Figure 2. Extraction of the wall point cloud and 3D CAD model generation; (a) wall extraction, and (b) 3D CAD model generation.

2.4 Interior 3D CAD Model Generation

Among the five segmented classes, the deviator block, top temporary PT bar anchor block, and bottom temporary PT bar anchor block were converted into CAD models.

For 3D model generation, this paper developed a sliced patch-based 3D reconstruction algorithm. This algorithm is based on the patch-based component matching algorithm proposed by Kim et al. [14]. In their approach, a 3D patch similar in size and shape to each

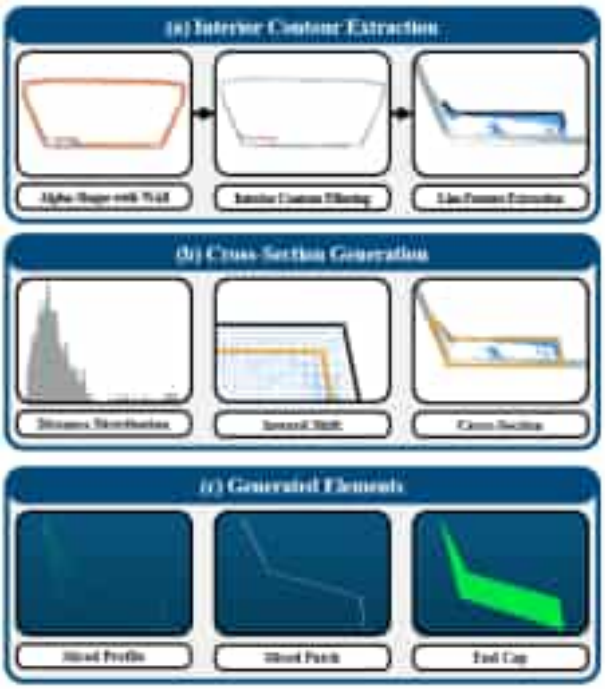


Figure 3. Generation of the three elements used in this study; (a) extraction of the interior contour, (b) cross-section generation, and (c) generated elements.

component is generated and traverses across the target region within the point cloud data. If the distance between points in the patch and the target point cloud data falls below a predefined threshold, the presence of the component is confirmed. This algorithm was originally proposed for detecting standardized components within scaffolding structures. In this paper, the algorithm was modified to a sliced patch-based approach.

Figure 3 illustrates the generation of patch elements used in the proposed algorithm, consisting of interior contour extraction, cross-section generation, and element generation. In Figure 3(a), the segmented point cloud of each component is projected onto the YZ plane and individual instances are separated using DBSCAN-based clustering. The wall point cloud is also projected, and the alpha shape algorithm is applied to the combined point cloud of the component and wall to extract the outer contour. Boundary edges for which either endpoint belongs to the wall point cloud are then removed, retaining only the edges corresponding to the interior surfaces. RANSAC-based line feature extraction [11] is applied to the remaining edges to extract parametric line segments Figure 3(b) shows the cross-section generation step, in which the distances from all points to the extracted contour are computed, and the line segments are inwardly offset to μ to align them with the actual surface location. The outermost line segments are then extended to the wall polygon to form a closed polygonal cross-section. Figure 3(c) shows the three elements generated from the refined cross-section: the sliced profile, created by extruding the contour with a thickness of 1 cm in the X-axis direction; the sliced patch, generated by sampling 4,000 points from the interior surface; and the end cap, generated by filling the interior of the contour with a mesh.

Figure 4 illustrates the procedure for generating 3D CAD models of components using the proposed algorithm. As shown in Figure 4(a), the segmented point cloud data of each class is projected onto the YZ plane along with its corresponding sliced patch to obtain a 2D point cloud data. For components arranged in two rows, such as the deviator block and top temporary PT bar anchor block, the left and right sides are separated using DBSCAN-based clustering (Figure 4(b)). Figures 4(c), (d), and (e) show the generated sliced patch for each component. The sliced patch is slid along the X-axis in increments of 1 cm, corresponding to the thickness of the sliced profile (Figure 4(f)). The sliding range is defined as the X-axis extent of the clustered segment with a 1 m padding on each side. At each position, the number of points $C(\tau)$ within a threshold distance τ between the sliced patch and the segmented point cloud data is calculated. $C(\tau)$ is defined by the following function:

$$C(\tau) = \sum_{i=1}^{N} 1[d(a_i, B) \leq \tau] \quad (1)$$

Where N is the number of points in the sliced patch, a_i is the i-th point in the sliced patch, B is the segmented point cloud, $d(a_i, B)$ is the distance from a_i to the nearest point in B, and $1[\cdot]$ is an indicator function that returns 1

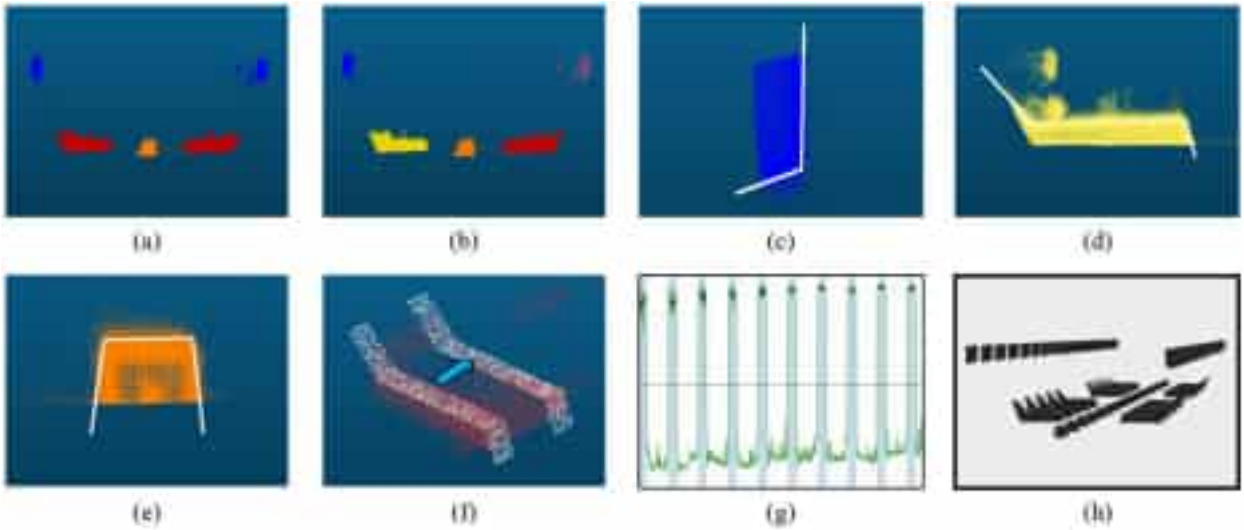


Figure 4. Sliced patch-based 3D reconstruction algorithm; (a) projection of the segmented point cloud onto the YZ plane, (b) left–right separation of the deviator block and top temporary PT bar anchor block using DBSCAN clustering, (c) generated sliced patch of the left top temporary PT bar anchor block, (d) generated sliced patch of the right deviator block, (e) generated sliced patch of the bottom temporary PT bar anchor block, (f) sliding of the sliced patch along the longitudinal direction, (g) binarization of the position– $C(\tau)$ graph using Otsu's algorithm, and (h) 3D reconstruction of individual components.

if the condition is satisfied and 0 otherwise. In this paper, τ was set to 4 cm.

This yields a position– $C(\tau)$ graph along the X-axis. As shown in Figure 4(g), the graph exhibits high point counts when the sliced patch passes through regions where actual components are present. To separate regions where components are present from those where they are absent, the graph is binarized using Otsu's algorithm. Sliced profiles are placed at positions along the sliding axis identified as containing components. Since components composed only of sliced profiles form an open shape at both ends, end caps are added to complete a closed solid geometry. Figure 4(h) shows the 3D reconstruction results of individual components. This algorithm is applied to the deviator block, top temporary PT bar anchor block, and bottom temporary PT bar anchor block. The final combined model is saved as a 3D CAD file.

3 Experiments and Results

3.1 Datasets

To validate the proposed method, interior point cloud data were acquired from three concrete box-girders. A total of 10 interior sections were scanned: 8 from Girders A and B, which contain 18 deviator blocks each, and 2 from Girder C, which contains 20 deviator blocks. In Girder A, deviator blocks are distributed with 10 on the east side and 8 on the west side relative to the cross-sectional center, while Girder B has the opposite distribution with 8 on the east and 10 on the west. Girder C has 10 deviator blocks on each side, totaling 20.

The acquired point cloud data underwent reprocessing. For each dataset, roll and pitch correction was performed, the longitudinal direction of the girder was aligned parallel to the X-axis, and wall points were removed. For datasets with complete boundary contours, wall removal was performed using the proposed wall extraction algorithm — specifically, two datasets from Girder A and one from Girder C. For the remaining datasets, wall removal was performed manually. All other preprocessing steps were performed manually.

For training SuperPoint Transformer, the 10 datasets were split into training, validation, and test sets. The training set consists of 6 datasets: 5 from Girder A and 1 from Girder B. The validation set includes 1 dataset from Girder A and 1 from Girder B. The test set consists of 2 datasets from Girder C, which has a different geometry, to evaluate generalization performance on girders with a different number of deviator blocks. Each point cloud data was manually labeled with the five classes.

Model training was conducted on a workstation running Ubuntu 20.04, equipped with an NVIDIA GeForce RTX 3090 GPU and an Intel Xeon Gold 6240M CPU. The parameters for SuperPoint Transformer were set as follows: voxel size of 1.5 cm, tiling of 2, and batch size of 2. Regularization strength values were set to (0.1, 0.2, 0.3). The maximum number of training epochs was set to 600.

3.2 Experimental Results

3.2.1 3D Semantic Segmentation Results

Table 1 presents the 3D semantic segmentation results. The evaluation metrics used are Intersection over Union (IoU), precision, recall, and F1 score, defined as follows. Here, TP, FP, and FN denote the number of true positive, false positive, and false negative points, respectively.

$$IoU = TP/(TP + FP + FN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

$$Recall = TP/(TP + FN) \quad (4)$$

$$F1\ score = \frac{2(Precision \cdot Recall)}{Precision + Recall} \quad (5)$$

The segmentation achieved a mean IoU (mIoU) of 94.1% and a mean F1 score of 96.9%. Figure 5 shows the segmented test data.

Table 1. Segmentation performance for concrete box-girder interior point cloud data.

Class	IoU	Precision	Recall	F1 score
Deviator block	82.5%	84.2%	97.6%	90.4%
Tendon	97.9%	99.8%	98.1%	98.9%
Top temporary PT bar anchor block	92.7%	92.7%	100.0%	96.2%
Bottom temporary PT bar anchor block	98.5%	99.2%	99.3%	99.2%
Cable tray	98.9%	99.0%	99.9%	99.5%
Average	94.1%	95.0%	99.0%	96.9%

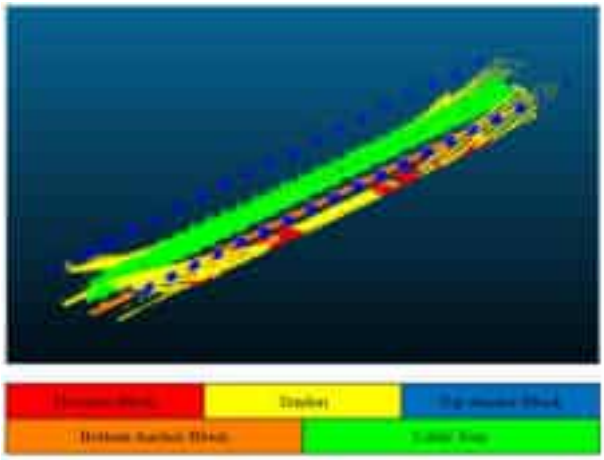


Figure 5. Segmentation result.

3.2.2 3D CAD Model Generation Result

Sliced profiles, sliced patches, and end caps for each component were generated from the segmented point cloud and applied to the 3D reconstruction. Table 2 presents the mean mesh-to-point cloud distance. For each point in the segmented point cloud, the distance to the nearest polygon of the generated mesh was calculated. The proposed method achieved mean distances of 0.025 m against the point cloud data. Figure 6 shows the 3D CAD model generation results.

Table 2. Mean mesh-to-point distance results for the generated 3D CAD models compared with the point cloud.

Class	Mesh-to-point cloud distance
Wall	0.007 m
Deviator block	0.002 m
Top temporary PT bar anchor block	0.080 m
Bottom temporary PT bar anchor block	0.010 m
Average	0.025 m



Figure 6. Generated 3D CAD model of the concrete box-girder interior.

4 Discussion

4.1 Point Cloud Segmentation

The point cloud segmentation results showed that the Top temporary PT bar anchor block, Bottom temporary PT bar anchor block, and Cable tray achieved high IoU values of 92.7%, 98.5%, and 98.9%, respectively.

These classes have minimal physical connections with other components in the concrete box-girder interior point cloud after wall removal, resulting in clear boundaries between classes. This is considered to have contributed to stable segmentation performance. Furthermore, the model demonstrated high performance on test data with a different deviator block arrangement compared to the training data, indicating potential for generalization.

In contrast, the deviator block showed relatively lower IoU values of 82.5%. As shown in Figure 7(a), tendons in contact with deviator blocks were found to be misclassified as deviator blocks. Since tendons structurally pass through deviator blocks and are physically in contact with them, tendon points were misclassified as deviator blocks. Additionally, Figure 7(b) shows cases where tendons were misclassified as cable tray. This is attributed to the tendons extending upward and approaching the locations where the cable tray is present.

4.2 3D CAD model generation

The proposed sliced patch-based 3D reconstruction algorithm achieved mean mesh-to-point cloud distances of 0.025 m against the point cloud and successfully generated 3D CAD models by separating components at the instance level from semantically segmented point cloud data. Notably, the proposed method reconstructed 3D CAD models using only the point cloud data.

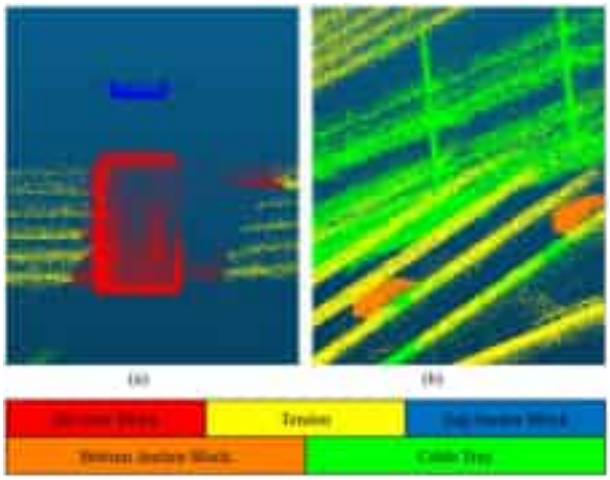


Figure 7. Examples of misclassification in point cloud segmentation results; (a) tendon misclassified as a deviator block, and (b) tendons misclassified as a cable tray.

However, cases were observed where the generated 3D CAD models differed from the actual geometry. Figure 8(a) shows a case where multiple 3D CAD models were generated for a single component. This occurred due to variability in the position– $C(\tau)$ graph constructed from nearest-neighbor relationships between the patch and point cloud, causing the Otsu-based binarization to segment a single continuous component into multiple fragments. Figure 8(c) shows a case where a 3D CAD model was generated for only a partial region of a component rather than the entire geometry. This is attributed to sparse points on all surfaces of the segment during scanning, resulting in missing points in certain regions.

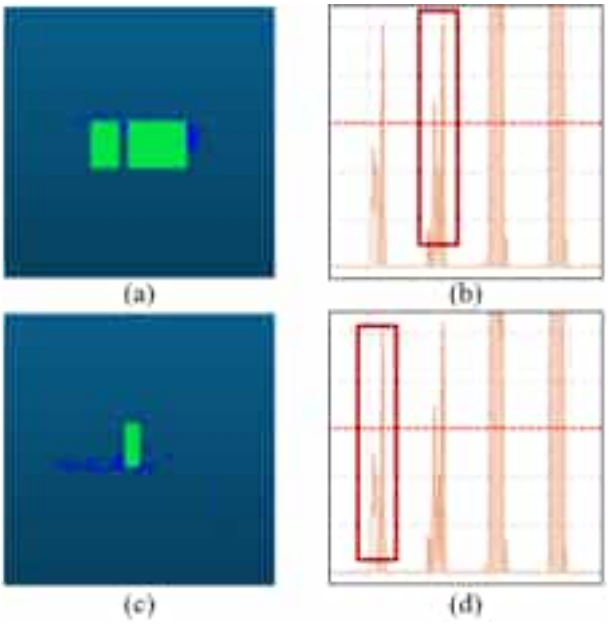


Figure 8. Examples of incorrect 3D reconstruction; (a) a top anchor block fragmented into two parts, (b) a segment split by Otsu algorithm-based binarization (highlighted by the red box), (c) a model generated only on the right-side surface, and (d) an excessively thin segment remaining after Otsu algorithm-based binarization (highlighted by the red box).

4.3 Future Development

This paper presented a procedure for segmenting concrete box-girder interior point cloud data using SuperPoint Transformer and reconstructing 3D CAD models from the segmented results. The proposed framework was applied to actual bridge girder data, demonstrating its potential applicability to concrete box-girder inspection by performing semantic segmentation and 3D reconstruction on girder data with different interior environments. However, several aspects require further improvement.

First, the proposed framework does not include a 3D reconstruction methodology for tendons. Considering the significance of a tendon in a concrete box-girder, it is necessary to develop a parametrization method for tendons and integrate it into the proposed framework.

Second, the validation of the proposed methodology was limited to data acquired from a single bridge. Therefore, it is necessary to verify the generalization performance on bridges with different internal structures. Additional concrete box-girder interior data from multiple bridges with varying interior geometries, component configurations, and scan qualities should be acquired, and performance should be validated by applying the same pipeline to enhance practical applicability.

Third, several steps in the proposed pipeline were performed manually. Roll and pitch correction and axis alignment, and diaphragm removal were conducted manually, and the parameters for SOR-based noise removal, the alpha shape algorithm, and RANSAC based line extraction were heuristically determined. Automating these procedures is necessary.

Fourth, the proposed sliced patch-based algorithm is applicable only to components that are linear and have a constant cross-sectional profile along their length. For components with varying cross-sections or curved geometries, alternative reconstruction approaches would be required.

5 Conclusion

This paper presents a framework for semantic segmentation of point cloud data acquired from the interior of a concrete box-girder and 3D CAD model reconstruction of individual components. The proposed methodology was validated on actual concrete box-girder data, achieving a semantic segmentation performance of 94.1% mIoU. The mean mesh-to-point cloud distance was 0.025 m against the point cloud data, demonstrating the feasibility of the proposed approach. By addressing the identified limitations and refining the methodology in future work, this paper is expected to contribute to the establishment of 3D data-driven inspection systems for interiors of a concrete box-girder.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. RS-2018-NR031077). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00457308). ※ MSIT: Ministry of Science and ICT. This research was conducted with the support of i-thetto Co., Ltd. and Korea Expressway Corporation.

References

[1] Seoul Metropolitan Government. Announcement of defect causes and PSM bridge inspection results for Jeongneungcheon Overpass On-line: https://news.seoul.go.kr/safe/archives/46885, Accessed: 14/01/2026.

[2] Seoul Facilities Corporation. Damage Investigation and Repair Progress of Cheongdam 1st Bridge. On-line: https://www.sisul.or.kr/open_content/main/introduce/innovation5.jsp, Accessed: 14/01/2026.

[3] Fang Y., Mitoulis S.-A., Boddice D., Yu J., and Ninic J. Scan-to-BIM-to-Sim: Automated reconstruction of digital and simulation models from point clouds with applications on bridges. *Results in Engineering*, 25:104289, 2025.

[4] Jing Y., Sheil B., and Acikgoz S. Segmentation of large-scale masonry arch bridge point clouds with a synthetic simulator and the BridgeNet neural network. *Automation in Construction*, 142:104459, 2022.

[5] Kim Y., Kim J., Kim J., and Kim H. Superpoint Transformer-Based Bridge Component Recognition Using UAV-Mounted LiDAR and Synthetic Point Cloud Generation. *Journal of Computing in Civil Engineering*, 39(4):04025043, 2025.

[6] Yang L., Lin Y.-C., Cai H., and Habib A. From Scans to Parametric BIM: An Enhanced Framework Using Synthetic Data Augmentation and Parametric Modeling for Highway Bridges. *Journal of Computing in Civil Engineering*, 38(3):04024008, 2024.

[7] Shu J., Zeng Z., Li W., Zhou S., Zhang C., Xu C., and Zhang H. Automatic geometric digital twin of box-girder bridge using a laser-scanned point cloud. *Automation in Construction*, 168(A):105781, 2024.

[8] Zhu Y., Diaw I., and Xiong W. Cable-stayed bridge all-round morphology identification based on 3D point cloud model. *Structure and Infrastructure Engineering*, 22(2):185–200, 2026.

[9] Xu W., Cai Y., He D., Lin J., and Zhang F. FAST-LIO2: Fast Direct LiDAR-inertial Odometry. On-line: https://arxiv.org/abs/2107.06829, Accessed: 14/01/2026.

[10] Wang Y. FAST_LIO_LC: The tight integration of FAST-LIO with Radius-Search-based loop closure module. On-line: https://github.com/yanliang-wang/FAST_LIO_LC, Accessed: 13/01/2026.

[11] Mahmoud M., Chen W., Yang Y., and Li Y. Automated BIM generation for large-scale indoor complex environments based on deep learning. *Automation in Construction*, 162:105376, 2024.

[12] Robert D., Raguet H., and Landrieu L. Efficient 3D Semantic Segmentation with Superpoint Transformer. On-line: https://arxiv.org/abs/2306.08045, Accessed: 13/01/2026.

[13] ASBI-Assoc. Design Manual for Concrete Segmental Bridges, 1st Edition. On-line: https://asbi-assoc.org/wp-content/uploads/2025/10/ASBI-Design-Manual-for-Concrete-Segmental-Bridges-1st-Edition-2.pdf, Accessed: 13/01/2026.

[14] Kim J., Kim J., Kim Y., and Kim H. 3D reconstruction of large-scale scaffolds with synthetic data generation and an upsampling adversarial network. *Automation in Construction*, 156:105108, 2023.

Incremental IFC updates with ‘why’ and ‘how’: A case of IFC extension for LLM agent-enhanced simulation-based optimization for building energy efficiency

Lingming Kong, Qianyun Zhou, Fan Xue

Department of Real Estate and Construction, University of Hong Kong, Hong Kong SAR, China

u3009509@connect.hku.hk qianyun8@connect.hku.hk xuef@hku.hk

Abstract -

Simulation-based optimization (SBO) for energy-efficient building design exemplifies a typical design-change process that remains disconnected from open BIM-enabled collaboration. The gap stems from IFC’s limited capacity to represent the information generated by SBO. To overcome the limitations, this study proposes a novel IFC schema extension for LLM Agent-enhanced SBO, named *IfcSboExtension*, that enriches incremental IFC updates with contextual metadata capturing not only what changed, but also why the change occurred and how it was derived. The proposed extension model enables exchangeable and traceable SBO iterations by embedding decision rationale and methodological provenance directly into incremental IFC updates. The feasibility and value of the extension are demonstrated through a case study on building energy efficiency optimization, in which an LLM agent enhances the interpretability of design changes.

Keywords –

Building information modelling (BIM), Industry Foundation Classes (IFC), Incremental IFC updates, Simulation-based optimization (SBO), Large language model (LLM)

1 Introduction

Building information modelling (BIM) is recognized as a digital integration of information throughout a building lifecycle, and increasingly mandated in the AEC (Architecture, Engineering, and Construction) industry [1]. BIM-based collaboration has progressively grown to level 2, namely “open BIM”, which signifies a fully integrated, collaborative environment for project stakeholders through federated discipline models and open exchange standards [2]. Industry foundation classes (IFC), and ISO 16739-1:2018 standard, are *per se* open BIM that provides a vendor-neutral data format for seamless data exchange between various BIM applications [3].

However, the scalability and adoption of IFC are constrained by the monolithic, file-based exchange paradigm. In practice, BIM stakeholders receive the modified model as a new monolithic file from collaborators and check manually to identify the potential impact on their own disciplines, which is intensive in labor and prone to errors.

IFC 5, the upcoming modular version, plans to implement incremental IFC updates to address these issues [4]. From a theoretical perspective, the transition toward incremental IFC updates aligns with the process-centric knowledge philosophy [5], which emphasizes that construction knowledge is dynamic, iterative, and context-sensitive regarding knowledge dissemination among project partners. Consequently, the modular architecture of IFC 5 is expected to better support both operational efficiency and effective knowledge management in construction projects.

Simulation-based optimization (SBO) is a typical design-change process in which design parameters are iteratively adjusted, leading to modifications to model design. Recent studies have leveraged large language models (LLM) to promote the interpretability of traditional black-box SBO workflows [6]. However, the information generated by the LLM-enhanced SBO remains neither exchangeable nor traceable within the context of sustainable, collaborative design based on open BIM. To address the gap between LLM-enhanced SBO and open BIM, this paper proposes an IFC extension for LLM-enhanced SBO, which explicitly answers ‘why’ and ‘how’ in addition to ‘what’ to enrich such incremental IFC updates.

The remainder of this paper is organized as follows: Section 2 reviews the state-of-the-art research on incremental IFC updates and simulation-based optimization. Section 3 introduces the proposed IFC schema extension for incremental update. Section 4 demonstrates the implementation of the extension

through a case study employing large language model (LLM) agents for building energy efficiency optimization. Section 5 provides the conclusion.

The primary contributions of this study are twofold. First, the proposed IFC schema extension emphasizes that the 'why' and 'how' dimensions are essential to the 'what' of incremental IFC updates. Second, the feasibility of the proposed extension was validated through a real-world project.

2 Literature Review

2.1 IFC and schemas for changes

IFC was originally conceived as a monolithic, object-oriented schema based on the EXPRESS data modelling language [7]. As an object-oriented data schema, the IFC schema contains several classes, namely "entities", and serves as "templates" to describe building information. The semantic meaning and the implementation of entities are generalized through inheritance mechanisms. Each entity declares a set of local attributes, in addition to the global attributes inherited from its supertype. There are two types of IFC entities, namely rooted and non-rooted entities, which contain or do not contain a globally unique identifier (GUID). In practice, the instantiated IFC entities are transformed into ASCII text rows in the text-based IFC-SPF (STEP Physical File) formatted file [8]. An IFC-SPF file is exchanged in a snapshot-based exchange paradigm, which means each IFC-SPF file represents a complete state of a model at a given time, regardless of actual changes. This exchange paradigm leads to inefficiencies in data transmission, version control, and change impact analysis for decision making in the collaborative, multi-disciplinary environment [9].

To overcome the limitations of holistic IFC file exchange, recent research and standardization efforts have shifted toward change-based IFC exchange approaches. In the literature, three categories of difference (or delta) computation methods have been developed to capture changes between two IFC models, which are (1) GUID-based; (2) flattening-based; and (3) graph-based approaches.

The GUID-based approach recognizes two IFC instances as identical if they share the same GUID, which is widely adopted by commercial BIM platforms such as Autodesk Revit, Navisworks and Graphisoft ArchiCAD. However, a significant limitation of this approach is that GUIDs may not remain consistent across different BIM authoring tools, rendering GUID-based comparison unreliable. The flattening-based approach eliminates randomly referenced identifier numbers (i.e., the "#"-IDs) in IFC files. The "flattening" process is to replace all the "#"-IDs with specific instant IDs, such as content-based hash ID [10][11]. Therefore, flattening-based approaches enable deterministic textual comparison.

Both the GUID-based and flattening-based methods are fundamentally text-oriented. In contrast, the graph-based approach leverages graph-based algorithms to compute semantic differences between two graph-represented IFC models. There are two main steps involved in the graph-based approach. First, the graph representation of IFC models is developed, such as RDF (resource description framework) graph and directed property graphs. Then, graph-based algorithms, such as graph transformation [12], are implemented to capture the changed part of two IFC models.

2.2 Knowledge engineering applications of changes in construction

The dynamic theory of knowledge creation developed by Nonaka and Takeuchi [13], namely the SECI model, posits that knowledge is created, transferred, and applied through the creative tension between tacit and explicit know-how. The SECI model reveals a significant source of knowledge, that knowledge is generated from continuously combined, refined, and reconfigured explicit assets. This definition aligns with the iterative development processes of construction projects [1]. Accordingly, changes in a construction project, such as design iterations, site conditions, and stakeholder feedback, are recognized as a knowledge source which has great influence on successful project delivery [14].

Knowledge engineering (KE) is the process of designing, building, and maintaining knowledge-based systems by identifying relevant domain knowledge, representing it in formal structures (such as ontologies, or knowledge graphs), and implementing reasoning mechanisms that can use the knowledge to make inferences, decisions, or recommendations [15]. A core KE approach is the development of domain ontologies that explicitly define change types, actors, causes, and impacts. Early work by Beetz et al. [16] transforms the EXPRESS schema of IFC into an OWL ontology, namely IfcOWL, and paves the way to change detection as ontology alignment. The following work of Oraskari and Törmä [17] introduces a signature algorithm for RDF diff computation. Furthermore, a domain ontology for pumped storage hydropower system was developed by Zhang et al. [18] to support design changes management.

IFC also can be serialized into various data structures, such as ifcJSON and labelled property graph, to support algorithmic reasoning for change detection. For example, A semantic differential transaction (SDT) approach and the following traceable SDT approach were developed based on the JSON-format IFC to capture the semantic changes through structure matching [10][11].

2.3 Simulation-based optimization as a typical change

Simulation-based optimization (SBO) is widely applied to building design by iteratively tuning design parameters towards sustainable objectives [19]. The generic SBO workflow comprises iterative simulations that are evaluated through objective functions defined in optimization algorithms. SBO constitutes a typical design-change process in which the design parameters vary across iterations, producing corresponding modifications in design. For example, energy performance optimization is a key task in SBO for early stage building design, aiming to reduce the environmental impact of future buildings. SBO involves tuning design parameters such as window-to-wall ratio and building materials. In each iteration, these parameters are adjusted to gradually improve energy performance.

Traditional SBO is challenged by high computational complexity and low interpretability. A recent study by Zhou and Xue [19] introduces an Automatic Information Gain-guided Convergence (AIGGC) method to refine critical design parameters through interpretable information gain analysis. Subsequent research integrated the large language model (LLM) to further enhance the interpretability and efficiency of complex SBO [6].

However, the information produced by LLM-Agent enhanced SBO is neither exchangeable nor traceable within the context of sustainable, collaborative design based on open BIM. Existing literature characterizes BIM-supported SBO as a one-way workflow that portions of the required parameter data are extracted from BIM [20], yet the SBO iteration results are not embedded back into BIM for collaborative use. The limitation arises for two reasons. First, contemporary SBO practices are predominantly implemented on visual programming platforms such as Grasshopper, which do not natively support IFC. Secondly, the IFC schema lacks the capacity to represent the information generated by LLM-Agent enhanced SBO, such as design parameters, constraints, and convergence analysis. The proposed IfcSboExtension uniquely encode the decision rationale, optimization provenance, and LLM interaction context to support change-based IFC exchange for traceable and interpretable collaborative design.

3 IFC extension for LLM agent-enhanced SBO

To extend IFC's capability of representing information of LLM Agent-enhanced SBO, an IFC extension model named '*IfcSboExtension*' was proposed. The extended schema illustrates 'what' data should be involved, 'why' the IFC schema should be extended, and 'how' to capture the incremental IFC changes within the LLM Agent-enhanced SBO. A general IFC extension workflow is illustrated in Figure 1, which constitutes three steps: (1) Information requirements in LLM Agent-enhanced SBO, (2) *IfcSboExtension* model development, and (3) verification and implementation for incremental IFC updates.

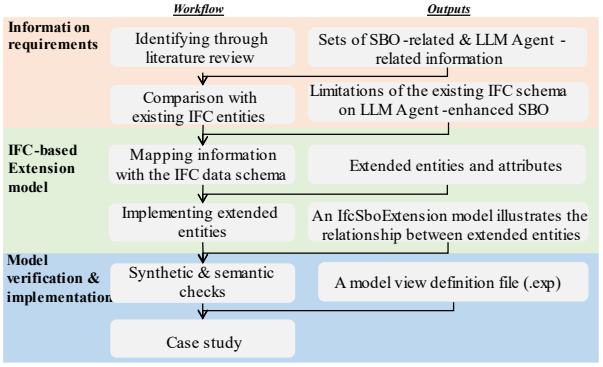


Figure 1. IFC extension workflow

The first step answers 'what' and 'why' the IFC schema should be extended to LLM Agent-enhanced SBO. Table 1lists the crucial data of two categories: (1) SBO-related data such as design parameters and optimization constraints, as well as (2) LLM Agent-related data such as prompts and structured responses.

A meticulous comparison with the existing entities within IFC4 ADD3 revealed that both SBO-related and LLM Agent-related information required extension to enable seamless data exchange. For example, *IfcProcess* is defined in existing IFC schema to represent one individual activity that has sequence relationships with other processes. Three subtypes of *IfcProcess*, namely *IfcTask*, *IfcEvent*, and *IfcProcedure* do not contain the semantics such as parameters, constraints, and objective functions. One compromise approach supported by IFC schema is identifying specific property sets, such as *Pset_SBO*. However, this approach converts SBO semantics to unstructured strings and leads to semantic ambiguity and semantic loss.

3.1 The *IfcSboExtension* model

According to the required data listed in Table 1, three rooted entities and two non-rooted entities were extended align with the IFC4 ADD3 schema. As Figure 2 shows, the extended entity *IfcSboWorkflow* and *IfcAiAgentInteraction* are two subtypes of the abstract *IfcControl*, and *IfcOptimizationIteration* is a subtype of the abstract *IfcProcess*. The reason lies in the semantic definition of *IfcControl* and *IfcProcess*. In IFC4 ADD3, *IfcControl* represents all concepts that control or constrain the utilization of products, processes, or

resource, whereas *IfcProcess* is defined as one individual activity or event. According to the definition, *IfcControl* and its subtypes are suitable for representing a workflow, whereas *IfcProcess* and its subtype focus on one specific process. In LLM Agent-enhanced SBO, SBO is a workflow that contains a set of 'simulation-optimization' iterations, and the interaction between stakeholders and LLM Agent is also recognized as a set of communications rather than single process. One optimization iteration aligns with the definition of *IfcProcess* which transforms input in output. *IfcObjectiveFunction* and *IfcSboOptimizationDirection-Enum* are two non-rooted entities extended at the resource layer. These two non-rooted entities cannot exist independently but can only exist if referenced by the rooted entity *IfcSboWorkflow*. Due to page limits, only the extended *IfcSboWorkflow* entity and relevant attributes is present in Figure 3, which illustrates the EXPRESS representation of *IfcSboWorkflow* to ensure both machine and human readability.

Table 1. Crucial data in LLM Agent-enhanced SBO

	Crucial Data	Data type (in IFC)	Example
SBO-related	SBO parameters	SET [1: *n*] of REAL	(0.45,0.45,1.1,2,5.0733)
	SBO constraints	STRING	$\left\|FAR_{ori} - FAR_{opt}\right\| \leq 0.05$ (FAR for floor area ratio)
	SBO objectives	STRING	EUI = 112.684 $kWh/m^2 \cdot yr$
	SBO results	STRING or BINARY	JSON or CSV
LLM Agent-related	LLM prompts	STRING	*"Please identify which design parameters most influence the f(S)?"*
	LLM responses	STRING or BINARY	Response text or Log file

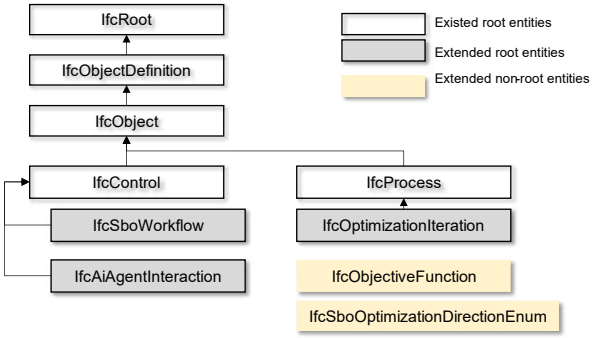


Figure 2. Extract of the IFC LLM Agent-enhanced SBO extension with new entities

3.2 Model verification

For verification, the official IFC documentation toolkit *IfcDoc* and a test software platform APSTEX were employed. The verification procedure involves two steps: syntactic verification and semantic verification. As summarized in Table 2, syntactic verification ensures compliance with the formal grammar of the EXPRESS language. The extended IFC entities were documented using *IfcDoc* to verify their compliance with EXPRESS language grammar. The exported use-case model view definition, designated as *IfcSboExtension*, incorporates 19 existing IFC entities and 5 newly defined IFC entities.

In the second stage, semantic verification was performed by compiling the extended *IfcSboExtension* model into Java classes using the Apstex IFC Framework. Additionally, accessor methods of explicit, derived, and inverse attributes of the extended IFC entities are developed. Critically, domain rule enforcement methods were developed, and corresponding unit tests were designed to verify the *IfcSboExtension* model based on the created Java classes.

```
ENTITY IfcSboWorkflow
  SUBTYPE OF (IfcControl);
    OptimizationType : IfcLabel;
    AlgorithmName   : OPTIONAL IfcLabel;
    StartTime       : OPTIONAL IfcDateTime;
    EndTime         : OPTIONAL IfcDateTime;
    Status          : IfcLabel;
INVERSE
    HasVariables    : SET [1:?] OF IfcPropertyBoundedValue FOR BelongsToWorkflow;
    HasObjectives   : SET [1:?] OF IfcObjectiveFunction FOR BelongsToWorkflow;
    HasIterations   : SET [0:?] OF IfcOptimizationIteration FOR BelongsToWorkflow;
    GeneratedFromLog : OPTIONAL IfcAiAgentInteraction FOR TriggersWorkflow;
END_ENTITY;
```

Figure 3. The extended IfcSboWorkflow entity and relevant attributes

Table 2. Model verification procedure, criteria, and tools

Verification procedure	Criteria	Tools
Syntactic checks	**Formal compliance with the grammar defined in ISO 10303-11(EXPRESS):**	*IfcDoc*

	• Notation conventions **EXPRESS-compliant declaration of** • Entities (“entity data types”) and types (“defined data types”), • Enumerations (“enumeration data types”), • Subtype and supertype constraints, • Explicit, derived, and inverse attributes of entities	
Semantic checks	**Compilation of the LLM Agent-enhanced SBO schema into Java classes for semantic checking:** • Java classes for entity, type, enumeration considering the inheritance hierarchy • accessor methods for explicit, inverse, and derived attributes • methods to check domain rules and uniqueness rules	Apstex IFC Framework

```
public void whereRuleCheck(WhereRuleCheckResultList list) {
    super.whereRuleCheck(list);
    List<IfcRelAssignsToSboWorkflow> rels =
this.getInverse("HasObjectives");
    boolean applicableObjectivesCheck = true;
    if (rels != null) {
        for (IfcRelAssignsToSboWorkflow rel : rels) {
            List<Object> relatedObjects = rel.getRelatedObjects();
            if (relatedObjects != null) {
                for (Object obj : relatedObjects) {
                    If
(!BUILTINFUNCTIONS.TYPEOF(obj).contains("IFCSBOMODEL.IFCSBOOB
JECTIVEFUNCTION")) {
                        applicableObjectivesCheck = false; break; } } } } }
    if (!applicableObjectivesCheck) {
        reportError(list, "IfcSboWorkflow: ApplicableObjectives");
    }
}
```

(a)

```
// create a new SBO workflow entity
IfcSboWorkflow ifcSboWorkflow = new IfcSboWorkFlow();
// create a relationship entity
IfcRelAssignsToControl ifcRelAssignsToControl = new
IfcRelAssignsToControl();
ifcRelAssignsToControl.setRelatingContnrol(ifcSboWorkflow)
SET<IfcObjectDefinition> RelatedObjects = new SET<>();
ifcRelAssignsToControl.setRelatedObjects(RelatedObjects);
// add SBO workflow components to related objects
RelatedObjects.add(new IfcObjectiveFunction());
// add module group to SBO workflow must cause an error
RelatedObjects.add(new IfcGroup());
// execute the domain rule check
WhereRuleCheckList list = new WhereRuleCheckList();
ifcSboWorkflow.whereRuleCheck(list)
```

(b)

Figure 4. Examples of domain rule checking method and unit test for "HasObjectives" of the IfcSboWorkflow entity

Figure 4 presents an example of a domain rule checking method and its associated unit test. For efficient compilation, so-called “lambda expressions” were used to implement the domain rule checks. The example ‘whereRuleCheck’ is defined to check the applicable objectives of the extended entity *IfcSboWorkflow*. Corresponding unit test is shown in Figure 4(b). “HasObjectives” of the *IfcSboWorkflow* entity is shown in Figure 3, which represents a set of *IfcObjectiveFunction* instances are assigned to one *IfcSboWorkflow* instance. In the unit test, the process of adding a module group to *IfcSboWorkflow* results in a ‘False’, which indicates an incorrect state of the IFC test model.

4 A case study of LLM agent-enhanced SBO

To illustrate ‘why’ and ‘how’ in addition to ‘what’ of IFC incremental updates, an energy consumption optimization scenario is selected as a representative design-change case study. The design optimization of a modular integrated construction (MiC) project for a university student hostel located in Hong Kong is selected as a case. The developed LLM Agent for efficient and interpretable SBO is presented in the published work of Zhou and Xue (2025).

4.1 Case settings

In this case, the optimization objective was to minimize the energy use intensity (EUI), defined as the total annual energy consumption divided by the gross floor area ($kWh/m^2 \cdot yr$). As shown in Figure 5, the project features a symmetrical layout with fixed modules and movable modules. The optimization process focuses on three types of movable modules which vary in length, namely M_A, M_B, and M_C. Thus, module layout, window size and material properties are involved as design parameters shown in Table 3.

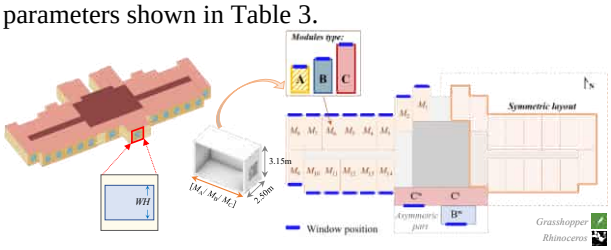


Figure 5. MiC case layout and schematic of module types (adopted from Zhou & Xue (2025))

4.2 Results

4.2.1 Schemas and sample changes

The implementation of the *IfcSboExtension* model using *IfcDoc* is presented in Figure 6 as an EXPRESS-G

diagram. Black rectangles represent rooted entities, whereas grey rectangles denote non-rooted entities. Entity's attributes – comprising own and inherited attributes - are separated by an ellipsis and categorized as either explicit or derived attributes (in black font) or inverse attributes (in light grey font). Relationships between the extended entities are connected through inverse attributes. For example, an *IfcOptimization-Iteration* is referenced by an *IfcSboWorkflow* through its inverse attributes '*HasIterations'*.

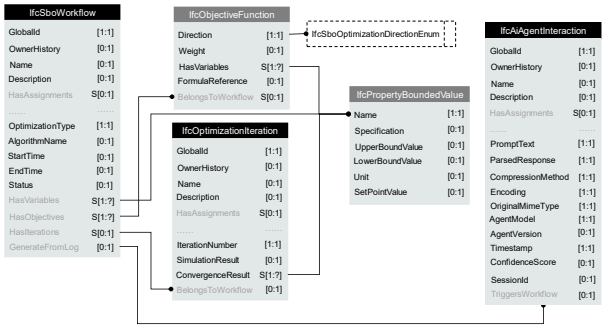


Figure 6. EXPRESS-G representation of *IfcSboExtension* model

The embedded LLM Agent-enhanced SBO semantics is illustrated in Figure 7. *IfcSboWorkflow* is assigned to a module through *IfcRelAssignsToGroup*, and *IfcAiAgentInteraction* is assigned to *IfcSboWorkflow* through *IfcRelAssignsToControl*. To illustrate the change of parameters during SBO iterations, parameters listed in Table 3 are selected. For instance, when an iteration increases the north window area from 4.33 m² to 4.49 m², embedded convergence semantics allow the owner to immediately identify this change as the primary driver for reducing Energy Use Intensity. This transparent exposure of the 'why' (energy optimization) and 'how' (sensitivity analysis) enables the owner to grasp the technical rationale instantly, facilitating an informed approval decision.

4.2.2 Sensitivity analysis

The text-based IFC-SPF file exhibits acute sensitivity to: (1) excessive line lengths, (2) special characters such as unescaped Unicode or JSON brackets, and (3) exponential file growth. Specifically, the extended *IfcAiAgentInteraction* records prompt and structured responses, which may comprise multi-megabyte strings, that can significantly decrease data exchange efficiency.

In this study, two mitigation strategies were evaluated: (1) compressing the structured response using the Gzip algorithm and encoding the compressed result as Base64 strings; and (2) storing LLM Agent prompts and structured responses in external files referenced via *IfcURIReference*. Gzip and Base64 are open standards for data compression and encoding, respectively. Metadata describing the compression and encoding algorithms are embedded in *IfcAiAgentInteraction* to ensure self-descriptiveness. URIs for external files should conform to standard URI schemes (e.g., HTTP and HTTPS), and the corresponding MIME type should be explicitly specified.

The evaluation was performed over 30 iterations of LLM Agent-enhanced SBO, with results presented in Table 4. The result indicates that the 'Gzip + Base64' strategy increased the IFC file size by a maximum of 2.47% and an average of 1.91%. The *IfcURIReference* strategy effectively eliminates the impact of LLM Agent responses on IFC file size.

Table 3. Design parameters for MiC parametric design

Design parameters	Parameter type	Unit	Parameter domain intervals	Precision
MiC Module Type	Categorial	-	$[M_A, M_B, M_C]$	-
Window-to-Wall Ratio (WWR_N, WWR_S)	Continuous	-	$0.10 \leq WWR_N, WWR_S \leq 0.60$	0.01
Window Height (WH_N, WH_S)	Discrete	M	$1.0 \leq WH_N, WH_S \leq 2.5$	0.1
Window U-value (U_w)	Continuous	$W/(m^2 \cdot k)$	$0.50 \leq U_w \leq 5.78$	0.01
Window Solar Heat Gain Coefficient (g)	Continuous	-	$0.100 \leq g \leq 0.775$	0.001
Window Visible Transmittance (τ_v)	Continuous	-	$0.300 \leq \tau_v \leq 0.881$	0.001

4.2.3 User acceptance and comments

The extended *IfcSboExtension* model was implemented on a real MiC project for a university student hostel located in Hong Kong. Stakeholders of the MiC project gave positive comments. For example, the owner of the project claimed that "The LLM Agent platform using extended IFC has transformed our design optimization. The black box SBO is interpretable, traceable and exchangeable through open BIM. This enables non-professional stakeholders to understand the design intentions and make decisions. Technical users emphasized that "Through the adoption of open BIM, we

can work smoothly between different software and disciplines with an open standard, which helps us streamline and automate many manual processes in LLM Agent and SBO. This has greatly increased our productivity by eliminating misunderstanding of design changes that happened in the black box SBO process.

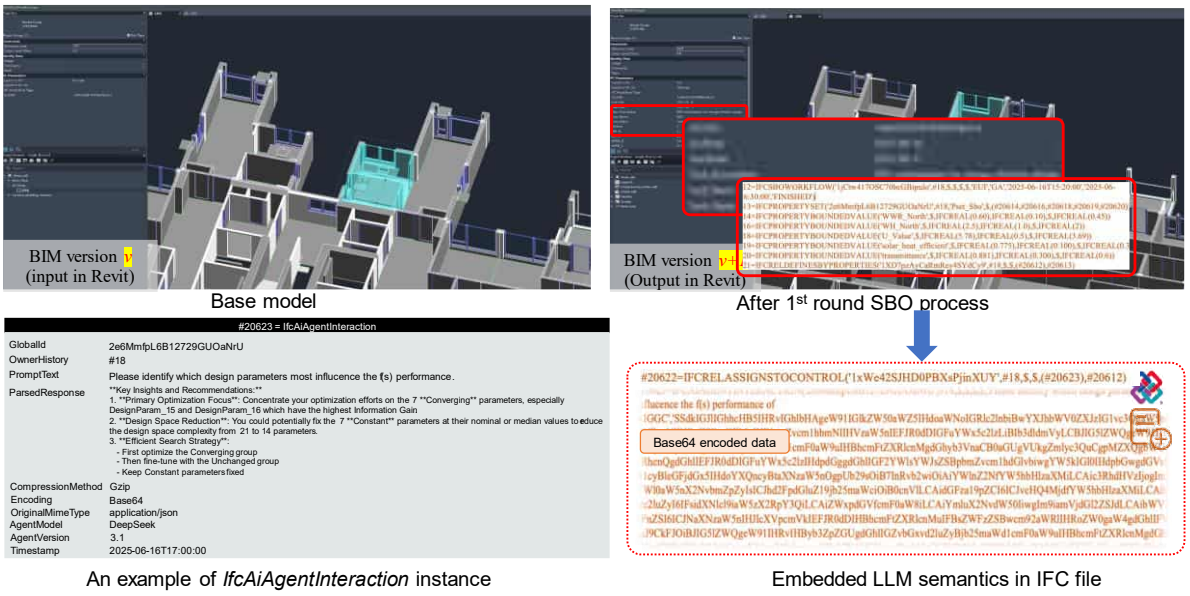


Figure 7. Embedded LLM Agent-enhanced SBO semantics in IFC

Table 4. File size comparison of two strategies

Strategy	Original	Gzip & Base64		External Ref.	
	(KB)	(KB)	Δ (%)	(KB)	Δ (%)
Minimal	1939	1946	+0.36	1940	+0.05
Maximum	1939	1987	+2.48	1941	+0.10
Average	1939	1976	+1.91	1941	+0.10

5 Discussion and Conclusion

The snapshot-based exchange paradigm of IFC introduces inefficiencies in data transmission, version control, and change impact analysis for making decisions in the collaborative, multi-disciplinary environment. To address these limitations, a modular version of IFC 5 is under development to support incremental IFC updates.

Simulation-based optimization (SBO) for energy-efficient building design exemplifies a typical design-change process which remains disconnected from open BIM-supported collaboration. The gap stems from IFC's poor capability of representing information generated by SBO. To bridge the gap and illustrate the full provenance of such design changes, including the 'why', 'how', and 'what' in incremental IFC updates, this study proposes a novel IFC schema extension for the LLM Agent-enhanced SBO.

The contributions of this study are twofold:

(1) An *IfcSboExtension* model was proposed and verified to enhance the IFC's capacity to represent the LLM Agent-enhanced SBO process.

(2) The feasibility and practical value of the proposed extension were validated through a MiC case study.

The case study yields two key findings:

(1) All information generated by the LLM Agent-enhanced SBO is embedded within the extended IFC entities, involving three rooted IFC entities: *IfcSboWorkflow*, *IfcAiAgentInteraction*, and *IfcOptimizationIteration*; and two non-rooted IFC entities: *IfcObjectiveFunction* and *IfcSboOptimizationDirectionEnum*.

(2) The changes of design parameters throughout the LLM Agent-enhanced SBO process can be captured as incremental IFC updates.

Nevertheless, several limitations remain that require further investigation:

(1) The extended *IfcAiAgentInteraction* faces a trade-off between semantic integrity and storage efficiency. If LLM Agent prompts and responses are fully encoded as structured data within *IfcAiAgentInteraction*, individual lines in the IFC file may exceed parser limits, resulting in increased file size and reduced exchange efficiency. Conversely, compressing the data (e.g., via Gzip) or referencing external files avoids file bloat, but incurs semantic opacity, requires stakeholders to employ additional tools to interpret the compressed string or external files.

(2) Semantic changes of the extended IFC entities

are not supported by the mainstream commercial BIM platforms (e.g., *Revit*) or open-source IFC parsing tools (e.g., *IfcOpenshell*). This limitation arises because the proposed IFC extension has not yet been submitted to buildingSMART for formal standardization – a necessary step for broader adoption.

Acknowledgement

The work presented in this paper was supported by the Hong Kong Research Grants Council (Nos. C7080-22GF and 17200221).

References

[1] C. Eastman. BIM handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors. John Wiley & Sons, 2011.

[2] M. C. Georgiadou. An overview of benefits and challenges of building information modelling (BIM) adoption in UK residential projects. *Construction Innovation*, 19(3):298-320, 2019.

[3] Y. Yu, S. Kim, H. Jeon, and B. Koo. A Systematic Review of the Trends and Advances in IFC Schema Extensions for BIM Interoperability. *Applied Sciences*, 13(23):12560, 2023.

[4] buildingSMART International. IFC 5 Core Consultation. Online: https://www.buildingsmart.org/ifc-5-core-consultation/, Accessed: 01/25/2026.

[5] G. Mentzas, D. Apostolou, R. Young, and A. Abecker. Knowledge networking: a holistic solution for leveraging corporate knowledge. *Journal of Knowledge Management*, 5(1):94-107, 2001.

[6] Q. Zhou and F. Xue. L-SCOPE: An LLM-assisted interactive platform for efficient building design optimization. *Proceedings of the International Conference on Smart and Sustainable Built Environment (SASBE)*. 96-105. https://isasbec.abc2.net/index.php/sasbe/article/view/2554/9

[7] International Organization for Standardization. ISO 16739-1:2024 Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries Part 1: Data schema. Online: https://www.iso.org/standard/84123.html, Accessed: 01/30/2026.

[8] H. Liu, G. Gao, and M. Gu. A parallel IFC normalization algorithm for incremental storage and version control. *30th International Workshop on Intelligent Computing in Engineering (EG-ICE)*, pages 511-520, 2023.

[9] K. Dietmair, S. Esser, and B. Harder. Incremental model updates in BIM: a critical analysis of the collaboration platform Speckle. *Proceedings of the 35th Forum Bauinformatik 2024*, 2024.

[10] F. Xue and W. Lu. A semantic differential transaction approach to minimizing information redundancy for BIM and blockchain integration. *Automation in Construction*, 118:103270, 2020.

[11] L. Kong, R. Zhao, C. Anumba, W. Lu, and F. Xue. Open BIM exchange on Blockchain 3.0 virtual disk: A traceable semantic differential transaction approach. *Frontiers of Engineering Management*, 12:510-528, 2025.

[12] S. Esser, S. Vilgertshofer, and A. Borrmann. Graph-based version control for asynchronous BIM collaboration. *Advanced Engineering Informatics*, 53:101664, 2022.

[13] I. Nonaka, and H. Takeuchi. *The Knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford university press, 1995.

[14] H. Wang, and X. Meng. BIM-supported knowledge management: Potentials and expectations. *Journal of Management in Engineering*, 37(4):04021032, 2021.

[15] E. Feigenbau. Knowledge engineering. *Annals of the New York Academy of Sciences*, 426(1):91-107, 1984.

[16] J. Beetz, J. Leeuwen, and B. de Vries. IfcOWL: A case of transforming EXPRESS schemas into ontologies. *Ai Edam*, 23(1):89-101, 2009.

[17] J. Oraskari, and S. Törmä. RDF-based signature algorithms for computing differences of IFC models. *Automation in Construction*, 57:213-221, 2015.

[18] S. Zhang, S. Zhang, H. Liu, X. Wang, Z. Zhao, C. Wang, and L. Yan. Integration of BIM and ontologies for pumped storage hydropower design change management in EPC projects. *Automation in Construction*, 175:106189, 2025.

[19] Q. Zhou, and F. Xue. Automatic information gain-guided convergence for refining building design parameters: Enhancing effectiveness and interpretability in simulation-based optimization. *Building and Environment*, 275:112788, 2025.

[20] Y. Shen, and Y. Pan. BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization. *Applied Energy*, 333:120575, 2023.

Automated Subway Tunnel Leakage Segmentation: A Data-Efficient Fine-Tuning Strategy Based on Large Vision Foundation Models

Longfei Dai[1], Zhiyao Tian[2*], Shunhua Zhou[1]

[1]Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, China
[2]Department of Architecture and Civil Engineering, College of Engineering, City University of Hong Kong, China

2310828@tongji.edu.cn, zhiytian@cityu.edu.hk, zhoushh@tongji.edu.cn

Abstract –

Leakage detection in subway tunnel linings remain critical yet challenging tasks. Conventional deep learning models typically require tens of thousands of training samples that are scarce in engineering practice. Large Vision Models (LVMs) possess impressive general-purpose capabilities but fall short when applied to specialized downstream tasks. To address this gap, this study proposes a tailored framework to fine-tune LVMs for leakage segmentation. Built upon Segment Anything Model 2 (SAM2), this framework comprises four dedicated components to address the complex visual characteristics of subway tunnels: (i) a tunable module for enhancing leakage edge delineation; (ii) embedding generators for projecting visual features into a leakage-adapted representation space; (iii) adaptors for learning leakage-specific representations; and (iv) a fully fine-tuned decoder for pixel-level segmentation. These components collectively activate SAM2's rich prior knowledge, facilitating high-accuracy leakage segmentation. Experiments demonstrate that the proposed method improves untuned SAM2's accuracy by 3 to 4 times. Remarkably, it achieves performance comparable to DeepLabV3+ (a state-of-the-art deep learning model trained on 2000 samples) using only 50 samples.

Keywords –

Data-Efficient Fine-tuning; Large Vision Model; Leakage Segmentation; Shield Tunnels; Intelligent Infra-structure Inspection

1 Introduction

With the rapid expansion of subway networks, water leakage remains a critical threat to operational safety [1-2]. Current manual detection is labor-intensive and subjective, creating a pressing need for efficient, high-precision, and fully automated solutions.

To overcome the limitations of traditional inspection methods, researchers have increasingly explored deep learning-based computer vision techniques for tunnel leakage detection. In recent years, convolutional neural networks (CNNs) have demonstrated remarkable advantages in automated leakage detection. For instance, Chen et al. [3] developed Leakage-YOLO, an enhanced version of YOLOv8, while Xue et al. [4] modified the Faster R-CNN model to further improve detection accuracy. However, these object detection methods cannot achieve pixel-level segmentation. To address this, Xue et al. [5] employed a Cascade Mask R-CNN framework to achieve segmentation of tunnel leakage. Zhao et al. [6] integrated ResNet-101 with Mask R-CNN for classification and segmentation of leakage images, while Feng et al. [7] enhanced the encoder structure of U-net and U-net++ to develop a U-shape model. These deep learning-based approaches have demonstrated promising results on specific datasets. However, these models require large annotated datasets (often comprising tens of thousands of images) for multiple rounds of training and validation [6]. Acquiring and labeling such large-scale datasets is labor-intensive and time-consuming, a difficulty compounded by the intrinsic scarcity of seepage data in engineering practice [8].

In recent years, Vision Transformers (ViTs) and Large Vision Models (LVMs) have attracted significant attention and have been widely adopted across a broad spectrum of visual domains [9-10]. Notably, one representative model, the Segment Anything Model (SAM) demonstrates strong zero-shot segmentation capability across diverse objects such as humans, vehicles, and animals [11]. Extending this potential to tunnel inspection, Wei et al. [12] extracted the encoder module of SAM and integrated it with an enhanced U²-Net to construct the DeU²-Net model for leakage segmentation. Nevertheless, as their method remains built upon traditional backbone architectures, it still requires extensive training data.

An emerging approach that has attracted increasing attention is the fine-tuning of LVMs with domain-specific knowledge, enabling these models to exhibit data-efficient learning behavior. However, tunnel leakage typically appears as a surface-attached, weak-edged, and low-contrast pattern on concrete, which can easily be mistaken for background noise and thus overlooked. Moreover, subway tunnels often exhibit complex visual conditions, including low illumination, texture interference, and strong visual clutter, making reliable segmentation even more difficult. Existing frameworks largely rely on uniform architectural fine-tuning strategies [13]. Fine-tuning an LVM for leakage segmentation, specifically considering the distinctive visual and structural characteristics of leakage patterns, remains largely unexplored.

To address this gap, we propose a specialized fine-tuning framework for tunnel leakage segmentation. Built on SAM2, this data-efficient learning strategy leverages pre-trained priors instead of complex training paradigm, ensuring high accuracy through a simpler, more robust training process practical for infrastructure maintenance. A Leakage Information Enhancement Module first amplifies leakage boundaries and suppresses background noise in the frequency domain. The Embedding Generator then projects general encoder features into a leakage-adapted latent space, enriching the model's semantic understanding of leakage patterns. These enhanced and projected features are integrated by Adaptors, which inject task-specific prompts into the SAM2 encoder to guide attention toward leakage-relevant regions. Finally, a fully fine-tuned Mask Decoder performs pixel-level segmentation, producing high-precision leakage masks with sharp boundaries and coherent spatial consistency.

2 Methodology

2.1 The Original SAM2 Architecture

SAM2 is a large vision foundation model specifically developed for image semantic segmentation [14]. Building upon the powerful segmentation capabilities of its predecessor SAM, SAM2 further enhances segmentation accuracy, boundary delineation, and computational efficiency. The model is based on the Vision Transformer (ViT) architecture and is specifically designed for Promptable Visual Segmentation, a segmentation paradigm that performs target segmentation through prompts such as points, boxes, or masks. As illustrated in Figure. 1 [14], Image Encoder, Prompt Encoder, and Mask Decoder constitute the core structure of SAM2.

Although the original SAM2 model possesses strong prior knowledge and demonstrates excellent generalization capacity, preliminary experiments (further discussed in Section 3.4) reveal several limitations when it is directly applied to tunnel leakage segmentation: (i) Dependency on user prompts. SAM2 requires user-provided prompts for each image, rendering it impractical for large-scale, automated workflows. (ii) Low leakage segmentation accuracy. Due to the lack of domain-specific training, SAM2 struggles with complex tunnel backgrounds, often misidentifying structural features (Figure. 1), while overlooking actual leakage regions.

2.2 The Proposed Framework

2.2.1 Framework overview

To address the suboptimal performance of SAM2 in tunnel environments, a fine-tuning framework is proposed (Figure 2) based on a parameter-efficient fine-tuning (PEFT) strategy. To minimize computational overhead, the heavy Image and Prompt Encoders are frozen; consequently, only 3.94 M parameters (representing approximately 1.76% of the 224.51 M total parameters) are updated. These trainable parameters comprise the fully fine-tuned Mask Decoder and three modules: the Leakage Information Enhancement Module, the Embedding Generator, and the Adaptors. This configuration ensures a lightweight architecture that maintains inference speeds and memory usage nearly identical to the original SAM2 while enabling effective domain knowledge learning.

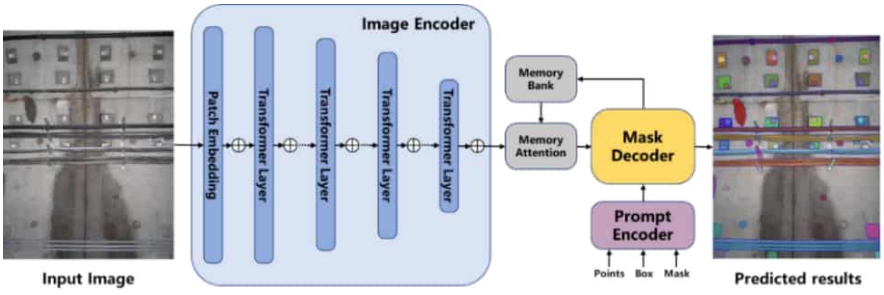


Figure 1. Architecture of the SAM2 large model and segmentation results on tunnel leakage.

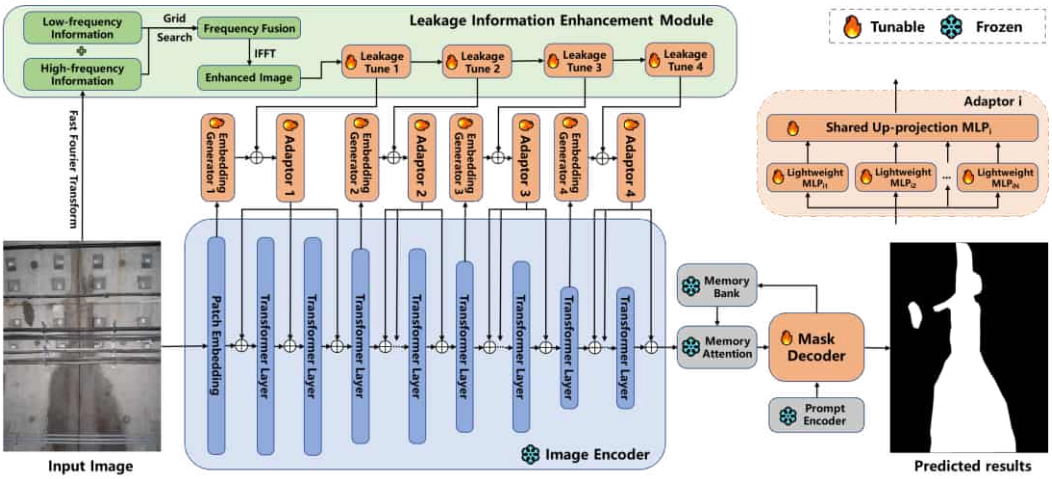


Figure 2. Architecture of the proposed fine-tuning framework.

2.2.2 Leakage Information Enhancement Module

The leakage recognition task in shield tunnels exhibits distinct challenges: (i) leakage boundaries are often blurred and gradual, lacking clear structural features; (ii) the tunnel background contains strong visual noise, including pipelines, bolts, grouting holes, and segment joints, all of which have sharp edges. Targeting these issues, an information enhancement module is designed.

The raw image I (of size $H\times W$) is first transformed into the frequency domain using the Fast Fourier Transform (FFT): z=FFT(I). The low-frequency components are shifted to the spectrum center (H/2, W/2), with frequency increasing radially outward. Two binary masks are constructed to extract different frequency bands:

(i) High-frequency mask $M_{\mathrm{h}}(\tau)$ for extracting rapidly varying edge and texture information:

$$M_{\mathrm{h}}^{i,j}(i,j,\tau)=\begin{cases}0, & \dfrac{4\,|(i-H/2)(j-W/2)|}{HW}<\tau\\ 1, & \text{otherwise}\end{cases} \tag{1}$$

(ii) Low-frequency mask $M_{\mathrm{l}}(\tau)$ for capturing slowly varying structural information and overall brightness:

$$M_{\mathrm{l}}^{i,j}(i,j,\tau)=\begin{cases}0, & \dfrac{HW-4\,|(i-H/2)(j-W/2)|}{HW}<\tau\\ 1, & \text{otherwise}\end{cases} \tag{2}$$

The corresponding high- and low-frequency components are then obtained as:

$$z_{\mathrm{h}}=zM_{\mathrm{h}}(\tau) \tag{3}$$

$$z_{\mathrm{l}}=zM_{\mathrm{l}}(\tau) \tag{4}$$

where the high-frequency component z_{h} captures sharp pixel variations, i.e., edges and texture details, including the boundaries of leakage, pipelines, and bolt holes. The low-frequency component z_{l} represents slowly varying regions, such as the main leakage body and background concrete surface. To highlight leakage characteristics while suppressing interference from irrelevant features, this module performs weighted fusion of the two components:

$$z_{\mathrm{e}}=\alpha z_{\mathrm{h}}+\beta z_{\mathrm{l}} \tag{5}$$

where α and β are learnable hyperparameters. After task-specific fine-tuning, the learned weights yield two key effects: (i) Leakage edges becomes clearer, while isolated non-target structures are attenuated; (ii) The gradient information of the main leakage region is retained, whereas smooth concrete backgrounds are weakened.

Finally, the enhanced image is reconstructed via the inverse FFT (IFFT): I_e=IFFT(z_e), and subsequently passed through four trainable Overlapping Patch Embedding layers, referred to as Leakage Tune modules. These modules are specifically designed to process enhanced leakage features corresponding to the four resolution stages of the SAM2 Image Encoder. This process generates multi-scale enhanced features F_{E1}, …, F_{E4}, which are then fed into the subsequent Adaptor modules for further feature learning.

2.2.3 Embedding Generator

Embedding Generators are further introduced to enhance the model's auto-prompting capability for leakage detection. Each Embedding Generator serves as an auxiliary input source for the Adaptor modules. Following the four-resolution stages of the SAM2 Image Encoder, four trainable Embedding Generator modules are constructed accordingly: When the raw image I enters the Image Encoder, it first passes through the Patch Embedding layer for patch extraction and embedding, resulting in the initial embedded feature I_p. Subsequently, Embedding Generator 1 projects this feature into a task-adaptive feature space:

$$F_{G1}=L_{EG1}(I_p) \tag{6}$$

The remaining three Embedding Generators follow a

similar process; however, their inputs are taken from the first Transformer layer of each corresponding resolution stage (see Figure 2). $L_{EGi}(.)$ (i = 1, 2, 3, 4) denotes a learnable linear transformation layer. By fine-tuning its internal weight matrices and bias terms, the generic visual features produced by the pretrained Encoder are reprojected into the leakage-specific task space. After projection, the resulting features F_{G1}, …, F_{G4}, are obtained sequentially and serve as additional inputs to the subsequent Adaptor modules.

2.2.4 Adaptor

Adaptors are further introduced to fuse the previously obtained leakage-enhanced features F_{Ei} and embedding features F_{Gi} (i=1, 2, 3, 4), guiding task-specific knowledge into the original SAM2 network based on the concept of prompt injection. This allows fine-grained adaptation and enhancement for downstream tasks without compromising the ability of the original model [15].

Four adaptors are constructed, corresponding to the four resolution stages of the Image Encoder (see Figure 2). For each Adaptor, F_{Ei} and F_{Gi} (i=1, 2, 3, 4) are fused to form the input embeddings F_{Ini}, which is subsequently processed multiple times within the internal structure of Adaptor. As shown in Figure 2, each Adaptor adopts a lightweight two-stage multilayer perceptron (MLP) architecture.

F_{Ini} first enters the initial linear layer of the first stage MLP_{i1}, passes through a *GELU* activation function, and is then fed into the shared projection layer MLP_i, producing the output prompt vector:

$$P_{i,1} = MLP_i(GELU(MLP_{i,1}(F_{INi}))) \tag{7}$$

Subsequently, F_{Ini} is sequentially processed by MLP_{i2}, …., MLP_{iN}. By fine-tuning the linear weight matrices and bias terms of the MLP layers, the model generates a set of leakage task-specific prompt embeddings $\{P_{i,1}, ..., P_{i,N}\}$, which are injected layer by layer into the outputs of the corresponding network layers in the i-th stage of the Image Encoder.

2.2.5 Fully fine-tuned Mask Decoder

The Mask Decoder is fully fine-tuned due to its pivotal role in mask generation and its parameter efficiency. Specifically, the Mask Decoder receives the embedding features through the previous framework, and processes them via a combination of multi-layer Transformers, upsampling modules, and Hypernetwork structures to produce multi-scale candidate masks along with their corresponding confidence scores. During fine-tuning, manually annotated leakage regions are used as supervision signals, and the model parameters are updated. Through the synergistic enhancement of the preceding modules, the model ultimately generates leakage masks with sharp boundaries, coherent structures, and high semantic consistency.

3 Experiment

3.1 Datasets

An extensive dataset was provided by Xue et al. [5], comprising 4755 images of shield tunnel leakages collected from Shanghai Metro tunnels. After verifying the one-to-one correspondence between images and labels and removing 196 invalid samples, we obtained a final set of 3359 annotated images for model training and 1000 annotated images to evaluate model performance. Representative examples from this dataset are shown in Figure 3.

3.2 Evaluation Metrics

To quantitatively evaluate the performance of the proposed method, two commonly used segmentation metrics are employed : Intersection over Union (*IoU*), Dice coefficient (*Dice*). The definitions of these metrics are as follows:

$$IoU = \frac{1}{N}\sum^{N}\frac{Int}{Uni} \tag{8}$$

$$Dice = \frac{1}{N}\sum^{N}\frac{2\times Int}{Int+Uni} \tag{9}$$

where, N denotes the total number of testing samples, *Int* and *Uni* refer to the intersection and union between ground truth mask and predicted mask generated by the model, respectively .



Figure 3. Representative samples from the used dataset, adapted from [5].

3.3 Experimental Settings

Among the four official SAM2 checkpoints, we adopted hiera-large as the backbone to fully leverage its strong prior knowledge and superior pretrained performance. During fine-tuning process, a joint loss function combining Binary Cross-Entropy (BCE) and *IoU* loss was employed to optimize the network. The batch size and number of training epochs were set to 2 and 40, respectively. To fine-tune the hyperparameters α, β, both defined within [0,1], a systematic grid search method was introduced. Candidate values for α and β were evaluated at intervals of 0.25. The grid-search results are presented in Figure 4, which identifies α=0.25 and β=0.75 as the optimal hyperparameter combination across all evaluation metrics; these values were therefore adopted for the final model configuration.

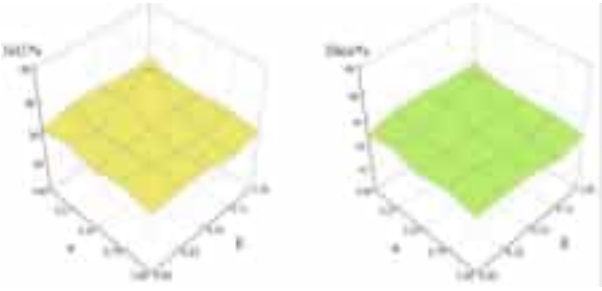


Figure 4. Grid search results for α and β.

3.4 Experimental Results

The validation set, i.e., the remaining 1,000 annotated images [5], was used to evaluate the performance of the fine-tuned model. For comparison, the original SAM2 model was also tested on the same dataset. Because SAM2 requires a manual prompt as an additional input, we simulated this process by randomly selecting one point from the ground-truth mask of each image as the prompt, which actually overestimates the original SAM2 model performance. The results are shown in Table 1.

The fine-tuned model substantially outperforms the original SAM2, achieving improvements by a factor of approximately 3–4. Specifically, the fine-tuned model attains an IoU of 87.01% and a Dice (F1-score) of 92.90%, with Precision and Recall reaching 92.34% and 93.89%, respectively. Representative segmentation results for the two models are shown in Figure 5.

Table 1. Performance comparison on the annotated validation dataset

Model	IoU(%)	Dice(%)
Proposed Framework	**87.01**	**92.90**
Original SAM2	21.49	30.75
UNet	83.07	90.75
SegFormer	85.88	92.42
Mask R-CNN	83.05	90.74
DeepLabV3+	85.96	92.45

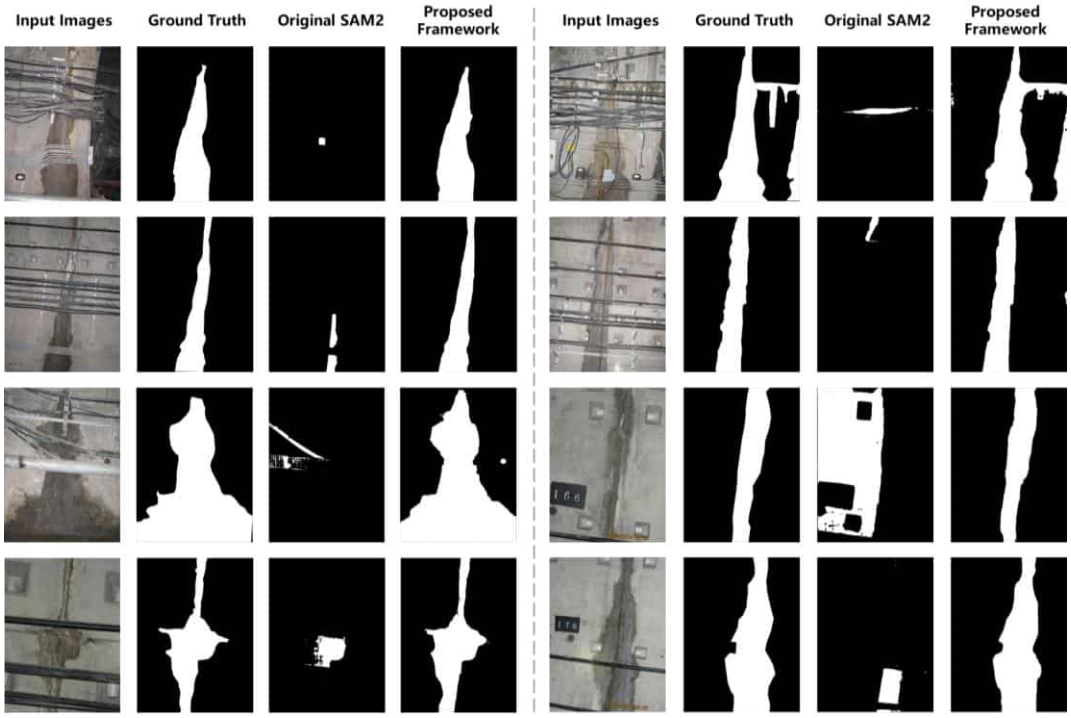


Figure 5. Representative leakage segmentation results on the annotated validation dataset.

4 Discussion

4.1 Comparison with the State-of-the-Art Model

The proposed approach is evaluated against state-of-the-art conventional deep learning models widely utilized for tunnel leakage segmentation. As summarized in Table 1, the framework is compared with several mainstream architectures, including UNet, SegFormer, Mask R-CNN, and DeepLabV3+, all of which are surpassed by the proposed method in terms of accuracy. Notably, DeepLabV3+ emerges as the strongest baseline and is therefore selected for a more rigorous comparative analysis under data-scarce conditions.

Consequently, DeepLabV3+ with an Xception backbone serves as the primary benchmark for evaluating model performance across increasingly constrained data regimes. To systematically investigate data efficiency and model robustness, experiments are conducted across eight training subsets , ranging from 3,000 images down to 10 randomly sampled images.

Table 2. Performance comparison between the proposed framework and DeepLabV3+

Training Set Size	Proposed Framework		DeepLabV3	
	IoU(%)	Dice(%)	IoU(%)	Dice(%)
3359	**87.01**	92.90	**85.96**	92.45
3000	**86.73**	92.72	**80.64**	88.28
2000	**85.37**	91.90	**71.31**	80.48
1000	**83.90**	90.85	**36.68**	47.05
500	**83.66**	90.73	**20.87**	30.44
100	**73.64**	83.71	**1.370**	2.540
50	**69.72**	80.95	**1.170**	2.140
20	**47.46**	63.15	**0.600**	1.160
10	**2.450**	4.780	**2.020**	3.740

The experimental results are summarized in Table 2, and the corresponding trend of IoU performance with respect to training set size is illustrated in Figure 6. The results show that when the training set contains only 10 images, both the fine-tuned SAM2 and DeepLabV3+ perform extremely poorly. Notably, the performance of the fine-tuned SAM2 improves steadily as the training set size increases. When the training set reaches 50 images, the IoU of the large model approaches 70%, which represents a reasonably competitive performance for applications with moderate segmentation requirements. In contrast, DeepLabV3+ shows almost no improvement at this scale, achieving an IoU of only 1.17%, suggesting that it requires substantially more data to learn effectively.

When the training set size reaches 500 images, the fine-tuned SAM2 achieves metrics exceeding 80% across all evaluation criteria, with Dice surpassing 90%, which corresponds to a high-accuracy segmentation performance commonly reported across applied imaging domains. As the dataset size increases further, the accuracy of the fine-tuned large model continues to improve, although the performance gains become less pronounced. Interestingly, DeepLabV3+ requires approximately 2000 training images to achieve a performance comparable to that of the fine-tuned large model trained with 50 images. This finding highlights the strong data-efficient learning capability of the proposed framework. When both models are trained with 3000 images, their performances become broadly similar; however, DeepLabV3+ still exhibits slightly lower overall accuracy (Figure6).

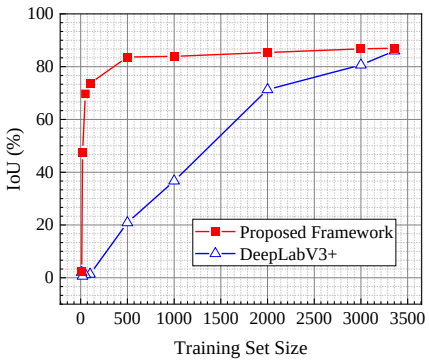


Figure 6. IoU performance versus training set size for the proposed framework and DeepLabV3+.

In summary, conventional deep learning models require large quantities of annotated images to achieve accurate shield-tunnel seepage segmentation. In contrast, although a large vision model initially lacks domain-specific knowledge, it exhibits a fundamentally different learning behavior. Once provided with even a small amount of task-specific supervision, the large model can rapidly activate its latent knowledge and extensive pretrained representations, enabling it to perform the previously unfamiliar task effectively. Importantly, the amount of data required for such fine-tuning is significantly smaller than that demanded by conventional deep learning approaches.

This property is particularly valuable for engineering applications, where annotated data are often scarce or difficult to obtain. For example, tunnel leakage images are typically controlled by transportation authorities or research institutions, resulting in inherently limited datasets. In such scenarios, the ability to fine-tune a large model is transformative, making effective machine learning workflows feasible even under limited data conditions. To further demonstrate this advantage, an engineering case study will be presented in the following subsection.

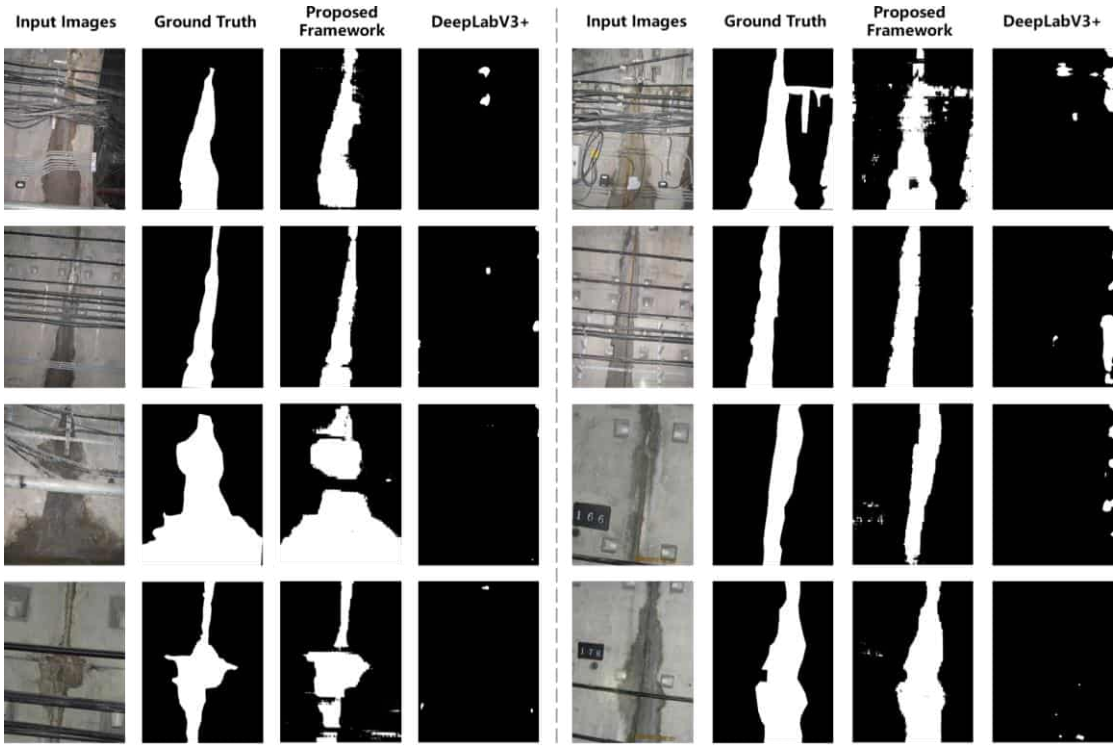


Figure 7. Representative leakage segmentation results on the annotated validation dataset using 50 training samples.

4.2 Ablation Study

In the proposed framework, the Leakage Information Enhancement Module and the Embedding Generator are architecturally coupled, as they generate the synergistic inputs (F_{Ei} and F_{Gi}) required by the Adaptors. These Adaptors serve as the central mechanism for injecting task-specific knowledge into the frozen SAM2 backbone through a prompt-injection strategy. Due to this functional interdependence, the Adaptors are consistently activated in our experimental configurations whenever either the Leakage Information Enhancement Module or the Embedding Generator is enabled to ensure the effective transmission of domain knowledge through the injection chain. To further investigate how this specialized fine-tuning transforms a general large model into a high-precision domain-specific one, four experimental configurations were designed to isolate the contribution of these key components. All configurations were trained on the same dataset and evaluated on the validation set, with the quantitative results summarized in Table 3.

It can be observed in Table 3 that fine-tuning only the Mask Decoder yields a remarkable improvement, increasing IoU from 21.49% to 78.47%. This substantial gain arises from the fact that, fine-tuned Mask Decoder refines the model's understanding of specific leakage cues, allowing it to distinguish weak-edged, low-contrast patterns from the generic objects emphasized in SAM2's pretraining. Nevertheless, the model still exhibits limitations in accuracy, falling short of the requirements for high-precision engineering applications.

Table 3. Results of the ablation study

Fine-tuning configurations	IoU(%)	Dice(%)
Original SAM2	21.49	30.75
Only Mask Decoder	78.47	87.51
Embedding Generator + Mask Decoder	85.89	92.17
Leakage Information Enhancement Module + Mask Decoder	85.61	92.01
Complete Framework	**87.01**	**92.90**

After introducing the Embedding Generator, the performance is further improved, achieving 85.89% IoU. The improvement arises from domain adaptation performed by the Adapter, together with the more effective prompt guidance provided by the Embedding Generator.

Similarly, adding the Leakage Information Enhancement Module produces gains comparable to those achieved with the Embedding Generator, attaining 85.61% IoU. Notably, the Leakage Information Enhancement Module strengthens the high-frequency boundary components of leakage regions in the frequency domain while suppressing background interference.

Finally, the full configuration achieves the optimal performance of 87.01% IoU.

5 Conclusion

This study adapts Segment Anything Model 2 into a specialized tunnel leakage segmentation model through a parameter-efficient fine-tuning framework. The main conclusions are as follows:

(i) The proposed framework overcomes SAM2's original lack of domain knowledge, transforming it into a high-precision tool for tunnel leakage detection.

(ii) The model exhibits remarkable data-efficient capability, where only 50 annotated samples achieve segmentation accuracy comparable to state-of-the-art model trained on 2,000 samples.

(iii) The framework's superiority stems from leveraging SAM2's pretrained priors through the synergy of the Mask Decoder and the proposed Leakage Information Enhancement Module, Embedding Generator, and Adaptors, which collectively enhance boundary delineation and generalization.

Preliminary tests on a Nanjing Metro dataset achieved 83.12% IoU through direct validation without retraining, demonstrating robust generalization. Future research will utilize multi-scale ablations and multi-city studies to further validate cross-domain adaptability and probe theoretical limits.

References

[1] Wang, L., Chen, H., Liu, Y., Li, H., & Zhang, W. Application of copula-based Bayesian network method to water leakage risk analysis in cross river tunnel of Wuhan Rail Transit Line 3. Advanced Engineering Informatics, 57, 102056, 2023.

[2] Tian, Z., Xu, P., Gong, Q., Zhao, Y., & Zhou, S. Health-degree model for stagger-joint-assembled shield tunnel linings based on diametral deformation in soft-soil areas. Journal of Performance of Constructed Facilities, 37(3), 04023019, 2023.

[3] Chen, C. & Liu, W. Leakage-YOLO: Real-Time Object Detection Algorithm for Crack and Leakage in Tunnel Scenarios. Journal of Computer Engineering & Applications, 61(6), 2025.

[4] Xue, Y., Gao, J., Li, Y., & Huang, H. Optimization of Shield Tunnel Lining Defect Detection Model Based on Deep Learning. Journal of Hunan University, 47(7), 137-146, 2020.

[5] Xue, Y., Cai, X., Shadabfar, M., Shao, H., & Zhang, S. Deep learning-based automatic recognition of water leakage area in shield tunnel lining. Tunnelling and Underground Space Technology, 104, 103524, 2020.

[6] Zhao, S., Shadabfar, M., Zhang, D., Chen, J., & Huang, H. Deep learning‐based classification and instance segmentation of leakage‐area and scaling images of shield tunnel linings. Structural Control and Health Monitoring, 28(6), e2732, 2021.

[7] Feng, S. J., Feng, Y., Zhang, X. L., & Chen, Y. H. Deep learning with visual explanations for leakage defect segmentation of metro shield tunnel. Tunnelling and Underground Space Technology, 136, 105107, 2023.

[8] Yu, X., He, B. G., Xu, X., Zhou, Y., Diaz, M. A., Chen, J., & Camacho, D. Application of Artificial Intelligence in Rock Tunnel Engineering: A Survey on Where and How. Expert Systems, 42(7), e70080, 2025.

[9] Wei, X., Li, J., Zhang, X., Gu, H., Van de Weghe, N., & Huang, H. An innovative framework combining a CNN-Transformer multiscale fusion network and spatial analysis for cycleway extraction using street view images. Sustainable Cities and Society, 106384, 2025.

[10] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022), 2021.

[11] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4026), 2023.

[12] Wei, Z., Ji, A., Xiao, Z., & Zhang, L. Segment anything model enhanced dual encoder segmentation for seepage quantification in operational tunnels. Advanced Engineering Informatics, 69, 104030, 2026.

[13] Zhou, Y., Sun, G., Li, Y., Xie, G. S., Benini, L., & Konukoglu, E. When sam2 meets video camouflaged object segmentation: A comprehensive evaluation and adaptation. Visual Intelligence, 3(1), 10, 2025.

[14] Ravi, N., Gabeur, V., Hu, Y. T., Hu, R., Ryali, C., Ma, T., ... & Feichtenhofer, C. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714, 2024.

[15] Chen, T., Lu, A., Zhu, L., Ding, C., Yu, C., Ji, D., ... & Zang, Y. Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. arXiv preprint arXiv:2408.04579, 2024.

[16] Wang, D., Hou, G., Chen, Q., Li, W., Fu, H., Sun, X., & Yu, X. Segmentation of tunnel water leakage based on a lightweight DeepLabV3+ model. Measurement Science and Technology, 36(1), 015414, 2024.

Experimental Design and Field Validation of a 3DGS-based Rebar Inspection Testbed for Infrastructure Projects

Kumar Adarsh [1,2] **Ashwani Jaiswal** [1] **and Nikhil Bugalia** [1]

[1]Department of Civil Engineering, Indian Institute of Technology Madras, India
[2]Heavy Civil Infrastructure-IC, Larsen & Toubro Construction

ce23m111@smail.iitm.ac.in, ce22d045@smail.iitm.ac.in, nbugalia@civil.iitm.ac.in

Abstract -

Conventional rebar quality inspection in construction relies heavily on manual measurements, which are time-consuming, error-prone, and impractical for dense reinforcement layouts. 3D Gaussian Splatting (3DGS) has recently emerged as a promising technique for vision-based automated rebar inspection, enabling dense 3D reconstruction under complex reinforcement configurations. Despite this potential, most existing approaches lack systematic parameter optimization and are primarily validated under laboratory conditions, which limits their applicability and reliability on real-world construction sites. This study proposes a structured experimental design to identify optimal parameters for a 3DGS-based automated rebar inspection pipeline, thereby reducing the need for trial-and-error during on-site deployment. A key contribution of this work is the identification of an optimal parameter configuration, comprising 10–20k training steps, a maximum splat count of 1M, 1920p downsampling, and UHD recording resolution. These parameters are subsequently validated on real-world U-girder reinforcement cages captured under constrained site conditions. Among the five field datasets, three demonstrate reconstruction and detection performance consistent with simulated predictions. Field validation further confirms that elevated camera viewpoints significantly improve reconstruction completeness and rebar detectability. Overall, the proposed virtual testbed enables efficient pre-deployment optimization, reducing calibration effort and enhancing the robustness of 3DGS-based rebar inspection in real-world construction environments.

Keywords -

Rebar Inspection; Computer Vision; 3D Gaussian Splatting; Point Cloud; Optimization.

1 Introduction

In the construction industry, conventional quality assessment of reinforcement remains heavily reliant on manual inspection methods, which are both time-consuming and inconsistent, and introduce errors [1]. On-site engineers typically measure rebar spacing, alignment, and layout using tapes and gauges, which requires direct physical access to dense reinforcement cages. Such practices not only compromise inspection efficiency but may also disturb the reinforcement arrangement, affecting construction quality [2]. With increasing project complexity and workforce shortages, the conventional inspection method is becoming more challenging [3]. In this context, digital and automated vision-based rebar inspection systems have emerged as a more effective solution to overcome the limitations of conventional assessment, offering the potential to standardize assessments, reduce rework, and enable continuous monitoring of structural compliance.

An automated rebar inspection pipeline typically consists of two sequential key processes: (i) data acquisition and (ii) computational processing for quality assessment [1]. Data acquisition aims to capture geometric information of reinforcement cages. It has been implemented using various sensing modalities, such as terrestrial or mobile laser scanners (TLS/MLS), RGB-D cameras, or monocular RGB cameras [4]. TLS and MLS systems directly generate dense, accurate point clouds and have demonstrated high precision in rebar spacing measurements, particularly under laboratory conditions [2, 3]. However, these systems are costly, require skilled operators, and involve significant setup time, which limits their scalability for routine on-site deployment. RGB-D cameras offer a lower-cost alternative and can directly provide depth information, but their effective range and robustness remain constrained in cluttered or large-scale construction environments [5]. Consequently, monocular RGB cameras have attracted growing attention due to their low cost, ease of deployment, and flexibility in real construction settings.

Unlike TLS/MLS and RGB-D sensors, monocular RGB cameras do not directly produce 3D geometry; therefore, they require an additional 3D reconstruction or preprocessing step to recover spatial information. Traditionally, this reconstruction has been achieved using photogrammetry or multi-view stereo (MVS) techniques [6]. While MVS is computationally efficient and widely adopted, it often struggles to reliably reconstruct dense

and repetitive geometries, such as closely spaced rebar cages, especially under occlusions and complex viewing conditions. These limitations significantly restrict the applicability of MVS-based pipelines in real-world rebar inspection scenarios.

Recently, 3-Dimensional Gaussian Splatting (3DGS) has emerged as a state-of-the-art technique for high-fidelity 3D reconstruction in visually complex scenes [7]. Compared to traditional MVS, 3DGS has demonstrated superior performance in handling repetitive patterns and complex occlusions. Recent studies have explored 3DGS-based rebar detection pipelines, typically developed and validated within a virtual reality (VR) environment, and have reported substantially improved geometric detail and reconstruction accuracy compared to MVS-based approaches [8]. However, the literature in this domain is nascent, is primarily focused on establishing a proof-of-concept experimental testbed, and does not systematically investigate the influence of 3DGS reconstruction parameter selection. Despite the demonstrated potential of 3DGS for reconstructing dense reinforcement geometries, its performance is susceptible to a range of reconstruction-level parameters [7, 9]. These parameters include the number of training steps, the maximum splat count, the image downsampling factor, and the selection of the best frames (training views). In addition, 3DGS performance is strongly influenced by field-level data acquisition parameters, such as the camera's trajectory, illumination levels, and recording resolution, which jointly affect view coverage and geometric quality [10]. However, existing studies employing 3DGS largely adopt fixed or empirically chosen parameter settings, without systematically examining their influence on reconstruction accuracy. This lack of parameter-level understanding represents a critical barrier to the reliable deployment of 3DGS-based rebar detection pipeline in real construction environments, where extensive trial-and-error optimization is impractical. Thus, there is a clear need for a structured experimental investigation that enables systematic parameter optimization prior to on-site deployment, thereby minimizing the effort required to tune and calibrate under real construction conditions.

Also, despite notable progress, the majority of existing automated rebar inspection studies are still evaluated primarily in laboratory or VR conditions. Among the reviewed literature, Yuan et al. (2023) present one of the few studies that conduct in-situ rebar inspection on a single-layer structural element, primarily focusing on relatively larger rebar spacing [2]. However, the study does not systematically account for variability in capture strategy, site constraints, or site conditions. In contrast, Kim et al. (2024) investigate multiple data collection scenarios to improve the generalizability of their approach [3]. However, their methodology is primarily validated in VR environments using synthetic datasets with single-layer rebar cages, and the framework lacks comprehensive validation on real-world data. More importantly, complex reinforcement geometries, such as multi-layered rebar cages in U-girders or I-girders, are underrepresented in existing field-based evaluations.

To address this research gap, this study builds upon a previously developed VR-based rebar inspection workflow featuring large-scale U-griders [8]. The study introduces a structured experimental design aimed at identifying optimal 3DGS reconstruction parameters (training steps, maximum splat count, image downsampling, and best frame selection). In parallel, key field-level data acquisition parameters (camera path, illumination, and recording resolution) are examined in the VR-based workflow to assess their impact on reconstruction quality. The resulting optimal data acquisition and reconstruction parameters are then field-validated on real on-site U-girders, where 3DGS-based rebar spacing measurements are compared with manual ground truth. The study makes significant contributions towards establishing systematic validation frameworks that bridge virtual (synthetic) or laboratory-based evaluations with real-world field validations.

2 Methodology

The methodology of this study builds upon the author's previous work on a 3DGS-based rebar detection pipeline. It extends it into a systematic framework for evaluating real-world scenarios and parameters to identify their optimal values.

2.1 3DGS-based Rebar Detection Pipeline

The developed pipeline follows a multi-step process. In the first stage, a synthetic environment is generated using the *Unity* engine, with a rebar cage designed in *Rhino* and a camera configuration to replicate realistic video capture conditions closely. The resulting environment is rendered as a video sequence and processed using *RealityCapture*. The synthetic video data are then processed using the 3DGS tool *Jawset PostShot* to generate splat cloud representations of the scene. These splat clouds are subsequently refined and denoised using *CloudCompare* to enhance data quality and remove outliers. Eventually, the cleaned point cloud is served as input to a custom analytical dashboard developed. Within this dashboard, K-means clustering is applied to segment the rebar into distinct layers, followed by a histogram-based projection method to identify individual rebar elements. Overall, this pipeline achieves a spacing error of less than 3% even for complex rebar configurations, such as U-girder (see

Figure 1). This end-to-end pipeline enables controlled experimentation while maintaining relevance to real-world conditions.

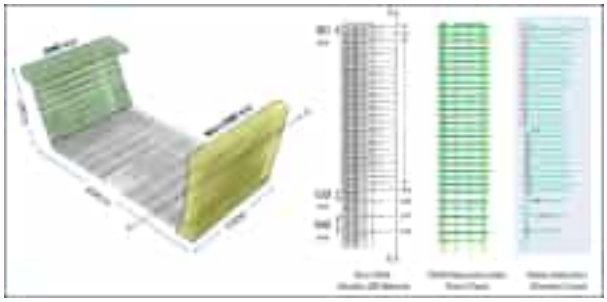


Figure 1. Overall process of rebar detection of non-OHE U-girder rebar cage.

2.2 Study Design and Evaluation Framework

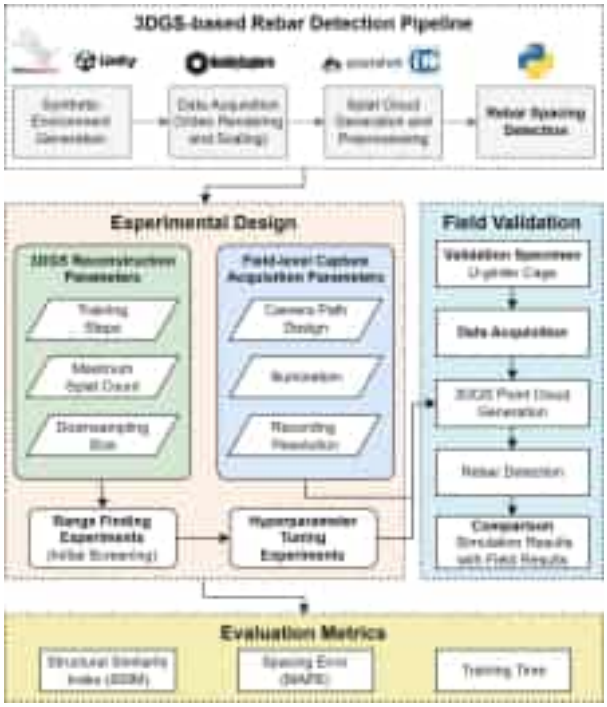


Figure 2. Schematic of the methodology of experimental design and field validation of the rebar detection pipeline.

The overall study design adopted in this research is illustrated in Figure 2. The study is structured into two main components: (i) controlled experiments conducted using synthetic data to investigate the influence of reconstruction and acquisition parameters, and (ii) field-level validation using real-world data captured from a representative U-girder specimen. To ensure consistent evaluation across all experiments and validation stages, a fixed region of interest within the U-girder rebar cage was selected as the reference layout (see Figure 1). The selected region corresponds to the non-OHE wall segment, which contains 28 closely spaced vertical rebars. This region was intentionally chosen for its high geometric complexity; the rebars are densely packed, partially occluded, and intersect with transverse reinforcement members, making both 3D reconstruction and rebar detection significantly more challenging than in single-layer rebar cage configurations.

3 Experimental Design

This section explains the experimental variables used to evaluate the 3DGS-based rebar inspection pipeline. The experiments were designed to separate reconstruction parameters from field-level acquisition parameters, enabling a systematic study of their individual and combined effects on inspection accuracy.

3.1 3DGS Reconstruction Parameters

The quality of 3DGS reconstruction is influenced by three key training parameters, namely training steps, maximum splat count, and downsampling size [7, 11]. These parameters were selected for systematic experimentation and varied individually using a one-factor-at-a-time approach, while all other conditions were kept constant. Specifically, during each experiment, a fixed orbital camera path, constant illumination (bright daylight), consistent recording resolution and identical preprocessing and reconstruction settings (other than the parameter under study) were maintained. **Training steps** control the duration of 3DGS optimization; increasing steps improves refinement of Gaussian position, shape, and colour up to a limit, beyond which gains are minimal and computation cost increases [11]. Hence, a range of training steps was tested to balance accuracy and efficiency. The **Maximum splat count** controls the number of Gaussian splats representing the scene. Higher values improve geometric detail for thin elements such as rebars but increase memory and training time; beyond a certain density, gains are marginal. Therefore, multiple splat limits were tested to identify a stable upper bound. Finally, the **Downsampling size** defines the resolution of input images during training. Higher resolutions preserve texture and edge detail but increase computational cost, so multiple levels were evaluated to determine the optimal setting. Range-finding experiments were conducted to establish suitable parameter bounds for subsequent hyperparameter tuning.

3.2 Field-level Data Acquisition Parameters

While 3DGS reconstruction parameters control the fidelity of the generated output, the overall rebar detection pipeline performance is also affected by field-level data-acquisition conditions. To investigate this effect,

a series of systematic experiments was conducted. These experiments were carefully designed to identify optimal conditions for achieving efficient and reliable reconstruction and detection performance. Among the capture parameters, camera path design, illumination, and recording resolution were selected due to their direct impact on image quality, coverage, and reconstruction [9, 10, 11].

3.2.1 Camera Path Design

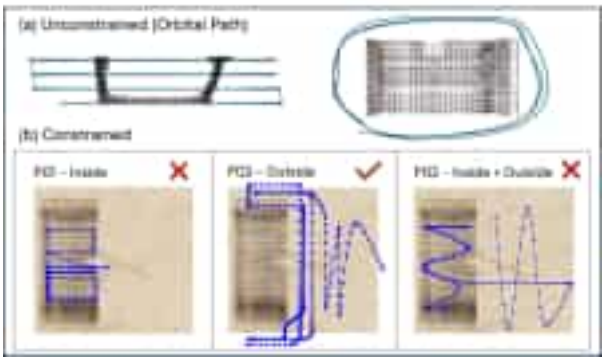


Figure 3. (a) Camera path design for unconstrained (orbital) and constrained scenarios (P01 to P03).

Camera path design determines how the camera moves around the rebar cage during data capture, directly influencing viewpoint diversity for reconstruction [10]. This experiment considers two practical site-based scenarios: unconstrained and constrained rebar cages. The unconstrained scenario represents cases where the rebar cage is fully accessible, allowing the camera to move and rotate freely around the structure. For this case, an orbital camera path was employed, in which the camera follows a closed-loop trajectory providing full 360° coverage of the cage at multiple fixed elevations. The constrained scenario reflects real construction conditions, where accessibility is limited, such as when rebar cages are embedded within U-girders or adjacent to formwork. To simulate these conditions, controlled camera paths were defined. Three configurations were evaluated: P01, where scanning is performed only from within the cage; P02, where scanning is performed only from outside the cage; and P03, which combines both inside and outside scanning. These paths were designed to assess the impact of accessibility constraints and camera placement on reconstruction performance.

3.2.2 Illumination

Illumination varies naturally throughout the day as the sun moves across the sky, resulting in fluctuations in on-site lighting conditions. For this reason, two illumination conditions were evaluated using experiments. First is a Bright daylight condition, representing typical outdoor precast yard environments, simulated at approximately 100,000 lux. Second is Low-light conditions, representing shaded or late-day site conditions, simulated at approximately 10,000 lux (see Figure 4a) [9].

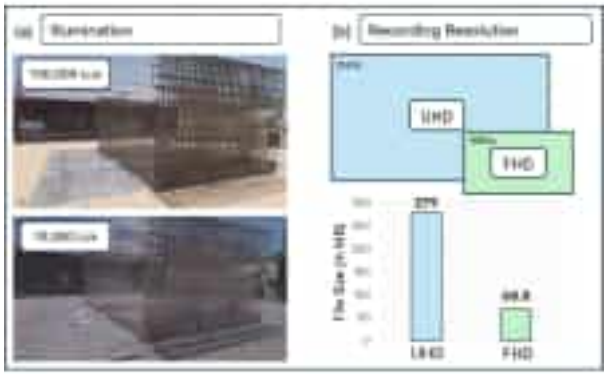


Figure 4. (a) Experimental variations for illumination, and (b) Recording resolution along with file size comparison.

3.2.3 Recording Resolution

The recording resolution of the raw video determines the finest pixel detail available before downsampling. Higher resolution offers more texture for feature matching, but at the expense of larger file sizes and longer processing times. The two most common recording resolutions were evaluated: FHD (1080p) and UHD (2160p) [12].

3.3 Evaluation Criteria

For evaluation purposes, an experiment was considered valid only if the rebar detection pipeline successfully detected all 28 rebars. Spacing errors were computed exclusively for these valid rebars using the Rhino model as ground truth. Three quantitative metrics were used consistently across all experiments to evaluate reconstruction quality, inspection accuracy, and computational feasibility. The first criterion is the Structural Similarity Index (SSIM). It captures similarity in structure, contrast, and luminance, and serves as an indicator of reconstruction stability rather than inspection accuracy. SSIM values were directly taken from *Jawset PostShot* after 3DGS reconstruction. The second criterion is Mean Absolute Percentage Error (MAPE), which measures the percentage deviation between detected rebar spacing values and the known ground-truth spacing obtained from the reference model, calculated using Equation (1).

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left| e_i^T - e_i^P \right|}{e_i^T} \tag{1}$$

where n is the total number of spacings analysed in a trial. e_i^P denotes the spacing estimated by the

workflow for the i-th bar pair, while e_i^T represents the corresponding ground-truth spacing obtained from manual calliper measurements (field data) or directly from the 3D model (synthetic data). The third evaluation criterion is training time, which represents the total duration required to complete the 3DGS reconstruction for a given configuration. This metric was recorded to evaluate the practical feasibility of the workflow and to balance accuracy gains against computational cost.

4 Results and Discussion

4.1 3DGS Reconstruction Parameters Results

This subsection discusses the influence of 3DGS generation parameters on reconstruction quality and spacing accuracy. The experiments were performed in two stages: range-finding and detailed hyperparameter tuning.

4.1.1 Range Finding Experiments

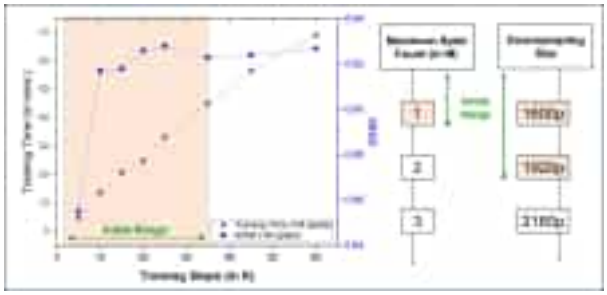


Figure 5. Training time and SSIM trends versus training steps, with highlighted initial range for maximum splat count and downsampling size.

Initially, a set of range-finding experiments was conducted to identify practical operating bounds for key 3DGS reconstruction parameters. A total of 19 experiments were performed using a fixed orbital camera path. Training steps varied from 5k to 60k, maximum splat count was set to 1M, 2M, and 3M, and input resolution was downsampled to 1600p, 1920p, and 2160p. Figure 5 illustrates the effect of training steps on SSIM and training time for a representative configuration (1M splats). While additional experiments with higher splat counts (2M and 3M) were conducted, they resulted in only marginal improvements in SSIM and no consistent reduction in MAPE. Therefore, these results are summarized here rather than explicitly plotted for clarity. Similarly, increasing the input resolution beyond 1920p results in higher training time, with negligible improvement in MAPE and only minor gains in SSIM, indicating diminishing returns beyond these thresholds.

4.1.2 Hyperparameter Tuning Experiments

Following the range-finding experiment, a structured hyperparameter tuning study was conducted to identify performance-efficient parameters for 3DGS reconstruction. A total of 60 experimental trials were conducted systematically by varying the reconstruction parameters listed in Table 1. The experimental results show a consistent trend across all resolutions. SSIM increases rapidly up to 20k training steps and plateaus thereafter for all resolutions (see Figure 6a). However, after reconstruction, the rebar detection pipeline with an 800p downsampling size failed to achieve complete rebar detection and was discarded (marked in red *). In addition, training with fewer than 10k training steps for resolutions 1600p and 1920p also failed. Thus, those samples were also discarded. The remaining samples demonstrated the best performance, achieving 100% rebar detection. Figure 6b compares the downsampling sizes of 1600p and 1920p, and for 1920p, the spacing error drops below 1% between 10k and 20k training steps.

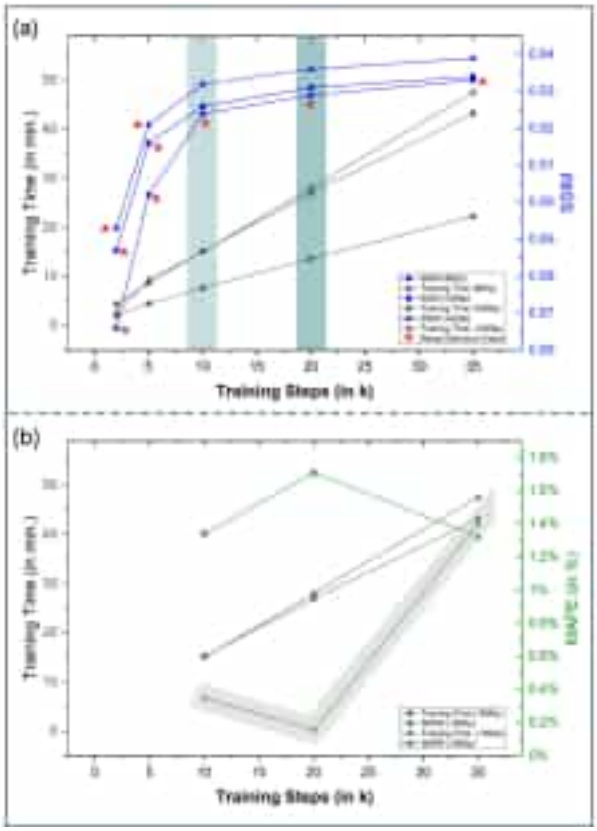


Figure 6. Training time versus training steps with SSIM, and (b) MAPE across different resolutions.

The maximum splat count strongly influenced detection reliability. Configurations with 0.1M and 0.2M splats frequently missed rebars, even at higher training steps. 0.5M splats improved detection but remained inconsistent,

whereas 1M splats consistently achieved 100% rebar detection across valid configurations. Increasing splats beyond this value did not yield further accuracy gains. Overall, based on these observations, the hyperparameter tuning phase identified 10k -20k training steps, 1M maximum splats, and 1920p downsampling as the most optimal and computationally efficient configuration. These values were fixed for all subsequent site-level parameter experiments.

Table 1. Hyperparameter variations and total number of experiments

Hyperparameters	Variation	#
Training steps	2k / 5k / 10k / 20k / 35k	5
Maximum splats	0.1M / 0.2M / 0.5M / 1.0M	4
Downsampling size	800p / 1600p / 1920p	3
	Total Experiment	**60**

4.2 Field-Level Parameters Results

A total of 24 field-level experimental trials were conducted to evaluate the influence of capture conditions on reconstruction quality (SSIM) and spacing accuracy (MAPE). The experiments systematically varied camera trajectory, illumination level, and recording resolution, with reconstructions evaluated at training steps of 10k and 20k. Under controlled acquisition conditions, the orbital camera path consistently achieved 100% rebar detection and exhibited minimal spacing error. For constraint scenarios, only path P02 enabled complete rebar detection across all trials. Therefore, subsequent analysis for constraint scenarios is limited to the P02 path.

Under bright daylight conditions, the orbital camera path consistently achieved low spacing error, with MAPE values remaining below approximately 2% across most configurations (Figure 7a). In contrast, the P02 path exhibited slightly higher, yet stable, error levels (4%), indicating reduced sensitivity to training duration and resolution. A rebar detection failure was observed for the orbital trajectory at 20k training steps using FHD input, highlighting the limitations of lower-resolution capture even under favourable illumination.

Moderate daylight conditions resulted in substantially different behaviour (Figure 7b). While controlled-path MAPE remained relatively constant (2-3%), orbital-path reconstructions showed pronounced error inflation for FHD recordings, with MAPE exceeding 12% at both 10k and 20k training steps. This degradation was mitigated primarily by UHD capture, which consistently reduced spacing error to below 4%, thereby increasing robustness to illumination-related feature loss. Across both lighting conditions, recording resolution emerged as the dominant site-level factor. UHD recordings consistently outperformed FHD recordings in all scenarios, maintaining less than 4% MAPE across all experimental trials. Overall, the experimental results indicate that site-level parameters can significantly impact rebar inspection performance, even when reconstruction parameters are optimized. Another point is that increasing the training steps from 10k to 20k roughly doubled the training time but produced only a limited improvement in accuracy, indicating that recording resolution is more critical than deeper hyperparameter optimization for 3DGS reconstruction.

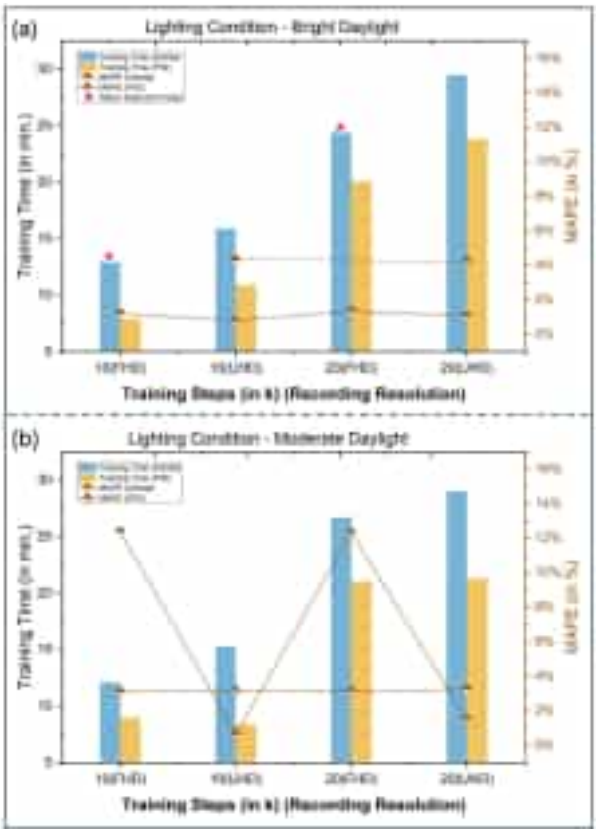


Figure 7. Training time and MAPE comparison for orbital and P02 camera path under (a) bright daylight. and (b) moderate daylight.

5 Site Validation

Site validation was conducted to verify whether the simulation-optimized parameters perform reliably under real-world precast yard conditions. A production U-girder reinforcement cage was evaluated under bright daylight, consistent with the simulated illumination setting.

5.1 Data Collection

For field validation, the capture strategies were intentionally designed to replicate the simulated strategies, including recording resolution, camera parameters, and

Table 2. Comparison across site validation, contrasting field data with its corresponding simulated data.

Trial	Recording Resolution	Training Steps (k)	Version	Training Time	SSIM	Detected Rebars (/28)	Detection Accuracy	MAPE (%)
1	FHD	10	S01	07:52	0.84	19	67.86%	–
			F01	10:25	0.65	27	96.43%	–
2	FHD	20	S02	20:03	0.87	23	82.14%	–
			F02	21:43	0.79	17	60.71%	–
3	UHD	10	S03	10:52	0.91	28	100.00%	4.31%
			F03	12:38	0.77	12	42.86%	–
4	UHD	20	S04	23:48	0.93	28	100.00%	4.23%
			F04	23:25	0.86	19	67.86%	–
			F04*	24:37	0.84	28	100.00%	15.03%

Note: S – Simulated, F – Field, F04* is data captured partially from the elevated ramp (shown in Figure 8b).

camera path. All recordings were captured from a smartphone device camera with fixed parameters across trials, matched to the simulation settings to ensure consistency with the underlying assumptions. Based on the experimental observations, both FHD and UHD recording resolutions, along with training steps of 10k and 20k, were selected, and the P02 camera path was fixed for data collection. All recordings were processed using the same multi-step rebar detection pipeline used in simulation. Evaluation was restricted to the same 28 rebars on the non-OHE wall to maintain consistency (See Figure 8a).

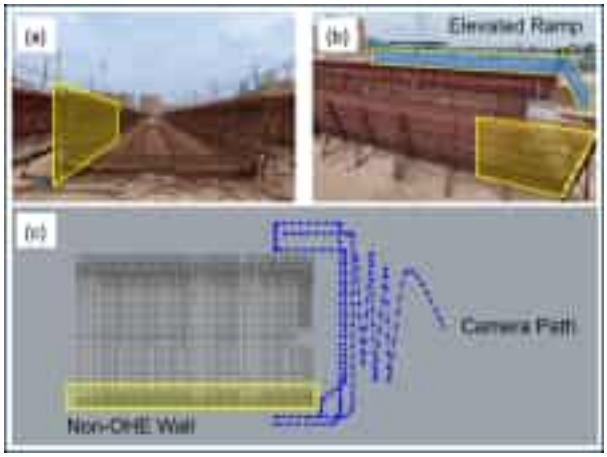


Figure 8. (a) U-girder reinforcement cage, (b) back-end isometric view highlighting elevated ramps, (c) schematic of the camera path used for site validation.

5.2 Field Validation Results

A total of five field samples were collected and processed using the rebar detection pipeline. The key parameters and their corresponding results are summarized in Table 2. Among the five samples, F01 and F02, captured in FHD resolution, failed to detect the rebars, identifying only 27 and 17 rebars, respectively. These results exactly match their corresponding simulated trials (S01 and S02), thereby classifying them as true negative cases. Although both cases resulted in incomplete detection, variations in detection accuracy between simulation and field trials indicate the influence of real-world factors, such as noise, illumination variability, and partial occlusions, that are not fully represented in the synthetic setup. Similarly, samples F03 and F04, which were captured in UHD, also failed to detect the rebars, detecting only 12 and 19 rebars, respectively. However, their replicated simulation cases (S03 and S04) were successful, classifying F03 and F04 as false positive cases.

In contrast to the four field trials that initially failed to achieve complete rebar detection, sample F04* was collected. The only variation in this case was that the capture was performed from a slightly elevated side view using an elevated ramp, as illustrated in Figure 8(b). In this sample, the rebar detection pipeline successfully detected all 28 rebars, with a MAPE of 15.03% in spacing estimation. Since the corresponding simulated case S04 also succeeded, this sample is considered a successful validation case and is treated as a true positive for outcome consistency, despite differences in capture geometry.



Figure 9. Confusion matrix summarizing agreement between simulation and field outcomes.

Figure 9 presents the confusion matrix derived from all evaluated samples. The key observations from the field trials indicate that under constrained scenarios, high-elevation capture is a decisive factor. Specifically, the introduction of a top-down perspective through elevated acquisition in F04* enabled complete

rebar detection, confirming that at least one elevated camera pass is essential for successful reconstruction in dense reinforcement layouts, consistent with the role of viewpoint diversity in camera path design (Section 3.2.1). Additionally, spacing errors in the successful field sample (F04*) were higher than those observed in simulation (approximately 15%), reflecting the impact of real-world noise, material variability, and uncontrolled occlusions. Finally, the consistent failure of F01 and F02 in both field and simulation confirms that the simulation framework accurately identified capture geometries unsuitable for reliable reconstruction.

6 Conclusion

Existing studies on automated rebar spacing detection are largely limited to controlled environments or simple geometries, with limited focus on complex configurations, such as U-girders, and on real-world feasibility. This study addresses this gap by investigating a 3DGS-based inspection pipeline previously developed by the authors using a structured, simulation-driven experimental design [8]. The experimental results established practical limits for key 3DGS reconstruction parameters (10-20k training steps, 1M splat count, and downsampling size 1920p) beyond which accuracy gains diminish while computational cost increases. Further experiments demonstrated that site-level capture conditions, particularly the camera path design and UHD recording resolution, were critical. While orbital trajectories performed best under unconstrained conditions, constrained scenarios required carefully designed paths, with outside-only scanning (P02) being the most reliable.

Field validation of metro U-girders demonstrated general correspondence between simulated predictions and real-world performance, particularly in identifying both successful and failed inspection scenarios. However, variations observed between simulation and field results, including differences within failure cases and the influence of elevated camera viewpoints, indicate that these relationships are not yet fully characterized. In particular, while elevated camera viewpoints were observed to improve reconstruction completeness in constrained scenarios, this factor was not systematically explored within the synthetic virtual environment and therefore requires further investigation. Similarly, differences in detection accuracy between simulation and field trials highlight the impact of real-world conditions, such as noise, occlusions, and illumination variability, that are not fully captured in the synthetic environment. Overall, while the proposed framework demonstrates promising practical applicability, further validation across diverse capture conditions, elevated viewpoints, and site scenarios is required to strengthen generalizability.

References

[1] M.-K. Kim, J. P. P. Thedja, and Q. Wang. Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3d point cloud data. *Automation in Construction*, 112:103077, April 2020. doi:10.1016/j.autcon.2020.103077.

[2] X. Yuan, A. Smith, R. Sarlo, C. D. Lippitt, and F. Moreu. Automatic evaluation of rebar spacing using lidar data. *Automation in Construction*, 131:103890, November 2021. doi:10.1016/j.autcon.2021.103890.

[3] B. Kim, S. P. K. R., Y. Natarajan, D. V, J. An, and D.-E. Lee. Real-time assessment of rebar intervals using a computer vision-based dvnet model for improved structural integrity. *Case Studies in Construction Materials*, 21:e03707, December 2024. doi:10.1016/j.cscm.2024.e03707.

[4] S. Li, B. Zhang, J. Zheng, D. Wang, and Z. Liu. Development of automated 3d lidar system for dimensional quality inspection of prefabricated concrete elements. *Sensors*, 24(23):7486, November 2024. doi:10.3390/s24237486.

[5] S. Wang, L. Deng, J. Guo, M. Liu, and R. Cao. Automatic quality inspection of rebar spacing using vision-based deep learning with rgbd camera. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC)*, Lille, France, June 2024. doi:10.22260/ISARC2024/0009.

[6] A. H. Qureshi, W. S. Alaloul, A. Murtiyoso, S. J. Hussain, S. Saad, and M. A. Musarat. Smart rebar progress monitoring using 3d point cloud model. *Expert Systems with Applications*, 249:123562, September 2024. doi:10.1016/j.eswa.2024.123562.

[7] S. Liu, M. Yang, T. Xing, and R. Yang. A survey of 3d reconstruction: The evolution from multi-view geometry to nerf and 3dgs. *Sensors*, 25(18):5748, September 2025. doi:10.3390/s25185748.

[8] K. Adarsh, N. Bugalia, and A. Pal. Development of a vr-based testbed for automated rebar inspection in infrastructure projects. In *ASCE Construction Research Congress*, San Antonio, Texas, March 2026. Accepted.

[9] D. Rangelov, S. Waanders, K. Waanders, M. Van Keulen, and R. Miltchev. Three-dimensional reconstruction techniques and the impact of lighting conditions on reconstruction quality: A comprehensive review. *Lights*, 1 (1):1, July 2025. doi:10.3390/lights1010001.

[10] D. Rangelov, S. Waanders, K. Waanders, M. V. Keulen, and R. Miltchev. Impact of data capture methods on 3d reconstruction with gaussian splatting. *Journal of Imaging*, 11(2):65, February 2025. doi:10.3390/jimaging11020065.

[11] G. Grubert, F. Barthel, A. Hilsmann, and P. Eisert. Improving adaptive density control for 3d gaussian splatting. In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 610–621, Porto, Portugal, 2025. SCITEPRESS. doi:10.5220/0013308500003912.

[12] A. Jaiswal, N. Bugalia, and Q. Phuc Ha. Optimizing structure from motion parameters for volume estimation of construction and demolition waste. In *Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC)*, Montreal, Canada, July 2025. doi:10.22260/ISARC2025/0127.

A structured Prompt Engineering Framework for Modular Construction Design Using Generative AI

Zaid Alwashah[1], Bo Xiao[2], Hexu Liu[1], Xiaoyun Shao[1], and Xiaoman Wang[3]

[1]Department of Civil and Construction Engineering, Western Michigan University, Michigan, USA
[2]Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, Michigan, USA
[3]Department of Educational Leadership, Research and Technology, Western Michigan University, Michigan, USA

zaid.alwashah@wmich.edu, boxiao@mtu.edu, hexu.liu@wmich.edu, xiaoyun.shao@wmich.edu, xiaoman.wang@wmich.edu

Abstract –
The modular construction industry requires early and tightly coupled reasoning across spatial layout, structural systems, module interfaces, and building services under fabrication, transportation, and assembly constraints. With recent advances in generative artificial intelligence (AI) and the emergence of large language models (LLMs), there is growing interest in leveraging these technologies to support design reasoning through natural language interaction. However, existing studies lack domain-specific methodologies that systematically translate modular design reasoning into effective prompts for generative AI systems. This study addresses this gap by introducing a reasoning-oriented taxonomy of modular construction design tasks and a structured prompt engineering framework tailored to modular design contexts. The proposed framework formalizes prompt construction through standardized blocks and demonstrates its application using representative prompts grounded in a modular residential case study. The proposed framework provides a methodological foundation for reasoning-aware human–AI interaction in early-stage modular construction design.

Keywords –
Modular Construction; Prompt Engineering; Large Language Models; Design Reasoning; BIM

1 Introduction

Modular construction has gained increasing attention as an alternative to conventional on-site construction due to its potential to improve productivity, quality control, and construction safety by shifting substantial portions of the building process from site environments to controlled factory settings [1], [2]. Prior research consistently indicates that off-site and modular construction can reduce schedule uncertainty, enhance manufacturing precision, and mitigate safety risks associated with labor-intensive site activities, particularly in residential and mid-rise building projects [3], [4]. Despite these advantages, modular construction introduces a distinct set of design and coordination challenges that differ fundamentally from traditional cast-in-situ construction. Modular projects require early and tightly coupled reasoning across spatial layout, structural systems, module interfaces, and building services to ensure manufacturability, transportability, and on-site assembly feasibility [1], [5]. Errors or inconsistencies introduced during the early design stages are difficult and costly to correct once production has commenced, highlighting the importance of structured, front-loaded design decision-making [6]. Building Information Modeling (BIM) has been widely adopted to support modular and off-site construction by enabling the integrated representation of geometry, components, and construction information across disciplines [2], [7]. However, prior BIM-based approaches largely rely on manual modelling, rule-based automation, or discipline-specific scripts, which remain time-consuming and require significant expert input [8], [9], [10]. As a result, BIM tools alone provide limited support for the reasoning-intensive nature of early modular design, where designers must iteratively explore layout configurations, assess structural logic, and anticipate inter-module and service coordination constraints.

Recent advances in generative artificial intelligence (AI), particularly large language models (LLMs), have demonstrated strong capabilities in interpreting natural language instructions, supporting design reasoning, and generating structured responses across complex problem domains [11], [12]. In parallel

domains, LLM performance has been shown to depend strongly on prompt formulation, with structured prompt engineering strategies significantly improving reasoning quality, consistency, and controllability [13], [14].

Within the construction and BIM context, emerging studies have begun exploring the use of LLMs for tasks such as automated compliance checking, design interpretation, and limited forms of reasoning support [11], [15]. However, current LLM research in the AEC and BIM space is still reported largely as isolated, task-level prototypes, such as BIM information search assistants, schedule or sequence generation, and chatbot-style query systems, rather than as reusable prompting methodologies that generalize across design reasoning tasks [16]. Moreover, recent AEC reviews explicitly note that outcomes are highly prompt-dependent, yet the prompt practices documented across studies vary widely (e.g., template-based, few-shot, role-based, chain-of-thought, iterative prompting), indicating limited consolidation into a consistent and transferable prompting approach for design reasoning [16], [17], [18], [19]. As a result, there is limited guidance on how to systematically translate reasoning-intensive modular design problems, such as the coupling of layout, structure, interfaces, and MEP systems under fabrication and transport constraints, into robust prompt structures that can be reused across modular design stages.

To address this gap, this study proposes a structured prompt engineering approach tailored to modular construction design. The research contributes (1) a task taxonomy that categorizes modular design activities based on their dominant reasoning requirements, (2) a structured prompt engineering framework that formalizes how modular design tasks can be expressed as prompts, and (3) representative prompt examples grounded in a realistic modular residential design scenario. Rather than evaluating model performance, the focus of this work is on clarifying the relationship between modular design reasoning and prompt structure, thereby providing a foundational framework for future generative AI applications in modular construction design.

2 Literature Review

Prior research on modular and off-site construction has primarily emphasized process optimization, production efficiency, and lifecycle performance. Reviews of prefabricated and modular construction management consistently highlight that modular projects require earlier design finalization and stronger cross-disciplinary coordination than conventional construction, due to the interdependence between manufacturing, transportation, and on-site assembly decisions [3], [4]. These studies emphasize that design decisions related to layout, structural systems, and service integration must be resolved up front to avoid costly rework once factory production begins, underscoring the reasoning-intensive nature of early modular design stages. Moreover, BIM has been widely adopted for providing integrated digital representations of building geometry, components, and construction information. Prior studies demonstrate that BIM facilitates coordination across design and construction disciplines in modular and off-site projects, enabling improved visualization, clash detection, and constructability analysis [2], [20]. Recent advancements in AI within construction engineering have introduced data-driven and learning-based approaches for design support, optimization, and automation. These applications commonly employ knowledge-based and machine-learning techniques for planning, monitoring, and decision support. These approaches often rely on predefined rules or structured project data (e.g., BIM/event logs), and several methods still face constraints such as limited adaptability and knowledge acquisition bottlenecks [21], [22]. These limitations are particularly pronounced in conceptual design contexts, where problem definitions are incomplete and reasoning requirements evolve dynamically.

More recently, generative AI and LLMs have demonstrated strong capabilities in interpreting natural language, generating structured outputs, and supporting reasoning-oriented tasks across engineering domains. Studies in engineering design report that LLMs can assist with concept generation, explanation, and design reasoning, particularly when problems are expressed in well-structured textual forms [23]. Within the construction and BIM domain, early investigations have applied LLMs to automated compliance checking and information retrieval tasks, showing that natural language interfaces can improve accessibility to BIM data and rule interpretation [15], [24]. However, these applications primarily focus on downstream validation or query tasks rather than upstream design reasoning.

Parallel research in prompt engineering has shown that LLM performance is highly sensitive to how tasks are framed. Studies indicate that structured prompt formulations, including explicit context definition, constraint specification, and reasoning objectives, significantly improve model reliability, interpretability, and reasoning depth [14], [25]. Design tasks differ from conventional NLP benchmarks in that they require spatial reasoning, constraint satisfaction, and multi-objective trade-off exploration. However, prompt engineering research remains largely domain-agnostic, offering limited guidance on how domain-specific reasoning tasks, such as those in modular construction design, should be translated into effective prompts. This limitation is critical for modular construction design, where early-stage decisions must simultaneously satisfy

manufacturability, transportation, and assembly constraints. Within modular construction research, computational design studies increasingly recognize the need for integrated reasoning across layout, structure, and services.

Taken together, the literature reveals a clear gap at the intersection of modular construction design, BIM-based workflows, and generative AI. While BIM provides an integrated data environment and LLMs offer powerful natural language reasoning capabilities, there is limited research on how modular design tasks can be systematically translated into structured, reasoning-oriented prompts that align with both domain constraints and principles of prompt engineering. This gap motivates the development of a domain-specific prompt engineering framework that explicitly connects modular design task characteristics with structured prompt formulations, enabling more effective use of generative AI during early modular design stages.

3 Research Methodology

This study adopts a design-oriented methodological approach to develop a structured prompt engineering framework grounded in modular construction design reasoning. The methodology systematically organizes modular design tasks and links design reasoning requirements to prompt formulation for generative AI systems. Figure 1 presents an overview of the research methodology.

The methodology consists of four sequential steps. First, modular design reasoning was analyzed using insights from modular and off-site construction literature and the representative case, identifying key design reasoning characteristics that distinguish modular construction from conventional building design, particularly in early design stages. These reasoning characteristics highlight the need for early coordination across spatial layout, structural systems, module interfaces, and building services under fabrication and transportation constraints. Second, building on the reasoning characteristics identified in Step 1 and the analysis of a representative modular residential scenario, modular construction design activities were decomposed and organized into a reasoning-oriented task taxonomy. For example, in the representative modular residential scenario (Section 6), the need to maintain vertical alignment of wet areas across stacked units, such as bathrooms and kitchens, was identified in Step 1 as a key reasoning constraint and formalized in Step 2 under layout and MEP coordination tasks. The taxonomy emphasizes core reasoning dimensions, such as spatial layout logic, structural continuity, module connection consistency, and building service coordination, rather than software tools or implementation-specific modelling techniques.

Third, the task taxonomy developed in Step 2 was used as a conceptual bridge to structure prompt formulation, drawing on established prompt engineering concepts such as context specification, explicit task definition, constraint articulation, and expected output structuring. This step resulted in a structured prompt engineering framework linking modular design task characteristics with prompt components and reasoning objectives. Finally, a representative modular residential case study was used to illustrate the application of the proposed taxonomy and structured prompt engineering framework. This step demonstrates how modular design tasks are translated into reasoning-oriented prompts, clarifying the coherence and applicability of the proposed approach, as detailed in Sections 4–6 corresponding to the task taxonomy, structured prompt framework, and case study components of the methodology (Figure 1).

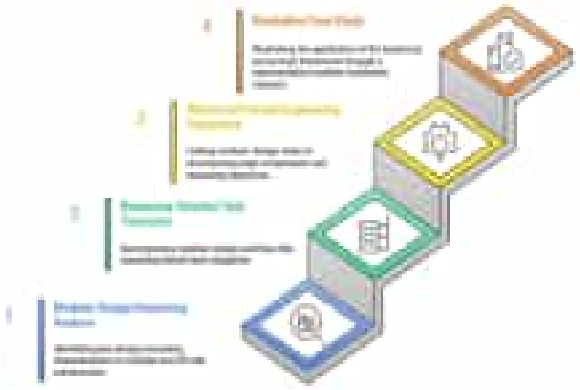


Figure 1. Research workflow for developing the prompt engineering framework.

4 Modular Construction Design Task Taxonomy

Modular construction design consists of a set of interconnected tasks that differ significantly from traditional on-site building design. This distinction stems from the production-oriented nature of modular construction, which requires off-site fabrication, transportation, and on-site assembly. Consequently, coordinated reasoning is needed for spatial layout, structural systems, inter-module connections, and building services, including zoning rooms within transport-constrained module envelopes, maintaining load-path continuity across stacked modules, aligning module interfaces for vertical stacking, and preserving vertical continuity of building services. As a result, design effort is shifted toward early-stage reasoning and decision-making rather than downstream detailing. Given this level of interdependence, the effective use of

generative AI in modular construction relies on understanding the types of design tasks and associated reasoning processes needed. In response to this need, this study presents a taxonomy of modular construction design tasks that organize these activities based on their primary reasoning requirements. The taxonomy is derived from a realistic modular residential design using a design decomposition approach, in which design decisions, constraints, and coordination requirements were identified and iteratively grouped according to their dominant reasoning characteristics. Task categorization was guided by the primary reasoning required to resolve each requirement, including spatial layout, structural continuity, module interface coordination, and building service integration. This approach ensures conceptual consistency across task categories and supports direct mapping between design reasoning and prompt formulation. To emphasize reasoning over implementation details, the taxonomy organizes modular design into reasoning-driven categories rather than software-specific modelling processes. Accordingly, Figure 2 presents the proposed modular construction design task taxonomy, derived through design decomposition of the case scenario, illustrating how design decisions, constraints, and coordination requirements are grouped based on their dominant reasoning characteristics and serving as the foundation for the prompt engineering framework introduced in the following section.



Figure 2. Reasoning-oriented task taxonomy derived from modular design decomposition.

Layout design tasks address spatial zoning, circulation organization, room adjacency logic, and compliance with transport envelopes and minimum dimensional requirements. In contrast, structural design tasks focus on framing logic, load-path continuity, and structural consistency across stacked modules, particularly under conditions of transport, lifting, and in-service use (e.g., maintaining vertical load transfer across module boundaries). Building upon these considerations, module connection tasks involve reasoning about vertical stacking alignment, horizontal joining strategies, continuity of load transfer surfaces, and fire or acoustic separation across module interfaces (e.g., alignment tolerances at vertical module interfaces). In parallel, MEP coordination tasks address service zoning, plumbing stack alignment, clash awareness, and continuity of building services across modules (e.g., preserving vertical service continuity across stacked units).

5 Structured Prompt Engineering Framework

Modular construction tasks differ in scope and complexity; therefore, their effective use of generative AI systems relies on how these tasks are converted into prompts. Unstructured prompts often lead to unclear or incomplete reasoning, particularly under modular fabrication and transportation constraints. To tackle this issue, this study proposes a structured prompt engineering framework specifically for modular construction design tasks. Figure 3 illustrates the proposed framework, showing how modular design tasks are translated into structured prompts through standardized prompt blocks and instantiated as different prompt types depending on the intended reasoning objective. The framework standardizes how prompts are created by breaking each prompt into a set of functional blocks that guide AI reasoning and minimize ambiguity. It is intentionally not tied to any specific tool and aims to support design reasoning across various generative AI systems, including general-purpose LLMs and BIM-integrated design assistants.

Each prompt is composed of five core elements. A Context Block defines the modular design scenario, including building type, modular system characteristics, and high-level constraints. A Task Definition Block explicitly states the intended action, such as generating, analyzing, checking, or revising a design aspect. A Constraints Block specifies relevant dimensional, structural, code-related, or modular limitations that must be respected. An Expected Output Block defines the desired response format, such as structured text, stepwise reasoning, or descriptive lists. Finally, a Validation and Guardrails Block requests explicit checks for known failure points or critical conditions, discouraging unsupported assumptions or hallucinated design elements. The five core blocks serve as a reference structure rather than a rigid template, allowing prompt configurations to be adapted, simplified, or extended in accordance with task complexity and specific reasoning requirements.

Using this structure, modular design tasks can be systematically linked to different prompt types, such as directive prompts for generation, analytical prompts for explanation, diagnostic prompts for checking consistency, and corrective prompts for refinement. Each prompt type emphasizes a different reasoning objective while retaining the same underlying block structure.

Reusable prompt templates are created by filling in the structured blocks with task-specific information and constraints, allowing for consistent prompt development across various modular design scenarios and supporting iterative prompt reuse as design objectives evolve.

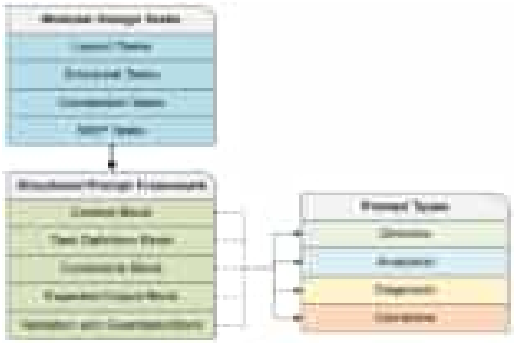


Figure 3. Structured prompt engineering framework for modular construction design tasks

6 Modular Case Study and Prompt Instantiation

This section presents a representative modular residential case study that demonstrates how the proposed task taxonomy and prompt engineering framework can be applied in practice. The purpose of the case study is not to validate the performance of automated modelling or generate construction-ready outputs, but rather to provide a realistic design context for prompt formulation and reasoning-based interaction with generative AI systems.

6.1 Scenario Definition

The demonstration scenario is based on a low-rise residential building composed of vertically stacked modular units, representative of common volumetric modular construction practices. The building configuration consists of two primary above-grade levels organized around a central circulation core, with repeated and aligned spatial layouts across floors. A sectional axonometric view of the case study is shown in Figure 4, illustrating the stacked modular arrangement, floor separation, and overall spatial organization. The internal program reflects a typical residential layout, including living, dining, and kitchen spaces on the lower level, and bedroom and bathroom spaces on the upper level. Wet areas and service zones are vertically aligned, reflecting common modular strategies that support service continuity and inter-module coordination. The scenario is constrained by module dimensions and the need for vertical alignment of structural and service elements across stacked units. Structural and MEP systems are represented at a conceptual level to highlight coordination logic, interface relationships, and service alignment, without implying finalized engineering specifications.



Figure 4. Modular residential case illustrating design stages aligned with task categories and prompt types

6.2 Example Prompts Based on Modular Design Tasks

Using the modular residential scenario illustrated in Figure 4, a set of representative prompts was created to demonstrate how the modular design task taxonomy (Section 4) can be operationalized using the structured prompt engineering framework introduced in Section 5. The examples presented in this section are designed to show how different modular design tasks can be translated into clear, reasoning-oriented prompts suitable for generative AI systems.

Although the representative prompts are expressed at an abstract level, they are implicitly grounded in the modular residential scenario described in Section 5.1, including vertically stacked units, aligned wet areas, centralized circulation, and transport-constrained module dimensions. These scenario characteristics inform the intent and constraints of each prompt without requiring explicit numerical parameterization.

Each example prompt follows a consistent structure, including a clearly defined context, an explicit task specification, relevant modular constraints, and an articulated expected output and validation intent. This structure enables a direct mapping between the modular design task categories (layout, structural, module connection, and MEP coordination) and the corresponding prompt types (directive, analytical, diagnostic, and corrective) defined in the proposed framework. In this mapping, the task category specifies the design aspect under consideration (e.g., layout or structure), the prompt type defines the intended reasoning action (e.g., generation or diagnosis), and the structured prompt blocks provide the mechanism for expressing this intent in a consistent and explicit form.

Table 1 summarizes the representative prompts used in this study, illustrating how each modular design task category is paired with an appropriate prompt type and

expressed as a concise, reasoning-oriented instruction. The examples in Table 1 demonstrate how abstract modular design tasks can be systematically translated into clear and reusable prompts suitable for generative AI systems, while maintaining alignment with the reasoning-driven taxonomy presented earlier.

Table 1. Representative prompts mapped to modular design task categories and prompt types (presented as concise, illustrative instructions)

Task Category	Prompt Type	Representative Prompt
Layout	Directive	Generate a spatial layout for a transport-constrained modular unit that minimizes corridor length while preserving privacy and maintaining vertical alignment of wet areas and service zones.
Structural	Analytical	Explain how load-path continuity can be maintained across vertically stacked modular units, including across module interfaces.
Connection	Diagnostic	Identify potential alignment issues at vertical module interfaces in a stacked modular configuration.
MEP	Diagnostic	Check whether the layout supports vertical continuity of plumbing and mechanical services across modules.

To further illustrate the added value of the proposed task taxonomy and structured prompt engineering framework, the following expanded example demonstrates how a modular design task can be systematically translated into an explicit, reusable prompt structure, rather than relying on ad hoc or user-specific prompt formulations, using a directive layout design task based on the modular residential scenario in Figure 4. For clarity, the representative prompts in Table 1 are intentionally expressed in concise form, whereas the expanded example below demonstrates how a single prompt can be fully instantiated using the same block structure defined in Section 4, without loss of generality across other prompt types.

Context Block: Consider a low-rise residential building composed of vertically stacked modular units with aligned floor layouts, centralized circulation, and vertically continuous wet areas, representative of the modular configuration shown in Figure 3.

Task Definition Block: Generate a conceptual spatial layout for a single modular unit that supports the overall building configuration and preserves the stated stacking and service-alignment logic across levels.

Constraints Block: The layout must respect transport-constrained module dimensions, preserve privacy between primary living and sleeping spaces, and maintain vertical alignment of wet areas and service zones across stacked units.

Expected Output Block: Provide a stepwise explanation of the proposed layout strategy, describing room zoning, circulation organization, and service alignment logic.

Validation and Guardrails Block: Verify that the proposed layout maintains vertical service continuity of wet areas and service zones, avoids introducing structural claims not supported by the stated scenario, and does not introduce spaces that violate modular transport or stacking constraints.

6.3 Implications of Structured Prompting for Modular Design Tasks

The examples presented in Section 6.2, grounded in the modular residential scenario illustrated in Figure 4 and summarized in Table 1, demonstrate how modular design tasks can be systematically translated into structured prompts, providing a common and repeatable way to express design reasoning that does not depend solely on individual user experience or personal prompt-writing skills by decomposing each task into explicit context, task intent, constraints, expected outputs, and validation objectives, the proposed framework transforms implicit design reasoning into more transparent, communicable, and reusable representation for interaction with generative AI systems.

From a modular design perspective, this structured prompting approach supports clearer separation between different types of design reasoning. Layout tasks emphasize spatial organization and adjacency logic, structural tasks focus on continuity and alignment across stacked modules, connection tasks highlight interface consistency, and MEP coordination tasks address service zoning and vertical continuity. The framework offers a consistent vocabulary and structure for expressing modular design intent, rather than relying on ad hoc or experience-driven prompt formulations.

More broadly, the framework contributes a task-driven method for prompt formulation in modular construction design. Rather than treating prompts as isolated or tool-specific inputs, the proposed approach frames them as reusable design artifacts that encode

domain knowledge and reasoning expectations. This supports knowledge transfer across designers, projects, and design stages, and offers a conceptual foundation for future integration with BIM-based workflows, design assistants, and decision-support tools, without requiring immediate automation or performance benchmarking. In a BIM-integrated context, the proposed framework can act as an intermediate reasoning layer between structured BIM data and generative AI systems. Inputs may be extracted from model properties, schedules, or IFC-based representations (e.g., space lists, module grids, dimensional envelopes, and coordination zones) and translated into prompt context and constraints. The resulting outputs represent structured reasoning responses, such as layout recommendations or coordination checks, which can be mapped back to BIM actions including model revision, alignment validation, and issue identification.

7 Conclusion

This study addressed a critical gap at the intersection of modular construction design, BIM-based workflows, and generative AI by proposing a domain-specific prompt engineering framework grounded in modular design reasoning. Modular construction demands early and integrated decision-making across layout, structure, module connections, and building services, yet existing AI and BIM-based approaches provide limited guidance on how such reasoning-intensive tasks should be translated into effective prompts for generative AI systems.

To respond to this challenge, the study first developed a reasoning-oriented taxonomy of modular construction design tasks, derived from a production-oriented residential modular scenario. By organizing modular design activities according to their primary reasoning requirements rather than implementation tools or modelling techniques, the taxonomy establishes a conceptual bridge between modular design logic and prompt formulation. Building on this foundation, a structured prompt engineering framework was introduced, defining a consistent set of prompt blocks and linking modular design tasks to distinct prompt types aligned with different reasoning objectives.

The Framework was illustrated through a representative modular residential case study and a set of example prompts, including a fully instantiated prompt demonstrating how contextual information, constraints, and validation objectives can be explicitly encoded. These examples illustrate how structured prompting can represent modular design reasoning as clear, reusable instructions suitable for interaction with generative AI systems, without requiring construction-ready outputs or performance benchmarking. The primary contribution of this work lies in its methodological framing rather than technical evaluation. By explicitly connecting modular design reasoning with structured prompt engineering principles, the proposed approach supports more transparent and controllable human–AI collaboration during early-stage modular design. Future research may extend this framework by applying it to additional building typologies, integrating it with BIM-based design assistants, or empirically evaluating its impact on design quality, efficiency, and user trust in generative AI–supported modular construction workflows. While the proposed framework provides a structured approach to prompt formulation, its impact on reasoning quality and performance has not been quantitatively evaluated in this study and remains an important direction for future research.

Acknowledgments

We gratefully acknowledge the financial support of the U.S. National Science Foundation (NSF) through its CyberTraining – Training program (Award # 2417804).

References

[1] R. M. Lawson, R. G. Ogden, and R. Bergin, “Application of Modular Construction in High-Rise Buildings,” *J. Archit. Eng.*, vol. 18, no. 2, pp. 148–154, Jun. 2012, doi: 10.1061/(ASCE)AE.1943-5568.0000057.

[2] X. Yin, H. Liu, Y. Chen, and M. Al-Hussein, “Building information modelling for off-site construction: Review and future directions,” *Automation in Construction*, vol. 101, pp. 72–91, May 2019, doi: 10.1016/j.autcon.2019.01.010.

[3] Z. Li, G. Q. Shen, and X. Xue, “Critical review of the research on the management of prefabricated construction,” *Habitat International*, vol. 43, pp. 240–249, 2014.

[4] M. Kamali and K. Hewage, “Life cycle performance of modular buildings: A critical review,” *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 1171–1183, Sep. 2016, doi: 10.1016/j.rser.2016.05.031.

[5] X. Zhai, R. Reed, and A. Mills, “Factors impeding the offsite production of housing construction in China: an investigation of current practice,” *Construction Management and Economics*, vol. 32, no. 1–2, pp. 40–52, Feb. 2014, doi: 10.1080/01446193.2013.787491.

[6] S. Gao, R. Jin, and W. Lu, “Design for manufacture and assembly in construction: a review,” *Building Research & Information*, vol. 48, no. 5, pp. 538–550, Jul. 2020, doi: 10.1080/09613218.2019.1660608.

[7] R. Volk, J. Stengel, and F. Schultmann, “Building Information Modeling (BIM) for existing buildings — Literature review and future needs,” *Automation in Construction*, vol. 38, pp. 109–127, Mar. 2014, doi: 10.1016/j.autcon.2013.10.023.

[8] S. Azhar, “Building Information Modeling (BIM): Trends, Benefits, Risks, and Challenges for the AEC Industry,” *Leadership Manage. Eng.*, vol. 11, no. 3, pp. 241–252, Jul. 2011, doi: 10.1061/(ASCE)LM.1943-5630.0000127.

[9] R. Sacks, C. M. Eastman, and G. Lee, “Parametric 3D modeling in building construction with examples from precast concrete,” *Automation in Construction*, vol. 13, no. 3, pp. 291–312, May 2004, doi: 10.1016/S0926-5805(03)00043-8.

[10] J. Zhang and N. M. El-Gohary, “Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking,” *J. Comput. Civ. Eng.*, vol. 30, no. 2, p. 04015014, Mar. 2016, doi: 10.1061/(ASCE)CP.1943-5487.0000346.

[11] Z. Alwashah, B. Xiao, H. Liu, S. T. Mueller, and X. Shao, “Generative artificial intelligence for construction: Use cases, trends, challenges, and opportunities,” *Journal of Building Engineering*, vol. 112, p. 113802, Oct. 2025, doi: 10.1016/j.jobe.2025.113802.

[12] P. Ghimire, K. Kim, and M. Acharya, “Opportunities and Challenges of Generative AI in Construction Industry: Focusing on Adoption of Text-Based Models,” *Buildings*, vol. 14, no. 1, p. 220, Jan. 2024, doi: 10.3390/buildings14010220.

[13] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing,” *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–35, Sep. 2023, doi: 10.1145/3560815.

[14] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, et al., “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 24824–24837, 2022.

[15] O. Iversen and L. Huang, “Leveraging large language models for BIM-based automated compliance checking,” *Automation in Construction*, vol. 182, p. 106707, Feb. 2026, doi: 10.1016/j.autcon.2025.106707.

[16] D. Onatayo, A. Onososen, A. O. Oyediran, H. Oyediran, V. Arowoiya, and E. Onatayo, “Generative AI Applications in Architecture, Engineering, and Construction: Trends, Implications for Practice, Education & Imperatives for Upskilling—A Review,” *Architecture*, vol. 4, no. 4, pp. 877–902, Oct. 2024, doi: 10.3390/architecture4040046.

[17] C. Li, T. Zhang, X. Du, Y. Zhang, and H. Xie, “Generative AI Models for Different Steps in Architectural Design: A Literature Review,” Oct. 23, 2024, *arXiv*: arXiv:2404.01335. doi: 10.48550/arXiv.2404.01335.

[18] M. H. G. Rad and M. Ilbeigi, “Generative AI in Lean Construction: a Scoping Review,” presented at the 33rd Annual Conference of the International Group for Lean Construction (IGLC 33), Osaka and Kyoto, Japan, Jun. 2025, pp. 953–964. doi: 10.24928/2025/0195.

[19] H. Wan, J. Zhang, Y. Chen, W. Xu, and F. Feng, “Generative AI application for building industry,” *arXiv preprint arXiv:2410.01098*, 2024.

[20] H. Liu, C. Sydora, M. S. Altaf, S. Han, and M. Al-Hussein, “Towards sustainable construction: BIM-enabled design and planning of roof sheathing installation for prefabricated buildings,” *Journal of Cleaner Production*, vol. 235, pp. 1189–1201, Oct. 2019, doi: 10.1016/j.jclepro.2019.07.055.

[21] Y. Pan and L. Zhang, “Roles of artificial intelligence in construction engineering and management: A critical review and future trends,” *Automation in Construction*, vol. 122, p. 103517, Feb. 2021, doi: 10.1016/j.autcon.2020.103517.

[22] L. Zhang, Y. Pan, X. Wu, and M. J. Skibniewski, *Artificial Intelligence in Construction Engineering and Management*, vol. 163. in Lecture Notes in Civil Engineering, vol. 163. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-16-2842-9.

[23] F. Chiarello, S. Barandoni, M. Majda Škec, and G. Fantoni, “Generative large language models in engineering design: opportunities and challenges,” *Proc. Des. Soc.*, vol. 4, pp. 1959–1968, May 2024, doi: 10.1017/pds.2024.198.

[24] S. Hellin, S. Nousias, and A. Borrmann, “Natural language information retrieval from BIM models: An LLM-based agentic workflow approach,” in *Proc. European Conference on Computing in Construction (EC3)*, 2025.

[25] L. Reynolds and K. McDonell, “Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm,” in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–7. doi: 10.1145/3411763.3451760.

Semantic Point-of-Interest–Driven IFC-Based Robot Navigation in Indoor Environments

Muhammad Shoaib Khan[1], Zhihao REN[2], Ho Jin Lee[1], Seonghyeun Kim[1], Woo-yong Jung[2], Jung In Kim[1]

[1] School of Civil and Environmental Engineering, Kookmin University, Seoul, 02707, Republic of Korea
[2] Department of Architecture and Civil Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR
[3] Dept. of Nuclear Power Plant Engineering, KEPCO International Nuclear Graduate School, Ulsan, 45014, Korea
muhammadshoaib5308@gmail.com, z.ren@my.cityu.edu.hk, leehojin3239@kookmin.ac.kr, tjd1103@kookmin.ac.kr, trustjung@gmail.com, jikim07@kookmin.ac.kr

Abstract –

Autonomous mobile robots deployed in building environments require comprehensive spatial semantics for navigation and operation. However, while current Industry Foundation Classes (IFC) standards provide both geometric and semantic information, they lack robot-operational attributes. This paper presents a Point-of-Interest (POI) framework that enriches IFC building models with robot-oriented behavioural intelligence through function-based mapping rules. The proposed semantic enrichment methodology utilizes IfcOpenShell to augment standard IFC entities with custom property sets that encode POI definitions, thereby enabling the integration of building geometry and robot semantics within Unity-based virtual reality simulation environment. The framework is validated through various scenarios, including doorway accessibility assessment, multi-robot coordination, and wall-following tasks, demonstrating its effectiveness and applicability. The approach establishes a foundation for proactive robot-compatible building design and supports the development of automated accessibility evaluation tools.

Keywords –

BIM-IFC; Point of interest (POI), robot-friendly design, semantic enrichment, BIM to robot, virtual reality

1 Introduction

The deployment of autonomous mobile robots in building environments has become increasingly prevalent [1]. For robots to navigate effectively, they require comprehensive spatial understanding. Robots must identify navigable pathways, detect and avoid obstacles, recognize transition points between spaces, locate service stations for charging and maintenance, and coordinate movements with other robots and human occupants [2]. These operational requirements necessitate rich semantic information about the built environment that encodes not merely what spaces and objects are, but how they function in relation to robotic navigation and task execution [3].

Navigation environment modeling for autonomous systems is conventionally categorized into three types based on spatial representation: metric-based models (grid or voxel discretization), topology-based models (graph structures with nodes and edges), and network models that combine geometry with topology [4–6]. Metric-based approaches tessellate space into cells with pathfinding algorithms such as A* and Dijkstra applied for shortest path analysis, but suffer from computational overhead and storage requirements [7]. Topology-based models provide compact graph structures but lack the geometric precision necessary for operational robot guidance [4]. Network models integrate geometric coordinates with topological connectivity through representations including Visibility Graphs, Voronoi Diagrams, and Medial Axis Transforms, yet these geometry-derived methods face scalability challenges in complex building environments and struggle with irregular floor layouts [8].

Building Information Modeling (BIM) has emerged as the dominant framework for digitally representing building assets throughout their entire lifecycle, from initial design conception through construction execution to long-term operational management [9,10]. The Industry Foundation Classes (IFC) standard, maintained by BuildingSMART International, provides the foundational data schema for BIM interoperability, defining standardized representations for building elements, spatial structures, material properties, and functional relationships[11]. The prior studies have used

IFC data to construct navigation models, such as voxel-based maps, which are generated using IFC geometry [6], and a graph-based network for building-level connectivity is developed using IFC semantics [12].

However, current IFC specifications, while comprehensive for human-centric building design and construction management, exhibit fundamental limitations when applied to robot navigation and operation contexts. IFC schemas define building elements: walls, doors, floors, and spaces, primarily through properties relevant to architectural design, structural engineering, and building code compliance. Attributes such as wall thickness, door fire ratings, floor load capacities, and room occupancy classifications serve architectural and engineering purposes effectively, yet these properties do not directly address robotic operational requirements. Critical robot-navigation parameters remain largely absent from standard IFC property definitions. Consequently, existing BIM-driven navigation studies cannot directly support robot deployment planning, as the authoritative building information models do not contain the behavioral and operational attributes necessary to inform robot path planning, obstacle avoidance, and service execution algorithms.

Addressing this gap requires a systematic approach to the semantic enrichment of building information models with robot-oriented spatial intelligence [13]. Recent research has introduced Point-of-Interest (POI) frameworks as a mechanism for redefining building elements from robotic behavioral perspectives. Unlike geographic POIs that mark locations of human interest, such as landmarks or commercial establishments, robot-oriented POIs represent spatial elements that directly influence autonomous navigation decisions and task execution. A door, for instance, functions not merely as a physical barrier with dimensional properties, but as a transition element that governs spatial accessibility under specific clearance and passage constraints. A wall serves simultaneously as an obstacle requiring avoidance and as a reference boundary for path guidance. By assigning such behavioral roles to building elements, POI frameworks transform static architectural representations into semantically rich spatial models that encode how robots should perceive, interpret, and interact with their operational environments.

This paper presents a comprehensive framework for semantic POI-based robot navigation that leverages IFC through systematic semantic enrichment. By embedding POI behavioral roles as native IFC metadata, the framework enables a semantics-first approach where building elements carry robot-operational intelligence within the authoritative BIM model, directly informing navigation algorithms without requiring separate POI annotation systems or geometry-derived topology extraction. Virtual Reality simulation in Unity validates that POI-enriched IFC models provide actionable navigation guidance through sensor-based accessibility testing and iterative design refinement before physical construction.

2 Method

The proposed framework integrates BIM-based building representation, semantic POI enrichment, VR simulation, and robot navigation through a systematic methodology comprising four components. Figure 1 illustrates the overall system architecture and data flow. Each component maintains interoperability with standard BIM workflows while introducing robot-specific semantic intelligence necessary for autonomous navigation.

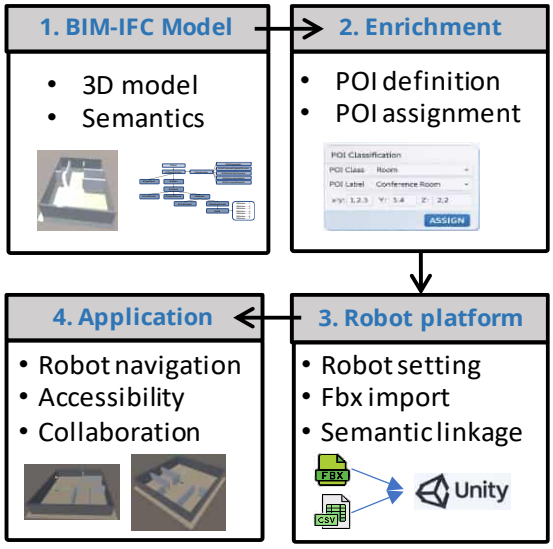


Figure 1. Proposed **F**ramework

2.1 BIM-IFC Building Model Development

Building information is represented as an IFC 4.3 file following standard BIM authoring practices, with building elements organized according to IFC hierarchical structure: IfcWall for boundary elements, IfcDoor for transition openings, IfcSpace for spatial zones (rooms, corridors), and IfcSlab for floor surfaces. IfcOpenShell, an open-source Python library for IFC file manipulation, is employed to parse the IFC model and extract both geometric and semantic information. Rather than developing custom IFC parsers, IfcOpenShell provides robust support for different IFC schema versions and enables straightforward extraction of element attributes through Python dot notation (e.g., IfcDoor.OverallWidth returns door width). The library's geometry processing capabilities convert implicit IFC

geometric representations (swept geometry, boundary representations, compound solids) into explicit mesh data comprising vertices, edges, and faces with absolute coordinate placement, eliminating the complexity of handling relative IFC placement hierarchies.

For each IFC element, the extraction process captures: (1) Global Unique Identifiers (GUIDs) for element tracking, (2) geometric parameters including mesh vertices, faces, edges, and spatial coordinates, (3) semantic attributes and element classifications, and (4) spatial hierarchies. This structured dataset serves as the foundation for subsequent POI semantic enrichment, with element GUIDs providing persistent links between IFC entities and their robot-operational metadata.

2.2 Semantic Enrichment with POI Classifications

2.2.1 Formal Mapping Rules

The semantic enrichment process employs a function-based classification framework that systematically assigns POI roles to IFC entities based on their spatial and functional characteristics. This approach ensures framework generalizability across the complete IFC schema, with mapping rules applicable to related IFC entities possessing the relevant functional properties. Five core mapping rules govern POI assignment based on entity function: spatial boundary provision, space connection, spatial containment, fixed object presence, and runtime constraint detection. Table 1 presents the formal mapping rules with representative example entities illustrating rule application across structural, functional, and object-based POI categories.

Table 1: Formal POI mapping rules

Rules	Functional Condition	POI Assignment	Examples
R1	Provides spatial boundary	Structural POI (A-1, A-3)	IfcWall, IfcColumn, IfcCurtainWall, IfcRailing, IfcBeam
R2	Connects two spatial zones	Structural POI (A-2, B-2 if clearance <threshold)	IfcDoor, IfcWindow, IfcRamp, IfcStair, IfcOpening
R3	Defines a spatial container	Functional POI (A-3, C-1, B-1)	IfcSpace (any predefined type or name)
R4	Fixed object or equipment	Object-based POI (A-1)	IfcFurniture, IfcEquipment, IfcDistribution element
R5	Runtime constraint detected	Dynamic role addition (B-2, C-2)	Any element with active sensor feedback

2.2.2 POI Role Taxonomy

This research demonstrates framework feasibility through three core POI categories essential for autonomous navigation: A-Mobility, B-Safety, and C-Coordination. Table 2 defines the complete role taxonomy for these categories, specifying seven operational roles that govern robot behavioral responses to building elements during navigation.

Table 2 POI role taxonomy

Category	Code	Role Name	Robot Behaviour
A-Mobility	A-1	Avoidance	Maintain clearance
A-Mobility	A-2	Transition	Verify clearance, control passage speed
A-Mobility	A-3	Path Guidance	Use as a trajectory reference
B-Safety	B-1	Hazard Zone	Proceed with caution
B-Safety	B-2	Narrow/Congested	Reduce speed
C-Coordination	C-1	Multi-Robot Zone	Apply the coordination protocol
C-Coordination	C-2	Robot-Robot Interaction	Maintain separation, adjust speed

2.2.3 IFC Property Set Extensions

Three custom property sets are programmatically attached to IFC entities using IfcOpenShell property manipulation capabilities, extending standard IFC entities with robot-operational metadata while maintaining full backward compatibility with BIM authoring and viewing software. Pset_POI_Classification (attached to all relevant entities) defines POI_Category {Structural, Functional, Object-based}, POI_Type {Boundary Element, Transition Element, Space Type}, POI_PrimaryRoles [array of assigned role codes], and POI_ConditionalRoles [runtime-assignable roles based on sensor feedback]. Pset_RobotAccessibility (attached to IfcDoor, IfcSpace) specifies MinClearanceWidth (meters),

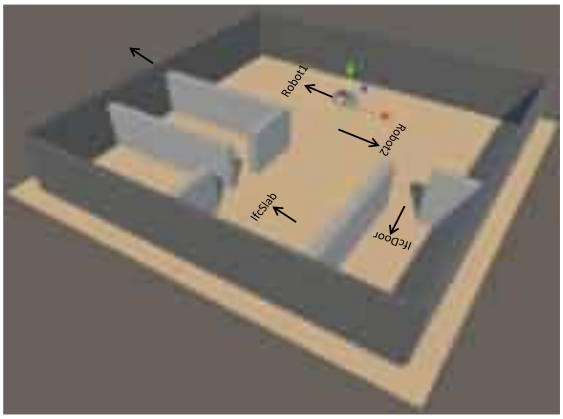
MinClearanceHeight (meters), AccessibilityStatus {FULL_ACCESS, RESTRICTED, BLOCKED, PENDING_VALIDATION}, ValidationMethod {SENSOR_BASED, GEOMETRIC_ANALYSIS, MANUAL}, and LastValidated (timestamp). Pset_RobotCoordination (attached to IfcSpace) establishes MaxSimultaneousRobots (integer), MinInterRobotSeparation (meters), CoordinationProtocol {YIELD, PRIORITY_PASS, SPEED_REDUCTION}, and AppliesTo {ALL_ROBOTS, SPECIFIC_TYPES}.

Following semantic enrichment, building data is exported in dual complementary formats: (1) geometric representations as FBX format preserving 3D mesh geometry, hierarchical structure, and material assignments for Unity import, and (2) semantic data as CSV tabular format containing element GUIDs, POI classifications, behavioural role codes, and property set values for runtime mapping within the VR simulation environment. This dual-export strategy ensures that both geometric fidelity and semantic richness are transferred to the VR platform while maintaining clear data provenance linking virtual building elements to their authoritative IFC representations.

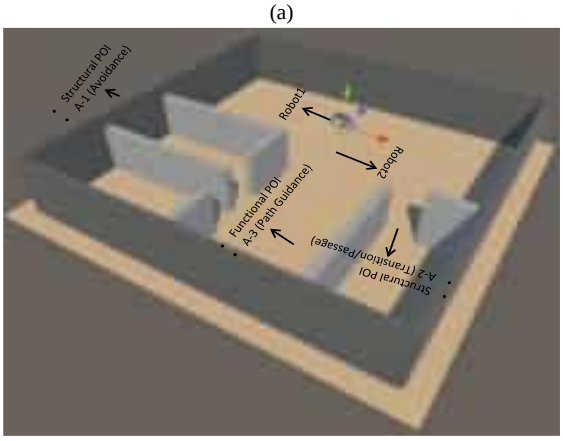
To validate the formal mapping framework, this research applies the enrichment methodology to a representative subset of IFC entity types covering structural elements, opening elements, and spatial zones essential for indoor robot navigation. Table 3 presents POI role assignments for the demonstration entities, illustrating systematic application of the formal mapping rules defined in Table 1. The framework is not limited to these six entities; rather, the function-based mapping rules (R1–R5) generalize to any IFC-compliant building model. Additional entity types, such as IfcColumn, IfcRamp, IfcStair, IfcFurniture, and IfcEquipment, can be enriched through identical rule application based on their functional characteristics. Also, Figure 2 presents an example of a visual POI assigned role to IFC classes.

Table 3 POI role assignments for demonstration

IFC Entity	Function	POI Category	Roles
IfcWall (exterior)	Building boundary	Structural POI	A-1, A-3
IfcWall (corridor)	Navigation boundary	Structural POI	A-1, A-3
IfcDoor	Space transition	Structural POI	A-2; cond. B-2
IfcSpace (corridor)	Shared navigation	Functional POI	A-3, C-1
IfcSpace (room)	Task destination	Functional POI	Destination
IfcSlab (floor)	Surface reference	Structural POI	A-3



(a)



(b)

Figure 2. Test Environment Configuration

2.3 POI-Driven Navigation Logic

The POI-enriched IFC data informs robot navigation through a systematic decision process that queries semantic properties at runtime and translates POI role assignments into behavioural responses. Algorithm 1 presents the navigation logic executed when robot sensors detect building elements within operational range. The algorithm demonstrates how POI semantic enrichment enables context-aware navigation decisions: sensor detections are matched to IFC entities via GUIDs, POI roles determine behavioural responses, and conditional roles are dynamically activated based on runtime spatial assessments.

Algorithm 1: POI-Driven Navigation Decision

Input: Robot position P, Sensor readings S, POI database D

Output: Navigation action A

```
Begin
    FOR each element E detected by sensors S:
        GUID = GetElementIdentifier(E)
        POI = QueryDatabase(D, GUID)
        roles = POI.PrimaryRoles UNION POI.ConditionalRoles
        IF A-1 (Avoidance) IN roles THEN
            MaintainClearance(E, minimum_distance)
        END IF
        IF A-2 (Transition) IN roles THEN
            clearance = POI.Pset_RobotAccessibility.MinClearanceWidth
            IF SensorClearance(E) < clearance THEN
                AddConditionalRole (E, B-2)
                SetAccessibilityStatus(E, RESTRICTED)
            ELSE
                SetAccessibilityStatus(E, FULL_ACCESS)
            END IF
        END IF
        IF C-1 (Multi-Robot Zone) IN roles THEN
            separation = POI.Pset_RobotCoordination.MinInterRobotSeparation
            IF DetectRobot(S) AND RobotDistance() < separation THEN
                AddConditionalRole (E, C-2)
                ExecuteProtocol(SPEED_REDUCTION)
            END IF
        END IF
        IF A-3 (Path Guidance) IN roles THEN
            UseAsTrajectoryReference(E)
        END IF
    END FOR
    RETURN ComputeNavigationAction(roles)
        End
```

2.4 Robot Navigation Platform in Unity VR

The robot navigation platform is implemented using Unity 6, a real-time 3D development engine providing physics simulation, sensor emulation, and interactive visualization capabilities. The FBX geometric model is imported into Unity, reconstructing the building environment with accurate spatial relationships and dimensional properties. Each imported building element is configured with physics components, including colliders for collision detection and rigid body properties where dynamic interactions are required. The CSV semantic data file is parsed within Unity using a custom POI Manager script that establishes mappings between Unity GameObjects (identified by name) and their associated IFC entities (identified by GUID). Each building element GameObject receives a POI Component attachment encoding its POI category, type, behavioural roles, and robot-navigation parameters extracted from the CSV data, enabling runtime queries by robot navigation algorithms to determine appropriate behavioural responses.

Two robot platforms are instantiated as autonomous agents within the Unity environment, each equipped with physics-based locomotion systems (mass: 5 kg, movement speed: 5.0 m/s, rotation speed: 100°/s), five-directional sensor arrays, and rotating brush mechanisms (500°/s). The sensor configuration projects rays at angular offsets of 0° (forward-center), ±30° (forward-angled), and ±90° (lateral) with 5.0m detection range, providing real-time visual feedback through color-coded visualization: Robot1 displays green (clear) and red (obstacle detected) sensor rays, while Robot2 displays green and blue rays for multi-robot visual distinction during simultaneous operation. Independent control schemes enable concurrent operation: Robot1 via arrow keys and Robot2 via WASD keys, with graphical interface buttons providing brush activation toggles. A dual-camera system (top-down and perspective views) switchable via keyboard input enables comprehensive observation of navigation scenarios from multiple vantage points. Figure 3 presents the Unity platform interface and robot physics configuration.

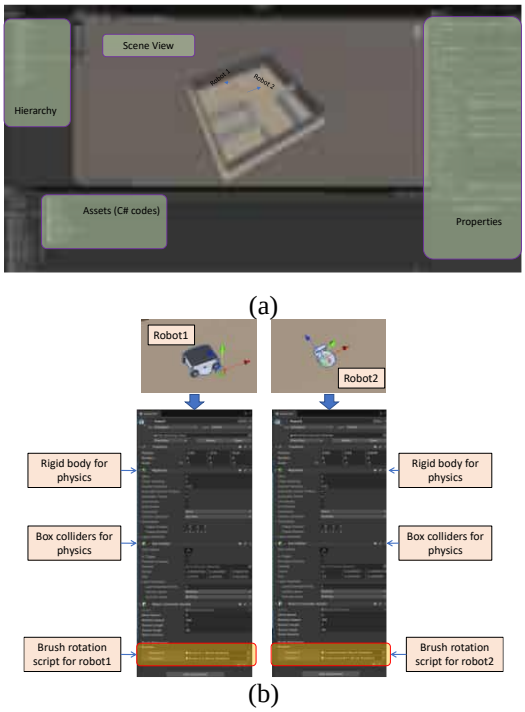


Figure 3. (a) Interface of the platform, and (b) example of physics properties assignment into the robot.

3 Validation Scenarios

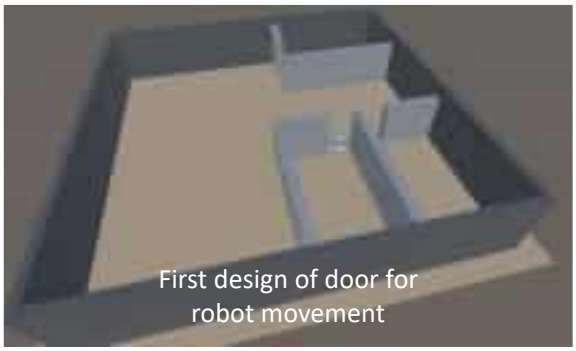
The POI-enriched framework is validated through

three sensor-based navigation scenarios demonstrating how semantic building information directly informs robot behavioural responses.

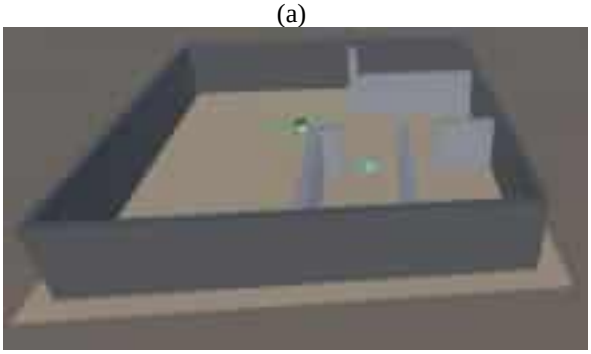
3.1 POI-Driven Doorway Accessibility Validation

The interior doorway connecting the central corridor to Room 2 is initially designed at 0.80m width, meeting standard architectural requirements for human passage but not validated for robot clearance constraints. The doorway's IFC entity (IfcDoor) is enriched with Pset_POI_Classification, assigning Structural POI category, Transition Element type, and A-2 (Transition/Passage) primary role, while Pset_RobotAccessibility specifies MinClearanceWidth = 0.90m as the robot platform's minimum safe passage requirement. As Robot2 approaches the doorway from the corridor at 5.0 m/s navigation speed, the five-directional sensor array evaluates spatial clearances against POI-defined constraints. The forward-center sensor (0°) maintains green status throughout approach, detecting unobstructed passage through the doorway centerline. However, both forward-angled sensors (+30° and −30°) transition to blue at 1.8m approach distance, detecting door frame edges within their angular projection cones at ranges below the 5.0m detection threshold. This sensor pattern indicates that while direct centerline passage is geometrically feasible, approach trajectories at ±30° encounter insufficient clearance margins. The simultaneous blue status of angled sensors triggers conditional POI role addition: the doorway classification updates from A-2 (Transition/Passage) to A-2 + B-2 (Transition with Narrow/Congested constraint), and Pset_RobotAccessibility.AccessibilityStatus property is set to restricted.

Based on sensor feedback indicating inadequate clearance at angled approach vectors, the building geometry is iteratively refined within the BIM authoring environment. The doorway width is increased from 0.80m to 1.20m (50% expansion), the IFC model is re-exported, and POI enrichment is reapplied with updated dimensional properties automatically reflected in Pset_RobotAccessibility parameters. Robot2's subsequent approach to the modified doorway produces distinctly different sensor responses: all forward sensors (0°, ±30°) now maintain GREEN status throughout the approach, indicating that door frame edges remain outside the ±30° detection cones even at proximity. The POI conditional role reverts to standard A-2 (Transition/Passage), and AccessibilityStatus updates to FULL_ACCESS. Figure 4 presents comparative sensor visualizations showing before-modification (BLUE angled sensors) and after-modification (GREEN angled sensors) states.



(a)

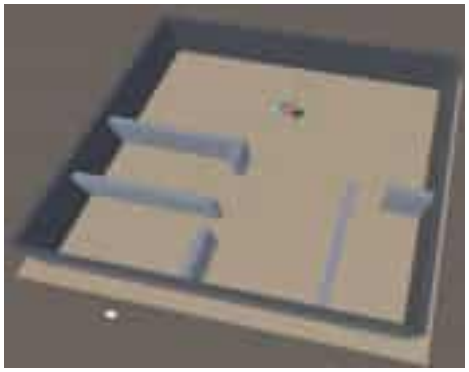


(b)

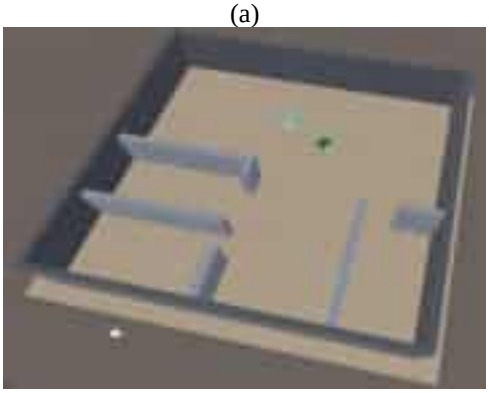
Figure 4. Doorway Validation Comparison

3.2 Multi-Robot Safe Separation Validation

Multi-robot coordination capabilities are validated through scenarios where both Robot1 and Robot2 navigate shared corridor spaces classified as Functional POI with C-1 (Multi-Robot Zone) role assignments. The central corridor IfcSpace entity is enriched with Pset_RobotCoordination specifying MaxSimultaneousRobots, MinInterRobotSeparation of 2.0m between them, and CoordinationProtocol set to SPEED_REDUCTION. As the two robots approach each other from opposite corridor ends, their sensor arrays continuously monitor inter-robot distances. When the separation distance decreases below the 2.0m threshold defined in the POI metadata, both robots trigger C-2 (Robot-Robot Interaction) conditional role activation and execute coordinated speed reduction from 5.0 m/s to 2.5 m/s (50% reduction), maintaining safe operational spacing while allowing continued navigation through the shared zone. Visual sensor feedback differentiates robot detection through color coding: Robot1 displays red rays when detecting Robot2, while Robot2 displays blue rays, enabling clear visualization of proximity detection and coordination protocol execution. This validation demonstrates how the Pset_RobotCoordination property sets embedded in IfcSpace entities directly govern multi-agent behavioural responses without requiring external coordination systems.



(a)



(b)

Figure 5. Multi-robot safe separation validation

3.3 Corridor Navigation Using Wall-Following Behaviour

Robot1 navigates along the central corridor by maintaining alignment parallel to the left wall boundary, classified as a Structural POI with A-3 (Path Guidance) role. The lateral sensor (+90°) continuously monitors the wall at 0.5m standoff distance, providing trajectory reference for corridor navigation. The wall POI classification enables the robot to use the corridor boundary as a navigation reference point, maintaining centered trajectory while moving toward the target room. This scenario validates how POI-enriched building elements provide spatial guidance for autonomous navigation beyond simple obstacle avoidance.

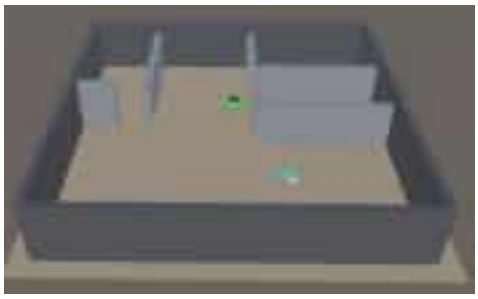


Figure 6. Corridor Navigation Using Wall-Following Behaviour

4 Discussion and limitations

The integration of POI classifications into IFC building models extends BIM utility beyond architectural documentation into autonomous robot operations. Traditional IFC entities provide geometric and material specifications adequate for construction management, yet lack behavioural semantics necessary for robot navigation algorithms to make context-aware decisions. The POI enrichment methodology transforms passive building descriptions into active navigation frameworks by augmenting standard IFC property sets with robot-operational intelligence accessible at runtime.

POI semantic enrichment operates at three levels: classification (Structural, Functional, Object-based categories organize elements by navigation role), behavioural (role codes encode specific responses such as deceleration near narrow passages or coordination in shared corridors), and validation (POI-defined thresholds provide quantitative criteria for automated accessibility assessment). The doorway validation scenario demonstrates proactive design optimization absent from conventional BIM workflows, rather than discovering accessibility constraints post-construction, the POI-driven simulation identified insufficient clearances during design, enabling cost-effective geometric modifications validated through iterative virtual testing.

The current implementation exhibits four primary limitations. First, the validation environment represents simplified geometry that does not capture real-world facility complexity, including furniture clutter, irregular layouts, and environmental variability. Second, validation relies exclusively on virtual Unity simulation rather than physical robot deployments in real buildings. Third, manual semantic enrichment becomes labour-intensive for large-scale projects; automated POI assignment algorithms would significantly improve scalability. Fourth, the POI role taxonomy has been validated only for a subset of role codes through demonstration scenarios, with comprehensive empirical validation across all role groups remaining necessary for full framework validation.

5 Conclusion

This research presents a framework for semantic POI-driven robot navigation through systematic IFC building model enrichment. By augmenting standard BIM entities with robot-oriented POI classifications and behavioural role assignments via three custom property sets (Pset_POI_Classification, Pset_RobotAccessibility, Pset_RobotCoordination), the framework transforms conventional architectural models into semantically rich spatial representations that directly support autonomous navigation while maintaining full compatibility with

existing BIM workflows. Unity VR-based validation confirms that POI-enriched building models provide actionable guidance for robot systems, with the doorway accessibility case study demonstrating sensor-based validation that identified and corrected spatial constraints during virtual design iteration before physical construction.

The framework's contribution extends beyond technical implementation to establish conceptual foundations for robot-compatible building design. By formalizing robot-operational requirements as standardized IFC metadata, the research provides a pathway toward automated accessibility assessment tools and building standards that explicitly accommodate autonomous robotic systems alongside human occupants. Future research directions include validation in complex real-world environments, integration with physical robot platforms, development of automated POI assignment algorithms, and engagement with BuildingSMART International to establish formal POI classification standards within the IFC specification.

References

[1] B.M. Tehrani, S. BuHamdan, A. Alwisy, Robotics in industrialized construction: an activity-based ranking system for assembly manufacturing tasks, Engineering, Construction and Architectural Management 31 (2024) 2264–2285. https://doi.org/10.1108/ECAM-02-2022-0143.

[2] H. Oyediran, W. Turner, K. Kim, M. Barrows, INTEGRATION OF 4D BIM AND ROBOT TASK PLANNING: CREATION AND FLOW OF CONSTRUCTION-RELATED INFORMATION FOR ACTION-LEVEL SIMULATION OF INDOOR WALL FRAME INSTALLATION, Journal of Information Technology in Construction 30 (2025) 352–374. https://doi.org/10.36680/j.itcon.2025.015.

[3] Z. Ren, H. Jin LEE, M. Shoaib Khan, X. Zeng, T. Yeon KIM, J. In KIM, Science Mapping for Advances in Integration of Domain Knowledge into Autonomous Construction Robots, (n.d.). https://doi.org/10.6106/ICCEPM.2025.0504.

[4] P. Boguslawski, S. Zlatanova, D. Gotlib, M. Wyszomirski, M. Gnat, P. Grzempowski, 3D building interior modelling for navigation in emergency response applications, International Journal of Applied Earth Observation and Geoinformation 114 (2022). https://doi.org/10.1016/j.jag.2022.103066.

[5] Y.H. Lin, Y.S. Liu, G. Gao, X.G. Han, C.Y. Lai, M. Gu, The IFC-based path planning for 3D indoor spaces, Advanced Engineering Informatics 27 (2013) 189–205. https://doi.org/10.1016/j.aei.2012.10.001.

[6] Q. Chen, J. Chen, W. Huang, Pathfinding method for an indoor drone based on a BIM-semantic model, Advanced Engineering Informatics 53 (2022). https://doi.org/10.1016/j.aei.2022.101686.

[7] X. Zhou, Q. Xie, M. Guo, J. Zhao, J. Wang, Accurate and Efficient Indoor Pathfinding Based on Building Information Modeling Data, IEEE Trans. Industr. Inform. 16 (2020) 7459–7468. https://doi.org/10.1109/TII.2020.2974252.

[8] A. Khalili, D.K.H. Chua, IFC-Based Graph Data Model for Topological Queries on Building Elements, Journal of Computing in Civil Engineering 29 (2015). https://doi.org/10.1061/(asce)cp.1943-5487.0000331.

[9] M.S. Khan, M. Adil, A. Khan, Assessment of structural design capability of building information modeling (BIM) in building industry of Pakistan, JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES 14 (2019) 385–401. https://doi.org/10.26782/jmcms.2019.04.00030.

[10] M.S. Khan, H.S. Cho, J. Seo, Voxel-Based Digital Twin Framework for Earthwork Construction, Applied Sciences 15 (2025) 7899. https://doi.org/https://doi.org/10.3390/app15147899.

[11] M.S. Khan, M. Khan, M. Bughio, B.D. Talpur, I.S. Kim, J. Seo, An Integrated HBIM Framework for the Management of Heritage Buildings, Buildings 12 (2022). https://doi.org/10.3390/buildings12070964.

[12] J. Zhu, M.O. Wong, N. Nisbet, J. Xu, T. Kelly, S. Zlatanova, I. Brilakis, Semantics-based connectivity graph for indoor pathfinding powered by IFC-Graph, Autom. Constr. 171 (2025). https://doi.org/10.1016/j.autcon.2025.106019.

[13] R. Sacks, L. Ma, R. Yosef, A. Borrmann, S. Daum, U. Kattel, Semantic Enrichment for Building Information Modeling: Procedure for Compiling Inference Rules and Operators for Complex Geometry, Journal of Computing in Civil Engineering 31 (2017) 1–12. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000705.

Process-Aware Reward Modeling for Automated Construction Planning

Akarsth Kumar Singh and Shang-Hsien Hsieh

Department of Civil Engineering, National Taiwan University, Taiwan
akarsth@caece.net , shhsieh@ntu.edu.tw

Abstract –
Construction planning remains error-prone because existing automated approaches primarily assess final schedules and lack mechanisms to evaluate intermediate planning decisions. As a result, early-stage errors in task decomposition, sequencing, or duration estimation can propagate unnoticed across planning stages. This study proposes a process-aware framework for automated construction planning that embeds planning-time supervision through a construction-specialized Process Reward Model (PRM). The framework integrates sequential multi-stage plan generation with rubric-based Small Language Model (SLM)-as-a-Judge evaluation to generate step-level quality signals, which are used to train the PRM to predict the quality of partial planning states. Experiments on real-world building construction data show that the PRM achieves 78% agreement with human-verified evaluations while reducing evaluation cost by approximately 50× and inference time by more than 5× compared to SLM-as-a-Judge assessment. The results indicate that process-level reward modeling provides scalable and efficient quality assurance for structured planning elements such as work breakdown consistency and resource feasibility, while serving as a pre-screening mechanism for complex dependency reasoning.

Keywords –
Automated Construction Planning, Process Reward Model, SLM-as-a-Judge, Process-Level Evaluation, Large Language Model

1 Introduction

Construction projects worldwide continue to experience significant delays and cost overruns, largely due to inadequate planning and scheduling, limited resource availability, and inefficient decision-making processes [1][2]. Developing a reliable construction schedule requires identifying activities from extensive contract documents and transforming them into structured planning elements. This process is typically performed manually, making it time-consuming, inconsistent, and susceptible to human error [3][4]. Although conventional project management software supports schedule generation, it provides little capability to assess planning quality during the progressive decomposition of complex construction scopes into executable work packages.

Recent advances in construction informatics have explored the use of large language models (LLM) and natural language processing (NLP) for activity extraction from textual documents [3][5], knowledge graphs for adaptive work packaging and reasoning [6], and machine learning techniques for schedule enhancement and optimization [7]. While these approaches improve automation and efficiency, they primarily focus on extracting information or validating final schedules. They do not explicitly assess intermediate planning decisions, including task decomposition, dependency formulation, and duration estimation.

In parallel, process reward models (PRMs) have recently shown strong potential in improving reasoning performance in domains such as mathematics [8], financial analysis [9], and question-answering systems [10]. Unlike conventional reward mechanisms that evaluate only final outputs, PRMs assess the quality of intermediate reasoning steps. Domain-specific PRMs, such as Fin-PRM, further demonstrate that embedding expert knowledge and procedural constraints leads to substantial performance gains [9]. Despite this progress, no PRM framework has been developed for construction planning, a domain characterized by sequential, interdependent decisions that must satisfy constructability rules, resource constraints, and contractual requirements.

This gap is particularly critical because construction planning errors tend to propagate across multiple stages, including work breakdown structure development, task sequencing, duration estimation, and resource allocation. General-purpose language models lack the domain knowledge required to reliably evaluate intermediate planning states against industry practices and project-specific constraints [11][12]. As a result, automated planning systems remain vulnerable to infeasible or low-quality schedules.

To address this limitation, this study proposes a process-aware framework for automated construction planning that integrates sequential plan generation, domain-calibrated LLM-as-a-Judge evaluation, and Process Reward Model (PRM) training. Construction planning is decomposed into four stages—work breakdown structure generation, task relationship identification, duration estimation, and resource allocation—evaluated using construction-specific rubrics to produce step-level quality signals that guide generation. Validation on real-world building projects shows that the approach (i) provides construction-specialized process-level supervision for planning, (ii) enables planning-time quality assurance rather than post-hoc validation, and (iii) supports transferable planning knowledge across similar projects. The study further introduces a hybrid evaluation framework in which the PRM performs planning-time screening of structured planning logic, while complex dependency reasoning is selectively validated using higher-fidelity evaluators.

2 Related Work

2.1 Automated Construction Planning

Construction planning remains largely dependent on manual expert judgment and semi-structured representations [4]. A three-decade review showed that although automated methods for scope definition, WBS generation, and sequencing exist, most projects still rely on manual workflows due to inflexible knowledge representations, dependence on manually maintained templates, limited learning from historical data, and weak integration between planning and optimization [7]. Earlier rule-based and optimization approaches, including genetic algorithms, require predefined task networks, limiting generalizability [13]. BIM-based 4D planning remains labor-intensive to maintain [14], while knowledge graph approaches improve semantic work packaging but rely on explicitly constructed graphs and do not infer implicit constructability logic from documents [6].

2.2 Large Language Models in Construction

NLP and LLMs have been applied to construction documents with high accuracy in activity and procedural information extraction [3][5][18], but these methods focus on retrieval rather than complete schedule generation. Learning-based approaches have modeled task dependencies and improved scheduling performance [7][12]. Multi-LLM pipelines with SLM-as-a-judge evaluation have also been proposed [11]; however, supervision is limited to synthetic WBS and construction description generation and post-hoc evaluation of final outputs, leaving intermediate planning stages unsupervised.

2.3 Process Reward Models

PRMs evaluate intermediate reasoning steps rather than only final outputs and have demonstrated notable performance gains in structured domains. Trajectory-aware PRMs improve mathematical reasoning by incorporating step-level supervision [8], while domain-specialized models such as Fin-PRM enhance financial reasoning through procedural and factual verification [9]. Generative PRMs further strengthen step-wise evaluation using chain-of-thought verification with limited supervision [16][10].Despite these advances, existing PRMs assume well-defined correctness criteria and structured reasoning spaces. Construction planning, by contrast, involves constraint satisfaction across constructability logic, resource feasibility, document alignment, and industry standards, which cannot be directly addressed using current PRM formulations.

2.4 LLM-as-a-Judge Evaluation

LLM-as-a-judge approaches enable rubric-based, domain-specific evaluation beyond generic accuracy metrics and correlate strongly with expert judgment when customized criteria are applied [17]. In construction contexts, rubric-guided evaluation has been used to assess planning completeness, hierarchical consistency, and task relationship correctness, including SLM-as-a-judge assessment [11][18]. However, these approaches operate post-hoc, requiring additional inference passes after plan generation, introducing computational overhead and preventing real-time feedback. In contrast, the proposed framework embeds a construction-specialized PRM directly within the planning loop, enabling planning-time guidance of intermediate decisions and mitigating error propagation across planning stages.

2.5 Research Gap

Despite advances in LLM-based planning, construction planning lacks a process-level supervision framework that validates intermediate decisions during generation. Existing approaches emphasize final-output evaluation, allowing decomposition and sequencing errors to propagate unchecked. Current PRMs are limited to domains with explicit correctness criteria, restricting their applicability to construction planning. This study addresses this gap by introducing a construction-specialized process reward modeling framework that incorporates explicit evaluation of intermediate planning decisions and integrates quality guidance throughout the reasoning process.

3 Methodology

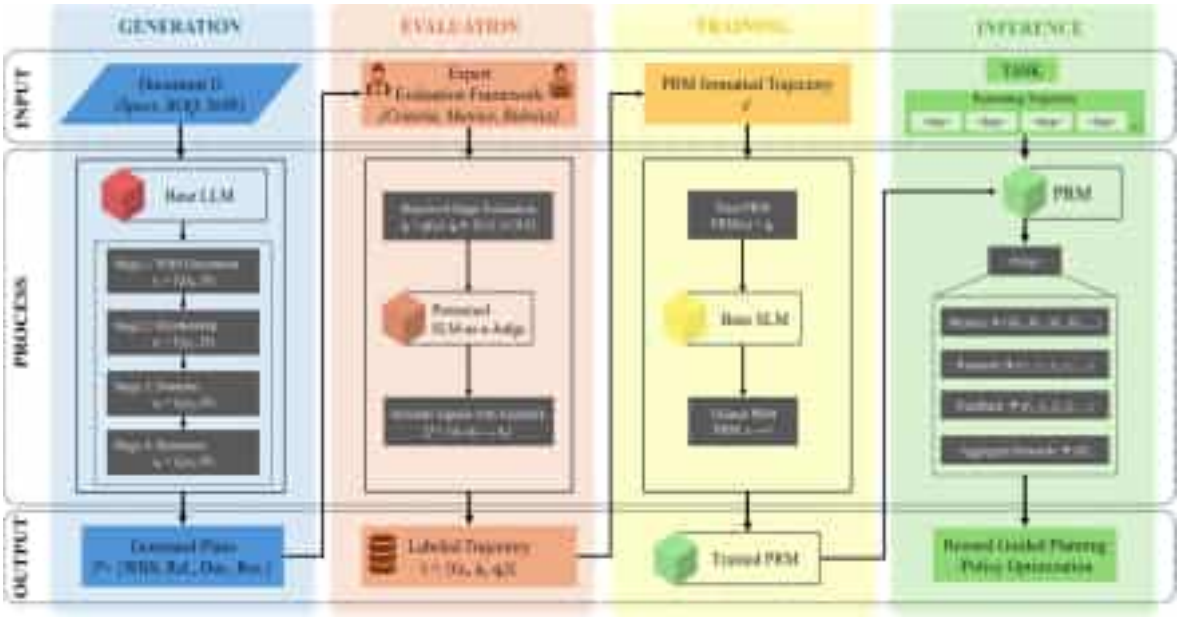


Figure 1. Overview of Methodology

This study proposes, as shown in Fig. 1, a process-aware framework for automated construction planning that integrates sequential plan generation, intermediate quality evaluation, and learning-based guidance. The framework operates in three stages: (1) sequential construction planning generation, (2) intermediate evaluation using a large language model acting as a judge, and (3) PRM training. Construction inputs are represented as the document set D defined in Eq. (1), while the generated construction plan P is defined in Eq. (2). Stage-level quality signals produced during planning are denoted by Q in Eq. (3). The PRM is formulated as a mapping from partial planning states to predicted quality rewards, as shown in Eq. (4).

$$D = \{\text{specifications,BOQ,scope}\} \quad (1)$$

$$P = \{WBS, Relationship, Duration, Resources\} \quad (2)$$

$$Q = \{q_1, q_2, \dots, q_n\} \quad (3)$$

$$\text{PRM}: s \rightarrow r \quad (4)$$

3.1 Sequential Construction Planning

Construction planning is formulated as a sequence of interdependent decision stages reflecting standard project workflows. Rather than generating a complete schedule in a single step, the framework incrementally transforms a partial planning state s_t into a refined state s_{t+1} using a stage-specific planning operator, as expressed in Eq. (5).

$$s_{t+1} = f_t(s_t, D) \quad (5)$$

This sequential formulation exposes intermediate planning states and enables targeted evaluation and correction across work breakdown formulation, task relationship reasoning, duration estimation, and resource allocation stages.

3.2 Intermediate Evaluation via SLM-as-a-Judge

Intermediate planning states are evaluated using a domain-calibrated LLM acting as a judge. For each planning stage t, the evaluator assigns quality signals using two instruments: binary scores $q_t \in \{0,1\}$ for structural validation checks (e.g., schema compliance, field presence), and 5-point Likert scores $q_t \in \{1,...,5\}$ for reasoning dimensions admitting partial credit. The stage-level signal is the mean across all applicable metrics within a step, with Likert scores linearly rescaled to [0,1], yielding a normalized composite reward $q_t \in [0,1]$ as defined in Eq. (6).

$$q_t = \text{g}(s_t), \text{q}_\text{t} \in \{0,1\} \text{ or } [0,5] \quad (6)$$

Each scale point is anchored to a domain-specific behavioral descriptor; for example, for Sequence Logic Explained: 1 = no sequencing rationale provided; 5 = explicit constructability justification for every task link (e.g., formwork must precede concrete pour as it defines the structural mould). In addition to the scalar quality signal, the evaluator generates structured textual feedback identifying logical inconsistencies, feasibility issues, and misalignment with the source construction context/documents; this feedback is used for analysis and interpretation but does not constitute the reward signal.

3.3 Process Reward Model Learning

Outputs from the evaluation stage are aggregated into labeled planning trajectories, where each trajectory captures the partial planning state, the corresponding planning action, and the assigned quality signal, as formalized in Eq. (7).

$$\tau = \{(s_t, a_t, q_t)\}_{t=1}^{T} \quad (7)$$

The PRM is trained to approximate the evaluation function by predicting the quality signal associated with a given partial planning state, as shown in Eq. (8).

$$\mathrm{PRM}(s_t) \approx q_t \quad (8)$$

During planning, the PRM guides decision-making by scoring candidate actions and selecting the action expected to yield the highest-quality planning outcome, as expressed in Eq. (9).

$$a_t^* = \arg\max_{a \in A_t} \mathrm{PRM}(s_t \oplus a) \quad (9)$$

Critically, this selection mechanism relies on ordinal ranking rather than precise cardinal scoring. The PRM is not required to predict absolute quality values, but only to consistently rank candidate actions such that an action a_1 leading to a higher-quality planning state is preferred over a_2. This distinction between relative comparison and absolute evaluation is central to the effectiveness of PRM-guided construction planning. Unlike SLM-as-a-Judge evaluation, which is applied after plan generation, the PRM is queried during the planning process to score candidate actions and guide selection toward higher-quality intermediate states. This guidance mechanism enables early suppression of low-quality planning paths and improves robustness and interpretability.

4 Experimental Setup

4.1 Dataset Construction

The experimental dataset is derived from a real-world high-rise residential building project comprising a reinforced concrete (RCC) framed structure of 22 floors above ground, with more than 2,500 scheduled activities spanning structural works, all internal and external architectural finishing, and full mechanical, electrical, plumbing, and fire protection (MEPF) systems. The majority of these activities represent repetitions of the same work types across different floors and locations. To avoid redundancy and prevent overfitting the PRM on structurally identical samples, 243 unique activity types were selected, capturing the minimum sufficient set of distinct planning scenarios across all domains. Selected activities were manually reviewed for completeness, clarity, and representativeness, producing a high-quality seed dataset for generation and evaluation.

4.2 Sequential Planning Generation

For each activity, synthetic Bill of Quantities (BOQ) detailed descriptions were generated following the procedure established in prior work [11]. These BOQ descriptions serve as the primary input documents for the multi-stage planning generation process.

At each planning stage, the system generates step-wise reasoning traces structured as ⟨Thought, Analysis, Decision⟩ triplets, followed by a stage-specific structured output that conforms to a predefined schema. This explicit reasoning decomposition exposes intermediate decision-making processes and enables targeted evaluation of planning logic at each stage. The complete generation workflow is formalized in Algorithm 1.

Algorithm 1: Multi-Stage BOQ Processing with Structured Reasoning
Input: Project data D, BOQ items B, prompt set Π, environment E
Output: Structured representations per BOQ item

1: **Initialize** LLM model and client using E
2: **Load** BOQ items B and project context Cproj from P

3: **for** each b ∈ B do ▷ Item level
4: Cb ← ComposeContext(b, C_{proj})
5: x ← ⊥ ▷ Initial state
6: **for** each stage s∈⟨WBS, REL, DUR, RES⟩ **do** ▷ Stage level
7: T ← GenerateReasoning (s, b, x, Cb) ▷ Thought
8: A ← Derive Intermediate (T, s, Cb) ▷ Analysis
9: x ← Select Structured Output (A, s) ▷ Decision
10: Store (b, s, x)
11: **end** for
12: **end** for

4.3 Step-Wise Evaluation with SLM-as-a-Judge

Each intermediate reasoning step produced during planning is evaluated using a rubric-based assessment framework calibrated for construction planning tasks. The evaluation combines binary validation metrics (0 or 1) with fine-grained Likert-scale scores (0–5). The framework quantifies reasoning quality across four dimensions applied in Algorithm 2 (lines 8–10 per reasoning step; lines 12–14 for structured output validation): (1) constructability logic via Likert metrics such as Sequence Logic Explained and Lag Time Justification; (2) domain knowledge correctness via binary checks verifying that task names, relationship types (FS/SS/FF/SF), and resource selections conform to industry norms; (3) hierarchical consistency via metrics such as L1–L7 Reasoning Shown and 100% Rule

Verified; and (4) document alignment via traceability checks linking extracted activities and quantities to the input BOQ.

Evaluation is conducted at three levels: (1) step-level assessment of individual ⟨Thought, Analysis, Decision⟩ triplets, (2) validation of stage-specific structured outputs against schema constraints, and (3) cross-stage consistency verification. For each evaluation instance, the judge model returns both a numerical score and explanatory feedback identifying specific strengths or deficiencies in the reasoning trace. The complete evaluation procedure is formalized in Algorithm 2.

Algorithm 2: Multi-Stage Reasoning Evaluation with SLM-as-a-Judge

Input: Stage outputs S, BOQ data B, metrics M, stage config Q, judge J
Output: Quality scores R per stage per sample

Load metrics M and quality config Q
for each stage σ ∈ ⟨WBS, REL, DUR, RES⟩ do
  Load stage outputs Sσ and BOQ data Bσ
  for each sample si ∈ S σ **do**
    Ci ← BuildContext (B σ , σ)
    for each reasoning step rj ∈ si **do**
      Tj ← Merge (rj.thought, rj.analysis, rj.decision)
      for each metric mk ∈ M[σ][rj] **do**
      R[σ][si][rj][mk] ← J.Evaluate (mk, Q[σ][rj], Ci, Tj)
      end for
    end for
    for each validation metric mv ∈ M[σ][val] do
      R[σ][si][val][mv] ← J.Evaluate(mv, Q[σ][val], Ci, si.output)
    end for
    PersistResults(R[σ][si])
  end for
end for
return R

The resulting evaluation dataset consists of labeled planning trajectories with dense, step-wise reward signals. These data support supervised training of PRM that estimate intermediate planning quality and guide future generations toward higher-quality, logically consistent, and feasible construction plans.

4.4 PRM Model Training

The labeled planning trajectories generated by the SLM-as-a-Judge evaluation are used to train the PRM. Each trajectory consists of partial planning states, corresponding ⟨Thought, Analysis, Decision⟩ reasoning steps, and judge-assigned quality scores. The PRM is trained to predict quality signals from partial planning states by minimizing the error between predicted rewards and judge-provided scores. Inputs to the model include the construction context and BOQ descriptions, the planning stage identifier, the accumulated planning state, and the current reasoning step. The model outputs scalar reward predictions aligned with the evaluation scale, using binary values (0 or 1) for validation metrics and discrete scores (0–5) for Likert-scale assessments.

Table 1. PRM training configuration

Item	Value
Model	Qwen2.5-1.5B-Instruct
Adaptation	LoRA (r=16, α=16, dropout=0)
Seq. length	2048
Optimizer	AdamW (lr=2e-4)
Batch / Epochs	2 / 3
Loss	MSE

The PRM is fine-tuned using parameter-efficient adaptation. Key training hyperparameters, including LoRA configuration, learning rate, batch size, and optimization settings, are summarized in Table 1. This setup enables efficient training while preserving the base model's general reasoning capabilities.

5 Results

5.1 Evaluation Setup

The PRM was evaluated against SLM-as-a-Judge reference scores across 251 matched samples spanning four construction planning stages: work breakdown structure generation, task relationship identification, duration estimation, and resource allocation. The SLM-as-a-Judge framework was used to generate reference scores, which were subsequently verified by the authors to ensure consistency and reliability. This combined automated and manual validation establishes a robust ground truth for PRM performance evaluation.

5.2 Overall Performance

Table 2 reports the overall agreement between PRM predictions and the manually verified SLM-as-a-Judge scores. The PRM achieved 78% accuracy with near-zero mean bias (−0.05), indicating it reliably captures relative quality differences between planning candidates without systematic over- or under-estimation. The moderate correlation ($r = 0.515$) confirms the model ranks alternatives consistently with human judgment.

For reward-guided generation (Eq. 9), this ranking

capability is more critical than absolute scoring accuracy: the PRM need only identify that candidate A is superior to candidate B, not assigning objectively correct scores to either.

Table 2. Overall performance metrics

Metric	Value
Samples (n)	251
Accuracy	78%
Pearson correlation (r)	0.515
MAE	0.887
RMSE	1.159
Within ±1.0	65.7%
Mean bias	−0.05

The minimal bias demonstrates the PRM fulfills this comparative function, enabling effective guidance away from low-quality planning paths during iterative generation.

5.3 Stage-wise Performance Analysis

Figure 2 presents a radar comparison of PRM and SLM-as-a-Judge scores across the four planning stages. The PRM shows strongest agreement for WBS generation and resource allocation, with correlations exceeding 0.5 in both stages. In WBS generation, the model performs reliably on scope interpretation and hierarchical structuring, indicating effective evaluation of structured decomposition tasks. Similarly, strong agreement in resource allocation reflects consistent assessment of labor crew composition and equipment selection.

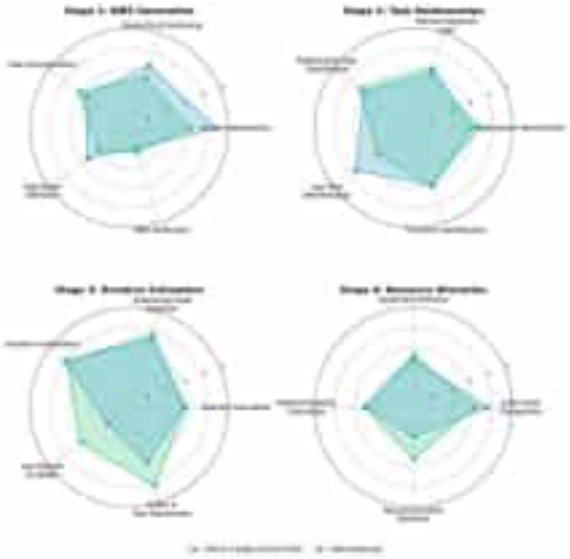


Figure 2. Stage-wise comparison of PRM and SLM-as-a-Judge evaluation scores.

In contrast, task relationship evaluation exhibits the weakest agreement, with near-zero correlation, highlighting challenges in assessing complex task relationship logic, predecessor–successor reasoning, and sequencing rationale. This stage requires multi-step causal reasoning and implicit constructability knowledge, which remains difficult for the current PRM formulation. Duration estimation shows moderate correlation, with the PRM exhibiting conservative scoring tendencies relative to the SLM baseline, particularly for buffer and risk-related considerations.

5.4 Cost–Accuracy and Cost–Time Trade-Offs

The accuracy–cost relationship between PRM and SLM-as-a-Judge is shown in Figure 3. While SLM-as-a-Judge achieves perfect agreement with the verified ground truth, it incurs substantially higher per-sample cost. In contrast, the PRM achieves 78% accuracy at a cost of approximately $0.003 per sample, representing an approximate 50× cost reduction. Cost estimates are based on online GPU pricing for an RTX 3090 (24 GB VRAM, 128 GB RAM) configuration.

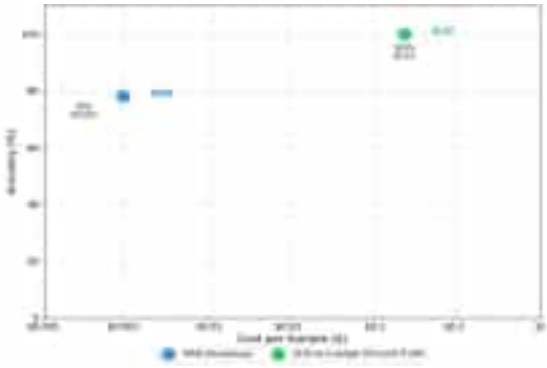


Figure 3. Accuracy–cost trade-off between PRM and SLM-as-a-Judge evaluation

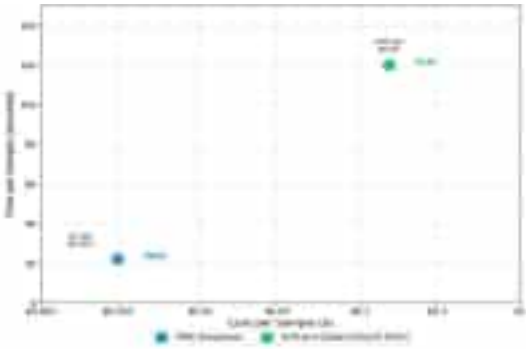


Figure 4. Cost–time efficiency comparison of PRM and SLM-as-a-Judge evaluation

Computational efficiency is compared in Figure 4. PRM inference requires approximately 22 seconds per sample, compared to 120 seconds for SLM-as-a-Judge, resulting in a 5.5× speed-up and an order-of-magnitude reduction in total evaluation cost under the same hardware assumptions. These results demonstrate the practical advantages of PRM for scalable construction planning assessment.

6 Discussion

The results indicate a deployment-aware performance profile rather than a fundamental limitation of the proposed approach. The Process Reward Model (PRM) demonstrates strong agreement for hierarchical planning tasks, including work breakdown structure generation and resource allocation, while exhibiting substantially weaker performance for task relationship evaluation. This divergence reflects differences in underlying reasoning requirements across planning stages.

Task relationship evaluation exposes a fundamental representational boundary of text-based PRMs. While linguistic representations support local and hierarchical reasoning, they do not explicitly encode the global network structure required to enforce precedence, concurrency, and constructability constraints across an activity graph. Multiple dependency configurations may appear locally plausible in text, but only a subset satisfies global feasibility. The near-zero correlation therefore reflects inconsistent scoring of locally valid but globally distinct dependency configurations due to missing structural information, rather than random model behavior.

This stage-wise performance profile motivates a tiered evaluation strategy in which the PRM is deployed as a first-stage screening and guidance mechanism rather than a final certification authority. The PRM's strength lies in comparative ranking of planning candidates based on hierarchical consistency and feasibility, enabling early filtering of low-quality plans. More complex dependency reasoning, which requires explicit topological constraints, can then be addressed using higher-fidelity evaluation methods capable of enforcing global network logic. In this role, the PRM enables focused application of detailed validation within large-scale automated planning pipelines.

Critically, within the iterative planning process of Algorithm 1, absolute correctness of intermediate evaluations is less important than relative improvement across candidate actions. The PRM guides selection by estimating which candidate is likely to yield a higher-quality planning state, rather than certifying global schedule correctness. Even when exact agreement on task relationships is limited, the PRM fulfills its role by consistently identifying inferior candidates and steering generation away from infeasible reasoning paths. This reliance on ordinal ranking rather than precise cardinal scoring, together with minimal systematic bias, explains why the PRM remains effective as a reward-guided planning signal despite imperfect dependency grading.

The PRM further supports this role by requiring substantially fewer computational resources and enabling significantly faster inference than SLM-as-a-Judge evaluation under identical hardware assumptions. This efficiency allows quality assessment to be embedded directly into the planning loop, supporting early suppression of low-quality reasoning paths and iterative refinement of candidate plans.

These findings directly inform future research directions. Improving task relationship evaluation will require incorporating graph-structured representations of task networks and explicit constructability constraints into the PRM input space. Additional opportunities include integrating PRM-based guidance with rule-based schedulability validators and applying active learning strategies to expand supervision in regions where PRM and high-fidelity evaluations diverge. Together, these directions point toward hybrid evaluation architectures that combine linguistic reasoning, structured representations, and deterministic constraints for robust automated construction planning.

7 Conclusion

This study introduces a process-aware framework for automated construction planning that embeds construction-specialized supervision through a Process Reward Model. The primary contribution lies in enabling planning-time quality assessment across intermediate planning stages, including work breakdown structure generation, task relationship formulation, duration estimation, and resource allocation, thereby supporting early identification and correction of planning errors before they propagate.

The framework's key contribution is the explicit alignment of evaluation strategy with representational requirements of different planning tasks. By recognizing that hierarchical decomposition and topological sequencing demand different forms of supervision, the proposed approach positions the PRM as an effective first-stage screening mechanism within a tiered evaluation architecture, while reserving dependency validation for methods capable of enforcing global network constraints.

Beyond immediate performance outcomes, this work clarifies both the strengths and boundaries of text-based intermediate supervision in construction planning and demonstrates how these boundaries can inform system design. Future work will focus on developing graph-augmented PRM architectures, expanding evaluation

across diverse construction domains, incorporating active learning strategies, and integrating learned rewards with deterministic schedulability validation in hybrid planning frameworks and extending the framework to support multi-domain schedule integration and project-level duration computation. Collectively, these directions advance the practical realization of reliable, domain-aware, AI-assisted construction planning systems.

Acknowledgements

The authors thank the National Science and Technology Council, Taiwan, for supporting this research through research grant Nos. NSTC 114-2119-M-002-034.

References

[1] M. O. Sanni-Anibire, R. Mohamad Zin, and S. O. Olatunji. Causes of delay in the global construction industry: a meta-analytical review. International Journal of Construction Management, 22(8):1395–1407, 2022.

[2] Y. J. T. Zidane and B. Andersen. The Top 10 Universal Delay Factors in Construction Projects. International Journal of Managing Projects in Business, 11(3):650–672, 2018.

[3] F. U. Hassan and T. Le. Extraction of Activities Information from Construction Contracts Using Natural Language Processing (NLP) Methods to Support Scheduling. In Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics - Selected Papers from Construction Research Congress 2022, volume 2-B, pages 773–781, 2022.

[4] Y. Hong, H. Xie, G. Bhumbra, and I. Brilakis. Improving the accuracy of schedule information communication between humans and data. Advanced Engineering Informatics, 53:101645, 2022.

[5] R. Ren and J. Zhang. Semantic Rule-Based Construction Procedural Information Extraction to Guide Jobsite Sensing and Monitoring. Journal of Computing in Civil Engineering, 35(6):04021026, 2021.

[6] X. Li, C. Wu, Z. Yang, Y. Guo, and R. Jiang. Knowledge graph-enabled adaptive work packaging approach in modular construction. Knowledge-Based Systems, 260:110115, 2023.

[7] F. Amer, Y. Jung, and M. Golparvar-Fard. Construction schedule augmentation with implicit dependency constraints and automated generation of lookahead plan revisions. Automation in Construction, 152:104896, 2023.

[8] J. Zou, L. Yang, J. Gu, J. Qiu, K. Shen, J. He, and M. Wang. ReasonFlux-PRM: Trajectory-Aware PRMs for Long Chain-of-Thought Reasoning in LLMs. arXiv preprint arXiv:2506.18896, 2025.

[9] Y. Zhou, S. Jiang, J. Zhu, J. Li, L. Guo, F. Chen, and C. Zhang. Fin-PRM: A Domain-Specialized Process Reward Model for Financial Reasoning in Large Language Models. arXiv preprint arXiv:2508.15202, 2025.

[10] S. She, J. Liu, Y. Liu, J. Chen, X. Huang, and S. Huang. R-PRM: Reasoning-Driven Process Reward Modeling. In Proceedings of EMNLP 2025, pages 13449–13462, 2025.

[11] A. K. Singh and S. H. Hsieh. Multi-LLM-based augmentation and synthetic data generation of construction schedules and task descriptions with SLM-as-a-judge assessment. Advanced Engineering Informatics, 69:103825, 2026.

[12] Y. Zhang and X. Yang. CONSTRUCTA: Automating Commercial Construction Schedules in Fabrication Facilities with Large Language Models. arXiv preprint arXiv:2502.12066, 2025.

[13] V. Faghihi, A. Nejat, K. F. Reinschmidt, and J. H. Kang. Automation in construction scheduling: a review of the literature. International Journal of Advanced Manufacturing Technology, 81(9-12):1845–1856, 2015.

[14] A. Pal, J. J. Lin, F. Amer, S. Hsieh, and M. Golparvar Fard. Automatic mapping of schedule activities and reality models for tracking construction progress. Engineering, Construction and Architectural Management, pages 1–27, 2025.

[15] M. Khalifa, R. Agarwal, L. Logeswaran, J. Kim, H. Peng, M. Lee, H. Lee, and L. Wang. Process Reward Models That Think. arXiv preprint arXiv:2504.16828, 2025.

[16] S. Kim, J. Shin, Y. Cho, J. Jang, S. Longpre, H. Lee, S. Yun, S. Shin, S. Kim, J. Thorne, and M. Seo. Prometheus: Inducing Fine-Grained Evaluation Capability in Language Models. In 12th International Conference on Learning Representations, ICLR 2024, pages 1–37, 2024.

[17] F. U. Hassan, T. Le, and C. Le. Automated Approach for Digitalizing Scope of Work Requirements to Support Contract Management. Journal of Construction Engineering and Management, 149(4), 2023.

[18] X. Zhao, K. Yeoh, and D. K. H. Chua. Extracting Construction knowledge from Project Schedules by using Natural Language Processing. In Lecture Notes in Mechanical Engineering, pages 371–381. Springer, 2020.

Towards Practical Digital Twin Construction Systems

Timson Yeung[1] and Rafael Sacks[1]

[1]Seskin Virtual Construction Laboratory, Technion – Israel Institute of Technology, Israel
timsonyeung@campus.technion.ac.il, cvsacks@technion.ac.il

Abstract

Building construction projects are socio-technical production systems in which variance, uncertainty and unpredictable events are the norm. Such systems need short-cycle Plan–Do–Check–Act (PDCA) loops to detect anomalies in production flow as early as possible so that remedial actions can be evaluated and well-informed decisions made. This is the goal of Digital Twin Construction (DTC). We define practical DTC as a minimal coherent system that captures a project's current status, compares the status to design and planning intent, analyses variance, predicts the outcomes of possible interventions, and thus informs specific managerial decisions at the required cadence. We introduce a 'thinking backward' decision-centric design approach for DTC systems that traces requirements from a target decision and its PDCA cycle through knowledge and information needs to data acquisition and enabling technologies. The approach characterizes managerial decisions by autonomy, fidelity, and scope, and proposes five degrees of implementation as system capability configurations for incremental development. This contributes an actionable basis for specifying and incrementally implementing DTC systems that deliver decision value early while maintaining a clear pathway toward more advanced implementations.

Keywords

Digital Twin Construction; Automation; Lean Construction; Decision-centric System Design; Plan-Do-Check-Act; Knowledge-Information-Data

1 Introduction

Building construction projects are socio-technical production systems where variance, uncertainty, and unpredictable events are the norm [1]. Production rates fluctuate, defects trigger rework, and small human errors can compound into major delays, while design changes, technical constraints, and supply-chain disruptions add further uncertainty. Managing such conditions requires short-cycle PDCA loops [2] to detect deviations early and support timely corrective decisions while recovery remains possible.

DTC [3] addresses this need. It is a data-centric mode of construction management that combines digital-twin information systems with lean construction thinking to close production control loops and support PDCA during the construction phase. DTC workflow continuously compares project intent with updated project status and feeds the resulting evaluations to people or algorithms that intervene on site and verify outcomes in the next cycle.

State-of-the-art research shows progress toward realizing DTC workflows, including interoperable information management layers, increasingly automated progress monitoring. Case studies have demonstrated closure of intent to status to intent feedback loops in live projects [4,5]. These efforts reveal that the key challenge in implementing DTC systems is not the availability of sensing, analytics, or cloud infrastructure, but the design problem of turning these components into coherent decision support pipelines that reliably improve planning and control outcomes. Whereas successful case studies begin from a clearly bounded use case and demonstrate measurable benefits, many other initiatives struggle to translate technological capability into repeatable closed-loop decision making.

A common pattern in both research and practice is a predominantly technology-driven starting point: teams begin from available technologies and only later ask what decisions the system should improve, so that digitalization often fails to deliver full decision value.

We propose a decision-centric perspective for designing practical DTC systems that support incremental adoption without sacrificing coherence. This is a "thinking backward" approach [6] that traces requirements from a target decision and its PDCA cycle back through knowledge and information needs to data acquisition and enabling technologies. It characterizes production planning and control decisions along three dimensions, autonomy, fidelity, and scope, and presents degrees of DTC implementation as coherent system configurations that help teams set realistic targets, avoid overengineering, and plan a stepwise pathway toward stronger closed-loop control.

2 State-of-the-Art in Digital Twin Construction

Sacks et al. [3] introduced DTC as a data-centric mode of construction management that combines digital-twin information systems with lean construction thinking to close production control loops and support PDCA cycles during the construction phase. In the DTC workflow, design and planning intent is formalized as project intent information (the as-designed product and as-planned process). Project intent is continuously compared with project status information (the as-built product and as-performed process), which is derived from raw data from site and supply-chain monitoring that is interpreted using data-fusion functions. Subsequently, the comparisons yield conformance evaluations that inform project status knowledge suitable for proposing possible corrective interventions. In parallel, simulations and analytics produce new project intent knowledge that supports performance prediction and the evaluation of decision alternatives, while bidirectional feedback connects these digital representations back to both planning/design updates and real-time production control on site. Archiving digital building twins that capture all this information enables learning across projects.

Practical DTC systems require a system architecture with explicit data layers and interfaces that integrate sensing, management, and analytics into a decision support pipeline. A common approach uses Semantic Web technologies to link heterogeneous product, process, and status data through open standards. Schlenger et al. [7] addressed the limited guidance on data management layers for DTC by proposing a plug-in-based reference architecture that couples multiple RDF graphs with specialized databases and introduces a Digital Twin Construction Ontology to compare planning information with monitored status, validated on the ConSLAM dataset. Complementing this platform view, Zheng et al. [8] proposed the DiCon-DT framework for construction workflow monitoring, where an ontology network integrates diverse data into a semantic DT data lake and supports simulation and contextual interpretation.

Among the technical building blocks of DTC systems, automated progress monitoring has been developed rapidly. Recent studies demonstrate that imagery and video can support inference of construction progress [9]. For geometry centric assessment, point cloud pipelines continue to improve scan to BIM alignment and association for element level completion inference [10]. Complementary nonvisual sensing provides additional evidence streams that can help connect captured evidence to coordination and compliance actions in management workflows [11].

Beyond monitoring, researchers have connected digital twin status synchronization with simulation and optimization to enable decision models to update assumptions and explore alternatives using current constraints. Work in this area typically targets two horizons: production planning that supports system configuration and medium-term plans, and production control that adjusts daily logistics and execution in near real time. For example, Yeung et al. [12] integrated an ontology based digital twin with parametric agent based simulation to support adaptive production system design, enabling real time data ingestion, status model instantiation, and scenario analysis. Jiang et al. [13] proposed a digital twin-enabled real time synchronization system for just-in-time precast on site assembly that uses real time resource status and construction progress to support planning, scheduling, and execution coordination and dynamic adjustment.

Large scale research and innovation initiatives have also sought to integrate the DTC building blocks into end-to-end prototypes and to validate them beyond isolated case studies. Under the EU Horizon 2020 Digital Building Twins call, a coordinated cluster of projects moved construction phase digital twin concepts toward cloud prototypes that combine data integration with decision support services. BIM2TWIN aligned closely with the DTC system concept by specifying a Digital Building Twin platform based on semantic-linked data from multiple sensor streams into a real time project status model along with data from BIM and project management data, and exposes it through APIs and dashboards for an extensible set of construction management applications. ASHVIN and COGITO complement this platform logic with toolboxes and vertical services for monitoring, planning, and governed workflows, including quality and health and safety functions and traceability mechanisms such as smart contracts. BIMprove focused on real time tracking and field interfaces such as AR and VR to keep BIM and digital twin views synchronized during execution.

Beyond component-level advances, recent full-workflow demonstrators illustrate how DTC concepts translate into closed-loop production control in live projects. Valinejadshoubi et al. [4] implemented a closed-loop production control workflow for a cement silo project by integrating BIM, IoT sensing, and GIS to enable near-real-time progress monitoring, planned versus actual comparisons, and deviation-triggered management actions. Their system coupled a BIM model with an ArcGIS-based digital twin environment and a cloud database, and it streamed formwork sensing data to support timely conformance checking and corrective intervention. They reported a 28% reduction in execution time and a 15% reduction in construction wastage, which they attributed to faster information updates and tighter coordination enabled by the integrated platform.

Martinez et al. [5] presented a mechanized tunnelling demonstrator in which the digital twin closed an intent-

to-status-to-intent loop for bracketry drilling design and installation. The team consolidated TBM and segment-manufacturing data into an as-built tunnel lining model and then used the updated status model to define no-drill zones, perform clash detection, and support bracketry design and execution. Automating this consolidation reduced the per-ring modelling effort from 15 to 20 minutes to about 1 minute of manual quality checks, cutting cycle time by more than 90%. The authors also noted persistent barriers such as imperfect interoperability across proprietary software and hardware systems and the lack of standardized BIM-IFC representations of tunnel geometry and information.

3 Decision-Centric System Design

The state-of-the-art review shows rapid progress toward realizing DTC workflows, including interoperable information management layers, increasingly automated progress monitoring, and end-to-end demonstrators that close intent to status to intent feedback loops in live projects. These efforts also reveal that the hardest part of DTC system implementation is often not the availability of sensing, analytics, or cloud infrastructure, but the design problem of turning these ingredients into a coherent decision support pipeline that reliably improves planning and control outcomes. Whereas successful case studies begin from a clearly bounded use case and demonstrate measurable benefits, many other initiatives struggle to translate technological capability into repeatable closed loop decision making.

A common pattern in both research and practice is a predominantly technology-driven starting point. Teams adopt tools such as cameras, LiDAR sensors, or monitoring services first and only later ask what decisions they improve and how they strengthen a closed-loop PDCA cycle. This opportunistic mindset is understandable. In research, it is often easier to trial novel sensing, algorithms, or system architectures than to define and validate end value across heterogeneous stakeholders. In commercial settings, technology vendors naturally lead with their product propositions and attempt to persuade project teams to adopt their products. As a result, many DTC implementations start from available technologies and then work forward toward decisions and learning loops.

We propose to flip this logic and take a decision-centred approach to DTC system design. DTC is framed as a management paradigm whose core purpose is to support and improve decisions through systematic learning, rather than as an umbrella label for digitalization technologies. The starting point is therefore the set of managerial decisions that are intended to be made within a closed loop PDCA cycle.

Two broad horizons are especially relevant in construction. Production control decisions govern near term execution, including daily coordination, constraint removal, sequencing, workface logistics, and the allocation of crews, equipment, and materials. Production planning decisions shape the production system itself, including adaptive production system design and the periodic reconfiguration of plans, control policies, and resource strategies in response to evolving constraints and performance. In both horizons, the central question is not what the digital twin can technically compute, but what decisions must be made, who makes them, how often they must be revisited, and what learning is expected from repeated cycles.

From an explicit target decision, the system designer can work down through levels of knowledge, information and data. At the knowledge level, a designer asks what knowledge about project intent and project status is needed to make and improve the decision over time, including what must be understood about conformance, deviations, causes, and consequences. At the information level, they ask what information must be synthesized to support those knowledge claims. For example, which state variables, constraints, progress states, resource states, and performance indicators must be explicitly represented. At the data level, they ask what combinations of raw observations can be used to derive that information. Multiple data sets can often support the same information requirements, and direct measurement is not always feasible. In such cases, indirect measurements, partial measurements, or complementary measurements from different tools may be combined through interpretation and data fusion to reconstruct the information of interest. Only once these requirements have been defined explicitly does it become rational to select monitoring solutions, data management strategies, and system architectures that implement a pipeline from data to information and then to knowledge that supports the decision defined at the outset.

This sequence helps avoid over engineering and over complication, because it focuses attention on the minimal sufficient evidence and transformations needed for the decision, rather than on maximizing the volume of captured data or assembling an attractive architecture that lacks a clear operational purpose.

A practical consequence of starting from decisions is that the use case must be characterized in a way that sets realistic expectations for the required system capability and complexity. As illustrated in Figure 1, production planning or control decisions can be described using three parameters with values arrayed along three axes: *autonomy*, *fidelity*, and *scope*. The autonomy axis describes the prospective agency of the human planner in the decision process, ranging from *human decide*, through *human decide with machine advice*, to *machine decide with human supervision*, and ultimately to

machine decide, while recognizing that responsibility, risk, and centralized versus distributed authority create intermediate nuances. The fidelity axis describes how close the information must be to reality for the decision to be valid. Fidelity can be specified through required *precision* or *resolution*, required *accuracy*, *acceptable latency* and *update frequency*. The scope axis describes the breadth and interdependence of the decision within the production system, including whether it is *localized or global*, whether it is a *single decision or a coupled set of decisions*, and whether it creates dependencies and cascading effects across other decisions and system states. Together, these axes help determine whether a simple single component solution is sufficient or whether the use case demands multiple modules, rich data fusion, and/or sophisticated predictive or prescriptive analytics.

We can apply this decision-centric approach to the silo case study [4] reviewed in the previous section. In slip-form construction, the team can first isolate the near-term production-control decisions that plausibly benefit from short-cycle PDCA. The silo workflow supports four such decisions: 1) whether to adjust formwork alignment to maintain verticality and rotation, 2) whether to intervene when curing conditions drift toward risk, 3) whether to change pouring rate and near-term resource allocation to maintain planned progress, and 4) what materials or embedded components to pull to the site next as progress advances. As a group, these decisions remain in the *human decide* autonomy class. They stay within a small and localized scope centred on the silo production-control loop. The use case demands high fidelity, with minute-level temporal resolution and high accuracy, and an hourly update frequency that matches the production-control cadence.

With the decision set fixed, the designer can then trace down the knowledge, information, and data pyramid to elicit system requirements. Figure 2 summarises this trace as a dependency map. At the knowledge level, the four decisions require knowledge of structural alignment relative to intent, curing condition status relative to thresholds, progress versus plan, geometric conformance for embedded elements, and material pull needs for the next work segment. At the information level, these knowledge requirements translate into a minimal set of information products that the system must maintain, including real-time construction parameters, trend histories that expose drift, visual status indicators on the model, progress metrics, alerts that flag excursions, and an as-built versus as-designed overlay that supports conformance assessment. At the data level, the same trace identifies a minimal viable data set and its sources, including distance and inclination sensing for geometry and stability, temperature sensing for curing conditions, BIM intent geometry as the reference for planned states, timestamps to support trend and auditability, and site spatial context where needed for situational framing.



Figure 1. Orthogonal axes characterizing a DTC system use case

This worked example shows how the thinking-backwards approach helps convert a case description into an explicit requirement specification for a practical DTC system. By linking a target decision and its cadence to concrete knowledge, information products, and data sources, the approach guides purposeful technology selection and, equally, prepares the production team to use the resulting information routinely to check conformance and act within the decision window.

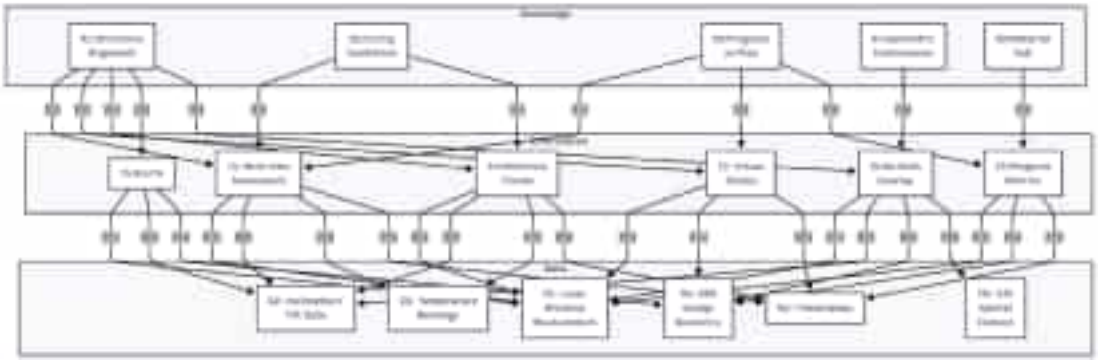


Figure 2. Derivation of required knowledge, information, and data for use cases in Valinejadshoubi et al. [20]

4 Degrees of DTC System Implementation

Building on the decision-centric framing, a key practical question is, "Given a target decision, its requirement for autonomy, fidelity, scope, knowledge, information, and data, what degree of DTC implementation is realistic, and what capabilities must be in place for the PDCA loop to run systematically?" The supportive and autonomous visions of DTC provide useful long-term reference points, but they can also appear overly ambitious when viewed against the realities of construction variability, fragmented responsibilities, and the cost of instrumentation and integration. At the same time, meaningful DTC is not an all-or-nothing proposition. Projects can achieve measurable improvement with partial digitization and partial automation, provided that the system delivers decision support that is traceable to the use case and enables learning across repeated cycles.

This section therefore proposes five degrees of DTC implementation (Figure 3) that describe coherent capability bundles, rather than a checklist of isolated technical features. The degrees presented in Figure 3 are not intended as a strictly linear maturity ladder. Organizations can adopt different degrees for different decisions, and a single project can exhibit mixed degrees across planning horizons and work packages. What changes across degrees is primarily the extent to which the system implements the knowledge-information-data chain for the target decision, and the extent to which monitoring, interpretation, and feedback become structured, repeatable, and timely. Higher degrees typically correspond to higher attainable autonomy, higher fidelity of the intent and status representations, and broader scope of the decision support.

The first degree is **unassisted**. Here, the PDCA loop exists in the traditional sense described in lean construction and production control practice, but it is largely manual and relies on managerial judgement. The only structured digital representation of intent is usually a schedule and associated planning artifacts, while project status awareness is formed through site walks, phone calls, ad hoc updates, and meetings. Teams check progress against the plan and revise near-term commitments, but the evidence base is fragmented and informal, so the loop's effectiveness depends on the availability, memory, and interpretation capacity of the management team.

The second degree is the **digital model**. Relative to the unassisted case, teams deploy monitoring technologies and generate richer project status data, for example through photogrammetry and computer vision, scanning, tagging, or other capture methods that observe both inputs and outputs of the production system. These observations can indicate where labour and materials have been deployed and what intermediate or final products have been installed in specific work zones. The key characteristic is that the data typically flows into the human planner's situational awareness model rather than into a structured digital interpretation pipeline. Project intent information still resides primarily in the schedule, BIM, and planning documents, and the comparison of intent and status remains a human task. This degree is therefore best described as data-augmented rather than systematically data-driven. Richer, more frequent, and potentially more objective observations support decisions, but they do not yet consistently produce consolidated project status information and actionable knowledge.

The third degree is **digital shadow**. This degree retains monitoring of inputs and outputs, but it differs from the digital model in that the system integrates multiple data sources and performs explicit interpretation and data fusion to generate consolidated project status information. Different sensing modalities provide partial views of the production system, and their combination makes it possible to reconstruct a more complete picture across resources, processes, and products. For example, beacons can provide evidence of labour presence and

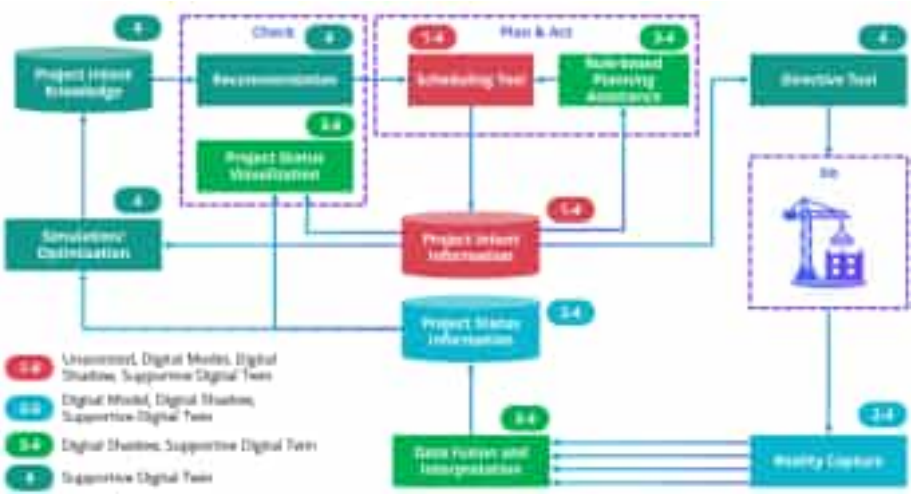


Figure 3. Degrees of DTC system implementation

movement, RFID can indicate material location and consumption, image-based methods can capture visible installation progress, and aerial or terrestrial scanning can characterize geometry for exterior or structural elements. The system uses these complementary observations to support consistent conformance evaluation, not only identifying deviations between intent and status but also helping explain why deviations occur, such as delays driven by missing inputs, rework, or disrupted workflows. Importantly, the feedback loop is increasingly explicit. Updated project intent information, including revised plans and constraints, is recorded and used as input for subsequent evaluations, enabling rule-based planning assistance and systematic iteration across PDCA cycles.

The fourth degree is the **supportive digital twin**. At this degree, the system extends beyond consolidated status to generate project intent knowledge through simulation, analytics, and optimization. The digital shadow foundation provides the evidence base for current state estimation, while predictive and prescriptive modules explore what-if scenarios, estimate consequences, and propose decision alternatives. This corresponds to predictive situational awareness, where the system supports not only understanding what happened and why, but also anticipating what will happen next and evaluating the likely impact of alternative actions. The system remains supportive because the human planner retains decision authority. It provides recommendations, explanations, and visualizations that help the planner decide whether and how to act, and the resulting plan updates are fed back into digital representations to support continuous improvement.

The fifth degree is the **autonomous digital twin**, which can be described as a digital agent governing decision selection. Here, the fused project status information and the project intent information feed an autonomous decision algorithm that evaluates alternatives and selects an action, which is then written back into the project intent representation and issued as directives, policies, or control inputs for execution. In the ideal form, the planner no longer participates in the loop. However, in practice, construction conditions, risk, and capital intensity will motivate a supervised autonomy arrangement in which human planners oversee, approve, or override algorithmic decisions. This makes explainability, traceability, and logging critical, both for operational assurance and for long-term learning through a digital repository of intent, status, and decision histories that can improve future models and algorithms.

Across these degrees, the central point remains consistent with the preceding section: the PDCA loop is not binary. Lower degrees can still run PDCA cycles, but with higher reliance on intuition and informal evidence. Higher degrees make the loop more systematic by formalizing what must be known, what must be inferred, how intent and status are compared, and how actions are selected and recorded. This typology therefore provides a practical bridge from a decision-centric system design approach to implementable system targets.

In relation to existing digital twin maturity models, these degrees align broadly with established capability progressions in the literature, particularly the transition from digital model and digital shadow toward predictive or prescriptive and autonomous twins described in models such as [14] and [15]. The similarity lies in the increasing formalization of data integration, interpretation, and feedback, and in the resulting shift from status visibility toward prediction, decision support, and eventually decision automation. The novelty of the proposed degrees is that they recast this progression for construction planning and control around the implementation of a target PDCA loop rather than around digital twin technological maturity. This shift is reflected first in the inclusion of an explicit unassisted degree, which recognizes that PDCA already exists in practice before a formal digital twin is introduced and therefore starts the framework from the real managerial baseline rather than from the first digital artifact. It is reflected again in the definition of the higher degrees, which are distinguished not simply by more advanced technology but by how systematically the knowledge-information-data chain is implemented to support a specific managerial decision.

5 Discussion

DTC initiatives frequently begin with a technology-first mindset, where teams adopt sensing and analytics tools and then attempt to fit management decisions around whatever data those tools produce. This approach often results in fragmented architectures and unclear value, because it optimizes technical capability rather than strengthening a specific PDCA loop. The framework proposed in this paper reverses that sequence. It starts from a decision that must be improved, and it treats system components as means to close that decision loop reliably.

The autonomy, fidelity, and scope characterization helps developers translate decisions into implementable requirements. Autonomy clarifies who decides and what the system is allowed to do, fidelity clarifies how accurate and timely the information must be, and scope clarifies how far the decision's effects propagate across planning and control. Framed this way, a project team can state what the system must deliver before selecting technologies, which reduces over-engineering and helps align organisational responsibilities, data ownership, and update frequency with the actual decision context.

Once requirements are clear, the knowledge–

information–data decomposition provides a practical route from decision needs to system design. It derives the minimal data capture and processing pipeline from what the decision maker must know about project intent and project status, rather than collecting data opportunistically. This logic supports lean implementation, including the use of partial measurements, proxies, and staged sensing deployments, provided they satisfy the fidelity requirements of the targeted PDCA loop.

Finally, the degrees of DTC implementation position DTC as incremental progress rather than a yes or no achievement. They enable teams to diagnose what a system can currently support and to prioritize the next feasible capability that materially strengthens the decision loop, without assuming that full automation or comprehensive sensing is necessary. The remainder of this discussion illustrates this practicality by applying the framework to three real-world demonstrators.

Table 1 details the proposed characterization with respect to three case studies and shows, first, how each implementation starts from a clearly bounded managerial decision and, second, how the system architecture is designed to deliver the specific status information needed to act on that decision.

Valinejadshoubi et al. [4] targeted concreting crew task assignment. While decisions remained human-centric (human decide), the use case demanded high fidelity with minute-level resolution. Consequently, the system required automated fusion of formwork sensing data to trigger threshold-based recommendations, placing it at the Digital Shadow level (Degree 3).

Similarly, the mechanized tunnelling system [5] supported bracketry design. It operated at a lower temporal cadence (daily) but required precise spatial data fusion. By consolidating TBM and manufacturing data into an as-built model to identify no-drill zones, it provided machine advice, also qualifying as a Digital Shadow.

As part of the EU BIM2TWIN research project, the authors implemented a DTC system for adaptive production system design rather than short-cycle production control [12]. The system integrated an ontology-based digital twin with agent-based simulation to ingest current project status, instantiate the status model, and evaluate alternative production configurations under changing project conditions. Relative to the first two case studies, it added predictive recommendation and supported updating the baseline production plan, but it did not rely on automated reality capture and only partially automated conformance evaluation. It therefore sits between Degree 3 and Degree 4, showing that the proposed degrees are better understood as practical capability regions along a spectrum than as rigid thresholds that a project must fully satisfy before it becomes useful.

Table 1. Classification of use case and degree of DTC implementation for three case studies

Study		**Valinejadshoubi et al. [4]**	**Martinez et al. [5]**	**Yeung et al. [12]**
Use Case		Concreting crew task assignment, conformance checking	Define no-drill zones, perform clash detection, and support bracketry design and execution	Adaptive production system design through adjusting resource, location, and work package parameters
Autonomy		Human decide	Human decide with machine advice	Human decide with machine advice
Fidelity	**Temporal Resolution**	Minute	Day	Day
	Update Frequency	Hourly	Daily	Weekly
	Accuracy	High	High	Medium
Scope		Small	Medium	Medium
Reality Capture		Yes, automated	Yes, automated	No
Data Fusion and Interpretation		Yes, automated	Yes, automated	Yes, automated
Conformance Evaluation		Yes, automated	Yes, automated	Yes, semi-automated
Simulation/Optimization		No	No	Yes, agent-based simulation
Recommendation		Yes, remedial action if monitored concrete condition below preset threshold	Yes, no drill zones per segment	Yes, alternative production system configurations
Report & Visualization		Yes, sensor data visualization and concrete pouring progress	Yes, 3D as-built IFC model with production information per segment	Yes, schedule, progress, and simulation results of alternative plans in visual dashboard
Rule-based Assistance		No	No	No
Directive Tool		No	No	Yes, update new baseline production plan to digital twin repository
Degree of DTC Implementation		3 - Digital Shadow	3 - Digital Shadow	Between 3 – Digital Shadow and 4 – Supportive Digital Twin

Taken together, the case studies show that practical DTC implementation is already feasible with existing technologies when the system is designed around a bounded managerial decision and its associated PDCA loop. The reported outcomes show that meaningful operational value can be achieved without an all-encompassing or highly autonomous twin, provided that the implementation delivers the specific status information, interpretation, and decision support required for that use case with sufficient reliability and cycle time. At the same time, the framework is most readily applied when the target decision and its PDCA loop can be defined clearly. In more highly coupled project settings, applying it may require decomposing interdependent decisions into smaller linked loops before coherent system requirements can be specified. Seen this way, the proposed degrees are not targets to be pursued for their own sake, but a framework for characterizing which capabilities have been implemented, which remain absent, and which additional capabilities are justified by the expected decision value, thus reframing DTC from an idealized end state to a practical system design problem of aligning scope, automation, and decision support with concrete managerial needs.

6 Conclusion

Short-cycle PDCA is needed to manage variance and uncertainty during construction production, yet DTC systems are often conceived based on available technologies and idealized end states rather than being based on the decisions they must support. This paper proposed a decision-centric design perspective and the notion of practical DTC as the minimal system capability required to improve a specific PDCA decision at its required cadence.

The paper introduced three decision axes, derived design principles that trace requirements from the target decision through information and knowledge needs to data acquisition and defined five degrees of DTC implementation to guide staged development. Applied to three real-world case studies, the framework separated the managerial use cases from the particular technologies and architectures used to implement them and clarified that PDCA improvement is achievable at multiple degrees of digitalization and automation. Together, these contributions provide a pragmatic basis for specifying and incrementally implementing DTC systems that prioritize decision value over technology completeness.

References

[1] M. Abdelmegid, A. Tezel, C.O. Sandoval, Z. Fang, W. Collinge, Conceptualizing Digital Twins in Construction Projects as Socio-Technical Systems, ISARC Proceedings 2024 Proceedings of the 41st ISARC, Lille, France (2024) 577–584.

[2] W.E. Deming, Out of the crisis, reissue, MIT press, 2018.

[3] R. Sacks, I. Brilakis, E. Pikas, H.S. Xie, M. Girolami, Construction with digital twin information systems, DCE 1 (2020) e14.

[4] M. Valinejadshoubi, R. Sacks, F. Valdivieso, C. Corneau-Gauvin, A. Kaptué, Digital Twin Construction in Practice: A Case Study of Closed-Loop Production Control Integrating BIM, GIS, and IoT Sensors, Journal of Construction Engineering and Management 151 (2025) 05025014.

[5] E. Martinez, F. Stauch, B. García De Soto, R. Sacks, Integration of building information modelling and digital twin in mechanized tunnelling construction, ECAM (2025).

[6] B. Flyvbjerg, D. Gardner, How big things get done: the surprising factors that determine the fate of every project, from home renovations to space exploration and everything in between, Crown Currency, 2023.

[7] J. Schlenger, K. Pluta, A. Mathew, T. Yeung, R. Sacks, A. Borrmann, Reference architecture and ontology framework for digital twin construction, Automation in Construction 174 (2025) 106111.

[8] Y. Zheng, A. Al Barazi, O. Seppänen, H. Abou-Ibrahim, C. Görsch, Semantic digital twin framework for monitoring construction workflows, Automation in Construction 176 (2025) 106301.

[9] F. Pfitzner, S. Hu, A. Braun, A. Borrmann, Y. Fang, Monitoring concrete pouring progress using knowledge graph-enhanced computer vision, Automation in Construction 174 (2025) 106117.

[10] Y. Zhang, L. Lu, X. Luo, J. Pan, Global BIM-point cloud registration and association for construction progress monitoring, Automation in Construction 168 (2024) 105796.

[11] M. Khazen, M. Nik-Bakht, O. Moselhi, Monitoring workers on indoor construction sites using data fusion of real-time worker's location, body orientation, and productivity state, Automation in Construction 160 (2024) 105327.

[12] T. Yeung, J.G. Martinez, J. Schlenger, A. Borrmann, R. Sacks, Integrating digital twin and agent-based simulation to support adaptive production system design in building projects, Automation in Construction 180 (2025) 106550.

[13] Y. Jiang, M. Li, M. Li, X. Liu, R.Y. Zhong, W. Pan, G.Q. Huang, Digital twin-enabled real-time synchronization for planning, scheduling, and execution in precast on-site assembly, Automation in Construction 141 (2022) 104397.

[14] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital Twin: Enabling Technologies, Challenges and Open Research, IEEE Access 8 (2020) 108952–108971.

[15] W. Kritzinger, M. Karner, G. Traar, J. Henjes, W. Sihn, Digital Twin in manufacturing: A categorical literature review and classification, Ifac-PapersOnline 51 (2018) 1016–1022.

Disaster Response Report Generation System

Ming-Lu Liu[1] and James Yichu Chen[1]

[1]Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan
m11205504@gapps.ntust.edu.tw, yie@mail.ntust.edu.tw

Abstract –

This study develops an automated disaster response report generation system that integrates large language models (LLMs), web scraping, and computer vision techniques. The system automatically retrieves data from 16 heterogeneous and distributed sources and, by combining Retrieval-Augmented Generation (RAG) with Chain-of-Thought (CoT) prompting, produces first drafts of disaster response reports that conform to the official reporting framework. Empirical results show that the system achieves a System Usability Scale (SUS) score of 85.8 and reduces report preparation time by approximately 73% for typhoon events and 64% for heavy-rain events. Expert evaluations indicate that the generated content attains acceptable levels of completeness, accuracy, and readability. To further validate content reliability, TLM trustworthiness scores of 0.89 and 0.95 confirm high factual reliability, while RAGAS Faithfulness scores of 0.99 (RAG–CoT) versus 0.43 (vanilla LLM) demonstrate the contribution of the proposed framework. These findings confirm the feasibility and practical potential of the proposed system for automated report generation in professional disaster management contexts.

Keywords –

Disaster Response, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Chain-of-Thought (CoT), Credibility Evaluation

1 Introduction

In recent years, extreme weather events have become increasingly frequent worldwide, and the associated economic and social impacts of natural disasters continue to escalate. In Taiwan, weather-related disasters cause average annual economic losses of approximately NT$15 billion (about US$490 million). According to statistics from the Water Resources Agency (WRA), the number of disaster response operations has increased markedly, from about 40–50 per year between 2015 and 2020 to 60–90 per year after 2020. The growing frequency of response operations has substantially increased the workload of frontline personnel. In particular, the preparation of post-disaster response reports—which must be completed within 30 days and serve as critical records for institutional learning—has become one of the most labor-intensive components of disaster response work.

At present, disaster response reports are still written largely by hand through a complex and time-consuming process. Taking the 2024 Typhoon Gaemi response report as an example, the preparation of this report required integrating 16 distinct data sources, compiling 11 tables, and producing 26 figures, resulting in a document of 163 pages. The entire process consumed approximately three full working days. Moreover, the sections describing meteorological evolution require substantial domain expertise; new staff members often need several months of training before independently drafting such reports, and during peak flood seasons, report writing must be conducted concurrently with ongoing emergency operations, further intensifying the burden on already-limited personnel.

In response to these challenges, this study proposes an automated disaster response report generation system that integrates LLMs, web scraping, and computer vision. The system primarily targets typhoon events and large-scale heavy-rain events, and seeks to improve report-writing efficiency through automated data acquisition, chart generation, and draft text generation. Furthermore, the system adopts a RAG–CoT framework and incorporates TLM-based evaluation to achieve the following objectives:

1. Reduce workload and improve efficiency by automating the labor-intensive processes of data collection and initial report drafting.
2. Integrate scattered heterogeneous data into standardized reporting templates, thereby lowering the technical threshold and facilitating the rapid onboarding of new staff.
3. Validate generative performance and content reliability through a hybrid evaluation framework combining expert assessment, TLM credibility scoring, and RAGAS faithfulness metrics.

2 Related Work

This section reviews literature relevant to the proposed system. First, it introduces the structure and

data integration challenges of disaster response reports in Taiwan. Second, it examines the automated report generation technologies adopted in this study, including web scraping, computer vision, and LLMs, and discusses hallucination issues and corresponding mitigation strategies. Finally, it surveys the current applications of LLMs in disaster response, which form the theoretical foundation of this research.

2.1 Disaster Response Report Standards and Architecture

Disaster response reports are key documents in the post-disaster recovery and mitigation cycle. In Taiwan, the Water Resources Administration of the Ministry of Economic Affairs has formulated structured writing standards for disaster response reports, requiring them to systematically present the background of the incident, meteorological and hydrological evolution, response operations, and the actual impact of disasters. These reports cover three core modules: event background and meteorological developments, hydrological status and disaster impact, and response operations and notification records.

However, producing such standardized disaster response reports relies heavily on cross-system data integration. In practice, the information required for the reports is distributed across multiple independent information systems that differ substantially in technical architecture, data formats, and access mechanisms. This study identifies 16 distinct data sources dispersed across 6 web platforms (Table 1), making cross-system integration a highly complex technical challenge. These data sources are maintained by various government agencies and include HTML tables, JSON data, image files (JPG/PNG), and unstructured text. Issues such as missing data and inconsistent time stamps are common. Under tight response timelines, specialists must repeatedly access multiple systems to retrieve and reconcile data, highlighting both the cumbersome nature of existing workflows and the urgent need for automation.

Table 1 Classification of data sources

Category	Primary Source Platforms and Formats
Water Conservancy Monitoring	Alert Issuance Log, Disaster Emergency Response System (TXT, HTML, JSON, DOCX)
Meteorological Information	Weather Analysis and Taiwan Climate Hybrid monitor system (WATCH), Typhoon Database (TXT, JPG, PNG)
External Warning	Public Open Data Platform (TXT)
Marine Conditions	Safe Ocean (XLSX)

2.2 Report Automation and LLMs-Based Technology

In the field of disaster management, automation has been widely adopted to handle multi-source heterogeneous data and real-time requirements. With respect to data collection, web scraping combined with API access has been used to automatically aggregate disaster-related information from government open data portals, alert platforms, and news or social media streams, thereby improving situational awareness[1, 2]. However, most of these studies focus on real-time monitoring dashboards rather than on the long-term event histories and multi-chapter structures required for formal reporting.

For data visualization and chart understanding, prior work has shown that directly summarizing charts using end-to-end language models can easily lead to numerical hallucinations[3]. A more reliable approach is to first recover intermediate structured data (e.g., tables or statistical matrices) from chart images and then apply data-to-text generation, which improves the consistency between textual descriptions and the underlying quantitative data[4]. This suggests that, in the context of formal disaster reports, converting meteorological and hydrological maps into machine-readable structured indicators is an important prerequisite for trustworthy text generation.

LLMs have been widely applied in text summarization, question answering, and automated report generation. Nevertheless, in high-risk professional domains such as disaster response, content accuracy and hallucinations remain major concerns[5, 6]. To address these issues, recent research has proposed strategies such as RAG and CoT[5]. RAG dynamically retrieves verifiable external information during generation, grounding model outputs in authoritative documents and reducing hallucination rates. CoT prompting encourages models to perform step-by-step reasoning, decomposing complex tasks into a sequence of intermediate reasoning steps, thereby improving the structure and interpretability of responses. Empirical studies have shown that combining RAG and CoT can enhance the reliability of generated professional texts while maintaining overall performance.

In the disaster response domain, automated reporting pipelines and disaster information systems have been developed to support situation awareness and documentation, while recent studies survey broader AI applications across multiple phases of disaster management, from early warning to post-event analysis. Recent interdisciplinary reviews indicate that AI and

LLM applications in disaster management span multiple phases—from detection and situation analysis to decision support and post-event documentation—with social media analysis and decision recommendation representing the most mature application areas[7]. For instance, DisasterResponseGPT integrates response guidelines and expert knowledge into prompts to accelerate the formulation of response plans[8]. Beyond individual LLM applications, knowledge graph-enhanced systems have been developed to provide evidence-based recommendations in emergency decision-making scenarios[9], and multi-agent AI frameworks have been proposed to integrate sensing, analysis, and communication functions into unified disaster management architectures [10]. Other studies employ LLMs to aggregate disaster-related texts and integrate multi-source information[11, 12]. However, existing approaches rarely integrate visual information, structured operational data, and standardized governmental reporting formats, and few quantitatively assess content credibility. This study addresses these gaps by proposing a RAG-CoT framework that automatically generates structured, traceable first drafts of Taiwan's water-related disaster response reports with quantitative credibility evaluation..

3 Methodology

This section describes the architecture and implementation of the proposed disaster response report generation system. The core design principle is human–AI collaboration: automated components handle highly repetitive data-integration tasks, while LLMs assist with drafting narrative sections, allowing domain experts to concentrate on high-value judgment and decision-making. The system supports the automatic generation of six sections of a complete disaster response report: Event Overview, Meteorological Situation, Hydrological Information, Emergency Response Team Operations, Disaster Impacts, and Conclusion and Annexes. Among these, the first three sections are highly narrative and require the integration of a large volume of meteorological and hydrological information; consequently, they are the primary focus of the LLMs-based generation strategies in this study.

3.1 System Architecture

The overall system architecture is shown in **Error! Reference source not found.**. It consists of six main modules: Data Collection, Data Acquisition, Data Visualization, Data Analysis, Text Generation, and Report Integration. Through the user interface, report compilers specify the event name and the time ranges for rainfall and water-level data. The Data Collection and Data Acquisition modules then automatically retrieve the raw data required for each chapter from multiple platforms. These data are passed to the Data Visualization and Data Analysis modules, which generate the figures and analytical results required in the report. The Text Generation module uses the structured datasets and analytical outputs to produce draft text for the corresponding sections. Finally, the Report Integration module assembles all generated texts, tables, and figures into a pre-defined Word template, thereby producing a first draft of the disaster response report for the selected event.

3.2 Clarification of Data Sources and Processing Pipelines

To ensure that the automated system can effectively support real-world reporting requirements, the content of the disaster response report was first analyzed and categorized according to information type and generation risk. The report includes both highly descriptive narrative texts and quantitative content—tables and figures—for which numerical accuracy is critical. Relying on a single generation strategy for all content would likely yield uneven quality and factual errors. Accordingly, this study adopts a diverted generation design, in which different sections are handled by different technical pipelines depending on their semantic and numerical requirements. The narrative sections—Event Overview, Meteorological Situation, and Hydrological Information—must synthesize a large number of weather bulletins, observational records, and response logs into coherent textual descriptions. For these sections, the system leverages LLMs with integrated RAG and CoT prompting. Automatically collected and pre-processed meteorological and hydrological data are converted into reference documents that serve as contextual inputs for the LLMs, which in turn generates draft paragraphs aligned with official reporting style and structure. In contrast, sections such as the charts in Hydrological Information, Emergency Response Team Operations, Disaster Impacts, Conclusion, and Annexes are primarily composed of numerical data, statistical tables, and visualizations. The key objectives for these sections are numerical correctness and consistent formatting rather than free-form narrative. For them, the system relies on automated data processing and visualization pipelines to perform data cleaning, computation, and figure/table generation. In practice, the pipeline consists of three steps: heterogeneous inputs from the 16 data sources are first converted into a unified, event-based schema with basic cleaning (e.g., missing-value handling and time alignment). The cleaned data are then aggregated into chapter-specific statistical indicators, which the visualization module uses to automatically generate standardized charts and tables and insert them into the predefined Word report template. This approach yields standardized charts and tables that conform to reporting specifications and ensures the reliability of the

quantitative content.

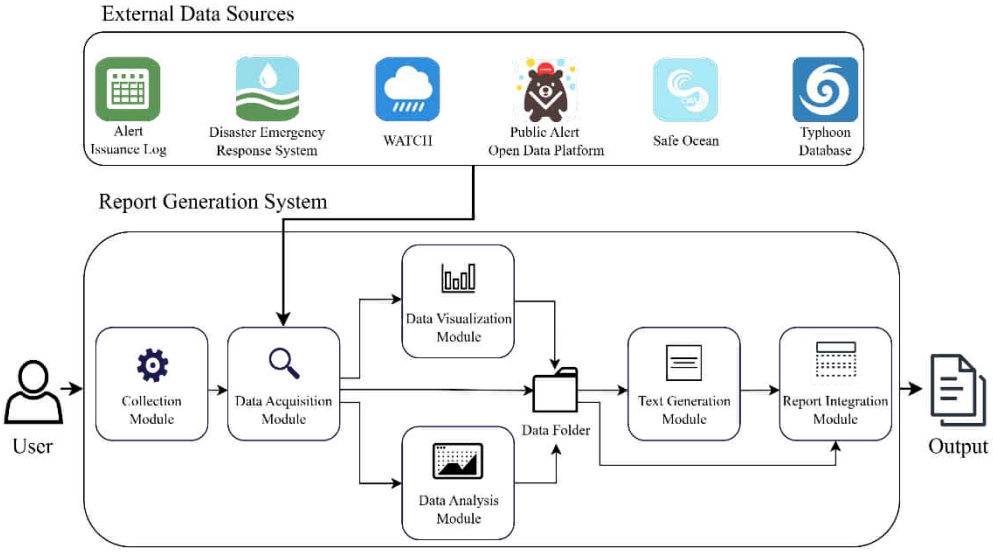


Figure 1 System architecture of the disaster response report generation system

3.3 RAG and CoT Implementation

To improve text generation quality and reduce hallucination risks, the system integrates RAG and CoT strategies, as illustrated in Figure 2. The multi-level prompting mechanism is designed to generate descriptive paragraphs for the Event Overview, Meteorological Situation, and Hydrological Information sections. For the RAG component, long-duration typhoon events can accumulate a large volume of textual data that exceeds the input length limits of current LLMs. To address this, the system first segments multi-day weather bulletins and related texts into smaller chunks, then encodes them into vectors using an embedding model and stores them in a FAISS vector database. During generation, the system performs similarity search based on the current prompt and retrieves the most relevant text fragments, which are supplied as contextual evidence to the LLMs. For the CoT component, prompts explicitly specify the reasoning steps required to construct each paragraph, decomposing complex generation tasks into sequential sub-tasks. Few-shot exemplars are included to demonstrate the desired writing style and paragraph structure of disaster response reports. For example, in the "Typhoon Development History" subsection, the prompt guides the model to chronologically summarize the typhoon track, key time points (e.g., warning issuance, landfall, departure, and cancellation), and to integrate rainfall and wind information at each stage, thereby producing narratives with clear timelines and causal relationships. In addition, hard constraints are embedded in the prompts to restrict the model to using only information retrieved by the RAG component, thereby discouraging unsupported speculation. Through the joint use of RAG and CoT, the system generates interpretable and well-grounded draft texts based on multi-source data, which are then subject to expert review and refinement.

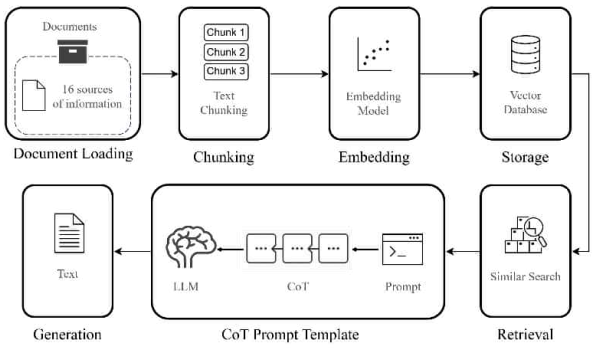


Figure 2 RAG - CoT architecture diagram

4 Validation

The system was evaluated across three stages: (1) system usability assessment, (2) expert evaluation of content quality and production efficiency, and (3) TLM credibility assessment. Test participants were personnel responsible for drafting disaster response reports at the Water Resources Administration of the Ministry of Economic Affairs of Taiwan.

4.1 Validation Methodology

Stage 1 – System usability.

The first stage examined whether the system effectively integrates dispersed data and reduces the

difficulty of report writing. Participants were asked to complete predefined tasks using the system without assistance. Their task completion time and operational difficulties were recorded. After the tasks, participants completed the SUS questionnaire and participated in semi-structured interviews to supplement the quantitative results. Six users with experience in disaster response report writing were recruited: two were classified as novice and four as experienced based on their years of practice, allowing the study to capture the perspectives of different user groups.

Stage 2 – Expert assessment of content quality and efficiency.

In the second stage, experts evaluated both the quality and efficiency of system-generated reports. One typhoon event and one heavy-rain event were selected as test cases, and experts with more than ten years of disaster response experience were invited to participate. For content quality, the experts rated each automatically generated paragraph on three dimensions—completeness, correctness, and readability—and compared the results with manually written versions to determine whether the generated content covered the necessary information and remained readable. All ratings used a five-point Likert scale. For production efficiency, experts recorded (i) the time required for the system to generate the first draft, (ii) the time needed to revise the drafts, and (iii) the time required to write the reports entirely by hand. The overall system-assisted production time was defined as the sum of generation and editing times and was compared with purely manual writing to calculate the percentage of time saved.

Stage 3 – Credibility assessment using TLM.

The third stage aimed to objectively and reproducibly evaluate the credibility of the system-generated texts, with particular attention to hallucination risks. Even when RAG is employed, LLMs may still produce misleading or incorrect statements; thus, an automated reliability check is necessary[13]. This study adopts the TLM proposed by Cleanlab[14] to evaluate the generated paragraphs in the Event Overview and Meteorological Situation and Hydrological Information sections. TLM quantifies the trustworthiness of responses based on the relationship among the user query, retrieved context, and generated output. The evaluation includes an overall trustworthiness score and four sub-indicators: context sufficiency, response groundedness, response helpfulness, and query ease. These metrics allow us to assess whether the generated content is sufficiently supported by retrieved evidence and internally consistent when synthesizing multi-source data, and the results also guide further system optimization. In addition, RAGAS is employed as a complementary evaluation tool to compare the RAG–CoT-enhanced configuration with a vanilla LLM using the same underlying model, quantifying differences in Faithfulness and Answer Relevancy and thereby demonstrating the contribution of the RAG–CoT architecture to overall content quality.

4.2 Sample Selection

To ensure representative samples for the Stage-3 TLM evaluation, disaster events that occurred in 2024 were screened using predefined criteria. Typhoon events were required to trigger at least a Level-1 response by the Ministry of Economic Affairs and to cause multiple flooding incidents across counties and cities. Large-scale heavy-rain events were required to trigger at least a Level-2 response or to result in 10 or more flooding locations. Based on these criteria, eight events were selected—four typhoon events and four heavy-rain events—as summarized in Table 2.

Table 2 List of selected typhoon and heavy-rain events

Event Name	Cumulative Rainfall (mm)	Screening Criteria
Typhoon GAEMI	1,944	Level 1 response: 3,308 floods in 15 counties and cities
Typhoon KRATHON	1,725	Level 1 response: 604 floods in 8 counties and cities
Typhoon KONG-REY	1,184	Level 1 response: 109 floods in 9 counties and cities
Typhoon USAGI	492	Level 1 response: 9 flooding in 3 counties and cities
0602 Heavy-Rain	223	Level 2 response: 12 floods in 5 counties and cities
0710 Heavy-Rain	135	Level 3 response: 55 flooding in 1 county and city
0806 Heavy-Rain	191	Level 3 response: 16 flooding in 2 counties and cities
0921 Heavy-Rain	436	Level 2 response: 83 floods in 6 counties and cities

4.3 Result

The SUS was used to evaluate overall system usability. According to Bangor et al. (2009), an average SUS score of 68 is generally considered the benchmark for acceptable usability[15, 16]. The proposed system achieved an average SUS score of 85.8 (Figure 3), which falls into the Excellent range, indicating a high level of usability in terms of both operation and user experience.

Further analysis revealed clear differences between novice and experienced users. Experienced staff scored the system between 92.5 and 100, reflecting strong recognition of its benefits for data integration and report preparation. Novice users, by contrast, assigned scores between 57.5 and 65, slightly below the benchmark. This suggests that users who have previously written large-scale disaster reports are more capable of perceiving the practical advantages of the system for multi-source data integration and draft generation, whereas novices tend to focus on interface intuitiveness and may have limited understanding of the complex disaster-response scenarios the system is designed to support. Nevertheless, during follow-up interviews, novice users generally agreed that the automatically generated Event Overview, Meteorological Situation, and Hydrological Information sections helped them quickly grasp the report structure and served as useful templates for writing.

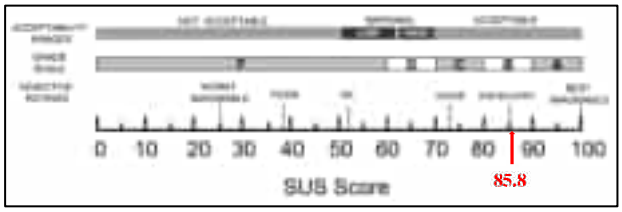


Figure 3 System Usability Scale score

Table 3 summarizes the expert evaluation results. For the specific content checklist, all chapters generated by the system obtained average scores of at least 3 out of 5, indicating that the model was able to cover the key information required in disaster response reports. For the overall quality dimensions—completeness, correctness, and readability—average scores were all at or above 4, suggesting that the generated content met the basic quality requirements of formal response reports. Experts noted that a small number of paragraphs could benefit from more precise wording and tighter narrative focus, but overall, only minor manual editing was needed before practical use. In terms of production efficiency, manually writing a typhoon event report required approximately 36 hours. With system assistance (automatic draft plus expert editing), the total time was reduced to about 9.60 hours, corresponding to a 73% reduction. For heavy-rain event reports, manual writing required about 2 hours, whereas the system-assisted process required approximately 0.73 hours, saving around 64% of the time. These results demonstrate that the proposed system can substantially improve report-writing efficiency for disaster events of varying scales.

TLM was employed to assess the credibility of the Event Overview and Meteorological Situation and Hydrological Information sections generated by the system. Scores range from 0 to 1; following Cleanlab's guidelines, values below 0.7 are considered low confidence, while values above 0.9 are regarded as high confidence. As shown in Table 4, the Event Overview and Meteorological Situation section achieved an average trustworthiness score of 0.89, close to the high-confidence range, whereas the Hydrological Information section achieved an average of 0.95, which clearly falls within the high-confidence range. These results indicate that the generated content is well supported by factual evidence and is overall reliable. Regarding the sub-indicators, average scores for context sufficiency, response groundedness, and response helpfulness all exceeded 0.9. In particular, response groundedness and response helpfulness both averaged 0.99, implying that the generated responses were highly dependent on retrieved evidence and exhibited virtually no hallucinations.

Table 3 Expert evaluation result table

Evaluation Dimension	Indicator	Evaluation Result
Completeness	Key Information Coverage	Average Score ≧ 3
Quality	Overall Quality Score	Average Score ≧ 4
Efficiency (Typhoon)	Report Generation Time	36 h → 9.60 h (73% Time Saved)
Efficiency (Heavy-Rain)	Report Generation Time	2 h → 0.73 h (64% Time Saved)

Slight differences were observed between disaster types: compared with typhoon events—where data continuity and publicly available information are more complete—heavy-rain events obtained a lower Query Ease score because the corresponding hydrological generation tasks must integrate heterogeneous evidence sources (textual alerts, numerical daily maximum rainfall data, and JSON-formatted rainfall-map analysis results) and follow prompts that require the model to infer daily weather patterns and affected areas during the response period, making these queries inherently more complex for the TLM to evaluate. Even so, both the Hydrological Information and Event Overview and Meteorological Situation sections remained at high credibility levels.

Overall, the TLM results demonstrate that the proposed system can consistently generate highly trustworthy and practically useful textual content in complex multi-source data settings.

Beyond assessing trustworthiness with TLM, this study further examined whether incorporating the proposed RAG–CoT framework improves the quality of LLM-generated reports. As shown in Table 5, both configurations achieved very similar Answer Relevancy scores (~0.82), indicating that the vanilla LLM is capable of generating topically relevant content. However, the large difference in Faithfulness (0.99 vs. 0.43) shows that vanilla LLM outputs frequently contain fabricated or unsupported statements, which is a critical risk in professional disaster reporting where factual accuracy is required. These results demonstrate that adopting the proposed RAG–CoT framework improves the factual grounding of automatically generated reports in this professional domain.

Table 4 Credibility assessment result table

Indicators/Chapters	Event Overview and Meteorological Situation	Hydrological Information
Trustworthiness	0.89	0.95
Context Sufficiency	0.98	0.93
Response Groundedness	0.99	0.99
Response Helpfulness	0.99	0.99
Query Ease	0.96	0.80

Table 5 vanilla LLM vs. RAG-CoT

Indicators/Configuration	Vanilla LLM	RAG–CoT	Improvement
Faithfulness	0.43	0.99	+ 0.56
Answer Relevancy	0.81	0.82	+ 0.01

5 Discussion

The disaster response report generation system developed in this study integrates web scraping, geospatial data analysis, and LLM-based text generation to automatically produce draft content for the Event Overview, Meteorological Situation, and Hydrological Information sections. Expert evaluations confirm that the generated content meets the basic quality requirements of formal disaster response reports, while the system reduces report preparation time by approximately 73% for typhoon events and 64% for heavy-rain events. Based on the 96 heavy-rain and 4 typhoon response events recorded in Taiwan in 2024, purely manual drafting would require an estimated 108 working days, whereas the proposed system reduces this workload to 17 working days, substantially easing the operational burden on response units during periods of intensive disaster activity. In addition, the SUS score of 85.8 (Excellent) indicates that the system achieves strong usability in practical settings.

Despite these promising results, several limitations remain. First, the system currently depends heavily on the availability and stability of external data sources and website structures; when source platforms change their interfaces or data formats, the scraping and parsing components must be updated accordingly. When critical data cannot be retrieved or are insufficient to generate a given section, the system automatically inserts a placeholder (e.g., “to be completed by responders”) into the draft report to clearly indicate that manual completion is required. Second, although the generated texts are generally factually correct, final review and refinement by domain experts are still required to ensure that the tone, wording preferences, and points of emphasis align with the conventions of specific agencies and with politically or socially sensitive issues. Third, in the case of small, localized heavy-rain events, the limited availability of official data can lead to insufficient context, causing the model to produce overly brief descriptions or explicit indications of missing data, which may reduce the perceived completeness and credibility of the final report.

In this study, the report structure is decomposed by analyzing each chapter’s content into three categories—narrative text, visual maps and charts, and numerical tables—and assigning each category to an appropriate technical module (LLM with RAG–CoT prompting for narrative sections, computer vision for map-derived indicators, and automated data processing and visualization pipelines for quantitative content). The same design principle can be applied in other regions or hazard types by re-analyzing the local report template,

classifying its required content into these three categories, and then adapting the data sources, computer-vision routines, and RAG document collections and prompt templates to local monitoring systems and reporting standards, while reusing the core RAG–CoT and report-integration architecture.

Future work will focus on three directions: developing an interactive query interface so that responders can use natural language to explore historical events and analogous cases, integrating historical reports and impact data to provide simple, context-aware response recommendations, and introducing AI agents and the Model Context Protocol (MCP) to enhance the system's ability to self-check, self-repair, and scale in real-world operational environments.

6 Conclusion

To address the practical need for responders to handle both real-time operations and formal report writing under high event frequency, this study proposes an automated disaster response report generation system that integrates LLMs, web scraping, and geospatial data analysis. The results show that, when coupled with a structured data-integration mechanism and interpretable generation strategies, LLMs can reliably support the production of first drafts of disaster response reports with acceptable credibility and practical utility. Overall, these findings confirm the practical feasibility of applying LLMs to governmental disaster-prevention documentation and public-sector reporting tasks, offering a modular and reusable framework adaptable to other professional domains.

Acknowledgement

This work was supported by the National Science and Technology Council (NSTC), Taiwan, under Grant **NSTC 115-2625-M-011-001**. The authors also thank the domain experts and practitioners for their insightful feedback and contributions during system evaluation.

References

1. Kusumasari, B. and N.P.A. Prabowo, *Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific.* Natural Hazards, 2020. **103**(3): p. 3415–3435.
2. Abid, S., et al. *Social Media Data: Challenges, Mitigation Strategies, and Opportunities for Disaster Management*. in *2024 International Conference on Frontiers of Information Technology (FIT)*. 2024.
3. Obeid, J. and E. Hoque, *Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model.* arXiv preprint arXiv:2010.09142, 2020.
4. Poco, J. and J. Heer. *Reverse-engineering visualizations: Recovering visual encodings from chart images*. in *Computer graphics forum*. 2017. Wiley Online Library.
5. Kumar, A., et al., *Improving the reliability of LLMs: Combining CoT, RAG, self-consistency, and self-verification.* arXiv preprint arXiv:2505.09031, 2025.
6. Naveed, H., et al., *A Comprehensive Overview of Large Language Models.* ACM Trans. Intell. Syst. Technol., 2025. **16**(5): p. Article 106.
7. Xu, F., et al., *Large language model applications in disaster management: an interdisciplinary review.* International Journal of Disaster Risk Reduction, 2025: p. 105642.
8. Goecks, V.G. and N.R. Waytowich, *Disasterresponsegpt: Large language models for accelerated plan of action development in disaster response scenarios.* arXiv preprint arXiv:2306.17271, 2023.
9. Chen, M., et al., *Enhancing emergency decision-making with knowledge graphs and large language models.* International Journal of Disaster Risk Reduction, 2024. **113**: p. 104804.
10. Li, B., et al., *Disaster Management in the Era of Agentic AI Systems: A Vision for Collective Human-Machine Intelligence for Augmented Resilience.* arXiv preprint arXiv:2510.16034, 2025.
11. Colverd, G., et al., *Floodbrain: Flood disaster reporting by web-based retrieval augmented generation with an llm.* arXiv preprint arXiv:2311.02597, 2023.
12. Chen, Z., et al., *Integration of large vision language models for efficient post-disaster damage assessment and reporting.* arXiv preprint arXiv:2411.01511, 2024.
13. Sardana, A., *Real-Time Evaluation Models for RAG: Who Detects Hallucinations Best?* arXiv preprint arXiv:2503.21157, 2025.
14. Northcutt, C.G., *Overcoming Hallucinations with the Trustworthy Language Model.* 2024.
15. Bangor, A., P. Kortum, and J. Miller, *Determining what individual SUS scores mean: Adding an adjective rating scale.* Journal of usability studies, 2009. **4**(3): p. 114–123.
16. Lewis, J.R. and J. Sauro, *Item benchmarks for the system usability scale.* Journal of Usability studies, 2018. **13**(3).

A BIM–IoT-Based Framework for Lifecycle Carbon Tracking in Modular Construction

Yu Wang[1], Yue Teng[1], Yucheng Guo[1], and Geoffrey Qiping Shen[1]

[1]Department of Building and Real Estate, The Hong Kong Polytechnic University, China

wangyu1.wang@connect.polyu.hk, yue.teng@polyu.edu.hk
yucheng.guo@connect.polyu.hk, geoffrey.shen@polyu.edu.hk

Abstract -

Modular construction (MC) has attracted increasing attention as a promising strategy for carbon emission reduction in the context of global low-carbon transitions. However, accurate estimation and continuous tracking of lifecycle carbon emissions in MC projects remain limited, as existing studies predominantly focus on individual stages and provide fragmented evidence. To address this gap, this study proposes a lifecycle-oriented carbon tracking approach for MC based on the integration of Building Information Modeling (BIM) and the Internet of Things (IoT). By combining BIM-based object representations with multi-source IoT-enabled sensing mechanisms, the proposed approach establishes a unified mechanism for carbon accounting and data integration at the module level, enabling consistent representation of carbon emissions within a unified framework. The approach is demonstrated through two real-world modular integrated construction (MiC) projects in Hong Kong, validated separately for the construction and operational stages. The results show that the approach supports continuous, time-resolved characterization of carbon emissions and effectively captures their dynamic patterns under different application contexts. Overall, this study provides a systematic methodological framework for lifecycle-oriented carbon tracking in modular construction, contributing to more data-driven carbon assessment practices.

Keywords -

Modular Construction, Lifecycle Carbon Monitoring, BIM, IoT

1 Introduction

The building sector contributes a substantial share of global energy-related carbon emissions, making it a key target for decarbonization strategies [1]. MC has been increasingly adopted to improve productivity and reduce environmental impacts, particularly in high-density cities such as Hong Kong [2]. However, existing studies have struggled to systematically characterize the carbon emission profile of MC at the lifecycle scale. This limitation is largely attributable to the lack of methods capable of supporting continuous lifecycle carbon data acquisition and tracking.

Existing MC carbon studies still rely predominantly on material-driven estimation approaches. Although BIM and IoT have been introduced to support carbon data acquisition and management, their applications are typically confined to individual stages or isolated systems. As a result, a lifecycle-wide data integration mechanism and unified accounting pathway have yet to be established, limiting further systematic analysis of carbon emission characteristics.

From the perspective of embodied carbon (EC), MC exhibits highly discrete and process-oriented characteristics. Conventional cast-in-situ construction is dominated by on-site activities, with carbon emissions largely localized in both spatial and temporal dimensions. Although prefabricated construction relocates part of component production to factories, substantial wet works remain on-site, and EC emissions are still primarily material-driven. MC adopts modules as highly integrated, standardized, and traceable construction units, resulting in EC emissions being distributed across multiple construction stages, including factory manufacturing, transportation, and on-site assembly. Due to highly dispersed production processes and a module-based object hierarchy, carbon emission data in MC are difficult to continuously acquire and consistently represent across construction stages, which weakens the applicability of traditional material-based estimation approaches at both module and building scales.

For operational carbon (OC), MC exhibits stronger early-stage coupling with EC. Module envelope design, connection configurations, and integrated building services are typically finalized during the manufacturing stage, thereby exerting persistent influences on building energy performance and operational carbon during long-term operation. However, existing OC studies in MC predominantly rely on design-stage energy simulations for independent estimation, which cannot capture dynamic variations during actual operation and lack effective linkage with construction-stage data, limiting deeper understanding of operational carbon behaviours.

To address these gaps, this study proposes a lifecycle-oriented BIM–IoT-based carbon tracking approach for MC. The proposed approach supports continuous carbon data acquisition and integration across lifecycle stages and enables unified carbon accounting across different stages, providing a systematic methodological framework for data-driven carbon assessment in modular construction. The remainder of this paper is organized as follows.

Section 2 reviews related literature, Section 3 introduces the proposed methodology, Section 4 presents the case studies and results, and Section 5 concludes the paper.

2 Literature review

In recent years, increasing research attention has been directed toward the use of digital technologies to address the complexities of lifecycle carbon assessment in MC [3]. Due to the fragmented nature of production processes and the hierarchical organization of building objects in MC, effective carbon assessment requires consistent object-level representation across different stages, and the capability to continuously capture time-resolved data from construction processes and building operation. In this context, BIM and IoT have emerged as key technological foundations for carbon data integration and analysis in modular construction.

BIM provides structured building semantics and supports automated quantity take-off, enabling multi-level representation of carbon emissions from components to modules and buildings [4]. It also facilitates identification of dominant material- and process-related emission sources [5]. However, existing BIM-based approaches largely rely on static databases and ex post accounting mechanisms, which limits their ability to capture temporal dynamics of carbon emissions [6].

Meanwhile, IoT technologies enable distributed sensing and real-time data acquisition, allowing process-level information across construction and operation stages to be continuously recorded, which is particularly critical for fragmented production processes in MC [7, 8]. Nevertheless, recent review studies indicate that IoT-based carbon management in the construction sector still faces challenges related to limited interoperability, fragmented system deployment, and lack of standardized integration frameworks, which constrain scalability and applicability for cross-stage carbon tracking [9].

Building on the above studies, recent research has explored BIM–IoT integration to establish links between digital building models and physical process data [10]. Existing studies indicate that such approaches can support process-level tracking of embodied carbon at component and module scales and can also be applied to operational energy monitoring and data acquisition [11]. However, compared with existing BIM–IoT-based studies that primarily focus on embodied carbon tracking during the construction stage, most applications remain focused on specific processes, individual lifecycle stages, or localized use cases, and a unified framework that enables lifecycle-wide data integration and consistent carbon accounting has yet to be established [12, 13].

In summary, although BIM and IoT provide important capabilities for carbon data acquisition in modular construction, systematic methods that support lifecycle-wide data integration and unified carbon accounting across stages remain insufficient. To address this gap, this study proposes a BIM-IoT-based carbon tracking framework for modular construction to support continuous lifecycle carbon data acquisition and integrated carbon accounting.

3 Methodology

3.1 Research design

This study proposes a BIM–IoT-integrated framework for carbon tracking in modular construction, in which modules are defined as the primary analytical unit and buildings as the aggregation level. The overall research framework of the proposed BIM–IoT-integrated approach is illustrated in Figure 1. By linking module-level BIM semantics with process- and operation-level data acquired through IoT, the framework enables consistent module-scale attribution of carbon emissions across lifecycle stages, providing a basis for the integrated analysis of embodied and operational carbon.

3.2 Data acquisition and preprocessing

This study employs a data acquisition strategy targeting lifecycle stages of modular buildings to support the quantitative analysis of carbon emissions during both construction and operation. In accordance with EN 15978 [14], and considering the research scope and data availability, the system boundary includes construction-related stages A1–A5, representing embodied carbon (EC), and the use stage B6, representing operational carbon (OC). Other lifecycle stages, including maintenance and replacement (B2–B5), end-of-life processes (C1–C4), and benefits beyond the system boundary (D), are excluded. The corresponding carbon emission sources, data types, and acquisition approaches are summarised in Table 1. For construction-related stages (A1–A5), data acquisition is organised around process-level activities, focusing on energy use associated with module production, transportation, and installation. In contrast, for the operational stage (B6), data acquisition is structured by end-use energy consumption, capturing subsystem-level demand such as HVAC, lighting, and plug loads.

The proposed method integrates three categories of input data: BIM semantic data, IoT-derived sensing data, and carbon emission factors. BIM semantic data define a structured hierarchy linking components, modules, and buildings, enabling consistent carbon attribution across lifecycle stages. IoT-derived data capture time-resolved activity states and energy-use patterns associated with module production, transportation, installation, and building operation. Carbon emission factors are obtained from established life cycle assessment databases, including Ecoin-

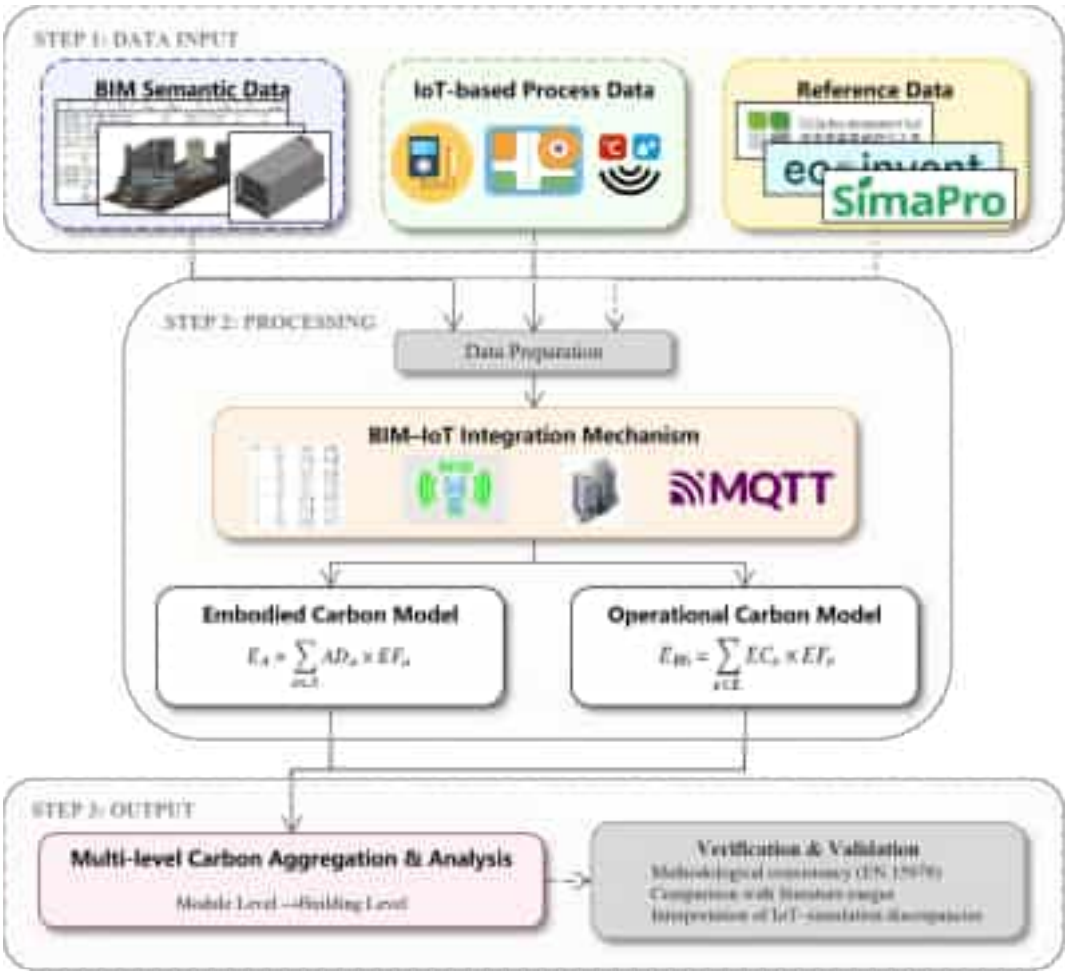


Figure 1. Research design of the proposed BIM-IoT integrated framework.

vent [15], SimaPro-integrated databases [16], ICE [17], and the CIC Carbon Assessment Toolkit [18].

Prior to integration, heterogeneous data from different lifecycle stages are preprocessed to improve temporal consistency and data reliability. This includes addressing differences in sampling frequency and temporal resolution, handling missing or incomplete data, and aligning time-resolved IoT data with BIM-defined modular objects and corresponding emission factors. These steps provide a reliable basis for subsequent module-level carbon quantification and lifecycle-scale analysis.

3.3 BIM–IoT integration mechanism

The integration of BIM and IoT introduces challenges related to object consistency and temporal alignment, as data collected from different lifecycle stages exhibit heterogeneity in structure, temporal resolution, and activity semantics. Without a unified reference framework, such data cannot be reliably linked to modular objects, thereby limiting the traceability of carbon emission quantification.

To address this issue, a module-oriented BIM–IoT integration mechanism is established, in which the module serves as the unified unit of analysis across lifecycle stages. Each module is assigned a unique identifier (Module ID) and linked to its corresponding BIM object, ensuring consistent identity representation within the defined system boundary. A one-to-one mapping between Module IDs and RFID tags enables physical identification of modules and supports the association of activity-related data during construction processes.

During module manufacturing, transportation, and onsite installation, RFID-based identification events are captured at key process nodes to link operational states and energy consumption data to individual modules. In this way, each construction activity can be directly associated with specific modules, allowing energy use to be traced back to corresponding processes. This activity-oriented integration approach establishes a clear connection between construction activities, energy consumption, and modu-

Table 1. Carbon emission sources and IoT sensors

Stage	Carbon Source Category	Representative Activities	Data Type	IoT Sensors
A1–A3	Material production; Module manufacturing	Raw material processing; Module manufacturing; Module decoration	Energy consumption; Equipment status	Smart meters; Accelerometers; Vibration sensors; etc.
A4	Module transportation	Off-site transportation	Travel distance; Transport status	GPS sensors
A5	On-site installation	Module lifting; Equipment operation	Energy consumption; Equipment status	Smart meters; Accelerometers; Vibration sensors; etc.
B6	HVAC systems	AC on/off switching; Temperature adjustment	Operating duration; Temp. & humidity	Environmental sensors; Current sensors
	Plug loads	Appliance usage; Device charging	Active power; Device status	Smart plugs; Smart switch
	Lighting	Light on/off switching; Active fixture count	Illuminance level; Operating status	Illuminance sensors; Smart switches
	Domestic water use	Water consumption; Usage patterns	Water volume; Flow rate	Smart water meters
	Gas consumption	Cooking; Hot water supply	Gas volume; Operating status	Smart gas meters

lar components, thereby enabling traceable module-level carbon quantification.

To ensure the temporal consistency and effective integration of multi-source IoT data, a publish–subscribe communication mechanism is adopted, with MQTT facilitating asynchronous data acquisition from distributed sensors. By linking module identifiers with activity types and corresponding data topics, heterogeneous operational data collected at different temporal resolutions can be systematically organized at the module level. This supports cross-stage data alignment and provides a coherent basis for subsequent carbon quantification.

3.4 Carbon emission accounting model

Based on the BIM–IoT data integration mechanism, carbon-related data generated at different lifecycle stages are consistently attributed at the module level. Within the defined system boundary, this study quantifies embodied carbon during the construction stages (A1–A5) and operational carbon during the use stage (B6) of modular buildings. All emissions are expressed in carbon dioxide equivalent (CO_2e), including major greenhouse gases such as CO_2, CH_4, and N_2O. Greenhouse gas emissions are converted into CO_2e using 100-year global warming potential factors (GWP100) from the IPCC Fifth Assessment Report (AR5) [19]. This choice ensures consistency with established LCA databases and prior studies commonly used in construction-related carbon assessment. Although updated characterization factors are provided in IPCC AR6 (2021), the differences are generally limited for the emission sources considered in this study and do not affect the methodological framework or comparative analysis.

For the construction stages (A1–A5), EC is quantified based on activity-related energy consumption captured through the IoT sensing system. Energy use associated with construction-related activities is converted into carbon emissions using corresponding emission factors, following an activity-based accounting logic. The general calculation form is expressed as:

$$E_A = \sum_{a \in A} AD_a \times EF_a \tag{1}$$

Where E_A denotes the total embodied carbon emissions of a module during the construction stages (A1–A5); AD_a represents the activity data associated with activity a, such as energy consumption, equipment operating time, or transportation distance; and EF_a refers to the emission factor corresponding to the activity and associated energy type.

For the operational stage (B6), OC covers emissions induced by energy use during normal building operation. Based on time-resolved energy consumption data collected by the IoT system, operational activities such as HVAC operation, lighting usage, and plug-load equipment operation are identified and aggregated over their operating periods. Operational carbon emissions are calculated as:

$$E_{B6} = \sum_{e \in E} EC_e \times EF_e \quad (2)$$

Where E_{B6} represents the operational carbon emissions associated with energy use during the use stage (B6); EC_e denotes the consumption of energy type e; and EF_e is the emission factor corresponding to the specific energy type.

After module-level quantification, carbon emissions are further aggregated across different object levels using the hierarchical structure defined in BIM. At the component level, emissions can be disaggregated based on component–module relationships, while at the building level, module-level emissions are aggregated to derive overall carbon indicators. This hierarchical aggregation ensures consistency across object levels and supports lifecycle-scale carbon analysis.

3.5 Verification and validation

The reliability of the proposed lifecycle carbon assessment approach is examined through methodological consistency checks and comparative interpretation of results. Rather than aiming to establish absolute accuracy, the evaluation focuses on whether the framework produces reasonable and interpretable carbon estimates under real operating conditions.

Methodological consistency is ensured by adopting the lifecycle stage definitions of EN 15978 [14] and applying a unified activity-based accounting logic across construction and operational stages. This provides a coherent basis for cross-stage comparison and aggregation of carbon emissions.

The plausibility of the results is assessed through two complementary perspectives. First, the magnitude of estimated carbon emissions is compared with ranges reported in existing modular construction studies, providing a reference for overall emission levels. Second, IoT-based monitoring results are interpreted in relation to simulation outputs, not as a direct validation benchmark, but as a means of examining deviations between assumed and actual conditions. In particular, discrepancies in temporal patterns are analysed to identify behaviour-driven variations, such as occupant-induced fluctuations in energy use, which are not captured by schedule-based simulations.

Given the absence of fully observed lifecycle carbon data, these comparisons remain indirect and do not constitute validation against ground truth. Instead, they provide supporting evidence on the interpretability and practical relevance of the proposed approach. This limitation reflects the inherent challenges of lifecycle carbon assessment under real-world conditions.

4 Case study

4.1 Embodied carbon emissions in the construction stage

Case I is a MiC student dormitory project in Kowloon Tong, Hong Kong, selected to evaluate the proposed BIM–IoT method for monitoring construction-stage embodied carbon (A1–A5). The project is a 15-storey steel-structured dormitory consisting of 867 prefabricated modules, with a gross floor area of approximately 50,200 m^2. A cast-in-situ portion of the project contributes about 33,042.82 tCO_2e during construction. As these emissions cannot be reliably attributed at the module level, they are excluded from the analysis. It should be noted that this exclusion represents a substantial portion of the total construction-stage emissions and therefore constitutes an important limitation of the module-level analysis. Within this scope, on-site emissions are limited to crane operations for module lifting, while other equipment energy use without consistent module-level attribution is omitted, in accordance with the defined module-based accounting boundary.

Figure 2 illustrates the composition of module-level embodied carbon emissions. Material-related emissions dominate across all module types, accounting for approximately 94.3% of the total, while contributions from production equipment, transportation, and on-site lifting remain relatively small. As summarised in Table 2, embodied carbon per module ranges from approximately 11.0 to 11.9 tCO_2e across different module types. This consistency, together with the dominant contribution of material-related emissions observed in Figure 2, indicates that, in a highly standardised MiC system, module-level embodied carbon is primarily governed by shared material compositions and uniform production arrangements, with limited influence from functional or spatial differences. These patterns are consistent with previous studies on modular construction [20, 21, 22]. Overall, Case I demonstrates that the BIM-IoT-based monitoring approach enables attributable and interpretable embodied carbon analysis during construction, particularly in scenarios where conventional inventory-based methods cannot provide traceable module-level attribution.

4.2 Operational carbon emissions in the operational stage

Case II investigates a MiC student dormitory project in Wong Chuk Hang, Hong Kong, to evaluate the proposed BIM–IoT-based approach for operational-stage carbon monitoring (B6). The project has been in operation since September 2023 and comprises two 17-storey MiC residential towers with 952 prefabricated modules. This case serves as an operational-stage validation of the proposed framework, complementing the construction-stage

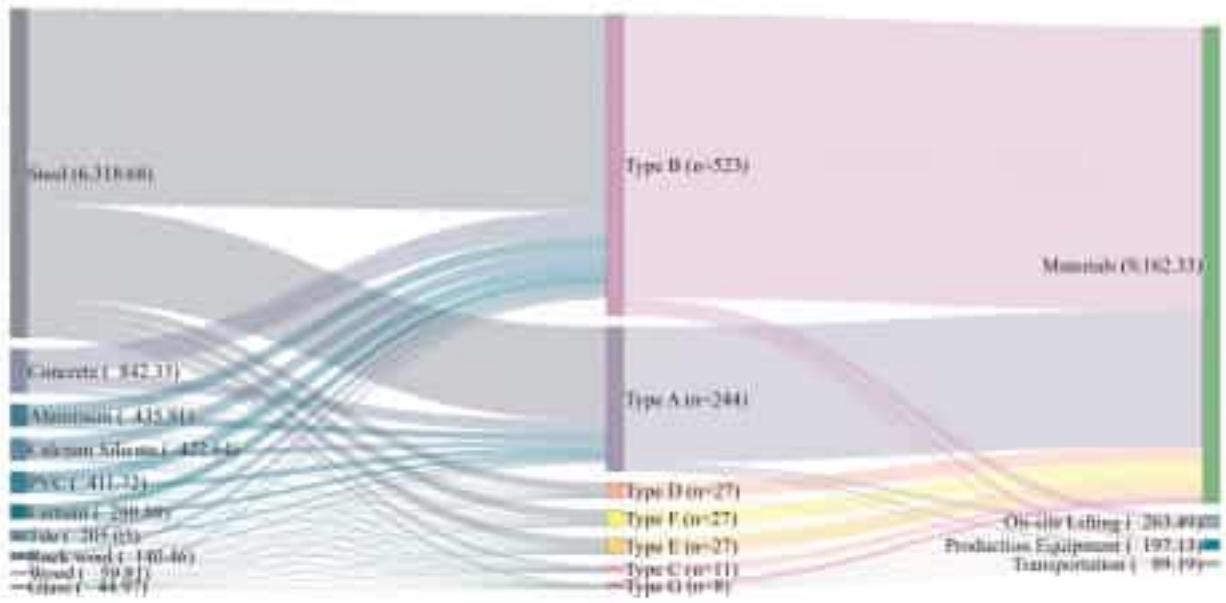


Figure 2. Carbon flow of module-related embodied carbon (tCO_2e)

Table 2. Carbon emission inventory of MiC modules by stage and type (kg CO_2e)

Module Type	Quantity	Production Equipment	Material	Transport	On-site Lifting	Total per Module	Total Emission by Type
Type A	244	230.04	10,744.96	104.17	306.32	11,385.49	2,778,060.18
Type B	523	224.73	10,431.34	101.12	300.62	11,057.81	5,783,235.75
Type C	11	220.51	10,430.61	98.52	300.08	11,049.72	121,546.93
Type D	27	228.89	10,638.78	104.91	309.77	11,282.36	304,623.65
Type E	27	237.47	10,959.92	110.92	318.42	11,626.72	313,921.57
Type F	27	246.20	11,250.80	117.02	327.64	11,941.67	322,425.02
Type G	8	224.73	10,412.25	101.37	302.11	11,040.46	88,323.72

analysis in Case I. Given the standardised layout and function of dormitory rooms, room-level energy use is primarily electricity-based. Multi-source IoT data, including HVAC energy consumption, plug loads, and lighting electricity use, were collected at the module level (room level) over a one-month period. These data provide time-resolved carbon emission profiles across different end-use categories and are compared with EnergyPlus simulation results.

Figure 3 presents the comparison between IoT-based monitoring and simulation at both total and end-use levels, based on a representative room selected from the monitored dataset. At the aggregate level, monitored carbon emissions exhibit pronounced temporal variability and noticeable deviations from simulation results. At the end-use level, lighting emissions show intermittent peaks associated with occupant-driven usage, plug-load emissions display irregular fluctuations not captured by simulation, and HVAC-related emissions deviate from simulated trends due to differences between assumed control conditions and actual indoor operation.

Overall, while simulation captures general trends, it fails to represent short-term variability and peak emissions observed in monitoring data. These results highlight the strong dependence of operational carbon emissions on occupant behaviour and demonstrate the advantage of the BIM–IoT-based approach in capturing dynamic, time-resolved carbon characteristics in modular residential buildings.

A combined view of Case I and Case II shows that the proposed BIM–IoT-based approach enables detailed module-level carbon analysis across both construction (A1–A5) and operational (B6) stages. Together, these results further indicate the broad applicability and versatility of the proposed framework across different lifecycle stages.

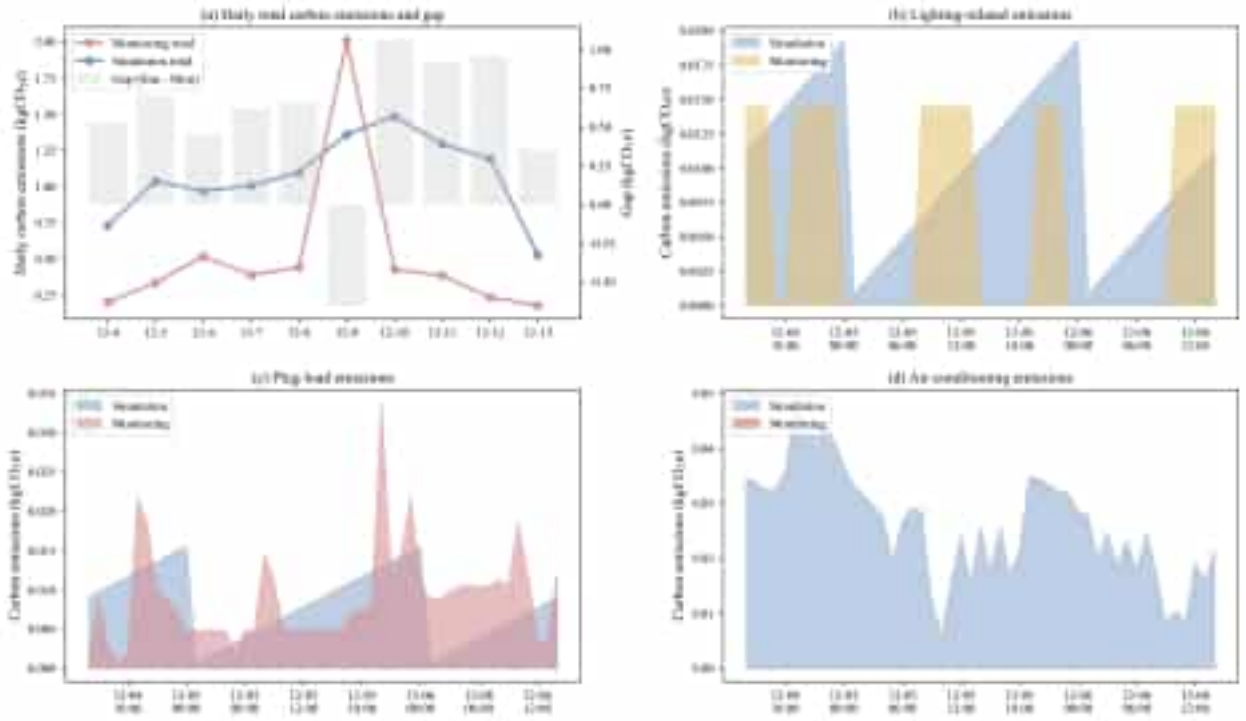


Figure 3. Operational carbon tracking using BIM–IoT integration at total and end-use levels.

5 Conclusion and discussion

Accurate and continuous lifecycle carbon monitoring in MC remains challenging, primarily due to fragmented production processes, heterogeneous data sources, and the lack of reliable mechanisms. To address these challenges, this study proposes a BIM–IoT-integrated approach in which modules serve as the fundamental analytical unit for carbon accounting, enabling consistent lifecycle carbon tracking at the module scale.

The results show that the approach supports consistent module-level representation of carbon emissions across both construction (A1–A5) and operational (B6) stages. By integrating BIM-based structural information with time-resolved IoT data, it enables attributable and traceable carbon analysis, providing clearer insights into carbon emission dynamics across activities and processes.

The findings also reveal key differences between simulation-based and measurement-based estimation of operational carbon. While simulation captures general trends, it overlooks short-term variability and peak emissions driven by occupant behaviour and operating conditions, potentially introducing systematic bias at finer temporal and spatial scales, particularly in highly standardised modular buildings.

Despite these contributions, several limitations should be acknowledged. The lifecycle scope is confined to stages A1–A5 and B6, excluding other stages, such as maintenance, and end-of-life processes, due to data constraints. Emissions from cast-in-situ components are also excluded as they cannot be reliably attributed at the module level, limiting completeness. In addition, embodied and operational carbon are demonstrated using separate projects rather than a single integrated case, due to the difficulty of obtaining continuous lifecycle data. Data uncertainty is also insufficiently addressed, including sensor noise, missing data, and their propagation in carbon accounting.

Future work will focus on implementing the framework within a single project to enable continuous data collection across construction and operational phases, extending coverage to additional lifecycle stages (e.g., maintenance and end-of-life) as data become available, and incorporating uncertainty-aware data processing and validation methods. These efforts are expected to enhance the robustness and practical applicability of data-driven lifecycle carbon management in MC.

Acknowledgements

This work is supported by grants from the Environment and Conservation Fund (ECF 146/2024) and the Guangdong Basic and Applied Basic Research Foundation (2026A1515011749).

References

[1] L. Chen, L. Huang, J. Hua, Z. Chen, L. Wei, A. Osman, S. Fawzy, D. W. Rooney, L. Dong, and P. Yap. Green construction for low-carbon cities: a review. *Environmental Chemistry Letters*, 21:1627–1657, 2023. doi:10.1007/s10311-022-01544-4.

[2] W. Pan, Y. Yang, and M. Pan. Implementing modular integrated construction in high-rise high-density cities: perspectives in hong kong. *Building Research & Information*, 51:354–368, 2022. doi:10.1080/09613218.2022.2113024.

[3] Z. Zhang, Z. Wei, S. Court, L. Yang, S. Wang, A. Thirunavukarasu, and Y. Zhao. A review of digital twin technologies for enhanced sustainability in the construction industry. *Buildings*, 2024. doi:10.3390/buildings14041113.

[4] D. F. R. Parracho, M. N. El-Din, I. Esmaeili, V. P. D. Freitas, L. Rodrigues, J. P. Martins, H. Corvacho, J. M. P. Q. Delgado, and A. S. Guimarães. Modular construction in the digital age: a systematic review on smart and sustainable innovations. *Buildings*, 15: 765, 2025. doi:10.3390/buildings15050765.

[5] S. Chen, Y. Zhang, and W. Pan. Systematic initial embodied carbon assessment of concrete modular high-rise residential buildings: a case in hong kong. *Building and Environment*, page 111917, 2024. doi:10.1016/j.buildenv.2024.111917.

[6] Z. Chen, L. Chen, X. Zhou, L. Huang, M. Sandanayake, and P. Yap. Recent technological advancements in bim and lca integration for sustainable construction: a review. *Sustainability*, 16:1340, 2024. doi:10.3390/su16031340.

[7] N. H. Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri. Internet of things (iot) and the energy sector. *Energies*, 13:494, 2020. doi:10.3390/en13020494.

[8] W. Yaïci, K. Krishnamurthy, E. Entchev, and M. Longo. Recent advances in internet of things (iot) infrastructures for building energy systems: a review. *Sensors*, 21:2152, 2021. doi:10.3390/s21062152.

[9] H. Pittri, S. Aklashie, G. Godawatte, K. Sackey, K. Agyekum, and F. Ghansah. The role of internet of things in managing carbon emissions in the construction industry: a systematic review. *Intelligent Infrastructure and Construction*, 1:8, 2025. doi:10.3390/iic1030008.

[10] X. Huang, Y. Liu, L. Huang, E. Onstein, and C. Merschbrock. Bim and iot data fusion: the data process model perspective. *Automation in Construction*, 152:104792, 2023. doi:10.1016/j.autcon.2023.104792.

[11] B. Loo and R. Wong. Towards a conceptual framework of using technology to support smart construction: the case of modular integrated construction (mic). *Buildings*, 13:372, 2023. doi:10.3390/buildings13020372.

[12] J. Xu, Q. Zhang, Y. Teng, and W. Pan. Integrating iot and bim for tracking and visualising embodied carbon of prefabricated buildings. *Building and Environment*, 236:110492, 2023. doi:10.1016/j.buildenv.2023.110492.

[13] C. Turner, J. Oyekan, and L. K. Stergioulas. Distributed manufacturing: a new digital framework for sustainable modular construction. *Sustainability*, 13, 2021. doi:10.3390/su13031515.

[14] European Committee for Standardization (CEN). EN 15978: Sustainability of construction works – assessment of environmental performance of buildings – calculation method. Standard, CEN, Brussels, Belgium, 2011.

[15] G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, and B. Weidema. The ecoinvent database version 3 (part i): overview and methodology. *The International Journal of Life Cycle Assessment*, 21:1218–1230, 2016. doi:10.1007/s11367-016-1087-8.

[16] PRé Sustainability. SimaPro database manual: Methods library, version 9.1. PRé Sustainability B.V., Amersfoort, The Netherlands, 2020.

[17] G. P. Hammond and C. I. Jones. Embodied carbon: The inventory of carbon and energy (ICE), version 2.0. University of Bath, Bath, UK, 2011.

[18] Construction Industry Council. CIC carbon assessment tool. On-line: `https://www.cic.hk/eng/main/carbon_assessment_tool/`, Accessed: 01/01/2024.

[19] IPCC. Climate change 2013: The physical science basis. contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013.

[20] X. Zhang and F. Wang. Assessment of embodied carbon emissions for building construction in china: comparative case studies using alternative methods. *Energy and Buildings*, 130:330–340, 2016. doi:10.1016/j.enbuild.2016.08.080.

[21] X. J. Li, W. J. Xie, L. Xu, L. L. Li, C. Y. Jim, and T. B. Wei. Holistic life-cycle accounting of carbon emissions of prefabricated buildings using lca and bim. *Energy and Buildings*, 266:112136, 2022.

[22] J. Jayawardana, M. Sandanayake, J. A. S. C. Jayasinghe, A. K. Kulatunga, and G. Zhang. A comparative life cycle assessment of prefabricated and traditional construction – a case of a developing country. *Journal of Building Engineering*, 72:106550, 2023. doi:10.1016/j.jobe.2023.106550.

From On-Site Video to Crew Balance Charts: A Computer Vision Approach for Worker Productivity Analysis

Sai Kumar Titti[1], Aritra Pal[1], Varun Kumar Reja[2] and Koshy Varghese[1]

[1] Department of Civil Engineering, Indian Institute of Technology Madras, India
[2] Department of Civil Engineering, Indian Institute of Technology Bombay, India
saikumartitti@gmail.com, aritrapal@civil.iitm.ac.in, varunreja@iitb.ac.in, koshy@iitm.ac.in

Abstract –

Construction productivity has remained stagnant for decades, underscoring the need for productivity improvement methods. Crew Balance Charts (CBCs) are an effective tool for analysing crew dynamics and identifying opportunities for productivity improvement. However, manual preparation of CBCs is time-consuming and limits scalability. This paper proposes a Computer Vision-based Crew Balance Chart (CV-CBC) framework for automated CBC generation using construction site videos. The framework comprises: Activity Definition, Data Preparation, Activity recognition model development, Inference pipeline, and CBC generation and Validation. The activity-level outputs are further aggregated into Crew-Based Work Sampling (CBWS) metrics for high-level labour monitoring. The framework is implemented and evaluated on masonry activity, with validation performed through per-second comparison between the automatically generated CBC and the manually prepared CBC. Results show that the activity recognition model achieves a macro-F1 score of 0.54 on the validation set across four activity classes, while the CBC achieves an accuracy of 60%. Activity predictions are aggregated into Value-Adding (VA), Non-Value-Adding (NVA), and Non-Value-Adding, but Necessary (NVAN) categories, enabling the calculation of CBWS metrics. The study introduces automated CBC generation and highlights its potential for scalable crew-level analysis in construction, while performance improvements remain an area for future work.

Keywords –

Computer Vision; Crew Balance Charts; Construction Productivity; Activity Recognition

1 Introduction

The construction industry is one of the largest sectors of the global economy, accounting for nearly 7% of global gross output, and is projected to reach USD 22 trillion by 2040 [1]. Despite this scale and anticipated growth, construction productivity has remained largely stagnant for decades. Between 2000 and 2022, global construction productivity increased by approximately 10%, whereas manufacturing productivity rose by roughly 90%, and the overall economy by 50% [1]. This highlights the need to improve productivity to meet future demands amid persistent labour shortages.

Labour accounts for approximately 30–50% of project costs, yet is often utilised at only 40–60% of its potential due to coordination inefficiencies [3]. Improving labour productivity is challenging due to the nature of construction work. Construction projects are temporary and executed in dynamic environments. Consequently, performance depends on interdependent and sequential activities, rather than isolated tasks [2][3]. At the operational level, work is predominantly executed by crews whose productivity is largely dependent on coordination, task sequencing and interaction among the crew members [3]. These underscore the need for crew-level analysis to quantify productivity losses and identify opportunities for improvement.

Lean Construction emphasises maximising value by eliminating waste, which requires systematic waste identification [4][5]. Crew Balance Charts (CBCs) and Crew-Based Work Sampling (CBWS) are established lean tools for analysing crew-level productivity. CBCs capture the temporal distribution of work among the crew across work cycles, highlighting crew interdependencies, exposing forced idleness, and work imbalance. By making such inefficiencies visible, CBCs facilitate the identification of opportunities for productivity improvement [6][7]. In contrast, CBWS quantifies waste by aggregating each worker's activities into Value-Adding (VA), Non-Value-Adding (NVA), and Non-Value-Adding but Necessary (NVAN) categories and can be applied beyond cyclic activities. While CBCs enable productivity improvements, CBWS enables effective labour monitoring.

Traditionally, CBCs are prepared through manual observation of crews across work cycles, requiring continuous observation of crew members across multiple

cycles. Task durations are recorded manually using stopwatches or time-lapse video and then translated into a graphical representation. The process is illustrated in Figure 1. While effective for identifying inefficiencies and supporting improvement decisions, this manual process is labour-intensive and limits the number of cycles that can be analysed [8]. To enable scalability and repeatability, this manual process needs to be automated. Similarly, CBWS requires repeated manual observations, making it time-consuming and challenging to scale.

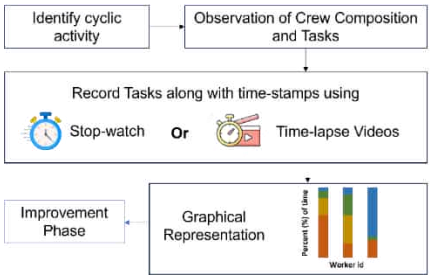


Figure 1 Traditional Crew Balance Chart Process

Prior studies have explored automating activity recognition using wearable sensors such as accelerometers [9]. While these capture detailed motion information, the hardware required for large-scale deployment is intensive and can be uncomfortable for workers. In contrast, recent advances in Computer Vision (CV) enable automated interpretation of construction site videos in a non-intrusive way by detecting, tracking and recognising worker activities [10]. This makes computer vision suitable for capturing crew interactions in labour-intensive construction activities.

Despite the potential of CV for automating construction site observation, existing applications remain fragmented and largely isolated. These approaches rarely translate the outputs into structured analytical tools such as CBC. Consequently, there is no integrated framework that uses computer vision to support crew-level analysis.

Accordingly, this study focuses primarily on CBC as the analytical tool, with CBWS concepts used only to aggregate recognised activities for high-level labour monitoring. Therefore, the objectives of this paper are to:

1. Investigate the feasibility of using CV to automate the CBC generation from construction site videos.
2. Develop an integrated framework, referred to as CV-CBC, that integrates worker detection, worker tracking, and activity recognition to convert video-based observations into CBC.
3. Evaluate the framework on real-world data and identify key technical limitations.

The remainder of this paper is organised as follows. Section 2 reviews related work on computer vision applications in construction productivity. Section 3 presents the proposed CV-CBC framework and its components. Section 4 describes the implementation and validation of CV-CBC framework using masonry activity. Section 5 presents the key results, Section 6 discusses ethical considerations, and Section 7 concludes by outlining future research directions.

2 Related Work

Computer vision applications in construction have largely focused on safety monitoring, while studies on productivity analysis remain limited [11]. The papers that address productivity are also focused on equipment rather than workers. Workers pose various challenges, as their movements are less well-defined and structured than those of equipment operations. Nonetheless, researchers have attempted to model the workers' actions, and a summary of these studies is presented below.

2.1 Equipment-Centric Activity Analysis

Several studies have focused on recognising equipment activities due to their structured, predictable motion patterns. Roberts and Golparvar-Fard proposed an end-to-end pipeline for detecting, tracking, and recognising excavator activities. However, the authors noted that extending such approaches to workers is challenging due to the complexity of human motion [12].

The multi-stage process can lead to error accumulation across modules, which affects the overall performance of the system. Ghelmani et al. developed a single-stage spatio-temporal activity recognition framework for excavator actions. This improved the performance compared to a multi-stage pipeline, highlighting the advantages of single-stage architectures [13]. Nevertheless, the study is limited to excavators and does not consider workers' activity recognition.

2.2 Worker-Centric Activity Analysis

Worker-centric analysis is slightly more challenging than equipment-centric analysis. Toledo et al. used YOLOv5 to detect worker postures and classify them into productive and contributory works, enabling the generation of CBC [6]. While promising, the approach was limited to single-worker scenarios and did not incorporate tracking or temporal modelling. Roberts et al. developed a 2D pose-estimation-based approach for CBC automation that achieved good accuracy but was limited to a single-worker scenario. As a result, the approach does not capture interaction among multiple workers, limiting its applicability for crew-level analysis [14].

Kim et al. proposed a generalised vision-based framework for construction productivity analysis, aiming to improve scalability. The framework addresses both

equipment and labour productivity analysis. However, it relies heavily on the synthetic and web-crawled images, rather than video data. In addition, limited emphasis is placed on fine-grained worker activity recognition, which is essential to developing CBC [15].

In summary, existing studies have explored vision-based productivity analysis. Despite these advances, current methods have largely focused on equipment, are limited to single-worker analysis and mostly rely on static images rather than continuous video data. To address these limitations, an integrated CV-CBC framework is developed to enable CBC generation.

3 Computer Vision-based Crew Balance Chart (CV-CBC) Framework

Based on previous studies, automated generation of CBCs requires three major CV techniques: object detection, tracking and activity recognition. Object detection enables the localisation of workers, object tracking maintains worker identity over time, and activity recognition classifies the activities performed by the worker [12] [10]. These techniques form the CV pipeline for generating CBC from site videos. The proposed CV-CBC framework comprises five components: Activity Definition, Data Preparation, Activity Recognition model development, Inference Pipeline, and CBC Generation and Validation. The detailed framework is illustrated in Figure 2 and each component is discussed in the following subsections.

3.1 Activity Definition

The first step is to identify the main activity for CBC analysis. CBC analysis is effective when applied to cyclic, labour-intensive, repetitive, and coordination-intensive work. Small improvements in cyclic activities can yield significant productivity gains over time. The selected activity should also be critical to the project in terms of time and cost.

In this study, the term *"main activity"* refers to the overall construction operation selected for CBC analysis. In contrast, *"activity"* refers to the specific actions carried out by workers within the main activity. Once the main activity is selected, it is further broken down into a set of activities that represent the work performed by the individual workers. This breakdown is carried out by observing on-site operations and understanding the crew composition. The objective is to identify all activities that may occur during one complete work cycle. Activities are classified into Value-Adding (VA), Non-Value-Adding (NVA), and Non-Value-Adding but Necessary (NVAN) categories to enable CBWS. To reduce ambiguity, activities are defined to be distinguishable based on visual cues, with visually similar activities grouped into broader categories.

3.2 Data Preparation

This stage converts the site videos into a structured dataset that enables model training in the subsequent stage. It includes Data Collection and Annotation, Clips Generation, Data Augmentation and Dataset Split.

Data Collection: Video data are collected from construction sites using standard camera-based methods. The primary requirement is that workers are clearly visible for a sufficient time to capture their activities. To improve robustness, videos are collected across different work cycles, camera viewpoints and site conditions.

Data Annotation: The videos are manually annotated using standard tools. For each visible worker, a bounding box (bbox) is drawn, and an activity label is assigned based on a pre-defined activity breakdown. A new bbox will be drawn each time the worker changes the activity. The bounding boxes are updated over time in response to the worker's movement. During annotation, the worker is preferably kept near the centre of the bbox, with a small margin, to capture more spatial context. The annotations are exported in a structured format, such as CVAT 1.1, which enables clip generation in the subsequent step.

Clips Generation: Using the annotations, videos are processed to extract worker-level tracks, which are then trimmed into shorter clips for training and validation.

Data Augmentation: It is used to improve the diversity of the dataset and the activity recognition model robustness. Both spatial (horizontal flipping) and photometric (brightness, contrast) augmentation techniques are applied. Data augmentation also helps eliminate the imbalance among activity classes.

Dataset Split: The dataset is split into training, validation, and test sets using a video-level splitting strategy to prevent data leakage and ensure fair evaluation. All clips that belong to the same video are assigned to a single subset.

3.3 Activity Recognition Model Development

This stage involves developing an activity recognition model to classify video clips into pre-defined activity classes, which requires learning both spatial and temporal information. Model development includes model selection, training and evaluation. Model hyperparameters are tuned using validation data to improve performance. Model selection is based on macro F1-score, where F1-score is defined as the harmonic mean of precision and recall, and the macro F1-score is computed by averaging the F1-scores across all classes giving equal importance to each class. This metric provides a balanced assessment of model performance. The output of this stage is a trained activity recognition model that is used in the subsequent inference pipeline.

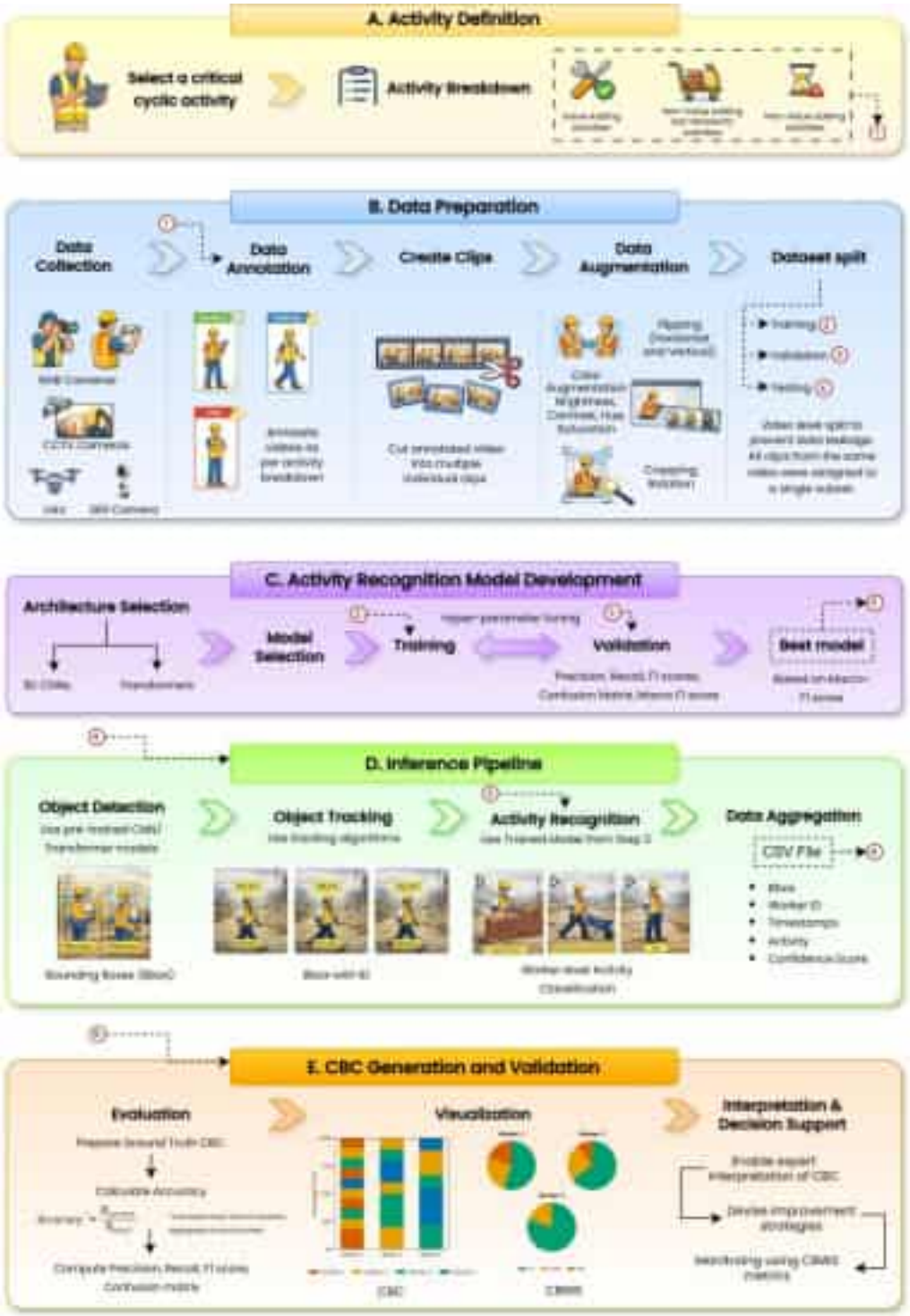


Figure 2 Computer Vision-based Crew Balance Chart Framework

3.4 Inference Pipeline

This stage represents the operational stage of the proposed CV-CBC framework. During inference, workers are detected in each frame, tracked across time with unique IDs, and converted into worker-level clips. The trained activity recognition model is used to predict the activity classes. These predictions are stored in a Comma-Separated Values (CSV) format along with temporal information. This serves as an input for CBC generation in the next stage.

3.5 CBC Generation and Validation

The final stage of the framework converts worker-level activity predictions from the inference pipeline into a CBC and evaluates its accuracy. Predicted activities are temporally organised into continuous timelines for each worker and aligned along a common time axis to generate the CBC. The accuracy is computed by comparing predicted activities with manually annotated ground truth labels at a per-second temporal resolution, aggregated across all workers, as shown in Equation (1)

$$Accuracy = \frac{N_{Correct}}{N_{Total}} \quad (1)$$

where N_{correct} and N_{total} denote the total number of correctly predicted and total activity predictions, respectively, aggregated across all workers.

Performance is evaluated using standard metrics, including accuracy, F1-score, precision, recall, and confusion matrices. Similarly, CBWS metrics are calculated by aggregating recognised activities into the Value-Adding (VA), Non-Value-Adding (NVA), and Non-Value-Adding but Necessary (NVAN) categories, enabling high-level crew monitoring.

4 Implementing CV-CBC Framework

4.1 Activity Definition

This section presents the implementation of the proposed CV-CBC framework using masonry as the main activity. Masonry was selected because it is a labour-intensive, repetitive process that requires close crew coordination, making it prone to work imbalance and enforced idleness. These characteristics make masonry well-suited for CBC analysis. Based on on-site observations of the masonry work, the main activity was broken down into a set of worker-level activities that occur during execution. These activities were defined to represent distinct and meaningful work states observed repeatedly across work cycles. The selected activities and their descriptions are summarised in Table 1.

Table 1 Description of selected activities

Activity	Category	Description
Applying Mortar	VA	Applying mortar between blocks and preparing bed
Idle	NVA	No productive or supporting work
Laying Blocks	VA	Placing and positioning masonry blocks
Shifting Materials	NVAN	Transporting/ Repositioning materials or tools

4.2 Dataset Preparation

Video data were collected from multiple construction sites under normal working conditions using a static camera. Most of the videos were not recorded specifically for computer vision analysis, but rather captured for manual observation. A total of seven videos were collected, with crew sizes ranging between three and four workers. The videos were annotated using the Computer Vision Annotation Tool (CVAT), and the XML files were exported in CVAT 1.1 format, which enabled clip generation. Worker-level clips were generated, each with a fixed duration of 8 seconds. Data Augmentation was applied using horizontal mirroring, brightness, and contrast adjustments. The details of the clips, class-wise, are presented in Table 2.

Table 2 Details of Clips

Class	Clips (No.)	Duration (min.)
Applying Mortar	305	42.62
Idle	357	47.75
Laying Blocks	263	35.15
Shifting Materials	318	40.81

The dataset was split using the following strategy: one video was reserved for validation, one for inference, and the remaining for training.

4.3 Model Training

A transformer-based architecture was selected for activity recognition, as transformer models are well-suited for learning spatio-temporal information from videos. In this experiment, the VideoMAE-Small model pre-trained on the Kinetics-400 dataset was used.

The small variant was chosen to mitigate overfitting given the limited dataset size. Each 8-second clip was uniformly sampled and resized to 224 × 224 pixels to match the pre-training input configuration. The proposed model was then fine-tuned on the annotated videos using supervised learning. A partial fine-tuning strategy was employed by unfreezing 6 of the 12 backbone layers, enabling greater adaptation to construction while preserving common spatial and temporal characteristics learned during pre-training. The trained model achieved a macro F1-score of 0.54, indicating balanced recognition performance across activity classes. During training, class-wise F1-scores were 0.44 for Applying Mortar, 0.56 for Idle, 0.49 for Laying Blocks, and 0.66 for Shifting Materials.

To validate the model choice, a 3D CNN-based I3D model was also evaluated, achieving a macro F1-score of 0.43, with class-wise scores ranging from 0.36 to 0.48. In comparison, VideoMAE achieved a higher macro F1-score of 0.54 and was therefore selected for subsequent experiments.

4.4 Inference Pipeline

During inference, the test set was used to generate the CBC using an end-to-end pipeline. The pipeline comprises object detection, object tracking, activity recognition and post-processing. The test video is 164 seconds long; a segment was trimmed to include only activity classes on which the model is trained, enabling fair evaluation. The crew size is three, and the work cycle is defined as the construction of one masonry course.

4.4.1 Object Detection

The YOLO11x object detection model, pre-trained on the COCO dataset, was used to identify workers in each video frame. In cluttered construction site environments, detection sensitivity and localisation accuracy were balanced using an Intersection-over-Union (IoU) threshold of 0.5 and a confidence threshold of 0.3. Frame-wise bounding boxes generated by the detector were used as inputs for tracking. High detection performance is achieved with a precision of 98.8% and a recall of 98.6%, indicating reliable worker localization.

4.4.2 Object Tracking

The BoT-SORT multi-object tracking algorithm with re-identification (ReID) enabled was used to track detected workers across frames. This was made to assign distinct worker IDs consistently. Only tracked IDs with a minimum continuous duration of 30 frames were retained for further analysis to eliminate workers not relevant to the work cycle. A total of two ID switches were observed across all workers, with Multi-Object Tracking Accuracy of (MOTA) of 97.4%. For better visualisation, tracking IDs were grouped for subsequent analysis.

4.4.3 Activity Recognition

For each tracked worker, bounding box–based crops were used to generate worker-centric clips. Activity recognition was performed using the trained VideoMAE model with an input resolution of 224 × 224 pixels. A 2-second sliding window, consisting of 16 frames, was applied with a stride of 1 second, enabling per-second activity predictions. For overlapping windows, the activity with the highest confidence was selected.

4.4.4 Post-processing

Before CBC generation, worker-level activity timelines were stabilised using post-processing rules. Idle predictions were filtered using a temporal persistence rule, assigning the idle label only if it persisted for more than two seconds and exceeded a confidence threshold of 0.75. This reduces false idle predictions caused by short pauses or low-motion intervals that occur during continuous productive work (such as laying blocks). The worker-level activity timelines were then combined to generate a CBC.

4.5 CBC Generation and Validation

Activity recognition performance was assessed using class-wise F1-score. The model shows limited generalisation compared to training performance. Laying Blocks achieved the highest F1-score (0.62), followed by Shifting Materials (0.38), while Applying Mortar and Idle recorded lower F1-scores of 0.18 and 0.09, respectively. The confusion matrix in Figure 3 shows that Applying Mortar and Laying Blocks are the most frequently misclassified activities, with many Applying Mortar instances incorrectly predicted as Laying Blocks. This confusion arises because both activities exhibit similar motion cues, such as bending and hand actions with a trowel or masonry blocks, making it difficult to distinguish between them.

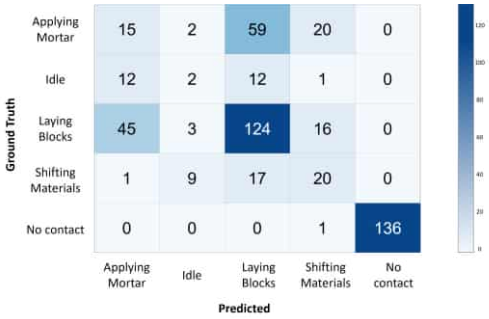


Figure 3 Confusion Matrix

The automated CBC is produced by aggregating worker timelines, whereas the manual CBC was prepared using the traditional process. To ensure ground truth reliability, inter-annotator agreement (IAA) was evaluated on the test set with two annotators and Cohen's κ, achieving 93.30% agreement ($\kappa \approx 0.90$), indicating almost perfect agreement. When the worker is out of the frame, it is termed "No Contact". Figure 4 illustrates the manually created Crew Balance Chart with automated CBCs generated by the CV-CBC framework.

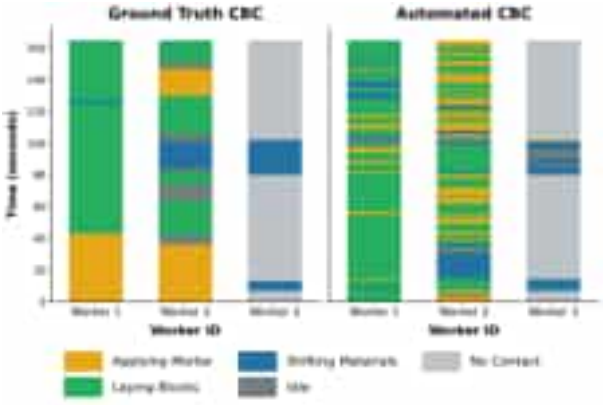


Figure 4 Ground truth CBC and Automated CBC

CBC accuracy was computed by comparing predicted and ground-truth activity labels at a per-second, per-worker resolution (Equation 1). The automated CBC achieved an accuracy of 60.0 % (297 accurate predictions out of 495 total) including No-Contact periods and 45% excluding them. This metric reflects the framework's ability to reconstruct the temporal structure of activities.

CBWS metrics were derived by aggregating activity predictions into VA, NVAN and NVA categories. For each worker, VA, NVAN and NVA time proportions were computed from ground truth and predicted activities and compared using absolute percentage deviation, as reported in Table 3. Workers 1 and 2 primarily engage in value-adding tasks, whereas Worker 3 is involved only in NVA and NVAN. This distribution is consistent with typical mason–helper role patterns and demonstrates the potential for identifying worker-level contributions to support high-level labour monitoring.

Table 3 CBWS comparison between ground truth and predictions

Worker	Category	Ground Truth (%)	Predicted (%)	Absolute Error (%)
1	VA	98.79	90.91	7.88
	NVAN	1.21	7.27	6.06
	NVA	0.00	1.82	1.82
2	VA	73.33	81.21	7.88
	NVAN	10.30	15.76	5.46
	NVA	16.36	3.03	13.33
3	VA	0.00	3.45	3.45
	NVAN	100.00	68.97	31.03
	NVA	0.00	27.59	27.59

5 Discussion

This study demonstrates that the proposed CV-CBC framework can automatically extract activity-level information to generate CBC. The activity recognition model achieved a validation macro F1 score of 0.54, while the automated CBC achieved an accuracy of 60%, comparable to prior work. Toledo et al. reported a 58% CBC accuracy under site conditions using a fixed camera, with contributory work not detected (0%) due to occlusions and viewpoint limitations [6]. The literature does not define strict deployment thresholds; however, prior studies suggest that accuracy levels around 78–85% are sufficient for practical use [14][15]. Accordingly, the results should be interpreted as a proof of concept.

The reported CBC accuracy of 60% includes no-contact periods, which inflates performance as it is easier to detect the absence of a worker; excluding them reduces accuracy to 45%, which better reflects the activity recognition performance. Given the high performance of detection and tracking stages, the overall pipeline performance is primarily constrained by activity recognition, which limits the CBC accuracy. Most errors arise from activity misclassifications, while detection and tracking contribute comparatively less. Although only two ID switches were observed, identity consistency remains critical for CBC generation, as even minor track fragmentation can affect worker-level timelines, suggesting the potential for post-processing strategies to maintain consistent identities.

These results should be interpreted in the context of the dataset. The dataset is small, and the framework is validated only on masonry activity, limiting generalizability. Extending to other activities or larger datasets may introduce class imbalance (e.g., overrepresented idle class), requiring resampling or augmentation. Due to the data constraints, the study was not extended to other activities.

At the activity level, significant confusion was observed between laying blocks and applying mortar because of their similar motion patterns. Table 3 shows that the mean absolute error for VA activities is within 8%, indicating higher reliability in detecting productive work. In contrast, NVA and NVAN activities exhibit higher error rates. The distinction between these two categories is subjective, as even human observers may classify the same activity differently during manual CBC generation. Higher error rates observed for Worker 3 can be attributed to the worker's limited presence in the video, which amplifies relative errors.

The generated CBCs reveal enforced idleness and work imbalance, supporting productivity improvement strategies during meetings. Extending activity level outputs to CBWS further enables high-level monitoring of worker contributions within a crew. Such monitoring enables supervisors to identify any sustained deviations from the expected value-adding work, allowing earlier intervention compared to post-hoc productivity reporting. Overall, the framework enables automation of Lean tools such as CBC and CBWS, supporting scalable crew-level productivity monitoring and continuous improvement.

6 Ethical Considerations

The video data used in this study were collected for manual observation of construction activities. To preserve worker anonymity, no images or visual data are presented in the manuscript. The analysis is limited to activity-level information, and no individual worker identities are used or reported.

The framework is aligned with Lean Construction Principles, where the activity classifications are intended for waste identification and process improvement. Periods labeled as "Idle" or "Non-Value-Adding" should be interpreted as indicators of workflow constraints (e.g., work front unavailability) and not as worker inefficiency.

A tolerance for short durations of idle activity should be accounted for allowing the natural rhythm of the worker. The framework is not intended for individual-level performance monitoring or punitive use, but for identifying recurring process-level inefficiencies, and outputs should be interpreted at the crew level.

7 Conclusions

This paper proposes a Computer Vision–based Crew Balance Chart (CV-CBC) framework for automatically generating CBCs from construction site videos by integrating worker detection, tracking, and activity recognition. The framework was implemented on the masonry activity and validated through per-second comparison with manually prepared CBCs. The results show that the activity recognition model achieves a macro F1-score of 0.54, while the automated CBC attains an accuracy of 60%. Analysis indicates that the primary source of error arises from confusion between temporally similar activities, Laying Blocks and Applying Mortar, which share similar motion cues. Despite these results, the study has several limitations. The dataset is relatively small and was not originally captured for computer vision–based analysis. Furthermore, the evaluation is restricted to a single construction activity, which limits the extent to which the results can be generalised. Moreover, the inclusion of No-Contact periods can inflate overall CBC accuracy, highlighting the need for more representative CBC-level evaluation metrics.

Future work will focus on implementing the framework on larger, more diverse datasets and on multiple construction activities. Incorporating vision–language models to capture richer spatial context, such as identifying the tools and materials workers handle, could enable zero-shot activity recognition. Given their strong ability to learn temporal information, video transformers can be ensembled with vision-language or keypoint-based models to improve robustness and better distinguish visually similar activities. Overall, the proposed framework highlights how computer vision can support crew-level productivity monitoring and productivity improvement.

References

[1] J. Mischke, K. Stokvis, K. Vermeltfoort, and B. Biemans. Delivering on construction productivity is no longer optional. On-line: https://www.mckinsey.com/capabilities/operations/our-insights, Accessed: 29/03/2026.

[2] P. E. Eriksson and H. Szentes. Managing the tensions between exploration and exploitation in large construction projects. *Construction Innovation*, 17(4):492–510, 2017.

[3] S. Loganathan and P. Forsythe. Unravelling the influence of teamwork on trade crew productivity: a review and a proposed framework. *Construction Management and Economics*, 38(11–12), 2020.

[4] A. S. da Barbosa and D. B. Costa. Productivity monitoring of construction activities using digital technologies: a literature review. In *Proceedings of the International Group for Lean Construction, pages 707–716, 2021.*

[5] W. Shou, J. Wang, P. Wu, and X. Wang. Value adding and non-value adding activities in turnaround maintenance process: classification, validation, and benefits. *Production Planning and Control*, 31(1):60–77, 2020.

[6] M. J. Toledo, M. Lorca, and M. Mora. Use of computer vision and object detection to measure construction productivity using crew balance charts. In *Proceedings of the International Group for Lean Construction,* pages 810–821, 2025.

[7] J. A. Kuprenas and A. S. Fakhouri. A crew balance case study—improving construction productivity. *CM eJournal*, 1(1):10–27, 2001.

[8] C. H. Oglesby, H. W. Parker, and G. A. Howell. *Productivity Improvement in Construction*. 1989.

[9] L. Joshua and K. Varghese. Accelerometer-based activity recognition in construction. *Journal of Computing in Civil Engineering*, 25(5):370–379, 2011.

[10] B. Sherafat et al. Automated methods for activity recognition of construction workers and equipment: state-of-the-art review. *J Constr Eng Manag*, 146(6):03120002, 2020.

[11] J. Li *et al.* A review of computer vision-based monitoring approaches for construction workers' work-related behaviors. *IEEE Access*, 12:7134–7155, 2024.

[12] D. Roberts and M. Golparvar-Fard. End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level. *Autom Constr*, 105:102811, 2019.

[13] A. Ghelmani, G. Torabi, A. Hammad, and C. Chen. Single-stage spatiotemporal activity recognition of excavators: a case study. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 800-807, 2024.

[14] D. Roberts, W. Torres Calderon, S. Tang, and M. Golparvar-Fard. Vision-based construction worker activity analysis informed by body posture. *Journal of Computing in Civil Engineering*, 34(4):04020017, 2020.

[15] J. Kim, J. Hwang, I. Jeong, S. Chi, J. O. Seo, and J. Kim. Generalized vision-based framework for construction productivity analysis using a standard classification system. *Autom Constr*, 165:105504, 2024.

BIM-LPS® Integration Case Study in the MEP Sector: Automating Phase Schedule Generation

Maroua Sbiti[1,2], Djaoued Beladjine [1], Abinandana Boodi [1], Lionel Chevalier [2], Karim Beddiar [1]

[1]CESI LINEACT Research unit, France
[2]EQUANS Organization, France

maroua.sbiti@viacesi.fr, dbeladjine@cesi.fr aboodi@cesi.fr lionel.chevalier@equans.com kbeddiar@cesi.fr

Abstract -

The Last Planner System® (LPS®) and Building Information Modeling (BIM) hold strong potential to improve construction productivity, yet their integration at the data-processing level remains limited. Information embedded in BIM models is often insufficiently structured and therefore not directly exploitable for LPS® implementation. This study proposes an approach that integrates LPS® with BIM while leveraging multiple sources of operational project data to automate phase planning. The work focuses on the mechanical, electrical, and plumbing (MEP) domain, which is particularly complex and has been less explored in prior research.

A digital tool was developed to automatically generate time–location schedules by combining tender-phase estimates, execution-phase BIM models, client-defined milestones, and expert knowledge of MEP work sequencing. System requirements were defined through a literature review and surveys with MEP professionals. Installation operations were standardized, and the solution was implemented as a three-tier web application supporting automated data extraction, processing, and visualization for project planning, using tools such as Dynamo Revit, MySQL, Java, and Angular.

Validation on a real project demonstrated clear benefits: the time required for initial planning is reduced by up to one week compared with traditional methods. In addition, automated schedule generation and updates yield a financial gain of 0.91% of project revenue - equivalent to nearly 30% of the planned profit margin - highlighting both the practical value and the economic relevance of the proposed approach.

Keywords -

LPS®, BIM, Lean construction, MEP, Construction Management

1 Introduction

Detailed production planning and control has been consistently emphasized in the Lean Construction (LC) literature as fundamental to overcoming the shortcomings of high-level project plans developed during the preconstruction stage (master plans). Such plans typically provide sufficient detail and time horizons for owner-contractor coordination, but are inadequate for guiding day-to-day construction activities and ensuring reliable project delivery [1, 2]. To address this limitation, LC advocates progressively translating project objectives into executable tasks closer to the point of delivery. The Last Planner System® (LPS®) operationalizes this principle by introducing phase planning as an intermediate level that bridges long-term strategies and short-term commitments, and by incorporating lookahead planning to identify and remove upcoming constraints. Combined, these components enhance planning reliability, stabilize workflow, and improve the predictability of project delivery [3, 4, 5].

Recent research shows that integrating LC and BIM, particularly through BIM-LPS® co-application, further extends these benefits, as BIM supports the rational use of resources (e.g., materials, crews, and equipment) in alignment with the schedule and facilitates improved information flow [6]. This synergy has motivated the development of so-called Lean IT-tools, which are designed to support one or more aspects of the LPS® while providing links to BIM models. Examples include LEWIS [7], which integrates 2D/3D plans, production data, and 4D visualization; KanBIM [8, 9], which enables short-term planning and monitoring with clear task-maturity visualization; and VisiLean [10], which supports multiple planning levels and resolutions, such as phase, lookahead, and weekly planning. VisiLean also provides a mobile interface allowing workers to access task information, update status, and synchronize progress with the BIM model in real time.

Despite these advances, BIM-based production management systems still lack data-level integration between BIM and LPS® [11, 12]. In practice, BIM information is not systematically structured for direct use in LPS®, and object–task links are often created manually during workshops, making the process slow and inefficient. These links also fail to enrich task attributes such as quantities or costs, leaving much of BIM's potential underused. This aligns with findings by Schimanski et al. [13, 14], who note that seamless BIM-LPS® integration requires advanced capabilities such as automatic task generation, cost estimation, and flexible data exchange. Reflecting this need, recent work shows that automating planning and control within LPS® supports 10 of the 11 Lean Construction principles [15].

Parallel to these observations, automated BIM-based

scheduling has been widely explored in structural engineering [16]. Early studies, such as De Vries and Harink [17], generated preliminary schedules using component relationships from 3D CAD models and duration estimates. Kataoka [18] expanded this direction by producing multiple schedule scenarios from early design models and construction knowledge templates. More recently, Wang and Lin [19] enhanced BIM with work-package information to address resource-constrained scheduling problems. Yet despite these contributions, a core scientific challenge persists: integrating heterogeneous datasets (e.g., estimates, BIM models, contracts, tenders, and task-sequencing knowledge) into a coherent automated planning process.

Field evidence gathered in this research reinforces this gap. While practitioners show strong proficiency in 3D modelling, a clear need remains to advance toward 4D BIM integration [20]. Tender-stage BIM models are often inconsistent (e.g., oversized vertical ducts or duplicated network elements) and frequently lack standardized codification (Omniclass, Uniclass), making them unreliable for workload estimation or schedule development [21, 22, 23]. As a result, professionals manually consolidate fragmented data (quantity take-offs, estimates, contracts) and create schedules in tools like MS Project or Primavera, relying more on personal expertise than on automated, data-driven methods. This process is time-consuming—about 0.91% of project revenue, a significant cost given a 3% profit margin [23]—and highlights the need for automation in MEP planning, where BIM-based scheduling remains underdeveloped [16].

2 Research objectives and methodology

In response to these gaps, this research aims to advance BIM-LPS® data integration by automating the generation of LPS® phase schedules. It focuses on MEP schedule automation, an area less explored than structural works due to its complexity and interdisciplinary nature [16]. The study moves beyond conventional BIM use by integrating multi-source, multi-phase data, combining project estimation information with execution-phase BIM outputs to detect discrepancies between tender assumptions and execution realities. This enhances decision-making and strengthens alignment between tendering and execution teams. Accordingly, the research investigates how integrating such operational data enables automated LPS® phase schedule generation in the complex context of MEP works, thereby laying the foundation for an optimized approach to managing building information, termed Lean BIM Management.

To ensure both practical relevance and theoretical contribution, the study follows a Constructive Research Approach (CRA). CRA develops innovative artifacts that address real-world problems while informing theory. It relies on abductive reasoning and close collaboration with practitioners [24]. Guided by the CRA framework (Figure 1), the research identifies the BIM–LPS® integration challenge through literature review, and an assessment of MEP scheduling practices and need within the industrial partner, Equans (Section 1: Introduction). Empirical insights were gathered through a survey of 100 professionals and semi-structured interviews with five project managers. The proposed automated planning solution was then designed (Section 3), implemented, and validated through a case study and qualitative feedback (Section 4), from which its theoretical contributions were derived (Section 5).

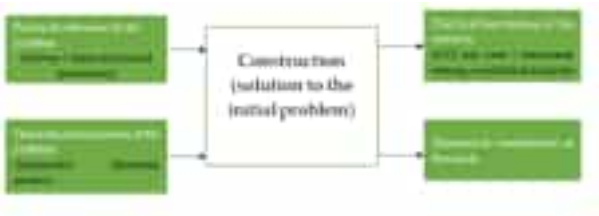


Figure 1. CRA framework and methods used

3 Design and development of the automatic scheduling solution

The system was developed from functional and operational requirements, as detailed in [20, 23], and implemented as a 3-tier web-based scheduling solution. Figure 2 presents the overall framework of the developed artefact. While the framework can guide automated planning in other MEP enterprises, consistent data requirements and algorithms notwithstanding, variations in input sources' structure and availability require careful analysis and adaptation to each organizational context.

3.1 Data sources

The first pillar outlines the primary data sources, including project cost estimates and the execution-phase BIM model, which provide material quantities at the tendering and execution stages, respectively. Cost estimate data are obtained from two sources: quantity takeoff sheets prepared by cost estimators and the final bidding document (client contract) generated using QuickDevis. An initial analysis of both sources reveals key limitations. The QuickDevis file lacks contextual information, such as the location of materials and their associated systems. Instead, it is organized into custom sections based on the client's global price decomposition structure (DPGF), which often results in the grouping of equipment and their associated installation durations without detailed breakdowns. Conversely, the quantity takeoff sheets provide

valuable contextual and technical information, including material quantities, locations (building and floor), MEP system assignments, and technical specifications (e.g., diameter, flow rate, pressure). However, they do not include installation durations. Although incomplete, this structured technical data could nonetheless serve as a basis for deriving installation times by matching each article to the corresponding cost database item, using material classification codes and technical characteristics. In case of ambiguity, item codes were assigned in consultation with superintendents. For example, a 315 mm flexible duct may correspond to GNF 0315 (aluminium), GNF I-315 (insulated aluminium), or GNF G-0315 (semi-rigid, non-insulated), with installation times of 0.2 h/m for the first two and 1 h/m for the third, illustrating the impact of incomplete specifications on planning.

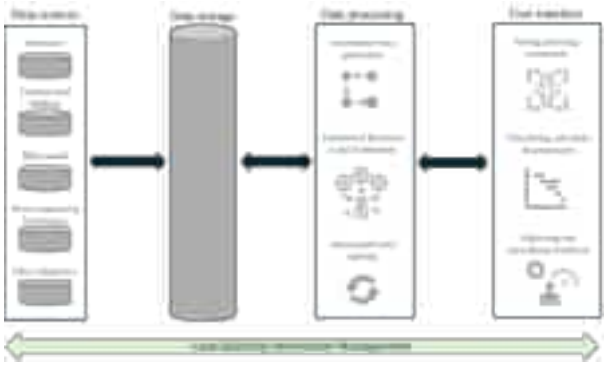


Figure 2. Architecture of the automated planning solution

A similar challenge arose when working with execution-phase BIM models, which often lack sufficient detail, including installation durations and technical specifications needed for direct cost database coding. To address this, a custom Dynamo–Python workflow was developed to link BIM materials to cost database categories, with item codes assigned in consultation with superintendents to resolve inconsistencies and ensure accurate mapping.

Beyond material quantities and durations, the contractual schedule provides the system with project milestones, organized by building level and, where relevant, by specific zones. Finally, the automated scheduling system required formalization of material installation sequences and time allocations, which were captured through expert workshops, converting informal know-how into structured planning knowledge. Previous work by the authors detailed the bill of process for the stainless black steel distribution pipe category (“TAD”) [20].

3.2 Data integration and database architecture

The second pillar consists of a MySQL database, whose conceptual model is illustrated in Figure 3. The database is structured around several entities (yellow blocks) interconnected by associations (blue blocks). For clarity, these entities and associations are grouped into four categories, each corresponding to a distinct data typology and visually delineated by dashed rectangles, within which the previously described data sources are stored. In terms of data integration, BIM quantity take-offs were automatically imported via a Dynamo script, whereas all other data were manually uploaded in Open Document Format.

3.3 Data processing algorithms and workflow

The third pillar is a backend server that represents the central and “intelligent” component of our automatic scheduling system. It handles data processing and employs advanced algorithms to transform scheduling input data into accurate and reliable time-location schedules. Developed using the Java programming language, this layer interfaces with both the database and the user interface, as shown in Figure 4. The data transformation process includes task instantiation by building, level, system type, and superfamily, followed by computation of durations, start, and end dates for the tendering and design phases. Execution dates account for working hours, level availability, and prerequisite tasks. Tasks are first transformed into a graph and topologically sorted using a depth-first search algorithm (DFS) to respect predecessor-successor relationships. The mathematical model described in [20] then computes task durations and dates while estimating required daily manpower.

3.4 User interface

The final layer is the user interface, implemented using the Angular framework, enabling intuitive interaction with the automatic planning tool. Results for the Rennes Surgical and Interventional Center case study are presented in the next section.

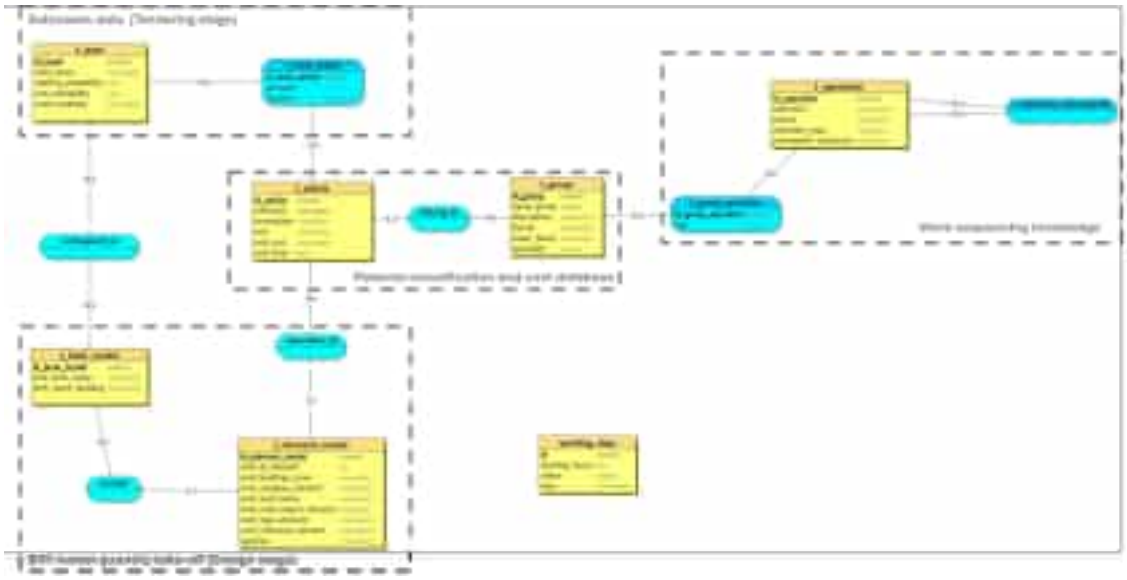


Figure 3. The conceptual model of the scheduling database

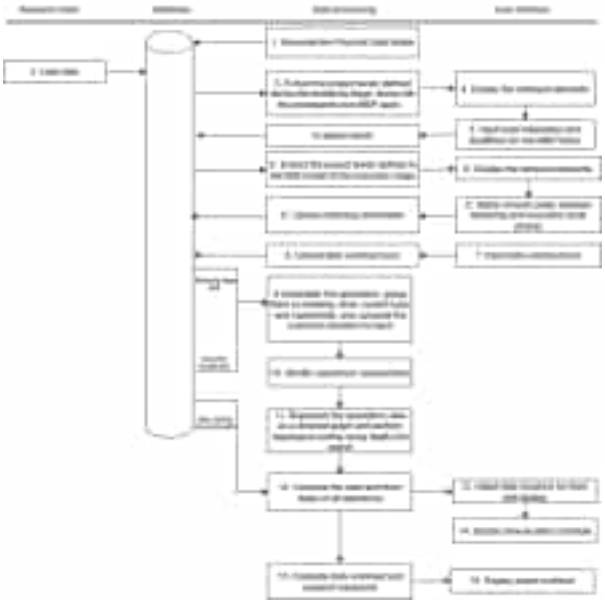


Figure 4. Algorithmic Workflow of the Automated Scheduling System

4 Case study

To assess the practical functionality of the solution in accordance with CRA guidelines, the proposed system was evaluated through a case study with Equans, using operational MEP data of the €20.8 M Surgical and Interventional Center in Rennes. The project involves the construction of a technical platform dedicated to surgery, interventional radiology, and intensive care, consolidating all surgical blocks and interventional radiology rooms of the Rennes University Hospital Center (CHU) in northwestern France. Figure 5 presents a 3D view of the project's BIM model. The evaluation comprised two phases: demonstration of core functionalities and validation of results through expert review.



Figure 5. Capture of the case study project 3D BIM model, featuring the properties of a circular flexible duct instance "GNF-G-0315"

4.1 Demonstration of the solution functionalities

The user interface offers functions for inputting scheduling constraints such as level availability dates, preliminary milestones from other disciplines affecting MEP activities, and working days and hours. Based on these inputs, the system automatically generates a time-location schedule, first using cost estimation data and subsequently updated with execution-phase BIM data. Results are visualized through a multi-level GANTT diagram (Figure 6), with cost-based schedules shown in green and BIM-based updates in orange. The "Start PRJ," "End PRJ," and "Du-

ration PRJ" columns show task dates and durations from costing data, while the "EXE Start," "EXE End," and "EXE Duration" columns report the corresponding values from the execution stage using the 3D BIM model. Unlike conventional LPS® representations, tasks are not displayed in a time-distance diagram, preserving detailed MEP sequencing and preventing overlaps.



Figure 6. Detailed overview of the tasks required to install hot water system terminals on the ground floor



Figure 7. Detailed task information. (a) General information tab. (b) Predecessors tab. (c) Resources tab. (d) Notes tab.

The interface further provides detailed task views through four windows (Figure 7):

- **General**: contains information on task duration, start, and end dates.
- **Dependencies**: lists all predecessors of the selected task.
- **Resources**: lists the equipment linked to this task derived from the quantity surveying data, along with their quantities. These correspond to the items checked in the list.
- **Notes**: lists the equipment linked to this task derived from the BIM model, including their quantities and properties.

Lastly, the system computes daily workload and provides estimates of the required human resources.

4.2 Expert validation results

To validate alignment with user requirements identified in semi-structured interviews, a workshop was conducted with six senior project managers ($\geq$ 20 years of experience, projects exceeding €15 M). The session presented the application's features and gathered feedback on functionality, gaps, and potential improvements. Figure 8 summarizes functional coverage and participant feedback.

A post-workshop questionnaire evaluated the solution's utility and reliability through seven key questions addressing problem relevance, current applicability, usefulness for construction planning, result accuracy, computation time, functional coverage, and willingness to test the solution on a pilot site.

Participants recognized the solution's relevance; however, two-thirds expressed concerns regarding calculation accuracy, potentially due to past negative experiences with automation, a preference for manual processes, or resistance to change. Eighty-three percent found processing times reasonable, and 67% indicated willingness to test the tool on live projects.

5 Discussion and analysis

Although planning automation is not a new issue, this study offers several significant contributions.

First, it addresses the automation of MEP work scheduling, an area that has received considerably less attention than structural works [16]. The relative neglect of MEP disciplines can be explained by their inherent complexity, as they involve highly interconnected systems installed in confined spaces and require coordination among multiple subcontractors. These challenges, combined with the dynamic and uncertain nature of construction projects, make the planning and control of MEP works particularly critical.

Second, the proposed solution capitalizes on operational data from multiple sources beyond the BIM model alone. This distinguishes it from previous research efforts, which often rely exclusively on BIM data [17, 18, 19]. However, the study also highlights key challenges encountered in construction projects, notably data incompleteness and heterogeneity. Data incompleteness stems from the lack of exhaustive equipment descriptions in both costing sheets and BIM models, which hinders accurate identification of corresponding equipment codes in cost databases. To improve the reliability and efficiency of the proposed au-



Figure 8. Coverage of User-Specified Functionalities and Project Manager Feedback (green: functionality fully developed, yellow: partially developed, orange: not developed, purple: new functionality identified)

tomated scheduling system, it is therefore essential to enhance the level of detail in these data sources. At present, this limitation restricts the full deployment of required functionalities and constrains the scalability of the solution. Consequently, rigorous data capture, codification, and preservation from the early project phases are required to fully realize the potential of the proposed planning approach. In addition, the adoption of standardized classification systems such as Omniclass or Uniclass could help reduce semantic incompatibilities between data produced by different actors and project phases.

Third, the research formalizes MEP planning know-how through expert workshops. However, the manual and explicit modeling of each material's operations and dependencies proves to be costly and labor-intensive, which may limit scalability when extending the system to new material categories. A promising direction for future work is the automated acquisition of planning knowledge from existing project records, such as historical schedules, BIM-based process data, technical specifications, and execution reports. Recent studies have demonstrated the potential of sequential machine learning to learn dynamic work templates from past schedules, of pattern-recognition methods to identify recurring process logic in BIM-based schedules, and of natural language processing techniques to extract structured knowledge from construction documents. Taken together, these advances suggest that part of the sequencing knowledge currently captured through expert elicitation could in the future be inferred automatically from accumulated project data, thus reducing setup cost and improving transferability across new material families [16, 25].

Fourth, the proposed automated scheduling application streamlines the project planning process by rapidly and effectively processing complex datasets. When input data are complete, the solution can save an average of one week of manual work per project manager, thereby freeing time for higher-value activities such as decision-making and problem-solving. Beyond these time savings, the economic impact is equally significant: automation generates a gain of approximately 0.91% of project revenue, corresponding to nearly 30% of the targeted profit margin (3%). These estimates were developed collaboratively with domain experts using average project data, reinforcing both their credibility and practical relevance. To further enhance its contribution, the application should be extended to the remaining phases of the LPS®, including lookahead planning, weekly work planning, and progress monitoring. The latter could itself be automated using advanced techniques such as unmanned aerial vehicles (UAVs) or real-time location systems (RTLS) for data capture, as well as 3D digital imaging and computer vision algorithms for progress measurement [26, 27].

Finally, this work aligns with broader objectives related to ensuring data continuity throughout the construction project lifecycle, thereby paving the way for more efficient building information management. A significant avenue for future research lies in expanding the automated planning tool to involve additional project stakeholders, including actors in design, procurement, manufacturing, logistics, and commissioning. This evolution entails harnessing large volumes of data from diverse operational sources, which requires the implementation of data warehouses (Figure 9). Such data warehouses are supplied with information generated by multiple operational soft-

ware systems across project stages, enabling reporting, data mining, and strategic decision-support analyses [28]. This approach introduces a new perspective on lean building information management, referred to as *Lean BIM Management*, and contributes to the emergence of a new era in Lean Construction characterized by intelligence and interconnection between machines, humans, and materials - an evolution described as the *Smart Lean era* by Demir [29].

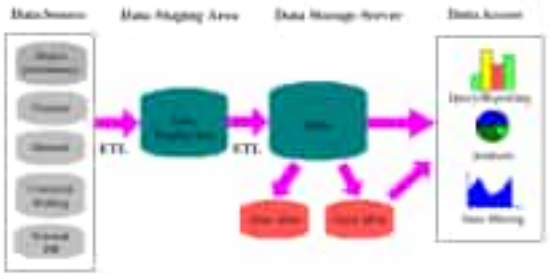


Figure 9. Architecture of a construction-project data warehouse [28]

6 Conclusion

This study aimed to enhance the integration between the Last Planner System® (LPS®) and Building Information Modeling (BIM) at the data-processing level. The main contribution of this work is the development of a web-based application that automatically generates time–location schedules for MEP trades by leveraging multiple data sources, namely cost-estimation data and BIM model information.

Following a Constructive Research Approach, the paper first identifies the planning problem and reviews its theoretical background, before detailing the architecture, input data, and user interface panels of the proposed automated scheduling solution. Real project data from the Rennes Surgical and Interventional Center case study are then used to demonstrate the practical functioning of the system. In addition, an analytical evaluation conducted through focus groups highlights the utility and relevance of the developed application, despite some initial and largely unfounded skepticism regarding the accuracy of automated planning calculations.

Finally, the study discusses the theoretical implications of the proposed artefact, laying the foundation for an effective approach to building information management, referred to as *Lean BIM Management*. Building on this foundation, the paradigm opens avenues for future developments through the integration of data from multiple construction stakeholders into centralized data warehouses. Such integration would ensure data continuity across the entire project lifecycle- from design and procurement to manufacturing, logistics, construction and commissioning- thereby significantly enhancing project management practices and data utilization within the construction industry. This progression ultimately supports the emergence of a Smart Lean Construction era, characterized by greater intelligence and stronger interconnections among data, actors, and processes.

References

[1] D. Bhargav. Developing a construction management system based on lean construction and building information modelling. On-line: `https://www.researchgate.net/publication/280840295_Developing_a_construction_management_system_based_on_lean_construction_and_building_information_modelling`, Accessed: 22/11/2025.

[2] R. Hannis, S. Sorooshiah, S. Bin, and G. Duvvuru. Lean construction tools. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pages 23–25, Detroit, Michigan, 2016.

[3] M. A. Obaidi and H. Karimi. Last planner system: An effective solution for waste reduction in construction projects. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 2025. doi:10.1007/s40996-025-02056-1.

[4] D. C. Hamerski, T. A. Saurin, C. T. Formoso, and E. L. Isatto. The contributions of the last planner system to resilient performance in construction projects. *Construction Management and Economics*, 42(4):328–345, 2024. doi:10.1080/01446193.2023.2262622.

[5] C. Görsch, O. Seppänen, A. Peltokorpi, and R. Lavikka. Task planning and control in construction: revealing workers as early and late planners. *Construction Management and Economics*, 42(5):431–450, 2024. doi:10.1080/01446193.2023.2270080.

[6] H. G. Bayhan, S. Demirkesen, C. Zhang, and A. Tezel. A lean construction and bim interaction model for the construction industry. *Production Planning & Control*, 34(15):1447–1474, 2023. doi:10.1080/09537287.2021.2019342.

[7] E. Sriprasert and N. Dawood. Multi-constraint information management and visualisation for collaborative planning and control in construction. *Journal of Information Technology in Construction*, 8:341–366, 2003.

[8] R. Sacks, L. Koskela, B. A. Dave, and R. Owen. Interaction of lean and building information modeling in construction. *Journal of Construction Engineering and Management*, 136(9):968–980, 2010. doi:10.1061/(ASCE)CO.1943-7862.0000203.

[9] R. Sacks, R. Barak, B. Belaciano, U. Gurevich, and E. Pikas. Kanbim workflow management system: Prototype implementation and field testing. *Lean Construction Journal*, 2013.

[10] B. Dave, S. Boddy, and L. Koskela. Visilean: Designing a production management system with lean and bim. In *Proceedings of the 19th Annual Conference of the International Group for Lean Construction*, 2011.

[11] M. Sbiti, K. Beddiar, D. Beladjine, R. Perrault, and B. Mazari. Toward bim and lps data integration for lean site project management: A state-of-the-art review and recommendations. *Buildings*, 11(5):196, 2021.

[12] A. Luna, R. F. Herrera, K. Castañeda, E. Atencio, and C. Biotto. Benefits and limitations of lean tools in the building design process: A functional and comparative analysis. *Applied Sciences*, 15(9):5137, 2025. doi:10.3390/app15095137.

[13] C. P. Schimanski, G. Pasetti Monizza, C. Marcher, and D. T. Matt. Development of a bim-based production planning and control system for lean construction. *Frontiers of Engineering Management*, 8(3): 429–441, 2021. doi:10.1007/s42524-020-0105-5.

[14] C. P. Schimanski, C. Marcher, G. P. Monizza, and D. T. Matt. The last planner® system and building information modeling in construction execution. *Applied Sciences*, 10(3):821, 2020. doi:10.3390/app10030821.

[15] A. K. Agrawal, Y. Zou, L. Chen, M. A. Abdelmegid, and V. A. González. Moving toward lean construction through automation of planning and control in last planner system. *Developments in the Built Environment*, 18:100419, 2024. doi:10.1016/j.dibe.2024.100419.

[16] F. Amer, H. Y. Koh, and M. Golparvar-Fard. Automated methods and systems for construction planning and scheduling. *Journal of Construction Engineering and Management*, 147(7):03121002, 2021. doi:10.1061/(ASCE)CO.1943-7862.0002093.

[17] B. de Vries and J. M. J. Harink. Generation of a construction planning from a 3d cad model. *Automation in Construction*, 16(1):13–18, 2007. doi:10.1016/j.autcon.2005.10.010.

[18] M. Kataoka. Automated generation of construction plans from primitive geometries. *Journal of Construction Engineering and Management*, 134(8):592–600, 2008. doi:10.1061/(ASCE)0733-9364(2008)134:8(592).

[19] H.-W. Wang, J.-R. Lin, and J.-P. Zhang. Work package-based information modeling for resource-constrained scheduling of construction projects. *Automation in Construction*, 109:102958, 2020. doi:10.1016/j.autcon.2019.102958.

[20] M. Sbiti, D. Beladjine, K. Beddiar, and B. Mazari. Leveraging quantity surveying data and bim to automate mechanical and electrical construction planning. *Applied Sciences*, 12(9):4546, 2022. doi:10.3390/app12094546.

[21] A. Baldrich Aragó, J. Roig Hernando, F. J. Llovera Saez, and J. Coll Bertran. Quantity surveying and bim 5d. *Journal of Building Engineering*, 44:103234, 2021. doi:10.1016/j.jobe.2021.103234.

[22] P. Smith. Project cost management with 5d bim. *Procedia - Social and Behavioral Sciences*, 226:193–200, 2016. doi:10.1016/j.sbspro.2016.06.179.

[23] M. Sbiti. *Interaction Lean-BIM en phase exécution de projet de construction*. PhD thesis, HESAM Université, 2023.

[24] K. Lukka. The constructive research approach. *Case study research in logistics*, pages 83–101, 2003.

[25] A. K. Agrawal, Y. Zou, H. Jin, M. Abdelmegid, and V. Gonzalez. Automated assembly sequence planning and scheduling in precast building projects using reinforcement learning. In *Proceedings of ISARC*, pages 1355–1362, 2025.

[26] M. Sami Ur Rehman, M. T. Shafiq, and F. Ullah. Automated computer vision-based construction progress monitoring. *Buildings*, 12(7):1037, 2022. doi:10.3390/buildings12071037.

[27] H. Bardareh and O. Moselhi. Automated generation of inspection reports for construction operations. In *Proceedings of ISARC*, pages 1347–1354, 2025.

[28] T. Rujirayanyong and J. J. Shi. A project-oriented data warehouse for construction. *Automation in Construction*, 15(6):800–807, 2006. doi:10.1016/j.autcon.2005.11.001.

[29] S.-T. Demir. The future of lean construction. Online: `https://leanconstructionblog.com/the-future-of-lean-construction.html`, Accessed: 02/01/2026.

AI-based Dynamic Quiz System: Disaster Knowledge King

Chen-Hao Tseng[1] and James Yichu Chen[1]

[1] National Taiwan University of Science and Technology, Taiwan
chtseng@9999gmail.com, yie@mail.ntust.edu.tw

Abstract

Amidst escalating global climatic threats, effectively disseminating disaster preparedness knowledge to the public remains a critical challenge. This study introduces Disaster Knowledge King (DKK), an automated, dynamic quiz system designed to modernize disaster risk reduction (DRR) education. While traditional educational games suffer from static, "dead" databases that require manual updates, DKK leverages Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) to facilitate real-time content creation.

The system features a fully automated pipeline that scrapes global news from sources like AP and Reuters, filters for recency, and utilizes the Gemini API to generate multiple-choice questions. By grounding responses in professional Ready.gov guidelines via RAG, DKK provides users with both situational awareness and actionable response strategies, and system is implemented as a LINE Bot to ensure high accessibility. While the DKK reduces the production cost to approximately $0.002 per question set, its G-Eval and UAT scores significantly exceed the median threshold. This study indicates that integrating AI-driven automation with gamified social media interfaces offers a scalable and efficient pathway for enhancing public disaster literacy.

Keywords

LLMs, RAG, automation, DRR, disaster literacy, education, LINE Bot, multiple choice questions.

1 Introduction

1.1 Background

For decades, humanity has confronted the escalating threat of anthropogenic global warming, characterized by an increase in extreme climatic events and natural disasters. While disaster management experts over the world have developed sophisticated mitigation strategies and response frameworks, a persistent challenge remains: the effective spread of “disaster preparedness knowledge” and “environmental literacy” to the general public.

The 2022 debut of OpenAI’s ChatGPT precipitated a revolutionary transformation in both professional productivity and daily life, signalling a new era of AI-augmented workflows. Consequently, the application of Large Language Models (LLMs) has expanded across diverse sectors, fostering unprecedented levels of innovation and generation. Within this context, experts have begun leveraging these models to enhance disaster education through the pedagogical adaptation of complex materials, the design of immersive practice scenarios, and the high-level analysis of learner performance.

1.2 Motivation

The classic mobile game “Knowledge King” is a landmark educational title that has achieved widespread popularity in Taiwan. The game utilizes a vast database of multiple-choice questions across diverse fields, challenging users’ knowledge in a competitive "contest mode." By presenting information in a thrilling and engaging format, Knowledge King has successfully demonstrated how gamification can enhance general knowledge. This study argues that “an educational tool modeled after Knowledge King is an ideal method for disseminating “disaster preparedness knowledge” and “environmental literacy” to the general public.”

However, this study identifies a critical limitation within the current Knowledge King model: the static nature of its database. Because the question pool is "dead," it requires labor-intensive and time-consuming manual updates to renew. This prevents the content from real-time information, so it’s a drawback that must be addressed.

1.3 Objective

This study aims to develop a dynamic quiz system titled "Disaster Knowledge King" (DKK). Designed as an advanced iteration of the Knowledge King framework, DKK focuses specifically on disaster prevention education. The system provides multiple-choice questions based on real-time disaster information, accompanied by comprehensive explanations and recommended response strategies. The most significant innovation of DKK is its ability to maintain a contemporary database; unlike traditional static systems, DKK utilizes Large Language Models to automate the

generation of questions. By leveraging the successful engagement strategies of its predecessor, DKK serves as a highly efficient educational tool for spreading “disaster preparedness knowledge” and “environmental literacy”.

2 Literature Review

As extreme weather and associated disasters are the biggest contemporary global concern, “disaster risk reduction (DRR)” is imperative that many papers have valued its integration into education and explored the better methods to practice it. Research argues that DRR in education is a psycho-social concept that the successful integration of DRR requires a transition in instructional design toward localized and participatory approaches, emphasizing individual specialized action and active engagement within local communities and environments[1]. Nahid Aghaei [2] et al. also focus on the DRR education strategy for raising knowledge, and the critical point is to transfer accessible information with high quality, especially through the community involvement to share knowledge.

However, the main challenge remains the deficiency in conceptual understanding regarding DRR, directly contributing to resource scarcity and the stagnation of social development in this area. As it is difficult to enhance public disaster literacy through traditional methods, utilizing social networks or media offers a viable and potential channel[1, 2].

Because of resource scarcity, utilizing Artificial Intelligence to assist material production is a promising method. Biancini et al.[3] have introduced a methodology for generating multiple choices question by LLMs. By providing the context and specific prompts, GPT-3.5 could be proficient in generating MCQs, which is evaluated through the non-parametric Friedman test. This study illustrates the viability of LLMs to produce MCQs and provides valuable insights for the future.

Additional research either demonstrates the capacity of LLMs to generate complex questions—specifically those involving analysis and synthesis—which highlights this model’s suitability for educational curricula. GPT-4 has exhibited the greatest versatility in constructing tasks of different formats and at multiple cognitive levels in alignment with Bloom's Taxonomy.[4] Concurrently, Shahab Sheikhalishahi et al. [5] compared the performance of ChatGPT-4o, ChatGPT-5, and Gemini 2.5 Flash on Persian internal medicine subspecialty board questions; the results reveal the superior performance of Gemini 2.5 in text-based clinical reasoning. Among existing LLMs, Gemini appears to possess the most robust textual understanding and Inference abilities.

Based on the concept above, content of DDK education should customize for diverse communities and environments, so the response strategies would be highly varied that LLMs can leverage Retrieval-Augmented Generation (RAG) to produce more ideal and accurate responses. Yunfan Gao et al.[6] thoroughly introduce RAG which effectively integrates parametric knowledge of LLMs with non-parametric external data sources, thereby mitigating hallucinations and inaccuracies; this establishes it as one of the essential frameworks for deploying large language models. The fundamental components of RAG are ”Indexing”, “Retrieve” and “Generation”, while indexing further consisted of data indexing, chunking and embedding. As a method combining retrieval and generation, RAG can offers multiple avenues for future research.

While early retrieval models rely on sparse representation (e.g., BM25 or TF-IDF), they frequently struggle with the vocabulary mismatch problem and complex semantic queries. Alternatively, by using pre-trained transformers to generate dense vector embeddings, these models can identify deeper conceptual connections between queries and documents. Consequently, dense retrieval offers a more robust framework for capturing context and improving overall retrieval precision[7].

The application of AI has emerged as a significant trend in disaster prevention, particularly within DRR education. There have been several examples, for instance, Ömer Cem Karacaoğlu et al.[8] explore AI-driven strategies in disaster education to determine their feasibility in fostering innovative and proactive preparedness. The research reveals that “Text classification guides students by identifying the topics covered in the educational material and helps focused learning. Question-answer applications enhance the learning experience by automatically extracting answers from texts.”

In addition to AI, DRR education can be integrated with others technology, such as chatbot, for the purpose of innovative pedagogies to enhance learning efficacy. Okan Yetişensoy [9] explores the effectiveness of AI-assisted chatbots in disaster education by comparing chatbot-integrated instruction against conventional pedagogical methods. The result reveals a significant positive impact on cognitive learning outcomes, with the experimental group demonstrating superior knowledge retention compared to the control group. Furthermore, semi-structured interviews suggested that learners favored the chatbot approach, citing its student-centered design and engaging pedagogical features as key drivers for more effective education. The research also suggests that “technology combining XR and chatbots would be helpful in enabling students to experience disaster situations firsthand and enhance the skills necessary for dealing with disasters.”

Squire’s research suggests that the immersive nature of computer and video games provides a comprehensive

framework for learning that surpasses traditional pedagogies in global reach and effectiveness[10]. Fletcher's research highlights the superiority of interactive tools, noting that students retain 75% of information through simultaneous seeing, hearing, and interacting, compared to just 40% of visual data and 20% via auditory instruction[11]. Therefore, gamification might be a suitable pedagogy for DRR education.

3 Methodology

This study constructs the Disaster Knowledge King (DKK), which is organized into three subordinate sections: the Generation System, the Database, and the Answer System, with several detailed modules under each section. The specific system architecture is illustrated in Figure 1. Specifically, this study develops a variety of functions into corresponding modules and then integrates associated modules into a primary section. Consequently, each section executes its respective tasks independently, allowing the DKK to stream results from all sections to form a complete, dynamic quiz system.

In the Generation System, the server executes a procedure at each interval to generate a series of intact question sets. First, the News Module collects real-time news sources concerning disaster events occurring globally. Second, the Filter Module ensures the news is fresh and removes expired question sets from the database. Third, the AI Manufacture Module utilizes a Large Language Model (LLM) to generate the question set, subsequently employing the RAG method with specific disaster response documentation to recommend response strategies. Lastly, the Packaging Module organizes the LLM output into a specific format and uploads it to the Database..

In the Database, this section consists of the "DB Calling Module" and the "Cloud DB". Its primary function is the storage of all question sets; it also serves as the connection between the Generative and Answer systems while transferring the DKK's essential properties: disaster preparedness knowledge and environmental literacy.

In the Answer System, this study adopts the LINE Bot as the operational medium for the DKK, substituting traditional application software with a near-ubiquitous social media interface. Upon user initiation, the Question Module retrieves and renders a question from the database. Subsequently, the Options Module presents four processed alternatives to the user. Following the user's input, the Answer Module displays the correct answer, an explanation, and the response method, using specific colour to indicate whether the user's response was correct.

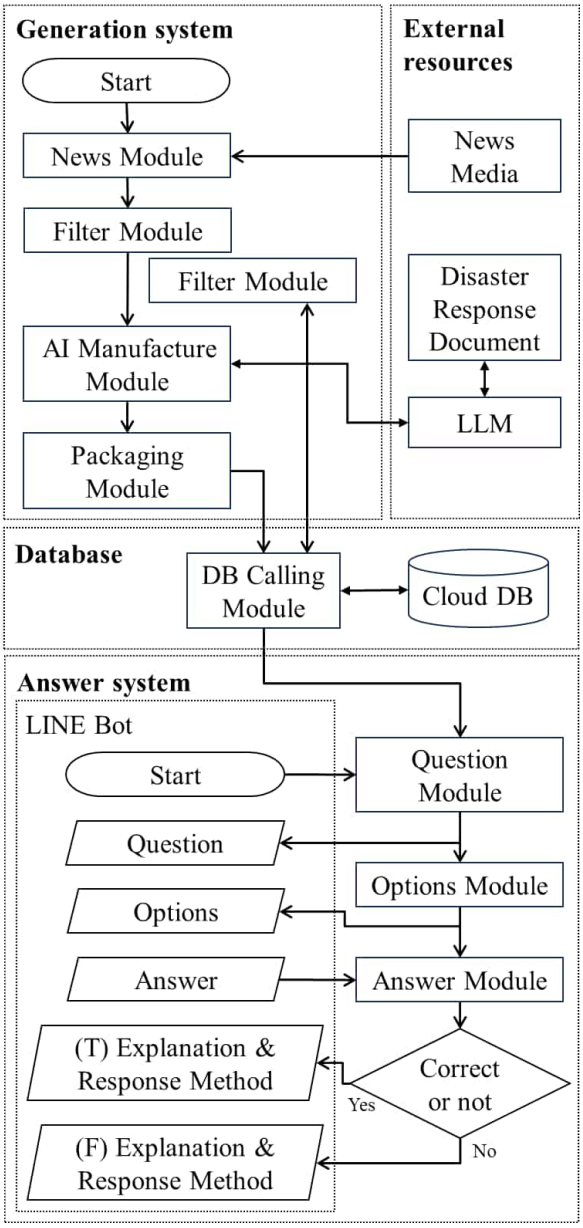


Figure 1 Disaster Knowledge King Disaster Knowledge system architecture diagram

4 Implementation

Diverse APIs and LLMs are integrated into DKK's framework, with key research outcomes focusing on prompt engineering, production pipeline design, and dynamic disaster prevention methodologies, and the whole system development is based on Python.

4.1 Generation system

The Generation system serves as the fundamental driver of overall operations. The system is comprised of

four distinct modules to establish a production pipeline: the News Integration Module, the Filter Module, the AI Manufacture Module and the Packaging Module.

4.1.1 News Module

To acquire global disaster information, the system integrates three primary news sources—the Associated Press, Reuters, and Channel News Asia—selected for their credibility and extensive regional coverage. The news integration workflow proceeds in three stages: first, defining extraction parameters (keywords, media outlets, and volume); second, utilizing the NewsAPI.org service to retrieve headlines, publication times, and URLs; and third, employing the BeautifulSoup library to scrape full web content from those URLs.

Once ingested, the News Module categorizes the data into specific disaster types, including “avalanche, earthquake, fire, explosion, extreme heat, drought, flood, hazardous materials, hurricane, typhoon, debris flow, pandemic, severe weather, thunderstorms, lightning strike, tornado, tsunami, volcano, wildfire, and snowstorm.”, and standardizes publication dates into a unified format (YYYY-MM-DD HH:MM:SS) for database consistency.

Finally, the module encapsulates this data into a structured schema—comprising news content, date, disaster type, and URL—before transmitting it to the Filter Module for processing.

4.1.2 Filter Module

The core characteristic of the DKK’s questions is their "dynamism," signifying that all information must maintain a high degree of real-time relevance and freshness. Consequently, the system must strictly manage the temporal validity of both newly imported news data and existing items within the database.

In the first step, the Filter Module receives data packets from the News Module. It compares the "publication date" of each entry against a pre-defined "update threshold" set in the system parameters, thereby filtering out data that exceeds the permissible latency.

In the second step, the system executes a stale data purging operation. Utilizing a data deletion function, the Filter Module applies a specific retention rule: "any data record with a publication date older than N days (N being the retention period defined by system parameters) shall be removed." If the deletion of records results in a specific disaster category containing zero items, the module invokes the table dropping function to remove the empty category table. This process ensures the structural integrity and operational stability of the Generation system.

In the final step, the Filter Module encapsulates all qualified data into a standardized format—"news content, publication date, disaster type, and news URL—and transmits it to the AI Manufacture Generation Module to serve as the foundational material for subsequent question generation.

4.1.3 AI Manufacture Module

The core of the DKK generation system is the AI Manufacture Module, which integrates the Gemini API, Retrieval-Augmented Generation (RAG), and structured data schemas to produce "news-dissemination-oriented" educational content. Gemini was selected as the primary LLM for its superior semantic parsing and reasoning capabilities. To ensure structural reliability, the module utilizes pydantic and google-genai libraries to enforce a predefined schema, guaranteeing that every output includes a high-quality question, four options, and a detailed explanation.

Prompt engineering serves as the study’s central research focus, guided by two principles: "multi-layered iterative development" and "contextual supplementation." The former breaks down complex tasks into high-success sub-tasks through iteration, while the latter applies directional constraints to ensure strict adherence to the desired content. Together, these principles enable the system to achieve superior and stable generation quality.

A distinctive feature of the DKK is the "response method" included in every explanation to meet Disaster Risk Reduction (DRR) goals. To ensure these methods are accurate, the research employs RAG by grounding the Gemini API in professional Ready.gov guidelines. This workflow involves converting guidelines to Markdown, indexing them via the google-genai SDK, and generating grounded responses. Ultimately, prompt engineering remains most vital, as it dictates the overall educational effectiveness of the module.

The following are the prompts used for each of the five production steps in AI Manufacture Module:

prompt1: f'You are now a global disaster intelligence expert. Please organize the article from link {news[“news_URL”]} into four key points, focusing on the damages caused.'

prompt2: 'Based on these four key points, create a multiple-choice question with four options(A, B, C, D), designed so that people can use reasoning to arrive at the answer and understand what happened. Please also provide the correct answer.'

prompt3: 'Then provide an explanation for the answer in under 70 words, and ensure that the text does not contain the words \'option\' or \'A, B, C, D\'." '

prompt4 : Next, based on the knowledge and content in "fileSearchStores", please tell everyone how to respond if they are personally involved in the case of this article, within 70 words.

prompt5: ‘Based on the content above, return the “Question”, “Option_A”, “Option_B”, “Option_C”,

"Option_D", "Correct_Answer", "Explanation", and "Response_Method". Do not change anything, especially the question and the options.'

4.1.4 Packaging Module

The Packaging Module processes source response from Gemini, which is received as a Python dictionary, by assigning a unique item ID, labeling options, and merging the "Explanation" and "Response Method" into a single string separated by a "&&&" delimiter. This concatenation strategy ensures the dataset can fit into the pre-defined database schema while allowing both data points to be stored in one block for future scalability.

To generate the unique ID, the module queries the database to identify existing entries and randomly selects a new value from a range of 1 to k, where k is an adjustable parameter tailored to table capacity. The finalized data is encapsulated into a standardized "questions dataset" having a schema: ('ID', 'Question', 'Option A', 'Option B', 'Option C', 'Option D', 'Correct Answer', 'Explanation & Response Method', 'Publication Date', 'News Title', 'News_URL').

Finally, the DB Calling Module uploads this dataset to the target database via a data-appending function to complete the run.

4.2 Answer system

This study adopts a chatbot as the primary vehicle for the answering system, with LINE selected as the implementation platform. Given its high penetration rate and convenience, the LINE platform allows users to access and engage with educational resources through a low technical barrier and at zero cost.

4.2.1 LINE Bot

The LINE Bot facilitates user interaction through message templates and Rich Menus. The LINE Bot's internal logic serves as the central controller for the entire answering system, as Figure 2 shows. Its development involved chatbot account creation, Vercel project establishment, API configuration, and UI design.

At the core of the system is the "Central Processor," a module built on the official LINE SDK. This processor governs the DKK-robot's operational modes and integrates database workflows to provide real-time question generation and feedback. The Central Processor operates via a five-step workflow: "initializing the Flask framework", "connecting to the LINE Messaging API", "monitoring webhook data", "parsing incoming information" and "executing module results".

Depending on the parsed input, the Central Processor triggers one of three functional modules—Question, Options, or Answer—each performing data operations via a DB Calling Module connected to a cloud database. To ensure a professional user experience, the interface was crafted using the LINE Bot Designer and Flex Message Simulator. The resulting JSON layouts are embedded directly into the bot's logic, providing users with an intuitive and clear educational interface.

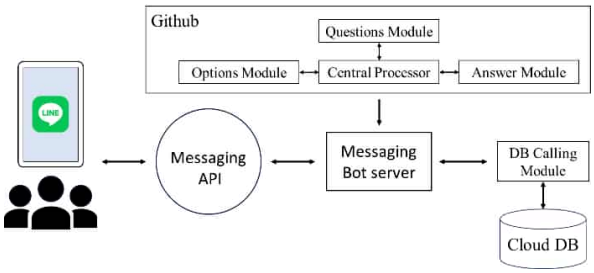


Figure 2 architecture framework of LINE Bot

4.2.2 Question Module

When receiving a response 'Disaster', the Central Processor invokes the Question Module to initiate the question generation process. The Question Module randomly fetches a record from the cloud database and deconstructs it into an 'information packet,' which is then returned to the Central Processor for subsequent process. Throughout its operation, the Question Module interacts with the database via the DB Calling module, utilizing several functions for connection configuration, table deletion, data deletion and data fetching.

The operational workflow of the Question Module designed in this study is as follows:

1. Data availability verification: The module first confirms the existence of question datasets within the database. If nothing is found, an error notification is transmitted to the Central Processor.

2. Random selection: If question dataset is available, the module retrieves all item IDs in database and utilizes the Python random library to select a single ID randomly.

3. Data extraction: Based on the selected ID, the module fetches the complete question dataset, including ID, Question, Option A, Option B, Option C, Option D, Correct Answer, Explanation & Response Method, Publication Date, News Title, News_URL.

4 .Message formatting: The module generates a text message formatted as 'disaster-[ID]: Question'.

5. Information packet encapsulation: The text message, question data are integrated into an information packet, which is then returned to the Central Processor.

Finally, once receiving this packet, the Central Processor triggers the Options Module, passing the information packet as an execution parameter. Notably, test message will be display later along with options.

4.2.3 Options Module

Immediately following the Item Module, the Options Module is activated to process the 'option content' within the item dataset. This module features two critical design

priorities: First, to mitigate potential spatial bias in LLM-generated outputs, an Option Shuffling Mechanism re-randomizes the choice sequence. Second, since the LINE Bot interface is inherently stateless and the complete interaction workflow (Category Selection → Item Generation → Answering → Feedback) requires three distinct exchanges, a cloud-based session memory mechanism was developed. This ensures that the context of the user's answering process is accurately tracked and retrieved. The operational workflow is as follows:

1. Extraction and shuffling: The module retrieves the four original options and applies a randomization algorithm to re-map the correct answer to a new position.

2. Session Management: To support concurrent multi-user access, a unique 8-digit Temporary ID is dynamically generated (e.g., 10125002), encoding the session status, category, and item number.

3. Data Persistence: A temporary dataset, including the shuffled options, metadata, and a start-time timestamp, is uploaded to the database's 'Memory' table.

4. UI Generation: The module constructs a Carousel Flex Message where each card serves as an interactive option containing the description and a postback trigger.

The Temporary ID serves as a primary key within the database's Memory table to uniquely identify each transient answering record. The 8-digit ID follows a specific encoding rule: 10,000,000 + (Category Code * 100,000) + (Item Number * 100) + Session ID.

In the end, Central Processor will display the question and options after receiving the Carousel Flex Message.

4.2.4 Answer Module

When receiving the user's response (containing a newline delimiter \n), the Central Processor invokes the Answer Module. The module retrieves the corresponding session record from the database's Memory table to compare the user's input with the correct answer. Given the high information density of the feedback, Flex Messages are adopted to provide a structured, customizable, and visually clear template for the user.

The operational workflow of the Answer Module is as follows:

1. Message parsing: The incoming string is deconstructed by using the \n and - delimiters to extract the item category, the Temporary ID, and the user's selected option.

2. Temporal validation: The module utilizes the datetime library to compare the current system time against the session timestamp. If the response exceeds the predefined time limit, an error message is returned.

3. Memory purging: To maintain database efficiency and minimize storage overhead, the module invokes a deletion function to remove the temporary record from the 'Memory' table once it is processed.

4. Conditional formatting: The system performs a logic check between the user's answer and the correct entry. The Flex Message adopts a green color scheme for correct answers and a red scheme for incorrect ones.

5. Flex message construction: The feedback's Flex Message is assembled into three distinct layers: upper block displaying the correct answer; middle block containing the explanation explanation; and lower block showing the publication date and the source URL.

6. Response method: After separating the response method from Explanation_Plue_Response_Method, the module displays it in another Flex Message template, which synchronize the pattern with feedback's template.

7. Final Dispatch: The finalized Flex Message of feedback and response method are returned to the Central Processor and transmitted to the user.

This structured interface allows users to quickly grasp the correct knowledge and background information, with the option to access the original news manuscript via the embedded hyperlink for deeper learning.

4.3 Database

To support the high-frequency and real-time data access requirements of the DKK, this study establishes a database. The database also serves as the connection for information transition between the independently operated Generation System and the Answering System. To ensure constant availability, the database is deployed on the Google Cloud.

4.3.1 Cloud DB (google sheet)

To mitigate the high operational costs of maintaining a MySQL instance on GCP, this study constructs an alternative storing framework by leveraging Google sheet, which is free under 15 GB. The database consists of only one single table 'disaster,' which contains 200 data slots (a capacity set by the DKK's Parameter Control Panel). Each data slot is in a uniform schema: 'ID', 'Question', 'Option A', 'Option B', 'Option C', 'Option D', 'Correct Answer', 'Explanation & Response Method', 'Publication 'Date', 'News Title' and 'Source URL.'

4.3.2 DB Calling Module

This study utilizes the distinct data source by leveraging specialized SDK "gspread". This library provides Python-based programmatic access, enabling this study to develop the DB Calling Module. This module is integrated into both the Generation system and Answering systems to serve as the primary interface for database interaction.

The DB Calling Module comprises seven functional functions: Connection Setting, Table Creating, Table Dropping, Table showing, Data Appending, Data Dropping and Data fetching. By adopting this modular architecture, functional components across the dual systems can execute database operations simply by

invoking the relevant function, regardless of the underlying data source.

5 Result

After registering a LINE account, users can scan the QR-Code (Figure 3) , click the “Add friend” block, then initial interaction with the DKK-robot which displays the comprehensive result.

Figure 3 QR-Code of DKK-robot

5.1 DKK Operation

The Generation System features fully automated, production capabilities. By utilizing a scheduled server task, the system initiates its operational pipeline daily at 13:00 (GMT+8), consistently performing news scraping, data filtering, item generation, and packaging.

Regarding economic efficiency, the DKK system operates with zero human labor costs. Infrastructure expenses include $1.67 per day for GCP MySQL and $0 for Google Sheets. With the Gemini-2.5-flash-lite model priced at $0.40 per million tokens and an average consumption of 5,000 tokens per item, the production cost per question is approximately $0.002.

5.2 DKK’s question set & AI generation

All question sets in DKK’s are made by AI, and generation at every steps are evaluated by G-Eval, a widely recognized benchmark in the NLP community. G-Eval is an open-source framework using LLM-as-a-judge with chain-of-thoughts (CoT) to evaluate LLM outputs,

"This study recorded 26 data points collected between March 20 and March 28, 2026. Based on the official default evaluation criteria shown in Figure 4, DKK’s average G-Eval scores for the first four production stages were 0.949, 0.950, 0.904, and 0.865, respectively, significantly exceeding the quality threshold (0.5). These results confirm the LLM integrated within DKK exhibits excellent robustness.

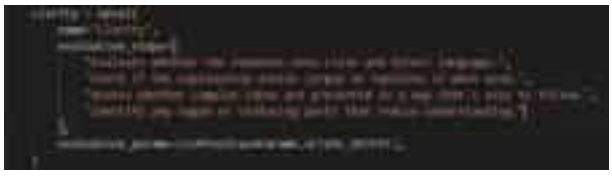


Figure 4 Official default evaluation criteria

"Table 1 presents an example of the DKK’s AI Manufacture Module generation, and it is based on the Source new: https://reurl.cc/ZldgVV

Table 1 Question set of DKK

Dataset Items	Content
Question	Considering the damages caused by the Wang Fuk Court fire, which of the following statements best summarizes the multifaceted nature of the disaster's impact?
Option A	The fire primarily caused minor property damage and a few injuries, with the main impact being a temporary disruption to local services.
Option B	The primary damages were limited to the immediate structure of the building, with minimal loss of life and no significant long-term economic consequences.
Option C	The fire resulted in a catastrophic loss of life, extensive structural destruction of multiple buildings, substantial financial strain on victims and the community, and profound psychological trauma.
Option D	The damages were solely economic, with significant insurance payouts compensating for the loss of personal belongings and the structural rebuilding efforts.
Correct Answer	C
Explanation	The correct choice encompasses the widespread fatalities, the severe destruction across multiple residential blocks, the considerable financial aid required for victims and rebuilding, and the lasting psychological impact on survivors. The other choices significantly underestimate the scale and breadth of the devastation
Response Method	In the event of a home fire, victims should immediately prioritize escape. If there's smoke, stay low and crawl to an exit. Before opening a door, feel it for heat; if hot, use an alternate escape route. Close doors behind you to slow the fire's spread. Once safely outside, call emergency services. If unable to escape, close the door, seal gaps, and signal for help from a window.

5.3 Users interface

Figure 5 illustrates the interface of the DKK-robot. By simply clicking the button “START” and “Select”, user is able to engage in the DRR educational gameplay.

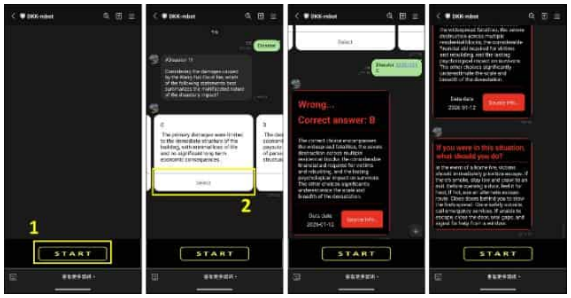


Figure 5 DKK-robot operation process

5.4 User acceptance testing (UAT)

This study utilized five customized questions to preliminarily evaluate user experience, each graded from 0 (disagree) to 5 (agree). 20 students and staff members from NTUST participated in the usability testing of DKK.

As illustrated in Table 2, the average UAT scores indicate that while system's usability is excellent, the tutorial content was rated as relatively moderate.

Table 2 UAT questions and scores

Questions	score
I find the operating process of this system very simple.	4.526
I believe most people can easily access and use this system.	4.579
I think the user interface of this system is clear and straightforward.	3.895
Compared traditional lectures, this system is more engaging for my learning.	3.684
The information presented by this system is fresh and interesting, and the knowledge I learned is very practical.	3.737

6 Future work

First, developing others more popular social media to be the interface of the DKK-robot, e.g. WhatsApp.

Second, while this paper prioritizes the execution framework, further research into prompt engineering is essential to refine the DKK questions' quality.

Third, various applications can be integrated into DKK, e.g. site safety induction for workers.

7 Conclusion

By integrating the LLMs with RAG, Chatbot, and automation, this study has "successfully developed a groundbreaking Disaster Risk Reduction education method". DKK is an advanced evolution of the classic mobile game Knowledge King, distinguished by its automatically and dynamically updating question database. Through a engaging gameplay process, users can effectively learn disaster-related news and practical knowledge from real cases.

While the average cost per question set in DKK is only $0.002, all G-Eval and UAT scores significantly exceed the median threshold. This study proposes a viable framework of DKK, so further research and development are needed to improve its practicality.

References

[1] W. Masocha, N. Takaidza, A. Manyani, and C. Mutseekwa, 'Disaster risk reduction integration into school curriculum: A global analysis,' 2025.

[2] N. Aghaei, H. Seyedin, and H. Sanaeinasab, 'Strategies for disaster risk reduction education: A systematic review,' *Journal of education and health promotion*, vol. 7, no. 1, p. 98, 2018.

[3] G. Biancini, A. Ferrato, and C. Limongelli, 'Multiple-choice question generation using large language models: Methodology and educator insights,' in *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, 2024, pp. 584-590.

[4] H. K. Ilarionov, I. Domanetska, and O. Fedusenko, 'Features of the practical use of LLM for generating quiz,' 2024.

[5] S. Sheikhalishahi, A. Haddadi, S. Sadeghipour, F. Rafiei, and H. Soltani, 'Comparative performance of ChatGPT-4o, ChatGPT-5, and gemini 2.5 flash on Persian internal medicine subspecialty board exams,' *Scientific Reports*, 2025.

[6] Y. Gao *et al.*, 'Retrieval-augmented generation for large language models: A survey,' *arXiv preprint arXiv:2312.10997*, vol. 2, no. 1, 2023.

[7] J. Genesis, 'Retrieval-Augmented Generation: A Comprehensive Survey of Advances and Applications in Deep Learning Models,' *Available at SSRN 5203514*, 2024.

[8] Ö. C. Karacaoğlu, D. Karacaoğlu, and N. Özbaltan, 'Artificial Intelligence Based Solutions in Disaster Literacy Training,' in *10th International Conference on Lifelong Education and Leadership for ALL (ICLEL 2024)*, 2025: Atlantis Press, pp. 28-45.

[9] O. Yetişensoy, 'AI-enhanced disaster education in middle schools,' *Educational Technology & Society*, vol. 29, no. 1, pp. 205-222, 2026.

[10] K. Squire, 'Video games in education,' *Int. J. Intell. Games & Simulation*, vol. 2, no. 1, pp. 49-62, 2003.

[11] J. D. Fletcher, 'Effectiveness and cost of interactive videodisc instruction in defense training and education,' 1990.

Permission Blockchain-BIM Framework For Construction Dispute Prevention

Abdelrahman Madkour[1], Elbethel Muluye[1], and Shiwam Singh[1], Hani Alzraiee[1]

[1]Civil Engineering Department, American University od Sharjah, UAE
b00087378@alumni.aus.edu, g00110271@aus.edu, b00088845@aus.edu, hanialzraiee@aus.edu

Abstract

Construction projects frequently face disputes caused by unclear documentation, conflicting design versions, and accountability gaps among stakeholders. While Building Information Modeling (BIM) enhances collaboration, it currently lacks the immutable audit trails necessary for legal evidence. This paper explores the integration of blockchain technology with BIM to create transparent, tamper-proof systems for dispute administration. This study proposes a Blockchain-Enabled BIM Dispute Resolution (B-BIM-DR) architecture using Hyperledger Fabric. The framework maps contractual dispute clauses to smart contract logic and defines a three-stage data flow Pre-construction, Construction, and Closeout compliant with ISO 19650 information management standards. Two illustrative scenarios, analyzing dispute contexts from the Dubai Creek Tower and Hong Kong International Airport, are utilized to verify the logical consistency of the framework. The proposed architecture demonstrates how automating the "Initial Decision Maker" and "Mediation" phases via smart contracts provides a legally robust mechanism for data-driven conflict management, theoretically reducing the time required for evidence assembly and arbitration.

Keywords –

Keywords: Blockchain, BIM, Dispute Resolution, Smart Contracts, Transparency

1 Introduction

The construction industry is undergoing rapid digital transformation, with Building Information Modeling (BIM) emerging as a central technology for collaboration, visualization, and project coordination. BIM enhances data integration across disciplines, enabling improved design accuracy and real-time decision-making [1]. However, despite these technological benefits, construction disputes remain a persistent challenge, often stemming from unclear documentation, conflicting design versions, and accountability gaps among stakeholders [2]. Globally, construction disputes significantly erode project value. According to the Arcadis Global Construction Disputes Report [3], the average dispute costs approximately USD 42 million and lasts over a year, emphasizing the critical need for transparent and tamper-proof information management systems. While BIM facilitates information flow, it lacks the inherent mechanisms to ensure that project data are immutable, legally verifiable, and traceable across multiple stakeholders [4]. Blockchain Technology (BCT), a decentralized and immutable digital ledger, offers a potential solution to this critical trust deficit. When integrated with BIM (B-BIM), blockchain can record every transaction, design change, and approval as a time-stamped, tamper-proof entry [5], [1]. This integration can revolutionize how project records are managed and how disputes are prevented or resolved [6].

2 Literature Review

The integration of blockchain with BIM, often referred to as B-BIM, is gaining recognition as a transformative approach for achieving transparency, data integrity, and accountability in construction project management. Blockchain's core properties, decentralization, immutability, and distributed consensus, make it ideal for ensuring that data shared across project participants cannot be altered retroactively [7].

In addition to transparency and immutability, various researchers argue that blockchain can be viewed as a trust infrastructure between BIM based collaboration, thus replacing traditional central authorities with cryptographic consensus developed among distributed nodes. As an example, [8] show that specifically, multiparty transactions, accountability requirements, and information asymmetries inherent in construction projects are the areas where a distributed ledger architecture fits best and, therefore, blockchain is a

logical way to handle BIM-related data transfer through the building life cycle.

The latest systematic reviews expand the theoretical framework in the field of BIM-blockchain integration. [12] enter this conversation in the Journal of Construction Engineering and Management with an article in the form of a review that is an introduction to the field of mixed method, called When BIM Meets Blockchain, identifying the main themes of integrations, i.e., information management, payment and procurement, contract administration, and collaboration. According to their findings, BIM and blockchain interaction is not a simple tool convergence, but rather the essential restatement of responsibilities, approval procedures and records keeping, where the latter are encoded as tamper-free transactions pinned on model-based directories.

The view is further supported by [9], who investigate the idea of blockchain integration into the construction sector in a systematic survey of blockchain use and blockchain implementation barriers, discovering that BIM-inspired scenarios can be seen as a specific type, especially in the environment where trust, traceability, and coordination among many stakeholders are essential.

2.1 Early Conceptual and Review Studies on B-BIM

The conceptual and review studies of B-BIM will be presented in the very first chapter of the work. The possibilities of blockchain to enhance the reliability of data and the confidence of stakeholders in construction management were initially described in the early conceptual research by [8]. Later on, [10] conducted a review of its development.

Based on the previous works, [11] provide one of the first systematic reviews of available literature on the adaptation of Building Information Modelling (BIM) and blockchain technology in the conditions of project delivery. Their trajectory of analysis shows that there is a spectrum of application cases, such as smart contracts, supply chain traceability, and records of immutable models, with much conceptual potential, but with a significant lack of standardization and actual application. [12] elaborate on this discussion with the introduction of bibliometric techniques with the help of qualitative analysis, showing that the discourse has shifted, leaving the theoretical to the applications fields that include payments, procurement, and digital twins, but most of the studies are in the proof-of-concept phase.

[13] presented an extensive overview of the implementation of blockchain in the BIM processes and highlights that despite the current academic interest, the prevalent adoption of it in the industry is impeded by legal, organizational and technical concerns. All such initial conceptual and review works provide the theoretical basis of Building-Blockchain Integration Modelling (B-BIM). They thus highlight the need to create more action oriented, dispute resolution-oriented frameworks, including the one presented herein.

2.2 Empirical Applications: Traceability, Procurement, Claims and Payments

More recent empirical studies have given importance to the role of blockchain in terms of enhancing project traceability and automation. As an example, [14] created a document tracking system that uses a blockchain to automatically document any changes in the construction documents and thereby verify any claim. In China, Xiong'an New Area, [15] presented a fund management model, which utilizes blockchain. Their work showed that this model improves financial transparency and auditability. In a similar effort, [16] designed a Blockchain-BIM Automated Procurement System (BBAPS) that eliminates procurement bias by recording every transaction immutably.

The note of the present study is particularly relevant since a corresponding model is the Block-Based Automated Procurement System (BBAPS) proposed by [16] according to which Building Information Modelling (BIM) elements are used to designate procurement items and then utilize blockchain technology and smart contracts to independently assess and award contracts. Every single decision is stored in the blockchain and thus provides a clear and immutable history of purchase operations. The latter is an example of the potential of B-BIM to eliminate conflicts based on the perceived unfairness or vagueness of the awarding criteria.

At the same time, blockchain-based claims and dispute management tools have been developed by the Korea Society of Civil Engineers (KSCE). In the journal of civil engineering, [14] presented an automated construction documentation tracker and extraction based on blockchain with a specific focus on claim and dispute solutions. The system tracks the document transfers via email, creates engine blockchain transactions associated with every relevant event and supports the historical tracking and the authenticity verification. With this contribution, it is represented that blockchain can provide chronological, tamper-evident records that directly help in resolving disputes and legal complaints.

Empirical studies on the automation of payments further highlight the potential of blockchain in the field of BIM to reduce the amount of dispute. [7] presented payment schemes that depend upon smart-contracts to

activate payments upon certified progress measurements using BIM and reality-capture data. These processes reduce one of the most typical bases of building arguments consisting of delayed or contracted payments, by dispensing funds dependent on objectively verifiable computer manifestations.

2.3 Governance, Distributed Common Data Environments (CDEs) and City-Scale Applications

From a governance perspective, [17] proposed blockchain-based frameworks for City Information Modelling (CIM) to enhance data reliability in urban development. [18] further demonstrated blockchain's role in Hong Kong's infrastructure projects, identifying significant opportunities in automated payments and contract enforcement.

[19] suggested a decentralized CDE architecture, which incorporates blockchain and IPFS in the domain of Common Data Environments (CDE). The infrastructure is structured so that storing voluminous BIM files that are controlled by IPFS is separated into the storage of a transaction log on the blockchain, establishing a scalable infrastructure where every model revision and the metadata associated with it can be tracked without unnecessarily burdening the blockchain. Such a strategy is particularly relevant to Building-BIM (B-BIM) dispute resolution since it provides a powerful platform of maintaining the historical record on design modification and approvals on a large scale.

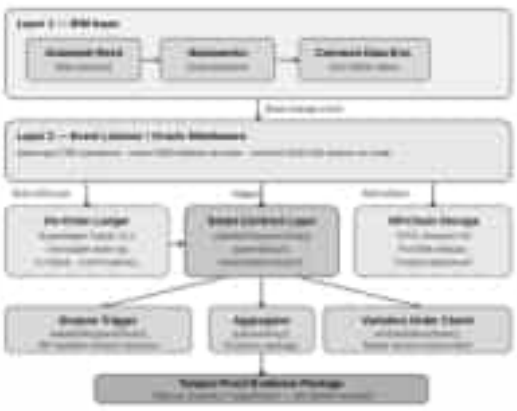


Figure 1. B-BIM-DR System Architecture

The diagram illustrates the three-layer structure: (1) As shown in Figure 1, BIM Input Layer (Autodesk Revit/Navisworks → CDE); (2) Event Listener/Oracle middleware routing large BIM artifacts to off-chain IPFS/S3 while committing SHA-256 hashes to the on-chain Hyperledger Fabric ledger; (3) Smart Contract Layer containing the Dispute Trigger Contract (*validateResponseTime()*), Evidence Aggregator Contract (*queryHistory()*), and Variation Order Verification Contract (*verifyVariationOrder()*), each operating on the immutable distributed ledger.

Besides that, [20] explore the concepts of using blockchain-based documentation systems and Common Data Environment (CDE) solutions to streamline design collaboration and contract administration. They claim that a distributed register can provide security as to the authenticity of the records of transactions carried out in the CDE operations. These architectures come in especially handy with large, multi-party projects, where ownership of data and the necessity to update information is often the source of controversy.

Blockchain-based City Information Modelling (CIM) frameworks at the municipal level can be used to show how these principles can be extended past the projects. According to [17], even a blockchain-based CIM can be used to create an immutable urban history, where approval of regulatory measures, maintenance work, and transfers of property are recorded safely. This type of structure provides a higher level of governance to project level B-BIM systems, especially where there is a dispute with government and long-life infrastructure ownership.

2.4 Dispute-Resolution–Oriented Frameworks and Remaining Gaps

With these developments, blockchain-based BIM is underdeveloped when it comes to resolving construction disputes. Most of the available research focuses on the technical viability, but there is no systematic approach of application. To improve the full potential of blockchain in solving construction disputes, [13] observes that the frameworks between technological capability and the legal and managerial practice do not exists.

Both [11] and [12] claimed that most of the studies on BIM-blockchain are based on the view that, instead of focusing on the dispute and claims workflows, the focus is on digital collaboration on the-large, payment systems, or integrating a supply-chain based solution, but not developing end-to-end workflows designed to serve this purpose. [9] also indicated that in most cases, validation of empirical validations is restricted and those based on piloting or lab-based, and with little information regarding the practical application of blockchain-authenticated records as a subset of arbitration, adjudication or negotiation.

In addition, works like [14], [16], and [19] provided strong backbone elements, which are document tracking underpinnings, automated procurement, and distributed common data ecosystems. However, these are usually provided to isolated functions rather than come up with a

comprehensive dispute-resolution framework. Therefore, it is observed that there still is a very discernible lack of Building-Information-Model-based (BBIM) systems that explicitly:

- Map BIM transaction, and document against legal contracts (marks of notices, instructions and variations).
- Contractual structure Blockchain records in compliance with traditional templates of contracts and dispute-resolution; and
- Prescribe transparent, reproducible workflow, which the project teams can replicate throughout the identification of a problem up to its solution.

Consequently, the academic discourse is narrowed down to the need of dispute-oriented, workflow-oriented B-BIM systems, situated beyond primitive technical prototyping. The Blockchain-Based BIM Dispute Resolution (B-BIMDR) paradigm developed in this project aims at filling this gap by incorporating the principles of BIM, blockchain, smart contracts, and common-data-environment into a consistent, logical approach to ensuring the transparent and impeccable prevention and resolution of disputes associated with actual construction scenarios.

3 Methodology

The proposed Blockchain-Enabled BIM Dispute Resolution (B-BIM-DR) framework is developed using a Design Science Research (DSR) methodology, operationalized through three interlocking DSR cycles: (1) the Relevance Cycle, which grounds the research in the construction industry's identified need for tamper-proof dispute evidence administration; (2) the Design Cycle, which iteratively specifies and refines the B-BIM-DR architecture artifact through conceptual design and logical evaluation; and (3) the Rigor Cycle, which ensures the design builds upon and contributes to the established knowledge base of BIM-blockchain integration [12][21]. This structured DSR approach ensures that the resulting artifact is both practically motivated and theoretically grounded. This section details the system's architectural design, its three-stage implementation model, the logical mapping of contractual rules to automated smart contract code, the technical network specification, and the conceptual validity boundaries of the framework.

3.1 Research Design and Framework Foundation

This study employs a conceptual framework development methodology, rigorously structured to align with the architectural specification protocols used by [21]. The conceptual approach is justified by the current state of the field, which recommends design specification following comprehensive reviews of BIM-Blockchain integration [12].

3.2 Framework Implementation Stages:

To ensure systematic management across the project duration, the B-BIM-DR architecture is structured into three discrete, chronologically linked stages, directly adapting the modular model proposed by [22]:

- **Stage 1:** Pre-construction (Rules Definition): The core contractual rules and project-specific Endorsement Policies are parameterized and translated into the initial smart contracts (chain code). This stage codifies the consensus mechanisms and establishes deterministic thresholds (e.g., RFI response deadlines) that will programmatically trigger potential breach events.
- **Stage 2:** Construction (Immutable Data Provenance): The B-BIM system operates as a secure logger for the Common Data Environment (CDE). Every critical transaction (RFI submission, Change Order, Model Version Update) is processed, hashed, and anchored to the ledger. This process ensures all digital artifacts adhere to the data transition states required by ISO 19650.
- **Stage 3:** Closeout and Dispute Resolution (Automated Adjudication): If a formal claim is initiated, the system automatically triggers the Evidence Aggregator Contract. This function instantaneously assembles the complete, timestamped, and tamper-proof evidentiary package from the ledger history, by passing the manual discovery process.

3.3 Architectural Components and Logic Mapping

The B-BIM-DR framework is built upon a hybrid on chain/off-chain data strategy to resolve the conflict between the large file size of BIM models and the scalability limitations of blockchain, while strictly adhering to legal admissibility standards.

3.3.1 Hybrid Data Flow Architecture (BIM Layer & Blockchain Layer)

The BIM layer (Input) leverages tools like Autodesk Revit and Navisworks as the data source. When information transitions within the CDE (Input), an Event Listener/Oracle intercepts the change. The Blockchain Layer (Processing and Immutability) then performs the following:

- Heavy BIM artifacts are sent to **Off-Chain Object**

Storage (e.g., IPFS/S3).

- The resulting **cryptographic hash (SHA-256)**, file Content ID, and minimal metadata (Timestamp, User ID) are committed to the **On-Chain Ledger**.

This system ensures that the hash acts as an immutable anchor, guaranteeing the file's integrity and authenticity [19].

3.3.2 Smart Contract Layer (Automation Logic)

The smart contract layer facilitates proactive dispute prevention by programmatically enforcing contractual clauses following the mapping methodology established by Faraji et al. [21]. This approach enables advanced automated functions similar to those demonstrated in automated procurement systems [16].The key functions within this layer include:

- Dispute Trigger Contract (*validateResponseTime()*): Automatically creates a formal dispute record if an RFI response exceeds a contractual deadline. The function accepts three inputs: *RFI_ID* (string), *IssueDate* (Unix timestamp), and *ContractualDeadline* (integer, days; default = 7 per FIDIC Red Book Clause 4.20) [25]. Processing logic: the function computes elapsed time by subtracting *IssueDate* from the current block timestamp. If elapsed time > *ContractualDeadline*, the function (1) writes a *BreachEvent* record to the ledger containing {*RFI_ID*, *IssueDate*, *BreachTimestamp*, *ContractorID*}; (2) sets the RFI status flag to '*DISPUTED*'; and (3) returns the *BreachTimestamp* as cryptographically immutable evidence. Given the assumed 5-second block time, breach events are anchored within one block interval, reducing evidentiary latency from days of manual discovery to seconds.
- Evidence Aggregator Contract (*queryHistory()*): Upon dispute initiation, the function accepts *DocumentHash* (SHA-256 string) as its sole input and returns an ordered array of all on-chain transactions linked to that hash. Each record includes: *BlockHeight*, Timestamp (Unix), *UserID*, *EventType* (e.g., RFI_SUBMITTED, BREACH_FLAGGED, MODEL_APPROVED), and the IPFS Content Identifier (CID) linking to the full off-chain BIM artifact. The resulting output constitutes a tamper-proof, court-admissible evidentiary package meeting BS 10008 authenticity requirement. With a 10 MB block size and 32-byte SHA-256 hash, the system can anchor approximately 327,680 document hashes per block, eliminating throughput bottlenecks on large projects [26].

Table 1 details the core Chain code modules designed to monitor and flag common construction challenges, establishing the necessary technical link to legal and information standards:

Table 1: Mapping Construction Challenges to Blockchain BIM Logic and Standards

Challenge	Contractual Role	Chaincode Function	Data Fields	Evidence Artifact	Legal/Evidence Standard
Late RFI Responses	Response required within 7 days (e.g., FIDIC Red Book).	validateResponseTime()	RFI_ID, IssueDate, RespondedID, Due_Date	Time-stamped 'Bench' Transaction Hash	ISO 19650-2 (Information Delivery Cycle)
Unverified Design Changes	Variations must be formally approved by the Project Engineer.	verify(VariationOrder)	ElementGUID, Status:APPROVED, SignerKey	Digitally Signed Block Header	BS 10008 (Evidential Integrity & Authenticity)
Tampered Audit Logs	Project records must remain unaltered for liability periods.	commitAnchorHash()	MerkleRoot, PreviousHash, Timestamp	Cryptographic Receipt (SHA-256)	BS 10008 (Electronic Identity & Integrity)

3.4 Technical Specification and Test Network Assumptions

To contextualize the B-BIM-DR framework, a permissioned blockchain environment simulating a Hyperledger Fabric deployment is assumed. The assumed network parameters and their logical implications are as follows. Platform: Hyperledger Fabric v2.x. Participants: four primary organizations (Owner, Engineer, Contractor, Subcontractor), each operating a peer node. Channels and Policies: a single Dispute Resolution Channel governed by a 2-of-4 endorsement policy, requiring signatures from at least two organizations before any transaction is committed. Performance Assumptions: 10MB maximum block size, 5-second block creation time. Chain code/SDK: Node.js SDK. These parameters support the following quantitative logical inferences.

- Evidentiary latency: breach events flagged by *validateResponseTime()* are anchored to the immutable ledger within 5 seconds (one block interval), compared to days or weeks of conventional manual discovery that is a reduction of at least four orders of magnitude.
- • Throughput capacity: a 10 MB block can anchor ~327,680 SHA-256 document hashes per block, confirming zero throughput bottleneck even on large, document-intensive projects such as the HKIA Expansion [23][26].
- Consensus integrity: a 2-of-4 endorsement policy ensures no single participant can unilaterally alter ledger state, satisfying the multi-party authentication requirements of BS 10008.

These inferences are grounded in the specified network parameters (platform: Hyperledger Fabric v2.x; block size: 10 MB; block time: 5 s; endorsement: 2-of-4;

SDK: Node.js) and are clearly labelled as assumed rather than measured, preserving scientific integrity.

3.5 Threats to Conceptual Validity

While the B-BIM-DR framework offers a detailed conceptual model, three specific threats to its conceptual validity are acknowledged. First, the Threat of Non-Empirical Data: the framework's efficacy is validated through illustrative scenarios rather than live deployments. Future work should pursue pilot implementations on actual projects to generate empirical performance data. Second, the Threat of Technical Simplification: the framework is intentionally scoped to objective, quantifiable contract clauses such as deadline-triggered RFI responses and model version endorsements. However, many construction contract clauses involve inherently subjective assessments, including judgements of "reasonable time", "fitness for purpose", or force majeure conditions, which cannot be encoded into deterministic smart contract logic without significant interpretive ambiguity. To address this, the B-BIM-DR framework adopts a hybrid governance model: clauses that are objectively quantifiable are enforced autonomously by the smart contract, while clauses requiring subjective interpretation are flagged by the contract and escalated to a designated human adjudicator within the Hyperledger Fabric channel, preserving legal oversight where automation is insufficient. Third, the Threat of Generalization: the framework is specified for a permissioned Hyperledger Fabric environment; adaptation to public blockchains or alternative DLT platforms would require re-evaluation of the endorsement policy model and confidentiality controls

4 Illustrative Scenarios and Conceptual Evaluation

4.1 Scenario 1: Dubai Creek Tower – Reactive Verification

The Dubai Creek Tower project, a complex, multi-billion dollar collaboration, frequently encountered high-stakes disputes related to the sequencing of design approval signoffs and resultant schedule delays. In a conventional CDE environment, proving accountability for these delays requires significant effort. The need for a robust blockchain solution for verifiable document tracking for claims is reinforced in literature.

1. Conventional Workflow Challenge

In a conventional CDE, when a contractor files a delay claim alleging a late RFI response, the legal process involves:

- **Manual Discovery:** Legal teams search through months of emails, server logs, and CDE audit trails (often weeks of effort).
- **Data Fragility:** Logs can be edited or lack cryptographic certainty. Establishing the exact moment the RFI was submitted and the exact moment the deadline was breached requires relying on human testimony or potentially inconclusive metadata.
- **Cost:** The preparation of evidence for arbitration typically costs tens of thousands of dollars and significantly extends the dispute resolution cycle.

2. Theoretical Application and Solution

The B-BIM-DR framework is conceptually applied to the RFI process. When the contractor submits an RFI, the Event Listener captures the action. The Blockchain Layer performs the following immutable transaction:

- **Transaction Data:** The transaction committed to Hyperledger Fabric includes the RFI document's hash (*H1*), the User ID, and the *IssueDate*.
- **Smart Contract Action:** The Dispute Trigger Contract (the function *validateResponseTime()* from Table 1) is activated. It creates a secondary ledger entry defining the immutable *Due_Date* (e.g., *IssueDate* + 7 days, as per the FIDIC Red Book standard).

3. Outcome: Post-Dispute Verification

- **Resolution:** When the dispute is formalized, the Evidence Aggregator Contract is triggered. Instead of launching a manual search, the contract instantaneously queries the ledger for all transactions linked to *H1*.
- **Admissibility:** The system provides an irrefutable package containing the initial RFI hash, the immutable *IssueDate*, and the automated "*Breach Event*" transaction timestamped after the *Due_Date*. Because every event is cryptographically anchored, the evidence satisfies the evidential integrity and authenticity requirements of BS 10008.
- **Impact:** This mechanism bypasses manual discovery, theoretically reducing the dispute resolution cycle from weeks to minutes [8]. The evidence is ready for adjudication, demonstrating the framework's superior reactive evidence verification capability.

4.2 Scenario 2: Hong Kong International Airport (HKIA) Expansion – Proactive Prevention

The $18 billion HKIA Expansion involves numerous

contractors and complex digital twin requirements, creating a chronic risk of "conflicting version" disputes where parties rely on outdated BIM models for execution and claims [23]. The goal is to enforce the Single Source of Truth (SSOT) proactively. The concept of using smart contracts for automated workflow validation is a key trend, including systems for automatic procurement [16].

1. Conventional Workflow Challenge

A contractor performs work based on an unverified or outdated BIM model (e.g., Model Version 5), then submits a payment claim.

- **Rework Risk:** When the engineer discovers the discrepancy against the official Model Version 6, the contractor is forced into costly rework, leading to a high-value Variation Order dispute.
- **Verification Time:** The verification process requires multiple stakeholders to manually compare model versions, which severely impacts the payment and project delivery schedules.

2. Theoretical Application and Solution

The B-BIM-DR framework operates as a unified, trust-based CDE. All model version updates must pass an Endorsement Policy (e.g., approval from the Design Manager and Project Engineer) before the new hash (*H6*) is anchored as *Status:PUBLISHED*.

- **Smart Contract Enforcement:** The system employs the Smart Contract Layer to enforce the ISO 19650 information delivery cycle. The contract includes the function *verifyVariationOrder()* (from Table 1), which is designed to prevent noncompliant transactions.
- **Proactive Veto:** When the contractor attempts to submit a claim referencing the outdated Model Version 5 (Hash *H5*), the smart contract automatically checks the ledger's state. It discovers that *H5* lacks the necessary *Status:PUBLISHED* endorsement.

3. Outcome: Risk Mitigation and Workflow Acceleration

- **Dispute** Prevention: Instead of allowing the claim to proceed to a dispute, the smart contract instantaneously rejects the transaction at the point of submission. The system informs the contractor that the reference hash is invalid for payment calculation. This action forces immediate compliance with the SSOT defined by the ledger.
- **Impact**: This demonstrates the system's shift from reactive dispute resolution to preventive governance. By programmatically enforcing the contractual requirement for the most current, verified data, the framework eliminates the opportunity for conflicting claims, resulting in a measurable acceleration of both verification and automated payment processes [18].

5 Discussion

The primary academic contribution is the detailed Conceptual Framework Specification (Section 4), particularly the Smart Contract Logic Mapping (Table 1). This mapping programmatically links contractual obligations to technical enforcement functions (e.g., *validateResponseTime()*), providing a rigorous model for digitalizing contract administration [21]. The resulting integration produces an unquestionable audit trail of time-stamped, unchangeable documentation of design modifications and milestone confirmations [1].

The conceptual evaluation via illustrative scenarios (Section 5) validated the framework's core utility: enabling a necessary shift from reactive, expensive evidence discovery to proactive governance. The framework was shown to enable near-instantaneous evidence verification in reactive scenarios (Dubai Creek Tower, supported by specialized claims tracking research, [24]. Furthermore, the integration of smart contract-based automated compliance procedures for claims and payment verification [16] was demonstrated to proactively avert disputes by enforcing the Single Source of Truth (HKIA Expansion, [23].

5.1 Compliance and Digital Governance Implications

The B-BIM-DR framework provides a critical mechanism for achieving compliance with international digital governance standards. By leveraging the ledger's immutable state, the system effectively maps the BIM workflow to the ISO 19650 information states: data only transitions from Work in Progress (WIP) to Shared/Published when its hash is immutably anchored and endorsed on the blockchain. Furthermore, blockchain features such as hashing, time-stamping, and non-repudiation ensure audit entries meet BS 10008 standards for integrity and authenticity, establishing B-BIM-DR as a trusted digital mechanism.

6 Conclusion

This study addresses data fragility and lack of transparency in construction dispute resolution by proposing the Blockchain-Enabled BIM Dispute Resolution (B-BIM-DR) framework. Using a Design Science Research approach, it introduces a three-stage architecture based on Hyperledger Fabric, with a hybrid on-chain/off-chain strategy that securely stores cryptographic hashes of BIM data to ensure tamper-proof evidence while maintaining scalability.

The framework's key contribution is mapping FIDIC

contract clauses into smart contract logic, demonstrated through scenarios of dispute verification and prevention. Results suggest significant efficiency gains, such as reducing breach detection time to seconds and supporting large-scale data handling. Limitations include its conceptual nature and focus on quantifiable clauses, with future work aimed at real-world testing and expanding contract coverage.

References

[1] Nawari, N. O., & Ravindran, S. (2019). Blockchain and the built environment: Potential applications and benefits. Journal of Building Engineering, 25, 100832.

[2] Wang, J., Li, Z., Sun, Y., & Wang, K. (2022). Research on multi-person collaborative design of BIM drawing based on blockchain. Scientific Reports, 12, 16312.

[3] Arcadis. (2024). Global construction disputes report 2024. Arcadis.

[4] Muhammad, R., & Nasir, A. R. (2022). Integrating BIM in construction dispute resolution: Development of a contractual framework. Buildings, 12(11), 1828.

[5] Celik, Y., Petri, I., & Rezgui, Y. (2023). Integrating BIM and blockchain across construction lifecycle and supply chains. Computers in Industry, 148, 103886.

[6] Mahmudnia, D., Arashpour, M., & Yang, R. (2022). Blockchain in construction management: Applications, advantages and limitations. Automation in Construction, 140, 104379.

[7] Hamledari, H., & Fischer, M. (2020). Construction payment automation using blockchain-enabled smart contracts and reality capture technologies.

[8] Turk, Ž., & Klinc, R. (2017). Potentials of blockchain technology for construction management. Procedia Engineering, 196, 638–645.

[9] Celik, B. G., Alalwan, M., Hameed, M., & Al-Qassim, A. (2024). A systematic review of applications and barriers of blockchain in construction. Buildings, 14(6), 1600.

[10] Kuperberg, M., & Geipel, M. (2021). Blockchain and BIM: Progress in academia and industry. arXiv.

[11] Chong, H.-Y., & Diamantopoulos, A. (2020). Integrating BIM and blockchain for project delivery: A systematic review. Automation in Construction, 118, 103–148.

[12] Xu, Y., Chi, M., Chong, H.-Y., Lee, C.-Y., & Chen, K. (2024). When BIM meets blockchain: A mixed-methods literature review. Journal of Civil Engineering and Management.

[13] Hany, A. (2023). Potential of adopting blockchain technology with BIM implementations in the construction industry ecosystem: A review paper. European Journal of Civil Engineering and Architecture. (Online ahead of print)

[14] Kim, E. W., Lee, H.-S., Lee, Y.-S., Lim, Y.-K., & Cho, Y. (2024). Blockchain-based automatic tracking and extracting of construction documents for claims/disputes. KSCE Journal of Civil Engineering, 28(5), 3707–3724.

[15] Song, W., Wu, H., Meng, H., Bian, E., Tang, C., Xi, J., & Zhu, H. (2023). A blockchain-based fund management system for construction projects: Case study in Xiong'an

[16] Yoon, J. H., Aurangzeb, I., & McNamara, S. (2024). BIM- and blockchain-enabled automatic procurement system (BBAPS) removing relationship bias: Proof-of-concept.

[17] Lawal, O., & Nawari, N. O. (2023). Blockchain and City Information Modeling (CIM): A new approach to transparency and efficiency. ITcon, 28, 711–734.

[18] Lu, W., Wang, H., & Zhang, Z. (2022). Application of blockchain-integrated BIM in Hong Kong infrastructure projects: Opportunities and challenges. Journal of Construction Engineering and Management, 148(4), 04022009.

[19] Tao, X., Das, M., Liu, Y., & Cheng, J. C. P. (2021). Distributed common data environment using blockchain

[20] Das, M., Tao, X., Liu, Y., & Cheng, J. C. P. (2022). A blockchain-based integrated document management framework for construction applications. Automation in Construction, 133, 104001.

[21] Faraji, A., Arya, S. H., Ghasemi, E., Rashidi, M., Perera, S., Tam, V., & Rahnamayiezekavat, P. (2023). A conceptual framework of decentralized blockchain integrated system based on building information modeling for digital administration of disputes in IPD contracts. Construction Innovation, 24(1), 384–406.

[22] Matarneh, S. T., Elghaish, F., Por Rahimian, F., & Dawood, N. (2022). Automated construction payment using blockchain-enabled smart contracts and BIM. In EGICE: Intelligent Computing in Engineering (Conference proceedings).

[23] Wang, H., Li, V., & Wu, P. (2021). Blockchain-based common data environment for Hong Kong International Airport expansion: A digital trust framework. Auto. in Construction, 129, 103827.

[24] Hyperledger Fabric documentation. (n.d.). Hyperledger Fabric docs.

[25] FIDIC, Conditions of Contract for Construction (Red Book), 2nd ed. Geneva, Switzerland: Federation Intern. des Ingenieurs-Conseils, 2017.

[26] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2008.

Robot-Oriented Lean Construction: A Toolkit Primer and the Pathway Forward

Rongbo Hu[1], Hiroki Yajima[2], Jiazhen Mao[3], Keiji Tanaka[1], Soungho Chae[1] and Thomas Bock[4]

[1]Kajima Technical Research Institute Singapore, Kajima Corporation, Singapore
[2]Kajima Kajima Overseas Asia (Singapore) Pte. Ltd., Kajima Corporation, Singapore
[3]Shanghai Weibuild Technology Co. Ltd., China
[4]School of Engineering and Design, Technical University of Munich, Germany
r.hu@kajima.com.sg, h.yajima@kajima.com.sg, jiazhen.mao@we-i-build.com, k.tanaka@kajima.com.sg, s.chae@kajima.com.sg, thomas.bock@tum.de

Abstract –

The construction industry faces persistent economic, social, and environmental challenges. Construction robotics and lean construction have emerged as potential solutions, yet their integration remains underexplored. This paper introduces the notion of Robot-Oriented Lean Construction (ROLC) – a toolkit that adapts lean principles to the characteristics of construction robots, emphasizing synergy between people, processes, and technology. Drawing on a case study of the research, development, and deployment of construction robots in Singapore, the paper demonstrates how lean methods such as Deming cycle, Last Planner System, 6S, Gemba Walk, and Total Productive Maintenance can guide the planning, development, deployment, and continuous improvement of construction robots. Findings with empirical evidence indicate that the toolkit enhances safety, efficiency, and collaboration, while reducing waste and highlighting respect for people. The paper concludes by outlining pathways for development, education, and standardization to advance the emerging research direction of ROLC.

Keywords –

Construction Robotics; Continuous Improvement; On-Site; Singapore; Toyota Production System

1 Introduction

The construction sector plays a crucial role in the global economy. Yet, it is constantly facing enormous economic, social, and environmental challenges. Economically, the sector shows slower productivity growth than any other industry, with low levels of digitization and lengthy planning and execution phases [1]. Socially, it faces a widespread shortage of skilled labor, worsened by hazardous working conditions and an aging population [2]. Environmentally, it is one of the most resource-intensive industries, making it central to efforts in combating climate change and achieving a circular economy [3]. These persistent challenges in the construction sector are largely caused by on-site production, the uniqueness of each project, and the high level of complexity involved [4]. As a result, the construction industry is in urgent need of innovative technologies and methods to tackle these challenges. Construction robotics and lean construction have emerged as promising potential solutions.

1.1 The Coming Ubiquity of Robots in the Construction Industry

Emerging in the late 1970s in Japan through the leading efforts of general contractors known as the "Big Five", construction robots began tackling major challenges in the industry by enhancing working conditions, safety, productivity, and quality [5]. Over the past decade, demand for construction robots has resurged, driven by government initiatives in regions like Hong Kong and Singapore that encourage the use of emerging technologies in the construction sector [6], [7]. Researchers have reported many advantages of robotic construction compared to manual work [8], [9]. However, there are still barriers to the full implementation of construction robots in the construction industry mainly due to fragmented nature of the industry, high initial investment, low level of standardization, resistance by workers and trade unions, lack of expertise, and lack of support from top management [10], [11].

1.2 The Evolution of Lean Construction

For the past fifty years, Western manufacturers have looked to Japanese manufacturing management practices as a benchmark, recognizing them as an innovative production paradigm widely known as lean manufacturing [4]. One of the most notable examples of lean manufacturing was Toyota Production System (TPS)

invented by T. Ohno who leveraged just-in-time and autonomation to eliminate waste and increase value for customers, while maintaining high quality [12].

Inspired by lean manufacturing, the construction industry expectedly adopted lean methods and formed the lean construction theory. It is generally believed that "lean construction" was coined by the International Group for Lean Construction (IGLC) in 1993 to describe the methods of planning and executing construction tasks that reduces waste in materials, time, and effort, while maximizing cost-efficiency [13]. Lean Construction Institute (LCI), an organization formed in 1997 by G. Ballard and G. Howell to develop and disseminate knowledge of lean construction, further refined its definition as "a culture of continuous improvement and collaboration among all project participants to maximize efficiency, minimize waste, and deliver exceptional project outcomes" [14]. Lean construction, ever since its advent, has continuously delivered a wide range of economic, social, and environmental benefits to the construction industry [15]. Over the years, many tools have been developed and successfully applied to lean construction including but not limited to Deming Cycle, 5S, Big Rooms, Daily Huddles, Kanban, Andon, Poka-Yoke, Heijunka, and Last Planner System [4], [16]. However, to continue advancing and benefiting the construction industry, lean construction must adapt to the transformations driven by Industry 4.0 while preserving its foundational focus on the synergy between people, processes, and technology [17].

1.3 The Crossroads

We are reaching a crossroads where two major emerging fields in construction, namely construction robotics and lean construction, need to synergize to further tackle the enduring challenges in the construction industry. Lean construction practices should consider the upcoming ubiquitous application of robotics in the construction industry [5]. In recent years, the Lean Construction 4.0 framework has emphasized integrating lean construction principles with advanced technologies like robotics [17]. Based on scientometric analysis, Cardenas et al. advocate for more synergies in the intersection of lean construction and construction robotics [18]. In a recent study, Hu et al. further reported that construction robots can potentially bring at least 17 lean related benefits to construction sites [7]. These studies indicate a strong potential for integrating lean construction with construction robotics.

As a highly developed city-state, Singapore has a strong economy and plays a unique role as a center of finance and innovation in Southeast Asia, bridging the East and West. The construction industry is one of its key sectors, employing over 450,000 workers [19]. Like many other countries, Singapore's construction sector faces many challenges such as stagnant productivity, construction waste, safety risks, labor shortages, and a heavy reliance on foreign workers [20], [21]. These issues were further intensified by the COVID-19 pandemic [19]. In response, Singaporean government agencies have introduced various incentives to foster both local and international innovation in the construction industry. This makes Singapore an ideal place for testing novel construction technologies including lean construction and construction robotics.

2 Methods

This paper aims to introduces the notion of Robot-Oriented Lean Construction (ROLC) and provide empirical evidence of how lean construction methods can be adapted to suit the characteristics of construction robots with a focus on the synergy between people, processes, and technology through a best-practice case study in Singapore. The study is primarily based on a case study of the planning, development, and deployment activities around construction robotics in the Construction Robotics Lab. Furthermore, a unique feature of this lab is its close link to several ongoing construction projects across Singapore, serving the ideal live testbed for the technologies developed in the lab environment or co-developed with partners. In this study, qualitative methods such as on-site observations, case study, and expert interviews are adopted. Specifically, various lean methods applied in the planning, development, and deployment of construction robots are shown in Figure 1.

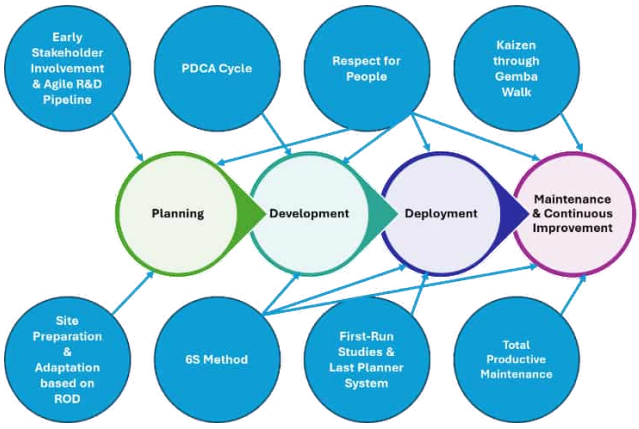


Figure 1. Non-exhaustive lean tools applied during construction projects' lifecycle

3 Results

The following sections will explain how these methods are applied in the case study in detail.

3.1 Planning

The lean methods applied in the planning phase of robots are reported as follows.

3.1.1 Early Stakeholder Involvement and Agile Research-Development-Deployment Pipeline

To begin the adoption process of robots on the construction sites, an agile research-development-deployment pipeline is adopted, involving all key stakeholders in the early stage (see Figure 2). The process begins with gathering requirements from the clients or general contractor affiliated departments, followed by assessing the availability of suitable technologies. If the target technology is already available and mature on the market, a market search is conducted, and collaboration with an external vendor is initiated to adapt the technology further to the site condition. If the technology is not available, R&D projects will be carried out in the Construction Robotics Lab to develop the technology (see Figure 3). In both cases, the developed or adapted technology is tested on affiliated construction sites. If the performance is found to be unsatisfactory, further improvements or iterations are made. Once the technology demonstrates satisfactory performance, it is deployed for use on real construction sites.



Figure 2. Research-develop-deploy pipeline for implementing construction robots



Figure 3. Construction Robotics Lab in Singapore

As shown in Figure 4, several construction robots including wall plastering robot, interior painting robot, floor grinding robot, mobile monitoring robot, and site inspection drone are deployed on the construction site of this case study. Methods such as Choosing by Advantages (CBA) workshops where key stakeholders quantitatively assess the advantages of all options by weighing the key criteria [22] are usually adopted to decide which technology to adopt for each task.



Figure 4. Examples of robotic solutions for various tasks tested for deployment on the real site

3.1.2 Site Preparation and Adaptation

The notion of Robot-Oriented Design (ROD) was first proposed by T. Bock, 1988, who emphasized that the final on-site construction operations shall already be considered at the design stage of building elements and construction methods in accordance with the characteristics of robots to improve their efficiency and usability [23]. This concept has also been applied to the site preparation process in this case study for better operation of robots. For example, a mobile monitoring robot is deployed on site to patrol in different rooms and spaces for site management. Barricades are modified so that the mobile robot's lidar sensor can detect which barricades are passable and which are not (see Figure 5). Also, slopes are added to the steps on the route of the robots to mitigate any height gaps.

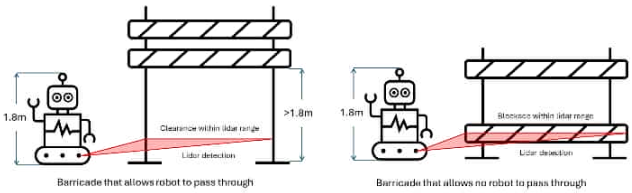


Figure 5. Example of site preparation and adaptation for implementing construction robots

Another common challenge for construction sites to adopt robots is network connectivity. Researchers have already discussed the strategy for deploying 5G infrastructure on construction sites [24]. However, connectivity may be limited in remote or underground locations where 4G/5G networks are unstable or unreachable. In such cases, it is necessary to assess network speed and signal strength in advance using tools

such as Opensignal. Based on the results, the site team may need to establish a wireless mesh network (WMN) to support the smooth operation of multiple robots.

3.1.3 Respect for People

As a key pillar of lean construction, respect for people is implemented throughout the construction phase to ensure smooth and harmonious collaboration between teams. For example, all signs on site including the ones on the robots are written in English, Chinese, Tamil, and Bengali, considering the demographic composition of the workers (see Figure 6).



Figure 6. Multilingual signage deployed on site

Furthermore, because Singapore is a multicultural "melting pot" where many ethnic groups coexist peacefully side by side including but not limited to Chinese, Malays, Indians, and other minority groups, the construction workers' faiths are also respected on the construction site. Figure 7 shows a Muslim prayer room and a Chinese shrine installed on the construction site for workers of different beliefs to practice their religions.



Figure 7. Prayer room and Chinese shrine on the construction site for workers of different faiths

3.2 Development

The lean methods applied in the development phase of robots are reported as follows.

3.2.1 PDCA Model for Developing Construction Robotic Applications

The development of construction robots typically involves various stakeholders from the construction sector, non-construction sectors, and investors. A domain-specific systematic design management approach helps integrate the needs and goals of different stakeholders and team members during the development process. Therefore, the construction robot development in this case study follows a procedure model derived from the Deming Cycle designed to provide comprehensive guidance on how to systematically conceptualize requirements and how to design, develop, implement, evaluate, manage, and refine the construction robots (see Figure 8).



Figure 8. Procedure model for developing construction robots (adapted from [25])

Specifically, its foundation follows the Deming cycle for continuous improvement (Layer ③), enabling the cyclic and iterative development process to begin at any stage and to be applied not only to new concepts but also to existing and even relatively mature construction robot designs. The central components of the procedure model are the four main process steps (Layer ②) located at its core, namely requirements engineering, development, implementation, and performance evaluation. Each of these steps consists of a defined set of activities, tasks, and specified requirements (Layer ④). The primary aim of the proposed process is the continuous advancement of the system architecture and progression of its Technology Readiness Levels (TRLs) (Layer ①). This development procedure model has been validated and proven feasible in multiple R&D projects. The procedure model's core principle allows it to evolve with each usage cycle. Furthermore, customized elements can be inserted into the model based on region, robot typology, and technology maturity [25].

3.2.2 6S Method Emphasizing Safety

Derived from 5S, 6S is workplace organization method which refers to "sort", "set in order", "shine", "standardize", "sustain", and "safety" [26]. In the lab, 6S is strictly implemented due to high risks of robots, where actions such as risk management implementation plan,

housekeeping, occupational health and safety (OHS) workshop are conducted periodically. On the construction site, 6S is also implemented with an emphasis on collaborating safely with robots on site. Workers must receive safety training and attend morning safety meetings. Moreover, incentives such as lucky draws for speaking up for safety and Why I Work Safe Cards are provided for workers on site to further strengthen their safety awareness (see Figure 9).

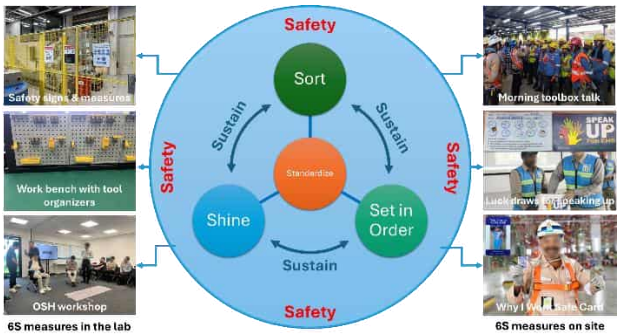


Figure 9. Use of 6S Method during R&D and deployment phases of construction robots

3.3 Deployment

The lean methods implemented in the deployment phase of robots are reported as follows.

3.3.1 First-Run Studies Before Robot Deployment

First-run study is a lean technique used to optimize a work process by conducting a controlled experiment to evaluate performance and uncover potential issues [27]. Because of high deployment cost, robotic technologies cannot be directly deployed before proven feasible by first-run studies.

For instance, the site team considered deploying a novel floor grinding robot. However, due to the robot provider's start-up nature, how different parameters of the robot settings will affect the task performance and quality was still unclear, especially for the concrete used on the construction site in Singapore (i.e., C40). With assistance from the R&D team, a first-run study was designed using a room of 2000 m^2 to evaluate the performance of the concrete grinding robot. As there are two main parameters in the robot's settings, namely robot moving speed (mm/s) and diamond grinding wheel's rotational speed (rpm), the first 500 m^2 was divided into 9 equal areas to test 9 combinations of both parameters. Two 3D scans were performed respectively before and after the test grinding to compare the height differences. After the test, the parameters of the A1 area were considered efficient, which was applied to the rest of the room as well as other areas for the floor grinding task (see Figure 10).

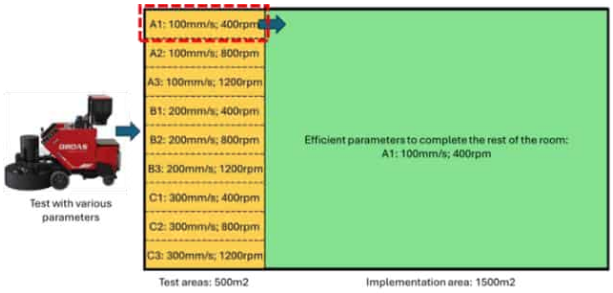


Figure 10. First-run study design for the floor grinding robot before implementation

3.3.2 Last Planner System for Robot Deployment

Due to their consistent performance, robots can help optimize construction processes and support the application of on-site lean construction. In this case study, robots were integrated into project planning through a flexible Last Planner System [28], with their deployment as key milestones integrated in the master schedule. Weekly work plans were developed using pull planning, supported by site preparation to ensure readiness. Daily coordination meetings between the robot operators and site team defined work scopes, while activity reports were submitted by operators and shared with stakeholders to enable systematic data collection and continuous improvement.

3.3.3 Evaluation Framework

After the first deployment, the R&D team developed a comparison and evaluation framework comprising quantitative and qualitative methods to measure the implications of the construction robots for various stakeholders on micro, meso, and macro levels based on the data collected in the deployment phase. In a case study, by evaluating the performance of wall plastering robot and interior spray painting robot using the evaluation framework, the R&D team reported that the robots have the potential to bring at least 17 lean related benefits that can reduce waste, increase value, and improve collaboration [7]. Specifically, for example, the plastering robot recorded a 462.5% productivity gain and 95.4% quality improvement (i.e., hollow rate reduction) compared to the conventional plastering method, whereas the spray painting robot achieved 28.6% productivity gain and 39.6% material saving [7].

3.4 Maintenance and Continuous Improvement

The lean methods used in the maintenance phase of robots are reported as follows.

3.4.1 Total Productive Maintenance

Total Productive Maintenance (TPM) is an equipment management method based on total

participation to maintain equipment to improve productivity, quality, and safety [29]. Due to its sophistication, all critical robotic equipment, whether in the lab (e.g., robotic arms, air compressors, automated guided vehicles, etc.) or deployed on site (illustrated in Figure 4), undergoes TPM periodically depending on its usage. TPM deeply connects to 6S methodology which ensures optimal total cost-efficiency, balancing between high preventive cost and high repair cost.

3.4.2 Kaizen Through Gemba Walk

Kaizen (Japanese term for "change for the better") is a lean philosophy that aims to make incremental, continuous improvements for long-term success. It is deeply rooted in the philosophy of Hansei (Japanese term for "self-reflection"), referring to deep introspection to identify own mistakes and pledge improvements. Gemba (Japanese term for "real place") Walk is a lean management term referring to physically going to the real site where leaders can identify areas for improvement and cultivate a culture of continuous learning [30]. On one hand, safety Gemba Walk by professional safety inspectors has already been implemented on a regular basis to ensure a safe working environment both in the lab and on site. On the other hand, Gemba Walk also plays a crucial role in the successful on-site implementation of construction robots.

For example, during a Gemba Walk session, the R&D team found two practical flaws of the spray painting robot: 1) Due to the inconsistent quality of certain brands of paint, solid impurities in the paint may clog the robot's nozzle, forcing the operator to stop and manually clean the nozzle before continuing to operate; 2) In some occasions, the cable carrier track of the vertical robotic movement which is slightly higher than the robotic arm may accidentally hit some fixtures hanging on the ceiling during operation, causing potential damage to building fixtures (see Figure 11). After the discovery of these two flaws, the R&D team requested the operator to provide feedback to the robot manufacturer, who decided immediately to improve the robot design in the next iteration by 1) adding a filter net in the paint feeding system and 2) flipping the cable carrier track to lower than the robotic arm.



Figure 11. Discovery of flaws during Gemba Walk for continuous improvement

4 Discussion

In this section, implications of this research are summarized. Furthermore, limitations are analyzed and future research directions are suggested.

4.1 Implications

This paper introduces ROLC as a practical guideline for integrating lean principles with the growing adoption of construction robotics. Through a case study in Singapore, it demonstrates how lean tools such as PDCA, Last Planner System, 6S, Gemba Walk, and TPM can guide the planning, development, deployment, and continuous improvement of construction robots while emphasizing respect for people, which resonates with the Lean Construction 4.0 initiative. The findings suggest that ROLC can enhance efficiency, safety, and collaboration, while reducing waste, eventually achieving an optimal balance between site preparation effort and robot development complexity (see Figure 12). As robotic applications become increasingly ubiquitous in the construction industry, further research, development, education, and standardization are needed to refine and expand this emerging interdisciplinary topic.

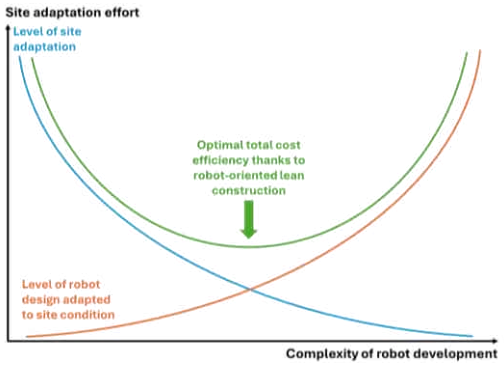


Figure 12. ROLC ensures an optimal balance between site adaptation and robot complexity

4.2 Limitations

Admittedly, limitations remain in this study. For example, the case study approach has several drawbacks, including challenges in generalization and replication (due to Singapore's unique policy environment and labor force structure), potential researcher bias, an intensive data collection and analysis process, and restricted scientific rigor [31]. Furthermore, more quantitative methods need to be employed to demonstrate evidence-based advantages and benefits of the proposed toolkit. In addition, other potentially promising lean methods and principles may not have been covered in this early-stage investigation. Further research is needed to refine and advance the topic.

4.3 The Pathway Forward

To overcome these limitations, the following aspects point to the directions and priorities for future research:

- Synergy between the two domains: Keep the upcoming ubiquitous adoption of construction robots in mind to optimize the process and flow of lean construction. Meanwhile, keep lean principles (e.g., respect for people, eliminating waste, optimizing process and flow, continuous improvement, increasing value for customers, etc.) in mind when developing robotic solutions.
- Education and training: Make training and education programs about the intersection of lean construction and construction robotics (i.e., ROLC) to construction managers, workers and students.
- Generalization and standardization: The feasibility of ROLC in other parts of the world needs to be assessed on a case-by-case basis. Furthermore, multi-case comparative studies in one region and cross-regional studies are necessary to further validate the effectiveness and performance of ROLC. In addition, the possibility of standardization especially related to construction robotics and lean construction needs to be investigated.
- More respect for people: The goal of implementing ROLC is not to replace human workers, but to empower the existing workers with emerging technologies to increase their work efficiency and enhance safety while mitigating labor shortages that are prevailing in the global construction sector. Therefore, the needs and fear of human workers towards AI and robotics need to be understood to increase the acceptance of emerging technologies.
- Continuous improvement of the toolkit: Admittedly, this paper only provides an introductory guideline for ROLC as a starting point. Further development is needed to enrich both theoretical and methodological aspects of the interdisciplinary field.

References

[1] McKinsey Global Institute, "Reinventing Construction: A Route to Higher Productivity," 2017. [Online]. Available: https://www.mckinsey.com/~/media/mckinsey/business functions/operations/our insights/reinventing construction through a productivity revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf

[2] M. Tender, J. P. Couto, A. Gibb, P. Fuller, and S. Yeomans, "Emerging Technologies for Health, Safety and Well-being in Construction Industry BT - Industry 4.0 for the Built Environment: Methodologies, Technologies and Skills," M. Bolpagni, R. Gavina, and D. Ribeiro, Eds., Cham: Springer International Publishing, 2022, pp. 369–390.

[3] G. L. F. Benachio, M. do C. D. Freitas, and S. F. Tavares, "Circular economy in the construction industry: A systematic literature review," *J. Clean. Prod.*, vol. 260, p. 121046, 2020.

[4] O. Salem, J. Solomon, A. Genaidy, and I. Minkarah, "Lean Construction: From Theory to Implementation," *J. Manag. Eng.*, vol. 22, no. 4, pp. 168–175, Oct. 2006.

[5] T. Bock, "The future of construction automation: Technological disruption and the upcoming ubiquity of robotics," *Autom. Constr.*, vol. 59, pp. 113–121, Nov. 2015.

[6] K. C. S. Ang, S. Sankaran, and D. Liu, "Sociotechnical considerations on developing human robot teaming solutions for construction: a case study," *Constr. Robot.*, vol. 8, no. 2, p. 23, 2024.

[7] R. Hu, J. Mao, S. Chae, and J. K. W. Yeoh, "Benefits of Construction Robots for On-Site Lean Construction: A Comparison and Evaluation Framework," *Eng. Constr. Archit. Manag.*, vol. 33, no. 2, pp. 914–936, 2026.

[8] C. Brosque, G. Skeie, and M. Fischer, "Comparative Analysis of Manual and Robotic Concrete Drilling for Installation Hangers," *J. Constr. Eng. Manag.*, vol. 147, no. 3, p. 5021001, Mar. 2021.

[9] C. Brosque and M. Fischer, "Safety, quality, schedule, and cost impacts of ten construction robots," *Constr. Robot.*, vol. 6, no. 2, pp. 163–186, 2022.

[10] F. G. Feldmann, "Towards Lean Automation in Construction—Exploring Barriers to Implementing Automation in Prefabrication," *Sustainability*, vol. 14, no. 19. 2022.

[11] A. E. Oke, J. Aliu, P. Fadamiro, P. S. Jamir Singh, M. S. Samsurijan, and M. Yahaya, "Robotics and automation for sustainable construction: microscoping the barriers to implementation," *Smart Sustain. Built Environ.*, vol. 13, no. 3, pp. 625–643, Jan. 2024.

[12] T. Ohno, *Toyota Production System: Beyond Large-Scale Production*. Productivity Press, 1988.

[13] R. F. Aziz and S. M. Hafez, "Applying lean thinking in construction and performance improvement," *Alexandria Eng. J.*, vol. 52, no. 4, pp. 679–695, 2013.

[14] Lean Construction Institute, "Lean Construction." Accessed: Aug. 01, 2025. [Online]. Available: https://leanconstruction.org/lean-topics/lean-construction/

[15] O. Babalola, E. O. Ibem, and I. C. Ezema, "Implementation of lean practices in the construction industry: A systematic review," *Build. Environ.*, vol. 148, pp. 34–43, 2019.

[16] S. Singh and K. Kumar, "Review of literature of lean construction and lean tools using systematic literature review technique (2008–2018)," *Ain Shams Eng. J.*, vol. 11, no. 2, pp. 465–471, 2020.

[17] F. Hamzeh, V. A. González, L. F. Alarcon, and S. Khalife, "Lean Construction 4.0: Exploring the Challenges of Development in the AEC Industry," in *Proc. 29th Annual Conference of the International Group for Lean Construction (IGLC)*, Lima, Peru, 2021, pp. 207–216.

[18] J. A. Cardenas, P. Martinez, and R. Ahmad, "Integrating lean and robotics in the construction sector: a scientometric analysis," *Constr. Robot.*, vol. 8, no. 1, p. 2, 2024.

[19] W. H. Gan and D. Koh, "COVID-19 and Return-To-Work for the Construction Sector: Lessons From Singapore," *Saf. Health Work*, vol. 12, no. 2, pp. 277–281, 2021.

[20] S. Y. Chia, "Foreign Labor in Singapore: Rationale, Policies, Impacts, and Issues," *Philipp. J. Dev.*, vol. 38, pp. 105–133, 2014, [Online]. Available: https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidspjd11-singapore.pdf

[21] B. C. L. Yin, R. Laing, M. Leon, and L. Mabon, "An evaluation of sustainable construction perceptions and practices in Singapore," *Sustain. Cities Soc.*, vol. 39, pp. 613–620, 2018.

[22] K. Parrish and I. D. Tommelein, "Making Design Decisions Using Choosing by Advantages," in *17th Annual Conference of the International Group for Lean Construction*, Y. Cuperus and E. H. Hirota, Eds., 2009, pp. 501–510. [Online]. Available: http://iglc.net/Papers/Details/663/pdf

[23] T. Bock, "Robot-Oriented Design," *Proceedings of the 5th International Symposium on Automation and Robotics in Construction (ISARC)*. International Association for Automation and Robotics in Construction (IAARC), pp. 135–144, 1988.

[24] S. Brell-Cokcan, "Enabling Construction Automation: Implementing Radio Frequency Communication Infrastructure on Construction Sites," *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*. International Association for Automation and Robotics in Construction (IAARC), pp. 355–363, 2024.

[25] T. Linner, R. Hu, K. Iturralde, and T. Bock, "A Procedure Model for the Development of Construction Robots," in *Innovation in Construction - A Practical Guide to Transforming the Construction Industry*, S. H. Ghaffar, P. Mullett, E. Pei, and J. Roberts, Eds., Cham: Springer International Publishing, 2022, pp. 321–352.

[26] M. Jiménez, L. Romero, J. Fernández, M. D. Espinosa, and M. Domínguez, "Extension of the Lean 5S Methodology to 6S with An Additional Layer to Ensure Occupational Safety and Health Levels," *Sustainability*, vol. 11, no. 14. 2019.

[27] O. Salem, J. Solomon, A. Genaidy, and M. Luegring, "Site Implementation and Assessment of Lean Construction Techniques," *Lean Constr. J.*, no. 2005, pp. 1–21, 2005.

[28] G. Ballard, "The Last Planner System of Production Control," 2000. [Online]. Available: https://lean-construction-gcs.storage.googleapis.com/wp-content/uploads/2022/09/08152942/the-last-planner-system-of-production-control-ballard2000-dissertation.pdf

[29] M. Wolska, T. Gorewoda, M. Roszak, and L. Gajda, "Implementation and Improvement of the Total Productive Maintenance Concept in an Organization," *Encyclopedia*, vol. 3, no. 4. pp. 1537–1564, 2023.

[30] G. D. Boca, "The Gemba Walk – A Tool For Management and Leadership," *Ovidius Univ. Ann. Econ. Sci. Ser.*, vol. 0, no. 1, pp. 450–455, 2015, [Online]. Available: https://ideas.repec.org/a/ovi/oviste/vxvy2015i1p450-455.html

[31] M. K. Glette and S. Wiig, "The Headaches of Case Study Research: A Discussion of Emerging Challenges and Possible Ways Out of the Pain," *Qual. Rep.*, vol. 27, no. 5, pp. 1377–1392, 2022.

Schedule-Driven Lightweight 4D Simulation for the Setup Phase: Enabling Practitioner-Led Construction Planning

Yuko Ishizu[1], Kota Yoshifuji[2], Shun Matsumoto[2], Muneyuki Ohigashi[2] and Hiroaki Yamasaki[2]

[1]GEL Inc., Japan
[2]Takenaka Corporation, Japan
yishizu@geometryengineeringlab.tech, yoshifuji.kouta@takenaka.co.jp, matsumoto.shun@takenaka.co.jp, ohigashi.muneyuki@takenaka.co.jp, yamasaki.hiroaki@takenaka.co.jp

Abstract –

In high-context construction cultures like Japan, the Setup phase—comprising *nemawashi* (consensus-building) and *dandori* (operational preparation)—is critical for reliable execution. While BIM models are increasingly available, site engineers often lack the modelling expertise to evaluate them independently. This reliance on specialists creates bottlenecks, forcing planning back to manual quantity take-offs and 2D drawing-based scheduling. Adopting a Design Science Research (DSR) approach, this study proposes a "schedule-driven lightweight 4D simulation workflow" to restore practitioner agency.

The workflow enables engineers to author schedules using a cloud-based tool, export data as CSV, and automatically link it to simplified Revit Box mass elements via native Key Schedules. Crucially, these Boxes act as spatial filters, automatically calculating quantities for specific structural components—including girders, beams, columns, sub-columns, and dampers—while utilizing Revit Assemblies to accurately count lifting units. The system simultaneously generates color-coded, flipbook-style 4D visualizations without specialist intervention.

A pilot workshop with five practitioners from a Tier-1 general contractor validated the approach. Baseline surveys revealed that 80% of participants previously relied on manual quantity take-offs. Following system use, participants rated its effectiveness for quantity verification and stakeholder communication at 4.40/5.00. A small comparison with conventional practice suggested that equivalent quantity verification and 4D output generation could be completed in 15–20 minutes, with minor revisions in 2–5 minutes. These findings suggest that lowering technical barriers to BIM access through lightweight Process Information Models (PIM) can enable practitioner-led construction planning and help bridge the gap between digital planning intent and on-site execution.

Keywords –

BIM accessibility; 4D simulation; Practitioner agency; Construction planning; Schedule-driven planning; Process Information Model (PIM)

1 Introduction

Production Planning and Control (PP&C) is a central yet challenging component of construction management due to the socio-technical nature of construction projects. To address uncertainty and variability in production, Lean Construction has advanced the Last Planner System (LPS), which emphasizes collaborative planning and reliable workflow as mechanisms for stabilizing execution [1,2].

Despite these advances, PP&C discussions have often been framed primarily around “planning” and “control,” while the transitional work that connects planning decisions to executable site actions remains under-specified. The importance of managing handoffs between specialists to achieve project goals has been highlighted [3]. However, intermediate coordination activities frequently lack formal structure, and assigning adequate staff to such work is known to be fraught with pitfalls [4]. In practice, this “in-between” work relies heavily on tacit knowledge and informal coordination, making it difficult to standardize processes or support them systematically.

To better conceptualize this gap, the Four-fold Framework for Process Information Management extends PP&C into four phases: Planning, Setup, Execution, and Control [5].

At the same time, BIM models from design and fabrication stages are increasingly available in contemporary projects and contain rich geometric and quantitative information that could support Setup-phase coordination. However, practical adoption remains limited when site engineers cannot directly utilize these models for production planning. A major barrier is the heavy reliance on BIM specialists to generate or modify BIM-based 4D outputs, which typically requires advanced modelling skills and familiarity with

specialized software environments. Prior studies have reported that such skill, workflow, and organizational barriers hinder both BIM implementation and the adoption of 4D modelling in practice [6,7]. As a result, iteration cycles during the Setup phase slow down, and planning frequently falls back to manual quantity take-offs and 2D-based coordination.

To address this problem, this study focuses on the accessibility of BIM-based planning during the Setup phase. Specifically, this paper proposes a schedule-driven lightweight 4D simulation workflow that enables site engineers without advanced BIM expertise to verify *dandori* and facilitate *nemawashi* by linking schedule information to existing models and generating time-based visual evidence for coordination. By reframing the difficulty of 4D simulation as an accessibility problem rather than a purely technical one, this research seeks to restore practitioner agency and support field-driven production planning.

2 Theoretical Framework: The Setup Phase and Process Information

2.1 The Four-fold Framework for PP&C

This study adopts the "Four-fold Framework for Process Information Management" proposed by Shigaki et al. [5], which extends the PP&C cycle into four distinct phases: Planning, Setup, Execution, and Control.

Within this framework, the Setup phase plays a pivotal role as the connective layer between abstract planning and value-adding execution on site. Rather than being a passive preparatory step, Setup is understood as an active phase in which planning assumptions are translated, negotiated, and validated before execution begins. The effectiveness of this phase directly influences workflow continuity, coordination efficiency, and the reduction of downstream disruptions. Table 1 summarizes the four phases and typical process information exchanged across them.

Table 1. Four phases and typical process information in PP&C (adapted from [5])

Phase	Representative information	Typical outputs
Planning	targets, constraints, work packages	plans (master/phase/lookahead)
Setup	zones, quantities, prerequisites, handoffs	nemawashi/dandori decisions
Execution	progress, actuals, issues	daily/weekly records
Control	variances, corrective actions	re-plans, improvements

2.2 Defining the Setup Phase: *Nemawashi* and *Dandori*

In high-context construction cultures such as Japan, the Setup phase is characterized by two interrelated practices: *nemawashi* and dandori. *Nemawashi* refers to the informal yet essential process of consensus-building among stakeholders, through which social, organizational, and contractual constraints are addressed before physical work begins. *Dandori*, in contrast, focuses on the technical and operational preparation required to realize planned activities. It involves verifying work-zone quantities, sequencing tasks, allocating resources, and confirming logistical feasibility.

Despite the availability of 3D BIM models in contemporary projects, both *nemawashi* and *dandori* often rely on manual, drawing-based methods. This disconnect occurs because BIM models exist but are not directly accessible to site engineers for independent evaluation. In current practice, engineers cannot independently verify quantities or test alternative construction sequences; instead, they depend on BIM specialists for iterative support. When digital tools fail to support independent exploration during this phase, the Setup phase becomes a bottleneck, disrupting workflow continuity and limiting the potential benefits of BIM-based planning.

Figure 1 illustrates the conventional Setup-phase workflow targeted by the proposed system. In existing practice, site engineers rely on manual quantity take-off for *dandori*, while 4D visualization for *nemawashi* requires manual task–element linking in tools such as Navisworks and Synchro. These workflows impose hidden coordination costs and often depend on BIM specialists.

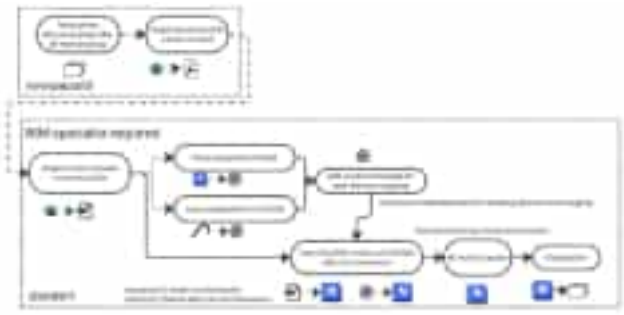


Figure 1. Conventional Setup-phase workflow showing manual quantity take-off and specialist-dependent 4D linking.

2.3 Process Information Model (PIM)

To rationalize the Setup phase, this study employs the concept of a Process Information Model (PIM) [12]. Unlike product-oriented BIM models that focus on "what" is being built, a PIM emphasizes "how" and "when" construction activities occur by associating dynamic process attributes with geometric elements. Typical PIM attributes include work zones, task identifiers, scheduling parameters, quantities, and labor requirements. By embedding such information into the model, PIM enables evaluation of spatial and temporal feasibility rather than purely geometric correctness.

In this research, a "lightweight" PIM is defined not merely by file size, but by its accessibility in three dimensions: (1) Cognitive: utilizing familiar planning concepts (e.g., zones) rather than complex BIM logic; (2) Technical: relying on standard software features (e.g., native parameter mapping such as Key Schedules) to avoid custom development; and (3) Operational: enabling rapid iteration by non-specialists. This contrasts with heavyweight PIM approaches that demand full BIM authoring capabilities and specialized knowledge.

The lightweight PIM is implemented by mapping schedule information to simplified volumetric elements (e.g., mass boxes). A scheduling tool is used as the planning interface, while mass elements are used to avoid complex geometric modeling, thereby maintaining the "lightweight" principle. This design enables the PIM to function as a bridge between scheduling logic and spatial validation, allowing site engineers to work within familiar concepts without specialist intermediaries. The lightweight PIM supports two key evaluations during Setup: (1) Spatial feasibility, confirming work-zone quantities and checking consistency for operational preparation (supporting *dandori*); and (2) Temporal feasibility, visualising sequences and communicating construction logic to stakeholders (supporting *nemawashi*).

3 Methodology（Design Science Research）

This study adopts the Design Science Research (DSR) approach presented by Hevner et al. [8] to address the heavy reliance on BIM specialists for generating 4D simulations during the Setup phase (*nemawashi* and *dandori*). Following the "build–evaluate" cycle described by March and Smith [9], we designed a practical artifact and evaluated it in a real project setting. The goal is to enable site engineers to verify quantities and explore planning scenarios independently, without requiring advanced modeling skills.

The artifact developed in this study is a schedule-driven lightweight 4D workflow. This workflow serves as an instantiation of the lightweight Process Information Model (PIM) interface described by Shigaki et al. [5]. Specifically, the system maps schedule data from GaNett (CSV) to simple Revit mass elements (Box) using standard Key Schedules and Dynamo Player automation. To ensure accessibility for non-specialists, we designed the workflow to be "lightweight" in three dimensions: (1) Cognitive: using familiar planning concepts (e.g., zones) instead of complex BIM logic; (2) Technical: using Dynamo Player as an execution interface to minimize setup and avoid custom plugin development for end users; and (3) Operational: enabling rapid iteration cycles to support dynamic site discussions.

To validate the utility and accessibility of the artifact, we conducted a pilot workshop with practitioners possessing diverse levels of BIM experience. Data were collected via post-workshop questionnaires (Likert-scale items and open-ended comments). In particular, the evaluation focused on accessibility, perceived usefulness, and practitioner agency (i.e., the ability to explore scenarios without specialist mediation) during Setup-phase coordination. We analyzed the results following DSR evaluation guidance recommended by Hevner et al. [8] and Peffers et al. [10] to examine whether the workflow effectively supports *nemawashi* and *dandori* activities.

4 Proposed System

4.1 System Overview and Design Rationale

The proposed system is a schedule-driven lightweight 4D simulation workflow designed to support Setup-phase coordination by site engineers without advanced BIM authoring expertise. In contrast to conventional 4D workflows that require detailed model authoring and specialist intervention, this system treats the BIM model as a verification and visualisation medium rather than the primary planning interface. The core design principle is to align the system with existing planning practices while minimizing cognitive, technical, and operational barriers.

The system focuses on two critical structural processes—steel erection and concrete placement—as representative high-impact activities where site engineers must verify quantities and communicate sequences under tight coordination constraints. By targeting these activities, the workflow supports simultaneous quantity verification and sequence visualisation, consistent with the lightweight Process Information Model (PIM) concept described in Section 2.3. Figure 2 illustrates the overall system architecture, highlighting the data flow from the scheduling tool to the 4D visualisation output.

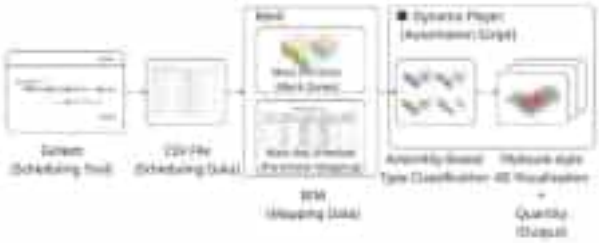


Figure 2. Overview of the schedule-driven lightweight 4D workflow (GaNett → CSV → Revit Key Schedule + Dynamo Player → flipbook output).

In contrast to the conventional 4D workflow illustrated in Figure 1 (Section 2), where manual task-element linking and specialist intervention are required, the proposed system eliminates these dependencies through spatial containment logic. This comparison highlights the reduction in cognitive and operational barriers achieved by the lightweight approach.

4.2 Schedule-Driven Workflow

The workflow begins with schedule authoring using an in-house cloud-based planning tool (GaNett) tailored for Japanese graph-based planning practices [11]. Site engineers define work packages using familiar concepts such as work zones, task identifiers, durations, and quantities. The schedule is then exported in CSV format, serving as the primary source of process information.

The Dynamo script includes a Python node that cleans and standardizes the exported CSV before Key Schedule mapping. Because this node is the sole adapter between the scheduling tool's export format and the Dynamo script's input schema, only this single node needs modification to support alternative tools (e.g., MS Project, Primavera P6); all downstream automation remains unchanged.

In the BIM environment, an existing structural model from design or fabrication stages is reused without modification. Instead of authoring detailed construction elements, site engineers define Box mass elements to represent work zones (*koku*). These volumetric elements act as lightweight spatial containers and are used as spatial filters to query structural and concrete elements located within their boundaries (e.g., girders, columns, braces, slabs, and walls). As shown in Figure 3, schedule attributes from the CSV file are linked to the Box mass elements using Revit's native Key Schedule functionality. This key–value mapping strategy associates each work zone with its corresponding task information, enabling consistent parameter population without custom plugin development.



Figure 3. Mapping schedule attributes to Box mass zones using Revit Key Schedules.

4.3 Automation and Execution

A pre-configured Dynamo Player script executes the core automation steps. The script reads the schedule CSV file, matches zone identifiers, and generates two key outputs—quantities and 4D visualisations—tailored to the targeted processes:

1. Automated quantity verification. The system applies calculation rules based on the material type found within the Box zones.

- Steel erection: The workflow calculates total weight and piece count, categorized by specific component types including Columns, Beams, Girders, Sub-columns, and Dampers. Piece counts are derived from Revit Assemblies, allowing users to pre-group multiple elements (e.g., a column and its brackets) into a single lifting unit. This supports detailed logistics planning for cranes and trucks (Figure 4).
- Concrete placement: The workflow calculates volume and area using geometric intersection between Box masses and concrete elements. This enables zone-based quantity take-off for slabs or walls split by zone boundaries (e.g., construction joints) without physically splitting BIM elements.

2. Flipbook-style 4D visualisation. Based on schedule timing information for steel and concrete tasks, the system generates time-indexed 4D outputs. To ensure intuitive comprehension, a simple color-coding scheme is applied to indicate three distinct states: Not Started, Under Construction, and Completed. These outputs are exported as sequential images, forming a flipbook-style representation of the construction process (Figure 5). This format was selected to support rapid review, comparison, and communication during Setup-phase discussions, prioritizing speed and clarity over high-fidelity animation.

Manual operations are limited to three actions: exporting the schedule CSV, defining Box mass zones, and executing the Dynamo Player script. All data linking, quantity aggregation, and visualisation generation are automated, maintaining the "lightweight" principle for end users.

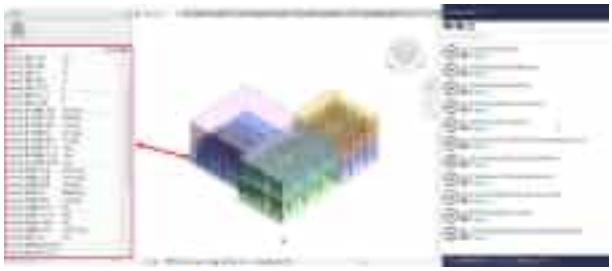


Figure 4. Automated Quantity Aggregation using Dynamo Player. The Dynamo Player (right) executes the spatial filtering logic, calculating quantities for each Mass element (centre). The system automatically aggregates piece counts and weights for specific element types—such as girders, beams, and columns—and populates the results directly into the Mass element's properties (left) for verification (*dandori*).



Figure 5. Example flipbook-style (*parapara manga*) 4D output for Setup-phase review (selected time steps).

4.4 Supporting Setup-Phase Activities

The proposed system directly supports the two core activities of the Setup phase defined in Section 2.2:

- Supporting *dandori* (operational preparation): Zone-based quantity aggregation enables site engineers to verify consistency between planned quantities and spatial work zones. By visualising quantities and sequences together, potential issues such as missing scope, incorrect zone definitions, or unrealistic sequencing can be identified early.
- Supporting *nemawashi* (consensus-building): The time-sequenced image outputs provide intuitive visual evidence that can be shared with stakeholders who do not use BIM tools. This facilitates discussion of construction logic, spatial constraints, and sequencing decisions without requiring specialist mediation. Through rapid re-execution of the workflow, practitioners can iteratively adjust schedules, regenerate visual outputs, and compare alternatives within short timeframes. This capability enables exploratory "what-if" discussions that are essential during the Setup phase.

4.5 Summary

By combining schedule-driven planning, lightweight spatial representation, and automated parameter mapping, the proposed system provides a practical implementation of a lightweight PIM for Setup-phase coordination. The system lowers barriers to BIM-based 4D reasoning and enables practitioner-led verification and communication, directly addressing the structural dependency on BIM specialists identified in Sections 1 and 2.

5 Evaluation and Results

5.1 Evaluation Design

The proposed system was evaluated through a pilot workshop conducted with a Tier-1 Japanese general contractor. Following the Design Science Research (DSR) approach described in Section 3, the evaluation assessed the perceived utility and accessibility of the artifact in an authentic Setup-phase planning context, rather than statistical generalization.

Five practitioners participated in the workshop, including site engineers and planning staff who are typically involved in *nemawashi* and *dandori* activities. Participants had diverse levels of experience with Revit and BIM-based workflows, ranging from limited exposure to routine use. No prior experience with BIM authoring or 4D simulation was required.

The evaluation focused on three criteria derived from the research objectives and theoretical framework: (1)**Accessibility:** whether practitioners perceived the workflow as usable without specialist support; (2) **Perceived usefulness:** whether the workflow supported quantity verification (*dandori*) and stakeholder communication (*nemawashi*) during the Setup phase; and (3) **Practitioner agency:** whether practitioners felt able to independently explore construction scenarios.

5.2 Data Collection

Data were collected using a post-workshop questionnaire consisting of Likert-scale items (five-point scale: 1 = strongly disagree, 5 = strongly agree) and open-ended questions. In addition, a short baseline questionnaire was administered prior to system use to capture participants' perceptions of current Setup-phase practices and challenges.

The questionnaire items addressed: (A) current challenges in quantity understanding and communication, (B) perceived improvements in stakeholder communication, (C) perceived improvements in evidence-based quantity takeoff, (D) perceived ability to independently explore scenarios, and (E) overall usability and intention to adopt the workflow.

5.3 Baseline Perceptions of Current Practice

The baseline questionnaire indicated that manual,

drawing-based work remains a major perceived bottleneck in Setup-phase planning. Specifically, 80% of participants reported that zone-based quantity understanding still relies on manual take-offs, and 80% reported difficulty explaining construction logic to stakeholders using static drawings alone.

These responses suggest that, even when BIM models are available, critical Setup-phase tasks related to *dandori* (quantity verification) and *nemawashi* (coordination) are widely perceived as difficult to support effectively using existing 2D-based practices.

5.4 Results After System Use (Quantitative)

After using the proposed schedule-driven lightweight 4D workflow, participants evaluated its contribution to Setup-phase activities. The mean rating for improving stakeholder communication (Question B) was 4.40/5.00, and the mean rating for enabling evidence-based quantity take-offs (Question C) was also 4.40/5.00. Participants rated the ease of operation of the workflow at 4.00/5.00, indicating that the system was perceived as accessible for practitioners without advanced BIM expertise. The intention to use the workflow in practice was rated positively (4.00/5.00), suggesting feasibility for real-world adoption. Crucially, participants evaluated the ability to independently explore construction scenarios (Question D) at 4.00/5.00, directly supporting the study's objective of restoring practitioner agency during the Setup phase. Table 2 summarizes the quantitative results.

Table 2. Summary of quantitative survey results (N = 5)

Metric	Related activity	Mean score (max 5.0)
Improvement in stakeholder communication (Q B)	*Nemawashi*	4.40
Improvement in quantity verification (Q C)	*Dandori*	4.40
Ease of operation (accessibility)	–	4.00
Ability to explore independently (agency)	–	4.00
Intention to use	–	4.00

5.4.1 Quantitative Efficiency Improvements

To further clarify the significance of the proposed workflow in relation to conventional practice, a small comparative time observation was conducted during the pilot workshop. Using one construction zone from a previously completed project prepared as workshop material, two participants were asked to count three structural element categories—columns, primary beams, and secondary beams—on 2D drawings using markers. The measured time per element category ranged from 3 to 9 minutes. This variation suggests that manual quantity take-off is partly operator-dependent, for example depending on whether the count is performed once or repeated for verification to avoid omissions.

Using the upper-bound value as a conservative estimate, counting three element categories for one floor would require approximately 27 minutes, and extending the same procedure across seven floors would require approximately 189 minutes. Although this was only a limited comparison, it indicates that the conventional workflow involves both substantial time burden and variability between operators.

By contrast, the proposed workflow enabled equivalent quantity verification and generation of 4D visual outputs in approximately 15–20 minutes. The time required for regeneration depended on the scale of the revision: minor modifications—such as resizing Box mass elements in Revit or changing the correspondence between schedule items and work zones—required approximately 2–5 minutes, whereas major modifications—such as changing the total number of work zones or restructuring the schedule in GaNett—required approximately 15–20 minutes.

In current practice, such full regeneration is rarely undertaken. The significance of the proposed workflow therefore lies not only in reducing task duration, but in making iterative scenario re-examination practically feasible for site engineers themselves.

5.5 Qualitative Observations from Open-Ended Responses

Open-ended responses further clarify why the workflow was valued. Participants noted that the system helped them interpret planning logic spatially and that the flipbook-style 4D outputs were easy to use when explaining plans to stakeholders who do not use BIM tools. Several participants also highlighted that rerunning the workflow after small schedule changes enabled "what-if" discussions during coordination meetings.

Importantly, multiple comments reflected a shift in perceived role: instead of waiting for BIM specialists to prepare visualisations, participants felt able to check scenarios by themselves and confirm whether the plan was reasonable at an early stage. These observations align with the quantitative ratings on independent exploration, supporting the interpretation that accessibility—rather than automation speed or visual fidelity—was the primary source of value.

5.6 Summary of Evaluation Results

The evaluation results indicate that the proposed schedule-driven lightweight 4D simulation workflow addresses perceived challenges in Setup-phase planning.

Participants reported that the workflow supports both *dandori*-oriented quantity verification and *nemawashi*-oriented stakeholder communication, while remaining accessible to practitioners without advanced BIM expertise. Taken together, the quantitative ratings, comparative observations, and qualitative responses suggest that lowering barriers to BIM-based 4D reasoning can restore practitioner agency and enable more repeatable and revisable planning cycles during the Setup phase. The workflow's value therefore lies not merely in automation speed, but in making practitioner-led iteration practically feasible.

6 Discussion

6.1 Restoring Practitioner Agency in the Setup Phase

The evaluation results indicate that the proposed schedule-driven lightweight 4D simulation workflow contributes to restoring practitioner agency during the Setup phase. Rather than positioning site engineers as passive recipients of BIM outputs generated by specialists, the workflow enables them to directly interact with spatial and temporal information and to explore construction scenarios independently.

This shift is particularly significant in the context of Setup-phase activities such as *dandori* and *nemawashi*, which require frequent iteration, negotiation, and validation. Participants valued the ability to "check by themselves" and to rerun the workflow after small schedule changes, supporting rapid "what-if" exploration. The quantitative results also suggest a positive perceived ability to explore scenarios independently (agency), consistent with this observation. These findings suggest that agency in this context is not defined by advanced modelling capability, but by the ability to access, interpret, and manipulate relevant process information at the moment decisions are made.

From this perspective, the proposed workflow does not merely automate existing tasks; it reconfigures the relationship between practitioners, BIM models, and planning decisions. By lowering barriers to engagement with BIM-based 4D information, the system allows agency to shift from specialists toward site-level practitioners who are directly responsible for coordination and execution.

6.2 Accessibility as a Design Principle for 4D Systems

A key contribution of this study is the articulation of accessibility as a central design principle for 4D simulation systems. The lightweight PIM operationalizes accessibility across three dimensions—cognitive, technical, and operational—providing a concrete framework for evaluating and designing practitioner-facing tools.

The evaluation results suggest that reducing these barriers can yield tangible benefits, even with a small-scale pilot. Participants reported high perceived usefulness and ease of operation, despite limited prior experience with BIM authoring. Importantly, the system's value was associated more strongly with accessibility and iteration speed than with visual fidelity or automation sophistication. This finding challenges the prevailing assumption that effective 4D systems must prioritize realism or optimization. Instead, for Setup-phase planning, the ability to quickly generate understandable visual evidence and to iterate collaboratively may be more critical than producing detailed simulations.

6.3 Limitations

This study has several limitations. First, the evaluation was a small-scale pilot (N = 5) in a single organization and therefore does not support statistical generalization. However, the workflow targets approximately 40–50 potential users in the organization, and the pilot cohort was selected to reflect variation in BIM usage frequency and understanding; thus, the findings should be interpreted primarily as evidence of feasibility and perceived usefulness.

Second, the study focused on Setup-phase activities and did not assess downstream impacts on Execution or Control (e.g., productivity or schedule reliability).

Third, the workflow was evaluated using simplified Box mass representations; future studies should test scalability with more complex models and coordination settings.

Fourth, although the system supports L-shaped and other non-rectangular mass geometries, zones containing interior voids (e.g., donut-shaped regions) still require manual workarounds. Moreover, spatial filtering is applied uniformly to all elements within a zone, and does not support element-type-specific selection (e.g., columns only), which limits the representation of phased erection sequences.

Therefore, semi-automated zone generation and element-type-based filtering remain important areas for future development.

6.4 Implications for Future Research and Practice

Despite these limitations, the lightweight PIM concept opens directions for comparative studies across projects and cultural contexts. For practice, embedding digital support within existing site workflows can expand practitioner access to BIM-based reasoning without

requiring adaptation to complex authoring environments. Following this formative pilot, we will conduct additional workshops to evaluate scalability across a broader range of roles and BIM literacy levels.

7 Conclusion

This study addressed a persistent gap between the availability of BIM models and their practical use in Setup-phase construction planning. While BIM and 4D simulation have been widely promoted, their adoption during *nemawashi* and *dandori* activities has remained limited due to heavy reliance on BIM specialists and the lack of accessible practitioner-facing tools.

To respond to this challenge, the study proposed a schedule-driven lightweight 4D simulation workflow grounded in a Design Science Research (DSR) approach. Building on the Four-fold Framework for Process Information Management, the system operationalized a lightweight Process Information Model (PIM) that prioritizes accessibility across cognitive, technical, and operational dimensions. By treating schedules as the primary driver and BIM geometry as a verification and visualisation medium, the workflow aligned digital support with how planning decisions are actually made in practice.

Evaluation through a pilot workshop suggested that the proposed workflow can effectively support Setup-phase activities. Participants reported improvements in zone-based quantity verification and stakeholder communication, as well as a strengthened ability to independently explore construction scenarios. These results suggest that the value of 4D simulation in this context lies not in high-fidelity visualisation or advanced automation, but in enabling practitioner-led iteration and decision-making.

The primary contribution of this research is the articulation and empirical grounding of restoring practitioner agency as a key objective for digital tools in construction planning. Rather than replacing human judgment, the proposed system repositions BIM-based 4D information as a resource that site engineers can directly access, interpret, and use to negotiate planning decisions during the Setup phase.

From a broader perspective, this study suggests that accessibility-focused design can play a critical role in bridging the gap between planning intent and executable site actions. By lowering barriers to BIM-based reasoning, lightweight systems such as the one presented here have the potential to support more resilient, field-driven Production Planning and Control practices. Future research can extend this work by examining longitudinal impacts on project performance and by exploring how lightweight PIM concepts transfer across different organizational and cultural contexts.

References

[1] L. Koskela. Management of production in construction: a theoretical view. In Proceedings of the 7th Annual Conference of the International Group for Lean Construction (IGLC 7), pages 241–252, Berkeley, California, USA, 1999.

[2] G. Ballard. The last planner system of production control. PhD thesis, University of Birmingham, Birmingham, UK, 2000.

[3] G. Ballard and G.A. Howell. An update on Last Planner. In Proceedings of the 11th Annual Conference of the International Group for Lean Construction (IGLC 11), Blacksburg, Virginia, USA, 2003.

[4] A. Laufer and R.L. Tucker. Competence and timing dilemma in construction planning. Construction Management and Economics, 6(4):339–355, 1988.

[5] J.S. Shigaki, K. Matsuda, K. Yoshifuji, T. Sugiyama, S. Someya, and N. Goto. Four-fold framework for process information management and the suitability of in-house developed tools in lean construction operations. In Proceedings of the 23rd CIB World Building Congress (WBC 2025), West Lafayette, USA, 2025.

[6] M. Swallow and D. Zulu. Benefits and barriers to the adoption of 4D modeling for construction site safety management. Frontiers in Built Environment, 4:86, 2018.

[7] C.T.W. Chan, J. Yang, J. Darko, and X. Shan. Barriers of implementing BIM in construction industry from the designers' perspective: A Hong Kong experience. Journal of Services Science and Management, 7(3):165–178, 2014.

[8] A.R. Hevner, S.T. March, J. Park, and S. Ram. Design science in information systems research. MIS Quarterly, 28(1):75–105, 2004.

[9] S.T. March and G.F. Smith. Design and natural science research on information technology. Decision Support Systems, 15(4):251–266, 1995.

[10] K. Peffers, T. Tuunanen, M.A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. Journal of Management Information Systems, 24(3):45–77, 2007.

[11] NTT Communications. Launch of "GaNett™, a cloud-based digital process application for the construction industry". On-line: https://www.ntt.com/about-us/press-releases/news/article/2024/0425.html, Accessed: 01/12/2026.

[12] S. Lee, S. Isaac, and T. Bock. An ontology for process information modeling. In Proceedings of the 32nd International Symposium on Automation and Robotics in Construction (ISARC 2015), pages 692–699, Oulu, Finland, 2015.

Activity-Flow–Based Modeling and Simulation for Construction Process Visualization

Diego Rojas1, Nelly Garcia1 and Luis Galvan1[1]

[1]Civil and environmental department, Universidad de los Andes, Colombia
d.rojasm@uniandes.edu, ne-garci@uniandes.edu.co.edu, l.galvan@uniandes.edu.co

Abstract

Construction activities depend on multiple interrelated flows, materials, labor, equipment, workspace, and precedence, which introduce significant complexity into planning and control processes. Current simulation approaches typically focus on activity-level information while overlooking how these flows interact and evolve within the project environment. As a result, project managers lack effective tools to visualize and understand on-site flow dynamics and their impact on activity execution.

This paper presents an Activity-Flow Model Simulation, a modeling and visualization approach that integrates activity- and flow-related data into a unified simulation environment. The method builds on the Activity-Flow–Based Model to capture essential information used in planning and control processes and employs an OpenBIM workflow combined with a Unity-based environment supporting agent-oriented flow simulation. The proposed model was tested using a pilot construction project to simulate and visualize on-site flows, including activity precedences, materials, labor, equipment, and workspace. Results show that the approach enhances the identification of flow interactions and reveals opportunities to improve activity sequencing and flow coordination. Overall, the study demonstrates the potential of activity-flow–based simulation as a complementary decision-support tool for construction planning and control.

Keywords –

Activity-Flow Based Model (AFM); Simulation; Planning and control; Building Information Modeling (BIM)

1 Introduction

The construction industry is widely recognized as one of the least productive sectors in terms of technological adoption and digital transformation, particularly when compared to manufacturing and other industrialized domains. Despite significant advances in construction management methods and digital tools, projects continue to suffer from inefficiencies related to poor planning, limited visibility of execution constraints, and inadequate coordination of resources and site logistics [1, 2].

Construction activities are inherently complex, as their execution depends on multiple interrelated variables and requirements, including labor, materials, equipment, workspace, and precedence relationships. The interaction and variability of these elements introduce uncertainty and make planning and control processes particularly challenging [3]. To manage this complexity, planning and control methods such as the Critical Path Method (CPM) and the Last Planner System (LPS) have been widely adopted to structure construction sequences and improve workflow reliability [4, 5]. While these methods provide a solid foundation for defining activity logic and commitments, they primarily focus on activity-level information and often treat resource and spatial constraints implicitly or in a fragmented manner.

More recent approaches have extended traditional planning paradigms by explicitly considering flows as essential elements for activity execution [6]. Flow-based perspectives emphasize that activities cannot start unless their required flows—such as space, labor, materials, and equipment—are available and properly coordinated [3]. Methods such as the Location-Based Management System (LBMS) incorporate spatial continuity and production flow into planning and control, improving workflow stability and reducing interference between crews [7]. Building on this perspective, the Activity-Flow Based Model (AFM) integrates multiple on-site flows into planning and control processes, enabling a more holistic representation of activity variability and construction complexity [8].

Despite these methodological advances, planning and control processes based on activity-flow concepts still lack adequate visual and simulation-based support. Building Information Modeling (BIM) has enabled the development of construction simulations, most notably through 4D models that link schedules to 3D geometry and support the visualization of activity sequencing over

time [9, 10]. However, conventional 4D simulations typically focus on activity precedence and timing, while overlooking the explicit representation of flows and their interactions within the construction site.

As a result, project managers and planners lack tools that allow them to visualize how activity-related flows move, interact, and compete for space and resources during project execution. This limitation reduces the ability to anticipate logistical conflicts, workspace interferences, and flow-related constraints that emerge during construction. Addressing this gap requires simulation approaches that integrate planning and control information with spatial and flow-based representations of construction processes.

Therefore, this paper proposes an activity-flow–based modeling, visualization, and simulation framework that builds on AFM principles and leverages OpenBIM workflows and a Unity-based simulation environment. The proposed approach enables the explicit representation and visualization of on-site flows—workspace, labor, materials, equipment, and precedences—within a unified simulation environment. The framework is demonstrated through a pilot construction project, illustrating its potential to support planning and control processes by enhancing the understanding of activity execution, flow interactions, and site logistics.

2 Theoretical Background

2.1 Planning and control processes

Planning and control processes are crucial in construction projects and have been extensively studied to improve project effectiveness [1, 10]. Project effectiveness is often hindered by multiple constraints or unmet requirements that prevent activities from starting as planned. Therefore, planning and control processes must be proactive and rely on iterative planning actions to identify, manage, and remove constraints in order to prepare tasks for execution [11]. The nature of construction activities involves multiple requirements which, if not properly planned, may become constraints during execution. These requirements are commonly referred to as flows [3] and can be managed and tracked using planning and control methods.

Production Control Plans rely on established planning approaches, such as the Critical Path Method (CPM) [4, 7] , to define construction sequences while considering multiple activity-related variables. These approaches attempt to represent the intrinsic complexity and variability of construction tasks. More advanced approaches have reformulated planning and control processes by incorporating not only activity-related information but also the flows required for execution. For instance, the Location-Based Management System (LBMS) integrates spatial flow considerations [7], whereas the Activity-Flow–Based Model (AFM) extends this perspective by explicitly including materials, equipment, labor, and workspace requirements [8]. Through a more holistic tracking of both activities and their associated flows, AFM seeks to capture the inherent variability of construction tasks and, consequently, their underlying complexity [8].

2.2 Activity Flow Based Model

The Activity-Flow–Based Model can be understood as an ontology that provides a robust conceptualization of the nature of construction activities—not only in terms of the activities themselves but also the flows that enable them, constrain them, and transition between them [12]. By analyzing flows, AFM allows for a broader understanding of project behavior during construction, moving beyond task-based logic.

Through the tracking of multiple variables, AFM strengthens planning and monitoring processes, supporting the transition toward more predictable projects with improved control over activity outputs. However, despite its robustness in capturing and structuring activities and flows, AFM does not provide a means to understand the interactions among flows.

Understanding the multivariable problem in which flows enable activities, and activities in turn release flows, remains complex. Although AFM captures this relationship conceptually, it lacks tools that facilitate the interpretation and interaction of these variables. For instance, the interaction between workspace and labor may result in trade stacking or labor collisions within the same workspace [13], which can be referred to as flow clashes.

Currently, AFM offers no visual support nor a simulation environment to complement planning and control processes. While traditional 4D simulations may partially represent sequencing, the effort invested in planning and tracking flows remains underutilized unless it is paired with flow-aware visualization. Therefore, a dedicated visualization and simulation framework is required to enhance the understanding of construction operations and address less predictable execution aspects.

2.3 Simulations in Construction

Simulation has become a valuable tool in the AECO industry to improve the understanding of complex construction tasks and to support and optimize planning and control processes [2, 14]. Building Information Modeling (BIM), through 3D models enriched with additional dimensions, has enabled the development of a wide range of simulation applications. Among these, 4D simulations are particularly effective in visually

representing construction logic by linking activity sequencing with spatial–temporal information [9].

Nevertheless, current 4D simulations typically consider only activity-based information—primarily precedence relationships—derived from the selected planning and scheduling method. As a result, they overlook the explicit representation of the flows required to execute each task.

On the other hand, numerous studies have explored the development of agent-based and discrete-event simulations to represent construction process logic [15, 16]. Agent-based simulations model individual agents (workers, equipment, or entities) as autonomous decision-making agents that interact within a virtual environment [16], while discrete-event simulations represent operations as a sequence of events that occur over time based on predefined rules and state changes. Although both approaches provide valuable approximations of flow behavior, they often remain disconnected from the activities that structure project execution and from one another, limiting their ability to capture the integrated nature of construction operations [15, 16].

To address this gap, this study proposes an activity-flow–based modeling process that integrates flow-related data into construction simulation and visualization frameworks.

3 Conceptual Model

During the planning phase, planners begin structuring a detailed schedule which, aligned with the Last Planner System, decomposes milestones into phases and progressively breaks them down into activities represented in look-ahead plans and weekly work plans [5]. In some cases, planning processes expand their scope beyond activities and precedence logic to incorporate limited flow-related considerations [12]. However, even when based on logically sound construction sequences, such plans cannot fully represent all required flows. As a result, planners lack certainty regarding flow behavior and whether plans are truly flow-clash-free, which increases uncertainty during project execution.

For this reason, the conceptual model begins with the development of a flow-aware schedule, using AFM to establish the principles governing activity–flow relationships. Activity information (precedences, planned start, and planned finish) and flow-related information (flow name and flow predecessor) are structured and embedded into 3D models through OpenBIM software, following IFC semantic export rules. The result is an IFC model enriched with AFM attributes, which is then instantiated within a Unity scene. This environment utilizes tools such as C#, semantic model classification, NavMesh pathing, and the Unity engine to guide activity execution and flow interactions.

In the simulation environment, flows are conceptualized as agents, resources, or spaces associated with activities. These representations are constrained through Unity-generated colliders, which regulate interactions between agents, between agents and resources, and between agents and activity elements. This enables the simulation of construction processes with spatial insight, uncovering interactions among flows and between flows and activities. In addition, the model incorporates performance metrics to support decision-making by identifying clashes that directly impact activity durations or the overall schedule. Initial metrics include agent–agent collision counts and travel distances.

Simulation outputs serve as iterative feedback during early planning, enabling progressively more realistic process definitions. Likewise, when applied to 4–8 week look-ahead windows, the model strengthens project control by supporting the weekly definition of work plans through a perspective that is flow-aware and, more importantly, flow-clash-aware as field conditions evolve.

3.1 Flow and activity visualization

The visual representation of flows constitutes the core pillar of the proposed conceptual model, where each flow is physically mapped and linked to activities within the Unity simulation environment. Figure 1 presents the interaction diagram illustrating how flows and activities are linked and represented within IFC models and the Unity simulation environment. Activities are visualized as IFC model elements—similar to traditional 4D simulations. Each element, and therefore each activity, is enriched with parametric data describing the required flows and the corresponding flow predecessors.

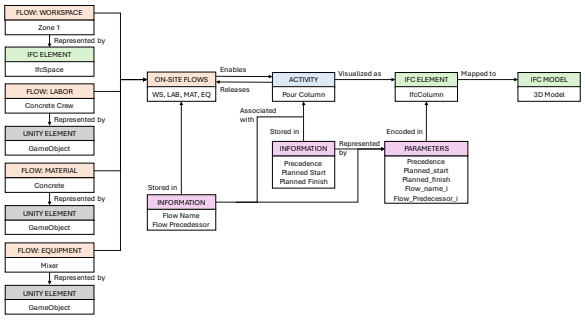


Figure 1. Flow and activity visualization process

Workspace flows are modeled as IfcSpace elements that define the volumetric boundaries of each execution zone. The remaining flows are instantiated as GameObjects inside the Unity scene—labor and equipment are modeled as autonomous agents, while materials are treated as resources consumed or transported by those

agents.

Finally, each flow is associated with its respective activity through the parametric information embedded in the IFC elements.

4 Methods

Figure 2 illustrates the overall workflow of the proposed activity-flow–based modeling and simulation framework. The process begins with AFM-based activity and flow definitions, followed by OpenBIM modeling and IFC-based data integration. The enriched model is then translated into a Unity environment to support agent-based simulation, enabling the visualization and analysis of flow interactions. The framework supports iterative planning refinement based on simulation outputs.

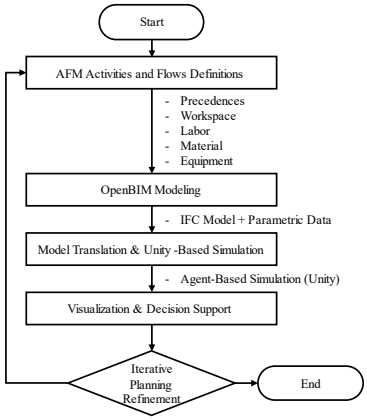


Figure 2. Workflow of the proposed activity-flow–based modeling and simulation framework.

4.1 AFM Information Tracking

The definition of the planning and control method is directly linked to the simulations to be executed, as the information established during planning and control determines what can be simulated and subsequently used to support decision-making processes. Accordingly, in this study, the Activity-Flow–Based Method (AFM) is adopted as the planning and control approach to enable the simulation of activity and flow-related information.

AFM allows tracking multiple variables associated with different off-site and on-site flows [8]. Although off-site flows may influence activity performance and execution, they are not considered within the scope of this framework. Instead, this study focuses exclusively on on-site flows, namely workspace, labor, equipment, and materials. For each flow, two key variables were considered: flow name and flow precedence. This means that, for each flow i, the model identifies an id and the activity that was using the flow prior to the start of the current activity (t–1) [17].

This structure enables the representation of both activity information and on-site flow interactions for workspace, labor, equipment, and materials, supporting the simulation of on-site logistics. Using this approach, information was extracted from a pilot project consisting of a sequence of 7 activities of 3 types, involving 3 crews and 3 workspaces. The modeled information is presented in Figure 3, Table 1 and Table 2, which illustrates the project layout, contains detail of the activity's sequences, and its flows.

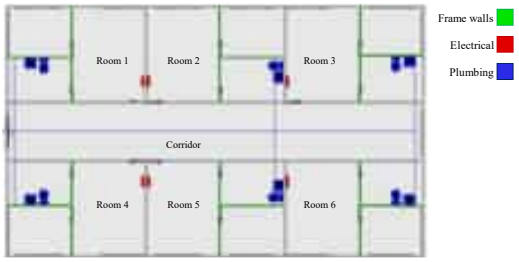


Figure 3. Project layout.

Table 1. Activity sequence and information.

ACT	ACT_PRED	ACT_T	P_ST	P_FIN
Frame walls R1	–	FW	09/22/25	09/23/25
Frame walls R2	–	FW	09/24/25	09/25/25
Plumbing rough-in R1	Frame walls R1	PR	09/24/25	09/24/25
Electrical rough-in R1	Frame walls R1	ER	09/25/25	09/25/25
Frame walls R3	–	FW	09/26/25	09/27/25
Plumbing rough-in R2	Frame walls R2	PR	09/25/25	09/25/25
Electrical rough-in R2	Frame walls R2	ER	09/26/25	09/26/25

ACT: Construction activity identifier; ACT_PRED: Activity precedence; ACT_T: Activity type (FW = Frame Walls, PR = Plumbing Rough-in, ER = Electrical Rough-in); P_ST: Planned start date; P_FIN: Planned finish date.

Table 2. Activity's sequence, flow information.

ACT	WS	WS PRED	LAB	LAB PRED	MAT	MAT PRED
FW R1	R1	–	DW_Crew	–	Studs+Gypsum	–
FW R2	R2	–	DW_Crew	FW R1	Studs+Gypsum	FW R1
PR R1	R1	FW R1	PL_Crew	–	Pipes+Fittings	–
ER R1	R1	PR R1	EL_Crew	–	Cables+Conduits	–
FW R3	R3	–	DW_Crew	FW R2	Studs+Gypsum	FW R2
PR R2	R2	FW R2	PL_Crew	PR R1	Pipes+Fittings	PR R1
ER R2	R2	PR R2	EL_Crew	ER R1	Cables+Conduits	ER R1

WS: Workspace where the activity is executed (e.g., R1 = Room 1); WS_PRED: Workspace-related precedence defining spatial availability constraints; LAB: Labor crew assigned to execute the activity; LAB_PRED: Labor-related precedence defining crew availability constraints. MAT: Materials required to execute the activity; MAT_PRED: Material-related precedence defining material availability dependencies.

4.2 Open BIM Modeling

Based on the planning results obtained using the

Activity-Flow–Based Method (AFM), the corresponding information is embedded into 3D models in accordance with the proposed methodology. The framework is grounded in OpenBIM principles, whereby parametric information is incorporated into Industry Foundation Classes (IFC) models for information exchange. This approach ensures that the framework remains independent of any specific authoring or modeling software used for simulation

The proposed workflow begins with the definition of activity and flow related key parameters: activity_start, activity_finish (actual_start and actual_finish, may be added for control processes), labor_name, labor_predecessor, and workspace_name, workspace_predecessor, material_name, material_predecessor. These parameters are embedded into the 3D models.

4.3 Unity Simulation and Visualization Development

The implemented methodology integrates the import of BIM/IFC models with an agent-based simulation system. The workflow begins with the semantic classification of IFC elements through rule-based algorithms, enabling the structured interpretation of model geometry and parametric information within the simulation environment. Subsequently, a precedence management system, derived from the Activity-Flow–Based Method (AFM), orchestrates task execution by dynamically assigning activities to autonomous agents. These agents are capable of navigating the virtual environment, executing construction tasks in real time, and interacting with other agents and site constraints. Scene assembly and simulation logic are implemented within the Unity engine, where IFC models are converted into FBX format and enriched with property data. Agent navigation and routing are defined using a NavMesh-based pathfinding system, allowing agents to move realistically across the construction site while respecting spatial constraints and task dependencies.

5 Results

Results are presented using both qualitative observations and basic quantitative indicators to demonstrate the analytical capabilities of the proposed activity-flow based simulation framework. The framework enables the representation of activity precedence relationships comparable to traditional 4D simulations, while explicitly visualizing workspace, material, labor, and equipment flows within a unified environment. In addition to qualitative observations, basic quantitative indicators were extracted from the simulation.

The simulation was implemented in a pilot project using a simplified activity sequence and predefined work trains. Despite this simplified setup, the simulation revealed logistical and coordination issues associated with each flow type that would not be evident from precedence-based schedules alone. Figure 4 illustrates the simulation output, including crew differentiation, worker movement paths, material and equipment locations, and associated indicators.

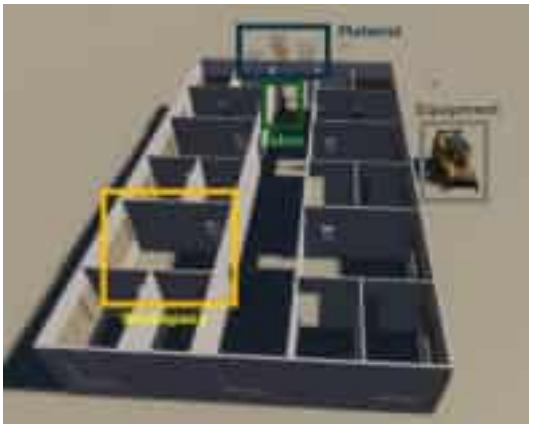


Figure 4. Representation of activity-flow elements (workspace, labor, material, and equipment) within the simulation environment.

To complement the visual analysis, quantitative indicators were extracted from the simulation. The model identified 14 labor collisions during the simulated execution of the pilot scenario. The average travel distance per crew was approximately 338 m, while the maximum distance traveled by a crew reached 374.4 m. These results confirm that the proposed framework enables the detection of flow-related inefficiencies that are not visible in traditional activity-based simulations. Figure 5 shows the simulation environment, highlighting the spatial representation of labor agents and their movement paths within the construction layout. The green lines represent the routes followed by workers as they navigate between activity locations. The interface also displays performance metrics, including collision counts, travel distances, and efficiency indicators, which provide quantitative insight into flow interactions. This visualization enables the identification of inefficient movement patterns, workspace interferences, and potential coordination issues that are not observable in traditional activity-based or 4D BIM simulations.

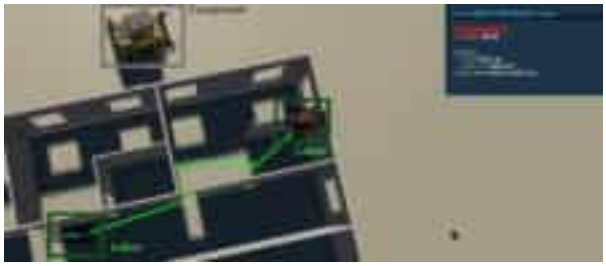


Figure 5. Visualization of labor movement and flow interactions with associated performance metrics.

5.1 Workspace Visualization

The simulation of workspace flows using the Activity-Flow–Based Method (AFM) enables the explicit definition of project sectorization. In the illustrative example, individual rooms (Room 1 to Room 6) and the corridor are clearly identified, along with the available area associated with each space.

In addition, the proposed framework incorporates the definition of activity construction areas, making it possible to explicitly represent the space required to execute each activity. This allows planners to assess whether multiple activities can be performed simultaneously within the same workspace without spatial conflicts.

Based on this information, AFM-based simulation strengthens planning and control processes by enabling project managers to make execution decisions aligned with available workspace conditions. This helps prevent schedules that overlook spatial constraints or rely on unrealistic activity durations derived from work crews that collide or interfere with one another, which can lead to delays for one or multiple crews.

5.2 Material Visualization

Material visualization enabled the identification of material storage areas, which determine both the location and the transportation distances of materials within the construction site toward each activity to be executed. In addition, it facilitated a clearer understanding of the material requirements associated with each activity.

This capability supports project managers in optimizing construction layouts, defining storage zones, and sequencing activities to reduce internal material transportation and establish material storage areas.

5.3 Labor Visualization

Labor visualization is complemented by workspace visualization, enabling the explicit representation of crew sizes and the observation of how labor moves across different zones to execute specific activities. This enhances the understanding of activity relationships, particularly start–start precedence relationships, which can be tracked through labor movements. For example, start–start relationships between structural and masonry activities can be defined through precedence logic and reinforced by visualizing crew movements. In such cases, minimum separation distances between crews can be enforced—for instance, masonry crews should remain at least two floors behind structural crews—and alerts can be generated when these constraints are violated, such as when masonry crews exhibit higher productivity rates.

Regarding finish–start precedence relationships, labor visualization also provides valuable visual support in scenarios involving multiple crews engaged in the same or interdependent activities, improving coordination and sequencing clarity.

In addition, the proposed framework incorporates a physical representation of labor within the Unity simulation environment. This enables the visualization of crew mobilization and movement paths and supports the identification of interferences or conflicts between circulation routes and activity execution areas using bounding or shape-based representations around workers.

Overall, this capability supports a clearer definition of activity and crew sequencing by facilitating the design of efficient internal circulation paths and the explicit definition of workspace buffers, particularly for start–start relationships. Similarly, for finish–start relationships, the simulation enables more effective planning of crew entry into the construction site and individual workspaces, reducing waiting times and minimizing labor-related inefficiencies and costs.

5.4 Equipment Visualization

Through the proposed framework, the simulation represents and facilitates the understanding of equipment transportation and movement for planning and optimization purposes, as well as equipment availability across activities. In addition, it enables the analysis of internal equipment circulation within the construction site.

Consequently, this capability provides a clearer perspective for defining activity trains that ensure the availability of the required equipment at activity start times, playing a relevant role in equipment resource allocation within the scheduling process.

6 Discussion

The primary contribution of this research is the extension of the Activity-Flow Model (AFM) through the explicit visualization of flows as they physically move on-site while being consumed and released by activities. By embedding activity- and flow-based planning logic within a spatial simulation environment, the proposed framework makes visible how workspace, materials,

labor, and equipment are mobilized, circulate, and interact with the physical layout of the construction site during execution. This shifts planning representations from abstract descriptions of activities and resource assignments toward a physically grounded understanding of how work is actually carried out in space.

From a Lean Construction perspective, this contribution directly supports Koskela's Transformation–Flow–Value (TFV) theory by operationalizing the flow view of production [18]. While Lean theory emphasizes the importance of managing flow continuity and reducing flow interruptions, planning methods often lack mechanisms to explicitly represent how flows manifest and interact on-site. The proposed framework bridges this gap by making AFM flows visible as physical phenomena, enabling planners to observe how transformations (activities) depend on and are constrained by the availability, movement, and interaction of flows within the construction environment.

The results show that the explicit representation of AFM flows provides significant value for planning. Even in a simplified pilot scenario, the framework reveals conflicts, circulation constraints, and infeasible sequencing conditions that are not apparent in precedence-based schedules.

Compared to traditional 4D BIM approaches, which primarily focus on activity sequencing, the proposed framework enables the explicit representation and analysis of flow interactions. This allows the identification of workspace conflicts, resource interferences, and flow inefficiencies that are not detectable in conventional 4D simulations.

By showing where flows originate, how they move, where they accumulate, and when they are released, the framework supports more realistic reasoning about workspace availability, internal logistics, and crew coordination. In this sense, visualization functions not merely as a communication aid, but as a planning instrument that externalizes assumptions related to flow availability, continuity, and interaction.

The framework is particularly valuable during phase and look-ahead planning, when sequencing, layout, and resource decisions can still be iteratively refined. At this stage, simulation outputs support informed trade-offs among activity sequencing, flow allocation, and spatial configuration, allowing planners to align execution strategies with Lean objectives such as reducing waiting, unnecessary movement, and interference among trades.

Beyond visualization, the framework establishes a foundation for deeper integration between simulation and Lean Construction research. While the present study focuses on making flows visible, the spatial representation of labor, equipment, and material as situated entities enables future exploration of how flow interactions give rise to emergent behaviors. From a TFV perspective, such behaviors reflect the dynamic coupling between transformations and flows, where local disruptions may propagate across the system and affect overall value generation. Future work will therefore extend the framework toward agent-based simulation of interacting flows, with the objective of understanding and mitigating flow disruptions and supporting the design of construction plans that are more robust, reliable, and aligned with Lean principles.

7 Limitations

This framework represents a pilot implementation for the visualization and simulation of activities and flows and, as such, its applicability cannot yet be generalized. The framework was demonstrated using a simplified activity sequence within a conceptual residential project, which limits the ability to extrapolate the findings to real projects or to different construction typologies and site conditions.

8 Conclusion

The proposed framework extends the Activity-Flow Model (AFM) by enabling the explicit visualization of construction flows as they move and interact within a spatial simulation environment. By linking activity and flow information with 3D geometry and agent-based simulation, the framework provides a physically grounded representation of construction processes, supporting a more realistic understanding of how activities are executed on site.

The results demonstrate that even in a simplified pilot scenario, the visualization of flows allows the identification of conflicts and inefficiencies that are not evident in traditional precedence-based schedules. In particular, the simulation revealed labor collisions, inefficient travel paths, and workspace interferences that would remain implicit in conventional planning representations.

Compared to traditional 4D BIM approaches, which primarily focus on activity sequencing, the proposed framework enables the explicit representation and analysis of flow interactions. This provides planners with additional insight into spatial constraints, resource coordination, and internal logistics, supporting more informed decision-making during planning and control processes.

From a Lean Construction perspective, the framework contributes to operationalizing the flow dimension of production by making flow behavior observable within the execution environment. This supports the identification of waste related to movement, waiting, and interference among trades, particularly during phase and look-ahead planning.

The framework is especially valuable during early planning stages, where simulation outputs can be used to iteratively refine activity sequences, flow allocation, and site layout. This enables planners to align execution strategies with actual site conditions and reduces the risk of flow-related disruptions during construction.

References

1. Formoso, C.T., Bernardes, M., Oliveira, L.F.: Developing a Model for Planning and Controlling Production in Small Sized Building Firms Proceedings. Presented at the (1998).
2. Zeng, N., Dichtl, M., König, M.: A scenario-based simulation framework of on- and off-site construction logistics. In: Proceedings of the 2017 Winter Simulation Conference. IEEE Press (2017).
3. Koskela, L.J.: Management Of Production In Construction: A Theoretical View. (1999).
4. Koo, B., Fischer, M., Kunz, J.: A formal identification and re-sequencing process for developing sequencing alternatives in CPM schedules. Autom. Constr. 17, 75–89 (2007). https://doi.org/10.1016/j.autcon.2007.03.005.
5. Ballard, H.G.: THE LAST PLANNER SYSTEM OF PRODUCTION CONTROL, (2000).
6. Tommelein, I.D., Singh, V. V, Coelho, R. V, Lehtovaara, J.: So Many Flows! In: Proc. 30th Annual Conference of the International Group for Lean Construction (IGLC). pp. 878–889. , Edmonton, Canada (2022). https://doi.org/10.24928/2022/0199.
7. Kenley, R., Seppänen, O.: Location-based management of construction projects: part of a new typology for project scheduling methodologies. In: Winter Simulation Conference. pp. 2563–2570. Winter Simulation Conference (2009).
8. Garcia-Lopez, N.P., Fischer, M.: Managing On-Site Production Using an Activity and Flow–Based Construction Model. J. Constr. Eng. Manag. 150, (2024). https://doi.org/10.1061/jcemd4.coeng-13643.
9. Salhab, D., Pourrahimian, E., Lindhard, S.M., Hamzeh, F.: Patterns, 4D Simulations, and Artificial Intelligence–Driven Insights: Redefining Construction Workspace Management. J. Constr. Eng. Manag. 151, (2025). https://doi.org/10.1061/jcemd4.coeng-16062.
10. Mi, Z., Li, J.: Maximizing project efficiency and collaboration in construction management through building information modeling (BIM). Applied and Computational Engineering. 72, 24–29 (2024). https://doi.org/10.54254/2755-2721/72/20240986.
11. Sacks, R., Brilakis, I., Pikas, E., Xie, H.S., Girolami, M.: Construction with digital twin information systems. Data-Centric Engineering. 1, (2020). https://doi.org/10.1017/dce.2020.16.
12. Garcia-Lopez, N.P.: AN ACTIVITY AND FLOW-BASED CONSTRUCTION MODEL FOR MANAGING ON-SITE WORK, http://purl.stanford.edu/jw488vx5801, (2017).
13. Tommelein, I.D.: Work Density Method for Takt Planning of Construction Processes with Nonrepetitive Work. J. Constr. Eng. Manag. 148, (2022). https://doi.org/10.1061/(asce)co.1943-7862.0002398.
14. Ismail, A., Srewil, Y., Scherer, R.: Integrated and collaborative process-based simulation framework for construction project planning. International Journal of Simulation and Process Modelling. 12, 42 (2017). https://doi.org/10.1504/IJSPM.2017.082789.
15. Rad, M.N., Maghrebi, M.: Bi-level integration of discrete-event simulation (DES) and social force modeling (SFM) for construction labor productivity. Autom. Constr. 160, 105317 (2024). https://doi.org/https://doi.org/10.1016/j.autcon.2024.105317.
16. Zankoul, E., Khoury, H., Awwad, R.: Evalutation of Agent-Based and Discrete-Event Simulation for Modeling Construction Earthmoving Operations. In: Malaska, M. and Heikkilä, R. (eds.) Proceedings of the 32nd International Symposium on Automation and Robotics in Construction and Mining (ISARC 2015). pp. 1–8. International Association for Automation and Robotics in Construction (IAARC), Oulu, Finland (2015). https://doi.org/10.22260/ISARC2015/0014.
17. Garcia-Lopez, N.P., Fischer, M., Alarcón, L.F.: Work Structuring for Flow. In: Proc. 27th Annual Conference of the International Group for Lean Construction (IGLC). pp. 311–322. , Dublin, Ireland (2019). https://doi.org/10.24928/2019/0140.
18. Koskela, Lauri.: An exploration towards a production theory and its application to construction, (2000).

Riverbed Depth Measurement Method Combining Visual SLAM-Based Camera Pose and Ultrasonic Slant Range

Norihito Kishi[1]

[1]Civil Engineering Research Institute for Cold Region, Japan
kishi-n22aa@ceri.go.jp

Abstract –

Heavy rains and flooding due to global warming are becoming more frequent, increasing the risk of scouring and riverbed deformation. Low-cost and simple techniques can be effective for detecting these damages at an early stage, however, water currents, turbidity, and waves make surveys difficult. Therefore, we devised a method to measure riverbed depth with spatial distribution by combining the device pose estimated by Visual SLAM with the slant range measured using underwater ultrasonic. This paper describes the results of land-based tests aim to integrate pose data and slant range data.

Keywords –

Riverbed Depth; Visual SLAM; Camera Pose; Ultrasonic; Slant Range

1 Introduction

In recent years, the frequency of heavy rainfall events has increased due to global warming in Japan [1]. Heavy rainfall that causes river flooding increases the risk of scouring around bridge piers and riverbed deformation, which is a global problem [2]. To mitigate these damages, it is important to understand the current conditions, and various survey techniques have been developed. For example, ALB (Airborne LiDAR Bathymetry) can efficiently survey wide areas with high precision, but it faces issues such as susceptibility to turbid water, the need for skilled operation, and high equipment costs [3]. A method that combines a remote-controlled boat and sonar, as in [4], can safely survey wide areas, but there are issues such as limited maneuverability in fast-flow environments. Many small, inexpensive remotely operated vehicles have been sold, and their use for infrastructure inspections has become more widespread, but underwater surveys in rivers with currents and turbidity remain difficult. A simpler and cheaper method is a float-type sonar with GNSS. In Japan, the Ministry of Land, Infrastructure, Transport and Tourism is working on applying it to disaster surveys, and its low cost and ease of use compared to other technologies are considered factors in its adoption. Furthermore, as an advanced version of this, research has been conducted on tethering the sonar using an unmanned aerial vehicle for bathymetry observation of inland water bodies [5]. We have also used a similar sonar and confirmed that it can measure water depth with an error of a few centimeters in flatwater. However, there are issues such as inaccurate depth measurement caused by surface waves, and poor GNSS reception near bridges or other structures.

For these reasons, we considered that a survey technique with "precision and data density comparable to float-type sonar, inexpensive, simple, and unaffected by water currents and waves" would be useful. Therefore, we devised a method to measure riverbed depth with spatial distribution by combining device pose estimated by Visual SLAM (Simultaneous Localization and Mapping) with slant range measured by underwater ultrasonic [6]. This paper describes the results of land-based tests that aim to integrate trajectory (a time-series of poses) and slant range data.

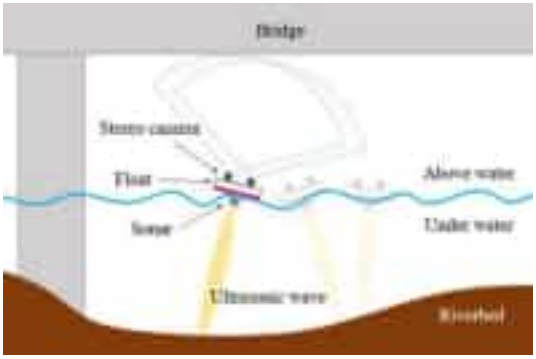


Figure 1. The schematic of proposed method

2 The Overview of Proposed Method

As mentioned above, the sway of device caused by water currents and waves makes surveys difficult. Therefore, we devised a method to measure riverbed depth spatially by swinging the ultrasonic beam transmitted from the sonar in multiple directions, taking advantage of the device's sway, and combining the device pose with underwater slant range data. A schematic of the proposed method is shown in Figure 1.

The measurement device is expected to have a simple configuration consisting of a stereo camera mounted on the top of the float, a SBC (Single Board Computer) inside the float, and a sonar attached to the bottom of the float. When surveying areas near structures such as under bridges, GNSS reception may be poor, but the structure is likely to be within the camera's field of view. Figure 2 shows an example of the underside of a bridge, which contains many feature points. For this reason, we considered that by specifically targeting environments containing structures that provide sufficient visual features, the camera pose obtained by Visual SLAM could serve as the device's pose.



Figure 2. An example of the underside of a bridge



Figure 3. The measuring unit

Table 1. Devices

Device	Specification
SBC	Raspberry Pi 5, 8GB
Stereo camera	ELP-USB3DGS1200P01-H120, angle of view 120 deg
Ultrasonic distance sensor	HC-SR04

3 Measuring Unit and Analysis

3.1 Measuring Unit

The measuring unit and devices are shown in Figure 3 and Table 1. The camera was fixed to the top of the SBC with its optical axis facing vertically upward, intended to capture images of the underside of the bridge. The ultrasonic distance sensor was fixed to the underside of the SBC with its axis facing vertically downward, intended to measure the riverbed depth. Since this was a land-based test to verify the feasibility of the method, we used an airborne ultrasonic distance sensor instead of an underwater sonar.

3.2 Measurement and Analysis

The SBC was used for data acquisition only, and the SBC and smartphone were connected via Wi-Fi, allowing the measurement application to be launched and operated on the smartphone. The data were post-processed on another laptop PC. The laptop PC ran Ubuntu 20.04, and ORB-SLAM3 [7] was used for Visual SLAM. The stereo_euroc program included in ORB-SLAM3 was used to perform the SLAM processing, and the depth was calculated using formula (1).

$$depth = camera_pos + R_q \cdot sensor_dir \cdot sr \quad (1)$$

where
camera_pos is the vector of camera position,
R_q is the rotation matrix obtained from the camera orientation quaternion [q_x, q_y, q_z, q_w],
sensor_dir is the vector of the ultrasonic sensor relative to the camera coordinate system,
as shown in equation (2), (3) and (4), respectively, and *sr* is the slant range.

$$camera_pos = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} \quad (2)$$

$$R_q = \begin{bmatrix} 1-2(q_y{}^2+q_z{}^2) & 2(q_xq_y-q_zq_w) & 2(q_xq_z+q_yq_w) \\ 2(q_xq_y+q_zq_w) & 1-2(q_x{}^2+q_z{}^2) & 2(q_yq_z-q_xq_w) \\ 2(q_xq_z-q_yq_w) & 2(q_yq_z+q_xq_w) & 1-2(q_x{}^2+q_y{}^2) \end{bmatrix} \quad (3)$$

$$sensor_dir = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \quad (4)$$

For camera calibration, OpenCV's pinhole camera model [8] was used.

4 Position Accuracy Validation

4.1 Experimental Conditions and Setup

In order to evaluate this measuring unit, we conducted simple tests. The schematic of the experimental setup is shown in Figure 4. The measuring unit was moved between Y = 0.00 m and Y = 1.00 m along the rail while recording images and slant range data. The ultrasonic distance sensor transmitted vertically downward so that the distance to the floor were measured. The distance between the sensor and the floor measured with a steel tape, was 0.77 m.



Figure 4. The experimental setup

4.2 Experimental Result

The position estimated by ORB-SLAM3 is shown as the red line in Figure 5. The actual traveling distance in the Y direction was 1.00 m, while the estimated distance was 0.898 m. The actual change in the Z direction was 0.00 m, whereas the estimated change was 0.036 m at Y = 0.898. The average absolute trajectory error of Stereo SLAM on the ORB-SLAM3 EuRoC dataset is reported to be 0.084 m [7], and our results in the Y direction in this experiment were similar. On the other hand, the estimated Z value tends to increase with increasing distance from Y = 0.00. This was considered to be due to the use of the pinhole camera model for camera calibration. Therefore, the estimated Z value was corrected using formulas (5) and (6). Coefficients a, b, c, d, e, and f were calculated using the least squares method. In this validation test, a, c, and d were fixed to zero because the measuring unit could not move in the X direction. The corrected Z value is shown as the blue line, with Z = 0.004 at Y = 0.898.

$$Z_{fit} = ax^2 + by^2 + cxy + dx + ey + f \quad (5)$$

$$Z_{corrected} = Z_{estimated} - Z_{fit} \quad (6)$$

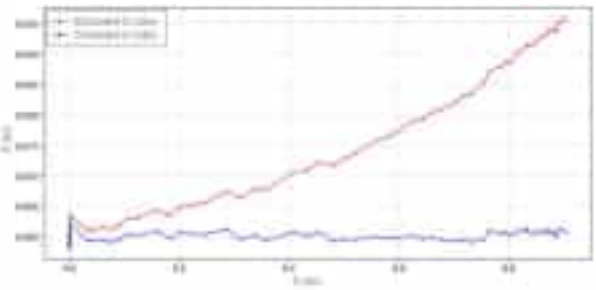


Figure 5. Changes of Z value

5 Depth Measurement

5.1 Experimental Conditions and Setup

This measuring unit was set on a flatbed cart, and



Figure 6. The experimental setup

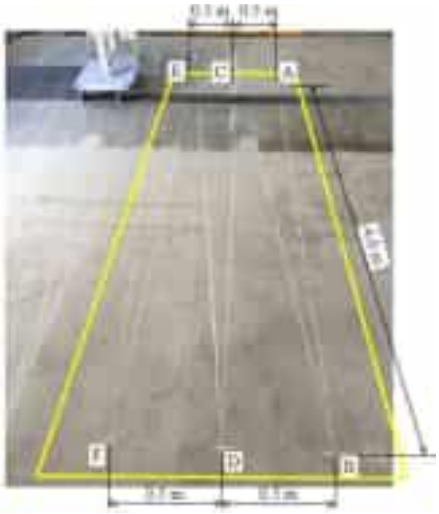


Figure 7. Marker lines

depth data (the distance between the ultrasonic distance sensor and the floor) was measured while the cart was moving in a laboratory. The experimental setup is shown in Figure 6. The measuring unit is expected to be used while floating on the water surface with waves. Therefore, we built a stand using an aluminum frame on a cart and suspended the measuring unit from the stand with polyester string and rubber bands to allow it to swing more easily. Marker lines were drawn on the floor of the laboratory as shown in Figure 7, and measurements were taken by moving the cart through points A to F in order, with the measuring unit passing over the mark lines. The floor was made of concrete and was flat but had small irregularities, and while the cart was moving, the measuring unit shook by a few centimeters. The resolution of the input images was 1280 × 720 pixels, with a frame rate of 10 fps. The distance from the bottom of the ultrasonic distance sensor to the floor was measured with a surveying staff and was 1.20 m.

5.2 Experimental Result

The results are shown in Figure 8. The blue lines represent the corrected trajectory of the measurement unit, the red lines represent the slant range from the ultrasonic distance sensor to the floor surface, and the green dots indicate the depth calculated using formula (1). The contour lines generated from the depth data are also

displayed. The points A' and F' are the positions of the measuring unit when the cart started and stopped moving, respectively, and points B' to E' are the positions of the measuring unit when the cart changed direction. Each point corresponds to point A to F in Figure 7. There are loop-shaped sections of the trajectory near points C' and E', which occurred when the cart changed direction by turning its wheels, and they closely represent the actual movement. The minor fluctuations in the trajectories represent the sway of the measuring unit. The Z value at point F', the farthest point from the starting point, was 0.01 m, and the distortion was significantly reduced through the correction process. The depth data had an average of -1.17 m, a minimum of -1.21 m, a maximum of -1.14 m, and a standard deviation of -0.01 m.

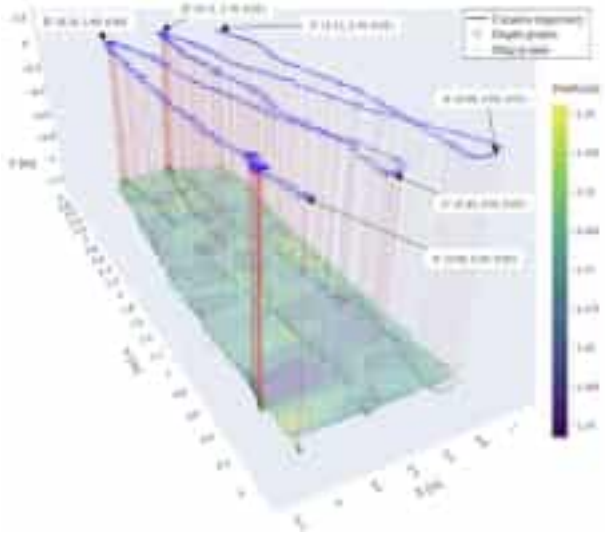


Figure 8. The trajectory and depth points

6 Conclusion

We devised a method to measure the riverbed depth around bridges and other structures in spatially by combining the pose data obtained through Visual SLAM with the underwater slant range data, and verified its feasibility through land-based tests. The result showed that even in images taken when the measuring device is shaking by a few centimeters, pose estimation by SLAM works, and depth can also be calculated from the slant range. Regarding horizontal accuracy, the error was approximately 0.1 m; however, since these results were obtained under limited indoor conditions, further verification is required to assess the generalization of the drift correction over a wider area and under actual bridges. Furthermore, by correcting the distortion of the SLAM output data using a two-variable quadratic polynomial, it became possible to estimate depth on the order of centimeters. However, since factors contributing to errors in the depth data (e.g., pose error, range noise, and mounting geometry) and the effects of actual underwater conditions (e.g., water waves and bottom return characteristics) remain to be investigated, we plan to conduct underwater tests using a underwater sonar to verify these factors. The sampling density along the trajectory provided by Visual SLAM is on the order of several centimeters. However, the effective spatial resolution of the depth map is constrained by the sonar's footprint. For instance, a 7-degree beam width used for a float-type sonar with GNSS, results in a footprint of 0.25 m at a depth of 2 m and 0.61 m at 5 m. Since the SLAM-based sampling density is finer than these footprint sizes, the system's overall horizontal resolution is dominated by the sonar's beam width.

References

[1] Ministry of Education, Culture, Sports, Science and Technology, Japan Meteorological Agency. Climate Change in Japan 2025. On-line: https://www.data.jma.go.jp/cpdinfo/ccj/2025/pdf/cc2025_gaiyo_en.pdf, Accessed: 04/12/2025.

[2] G. Zhang, Y. Liu, J. Liu, S. Lan and J. Yang. Causes and statistical characteristics of bridge failures: A review. Journal of Traffic and Transportation Engineering (English Edition), 9(3): 388–406, 2022.

[3] R. Tsubaki, T. Moramarco, G. Maniatis, J. L. Coz and M. Muste. Advances in river measurement technologies. New Waves in Hydrometry, Vol.1, 36-59, 2025.

[4] R. Shuchman and G. Meadows. BathyBoat. On-line: https://www.mtu.edu/mtri/research/project-areas/environmental/water/bathyboat/factsheet.pdf, Accessed: 10/12/2025.

[5] F. Bandini, D. Olesen, J. Jakobsen, C. M. M. Kittel, S. Wang, M. Garcia and P. Bauer-Gottwein. Bathymetry observations of inland water bodies using a tethered single-beam sonar controlled by an Unmanned Aerial Vehicle, Hydrology and Earth System Sciences, 22, 4165–4181, 2018.

[6] N. Kishi. Study of Riverbed Depth Measurement Method by Integrating Camera Posture Data Obtained by Visual SLAM and Slant Range Data Obtained Underwater, The 23th Symposium on Construction Robotics in Japan, P1-27, 2025 (in Japanese).

[7] C. Campos, R. Elvira, J. J. Gómez Rodríguez, J. M. M. Montiel and J. D. Tardós. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM, IEEE TRANSACTIONS ON ROBOTICS, VOL.37, NO.6, pp. 1874 – 1890, 2021.

[8] OpenCV. Camera Calibration. On-line: https://docs.opencv.org/4.x/dc/dbb/tutorial_py_calibration.html, Accessed: 10/12/2025.

Trial of Point Cloud Registration Method at Excavation Sites and Implementation of Machine Guidance System

Kai Sato[1], Takashi Yokoshima[1], Yuki Miyashita[1], Aoi Tarutani[1], Keisuke Imoto[1], Fuku Himuro[1], Yutaro Fukase[1], Yusuke Nishida[2], Yusuke Shibata[3], and Toshiki Aizawa[3]

[1]Institute of Technology, Shimizu Corporation, Japan
[2]Tokyo Civil Engineering Division, Shimizu Corporation, Japan
[3]East Nippon Expressway Company Limited, Japan
sato_k_213279@shimz.co.jp

Abstract–

Machine guidance using total stations (TSs) and global navigation satellite systems (GNSSs) is implemented to reduce labor and improve productivity in excavation work. However, a challenge with existing methods is that obstacles can interfere with guidance. Direct measurement of excavation surfaces using LiDAR can solve this problem, but it requires a point cloud registration method capable of functioning even in feature-poor environments, such as tunnels. We developed a registration method that can be used feasibly in excavation sites and built a system that visualizes the excavation status using LiDARs. The proposed robust registration method utilizes a specialized jig and a TS, specifically designed for feature-poor excavation sites. The system achieved a measurement accuracy of ±20 mm in field trials, satisfying the requirements set by site staff (±50 mm in every 250-mm pitch and a refresh rate of <30 s) and overcoming the limitations of conventional TS/GNSS methods in tunnel environments.

Keywords

LiDAR; Point cloud; Machine guidance

1 Introduction

To improve productivity in response to construction labor shortages [1], researchers have proposed machine guidance (MG) for presenting the excavation status to operators and machine control (MC) for autonomous excavation [2]–[4]. Systems detecting heavy machinery positions using global navigation satellite systems (GNSSs) or automatic-tracking total stations (TSs), and systems estimating boom and bucket postures via tilt sensors, are already being used in practice as MC and MG methods. In actual construction sites, especially open-cut sites, GNSSs may be unusable and struts or existing structures may obstruct optical signals, compromising MG/MC implementation. In addition, construction management is difficult because only the operator can view the MG/MC screen, so site staff cannot confirm the construction status in real time. A potential solution is to measure the excavation surface directly using dense point



Figure 1. System configuration



Figure 1. Tokyo Gaikan Expressway Keiyo Junction G ramp

cloud data obtain from 3D LiDAR (hereinafter "LiDAR") and transmit the data to the cloud in real time instead of estimating machine posture by tracking prisms. Although previous studies have proposed point cloud registration methods utilizing feature points and BIM, there are very few examples of these methods being applied to MG or autonomous construction [5]–[7]. Because point cloud registration may fail in feature-poor sites, such as tunnels, we propose a method for transforming point clouds from multiple LiDARs into the site coordinate system using a jig and TS measurements. We constructed a point cloud measurement system using this method in an actual open-cut tunnel site and performed measurements. This paper describes the proposed system and the evolution of measurement conditions.

2 System Configuration and Site Overview

Figure 1 illustrates the configuration of the proposed system. It comprised a LiDAR (Livox HAP [8]) for point cloud measurement, a PC (ASUS NUC) for data processing, and a Power over Ethernet hub and mobile router for data transmission to the cloud (AWS S3 server). The field of view of the LiDAR was 120° (horizontal) × 25° (vertical), and its detection range was 150 m. The system was installed at the Tokyo Gaikan Expressway Keiyo Junction G ramp construction site in Ichikawa City, Chiba Prefecture. In this open-cut tunnel construction site, excavation proceeds within diaphragm walls supported by struts. Figure 2 presents an overview photograph of the site. Measurements were conducted at three locations (Sections 1–3) according to the construction progress. The basic system configuration was consistent across all measurements, but the LiDAR installation and registration methods differed slightly between sections, as described in the subsequent subsections.


Figure 3. LiDAR and type I jig

3 Measurement Method and Result

3.1 Measurement in Section 1



Figure 4. LiDAR installation in Section 1

Measurements in Section 1 were conducted in July 2024. Point clouds were acquired using three LiDAR units (#110, #111, and #113) installed within an area of approximately 10 m × 10 m. Figure 3 shows the appearance of the LiDAR and the jig (type 1), and Figure 4 shows the LiDAR installation status. The LiDAR-obtained point clouds, which were in a local coordinate system centered on the LiDAR's optical origin, were transformed into the survey coordinate system. Mainstream methods for point cloud registration, such as RANSAC and ICP matching [9][10], estimate LiDAR position and orientation based on feature points. However, construction sites commonly have identical, equally spaced pillars or feature-poor environments, such as in tunnels.

We address this problem using the method illustrated in Figure 5. First, the TS instrument point was determined using prisms #1-1 and #1-2, which were placed at two or more known points. Next, three or more prisms or reflective sheets (#2-1 to #2-3) attached to the jig were measured in the site coordinate system. Finally,

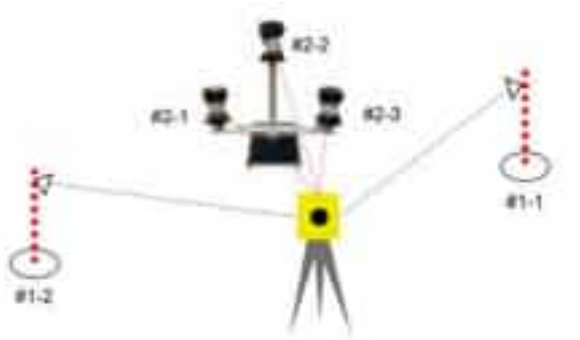


Figure 5. LiDAR position estimation

we estimated the LiDAR's optical origin and yaw, pitch, and roll angles by calculating the rigid-body transformation matrix required to align the coordinates of the prisms on the jig. Rigid-body transformations were used to determine the LiDAR origin and orientation from the three prisms. Let p_i denote the coordinates of the three prisms in the LiDAR coordinate system (where i is the prism number; i = 1, 2, 3), and let q_i denote the coordinates of the prisms in the common coordinate system. The relationship between p_i and q_i is given by Equation (1):

$$q_i = Rp_i + t \tag{1}$$

where R is a 3 × 3 rotation matrix and t is the translation vector of the LiDAR origin. R is obtained using the covariance matrix and singular value decomposition; further details are available in a previous study [11].

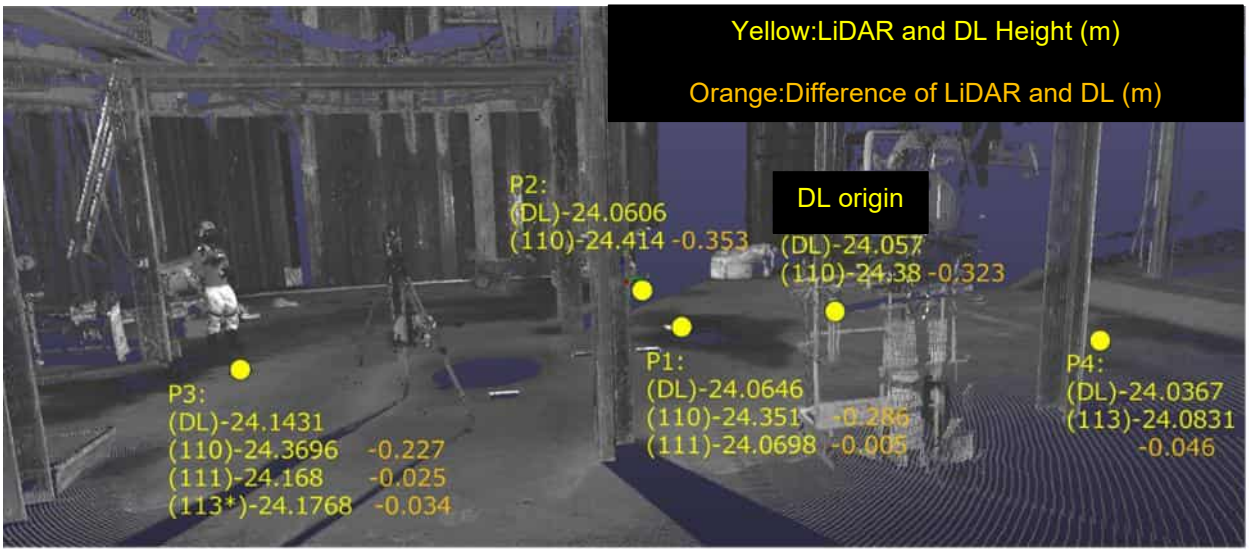


Figure 6. LiDAR and DL heights

From the obtained rotation matrix R, quaternions were calculated and the roll, pitch, and yaw angles were estimated [12][13]. Figure 6 compares the height of the acquired point cloud with the height measured by a digital level. The background in Figure 6 was obtained using a terrestrial laser scanner. The error was ~5 mm in high-accuracy areas but reached 400 mm in others. The excavation depths at this site were currently monitored at fixed points approximately 10 m × 5 m grids with ±50 mm accuracy by level gauge, and similar or higher accuracy and resolution were expected from the MG system. The TS used for prism surveying (TOPCON GT-1003) had an error of 2.1 mm when measuring distances in the nonprism mode from 50 m away; hence, the LiDAR origin and orientation were calculated using coordinate values incorporating this error. Therefore, each coordinate in Equation (1) contained this error, potentially causing position and orientation estimation errors. Furthermore, for example, acquiring point clouds from a distance of 10 m with an accuracy of 20 mm or less requires an accuracy of 2/1000 ~ 0.11° or less. Consequently, the coordinate values of the prism, which were in millimeters, would not have enabled angle estimation with the necessary precision. Given the difficulty of estimating the LiDAR position and angle solely using the prisms attached to the jig, we concluded that sufficient MG/MC would not be feasible under these conditions.

Furthermore, installing the LiDAR at a near-horizontal angle relative to the ground surface resulted in significant point cloud occlusion by obstacles and required frequent adjustments to the device angle and position as the excavation deepened. Therefore, we explored an installation method that would enable continuous measurement of the same area to the greatest extent possible.

3.2 Measurement in Section 2

Measurements in Section 2 were conducted from November 2024 to August 2025. The measurement area was approximately 10 m wide and 30 m long. One LiDAR unit was installed between each strut, for a total of seven units (#1–#7). For LiDARs #1–#5, the fabricated aluminum frame jig (type 2) differed from that in Section 1. Figure 7 shows this LiDAR jig, and Figure 8 shows the LiDAR installation status. The measurement coverage in Section 2 was improved relative to that in Section 1 by positioning the LiDAR near ground level and directing it toward the excavation surface; in Section 1, it was placed at a near-horizontal angle near the excavation surface.

For point cloud registration, in addition to calculating the LiDAR position and orientation from three or more prisms or reflective sheets (Section 1), we placed reflective sheets on the excavation surface during installation and measured their coordinates using a TS. We then fine-tuned the numerical values such that the LiDAR and TS-measured coordinates matched. Figure 9 illustrates the fine-tuning of the LiDAR orientation. The measured point cloud data were processed using a PC installed inside an on-site distribution board and transmitted to the cloud via an LTE connection. When performing MG, it is necessary to display point clouds of the excavation surface to the operator in real time and with the highest possible resolution. However, attempting to display the point cloud data as is presents challenges such as noise in the point clouds and the sheer volume of data. For this system, field construction staff requested a resolution of 250-mm intervals and a refresh rate of 30 s or less. Therefore, rather than displaying the point cloud data directly, we decided to accumulate the measured point cloud data within a 250-mm square grid for 30 s and display the average value, resolving the issues of point cloud noise and data transmission volume. The cloud data were visualized as contours and displayed on tablets in the heavy machinery operator's cab and on monitors in the site office every 30 s. Additionally, LAS data were uploaded to the cloud once a day for viewing in KOLC+ [14], balancing utility for MG/MC and convenience for daily construction management. Figure 10 shows the MG screen, and Figure 11 shows the visualization in KOLC+. Parts of the area are occluded by elements such as the horizontal struts. Figure 12 shows the progress of the final excavation in the measured area. The contours represent the difference between the measured data and the BIM model data; purple indicates high areas, yellow-green/green indicates areas within the control values, and blue indicates low areas. Evidently, the excavation surface lowers over time. Additionally, some areas appear high due to temporary soil stockpiles or heavy machinery. Activities occurring in parallel with excavation, such as the spreading of crushed stone, are also visible.

Regarding measurement accuracy, Figures 13 and 14 show a comparison of LiDAR and TS data, which was conducted during LiDAR installation, and a histogram of the errors relative to the TS data (data for #5 and #7 are missing), respectively. The errors were generally approximately ±20 mm, falling within 40 mm at most. For #6 and #7, relatively larger errors likely occurred because the jig was the same as in Section 1 and the installation dates were different. LiDAR accuracy may have varied due to the irradiation distance, strut displacement, and vibrations caused by the heavy machinery or wind. Moreover, manual fine-tuning was time consuming, and registration accuracy relied on operator subjectivity. These limitations highlighted the need for a method that would determine positions and angles with high precision, eliminating operator dependence.



Figure 7. LiDAR and type 2 jig

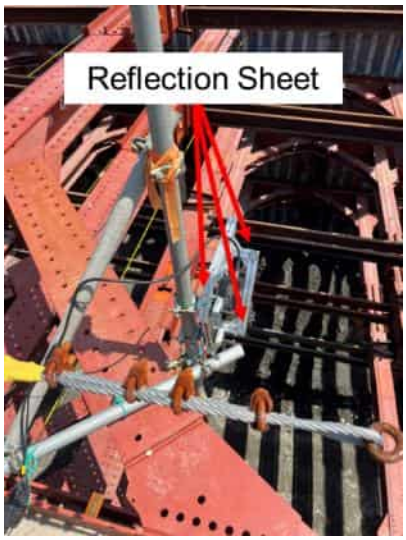


Figure. 8. LiDAR installation LiDAR in Section 2

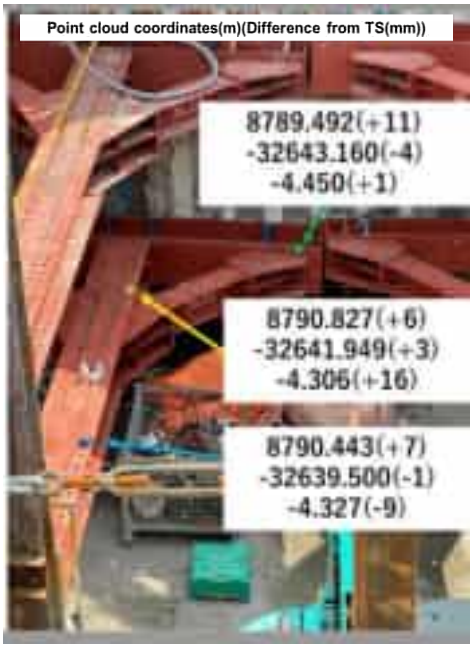


Figure 9. Adjusting of LiDAR position



Figure 10. MG display

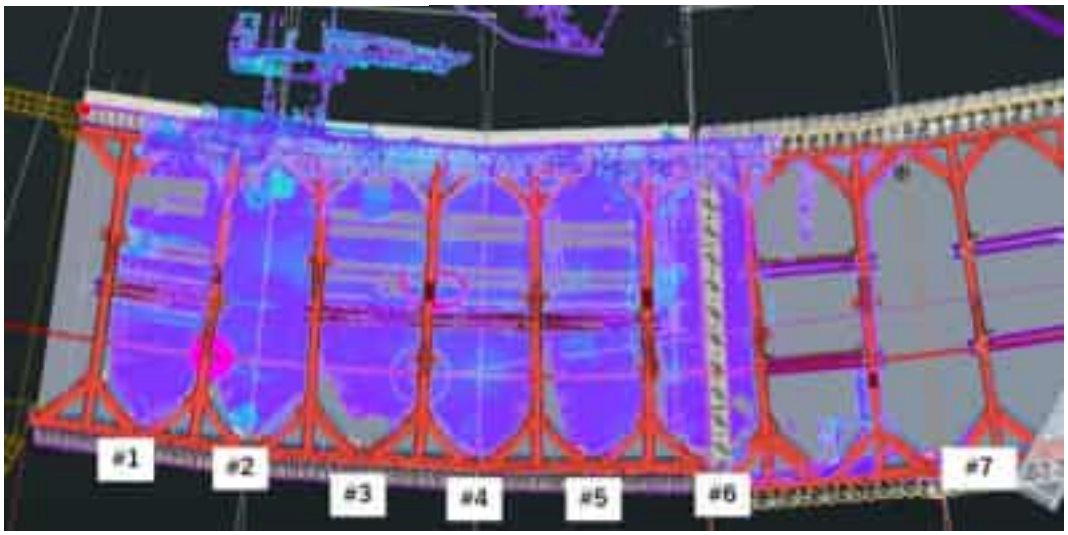


Figure 11. Point cloud visualization in KOLC+

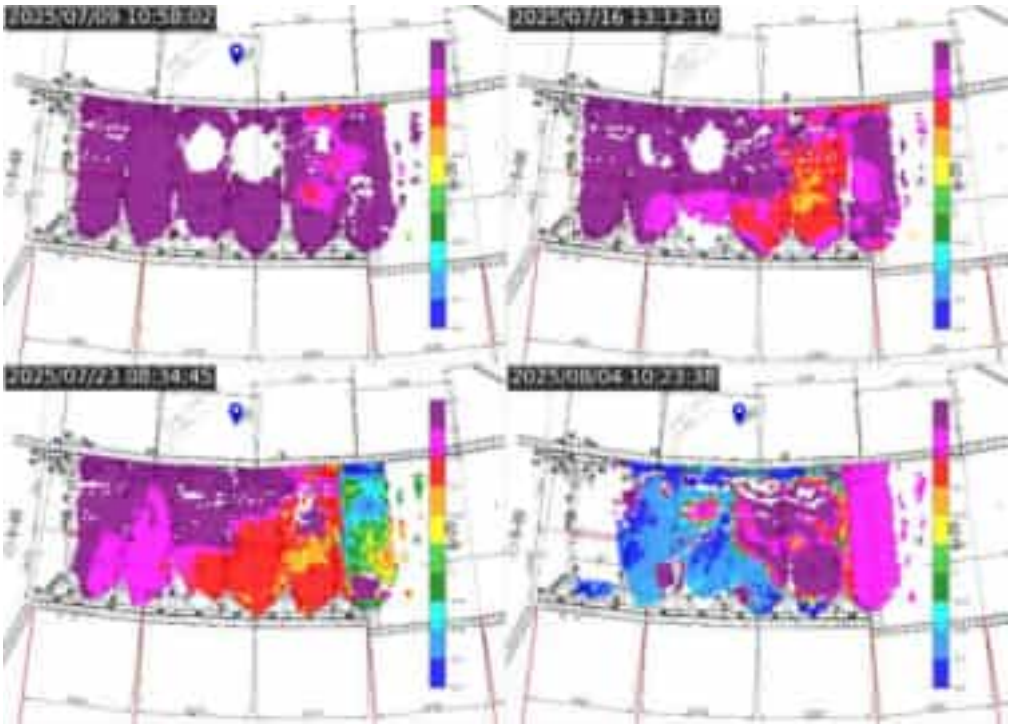


Figure 12. Progress of excavation

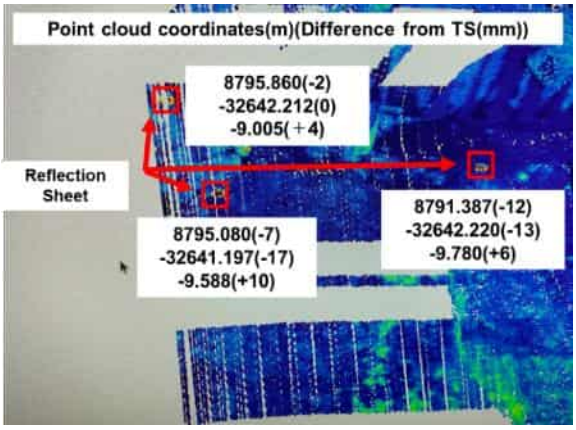


Figure 13. Accuracy verification

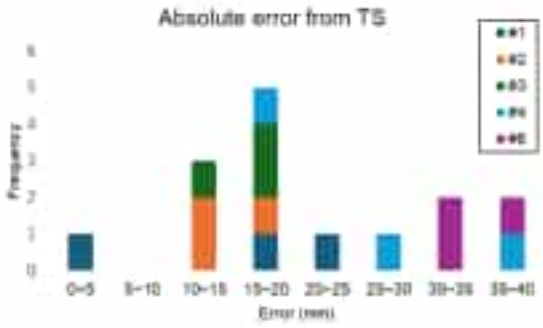


Figure. 14. Error histogram

3.3 Measurement in Section 3

Measurements were performed in Section 3 from October 2025 to January 2026. In Section 2, workability problems occurred due to the placement of reflective sheets on the excavation surface during installation to fine-tune the LiDAR orientation based on their coordinates. Therefore, we fabricated a type 3 jig (Figure 15) from sheet metal. It had a tribrach at the bottom, a circular bubble level on top, and two reflective sheets attached to its top and side. The roll and pitch angles were precisely set to 90° and 45°, respectively, by leveling the jig using the bubble level. Once the coordinates of the two reflective sheets were determined, the LiDAR's optical origin and yaw angle were calculated. The origin and angles (X, Y, Z, θ, φ, γ) of this LiDAR are expressed as shown in Equation (2) using the coordinates (X_1, Y_1, Z_1) of the upper reflective sheet and the coordinates (X_2, Y_2, Z_2) of the lateral reflective sheet.

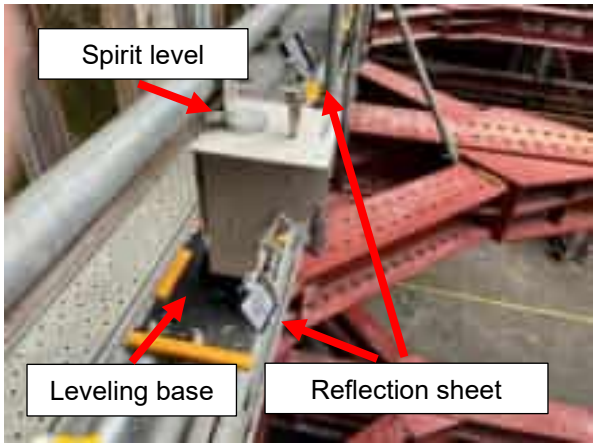


Figure 15. LiDAR and type 3jig

$$
\begin{aligned}
X &= X_1 - 0.005\ sin\gamma \\
Y &= Y_1 - 0.005\ cos\gamma \\
Z &= Z_2 - 0.005, \\
\theta &= -\frac{\pi}{2} \\
\varphi &= \frac{\pi}{4} \\
\gamma &= atan2((Y_2 - Y_1), (X_2 - X_1)).
\end{aligned} \tag{2}
$$

An example of the results is shown in Figure 16. Without manual fine-tuning, the height was generally within 20 mm, but the error exceeded 100 mm in the planar direction. This large error may have been resulted from the difficulty of determining the yaw angle using a prism (Section 3.1). In future, we plan to investigate methods for estimating yaw angles without using a prism.

4 Discussion and Future Work

Unlike conventional registration methods in general robotics applications, point cloud registration often relies on feature-based matching algorithms, such as RANSAC and ICP [6][7]. However, the environments in open-cut tunnel construction sites have repetitive geometric features (e.g., equally spaced earth-retaining struts) and feature-poor surfaces (e.g., flat excavation bottoms or shotcrete walls). In such environments, conventional algorithms tend to fall into local minima, leading to significant registration errors. Furthermore, construction management requires not only relative-map merging but also transformation of point clouds into the absolute site coordinate system for comparisons with BIM model data. Our proposed method, which utilizes a jig and TS measurements, ensures robust point cloud transformation into the site coordinate system, regardless of the geometric features of the environment. This reliability is a critical requirement for practical construction management and MG.

The measurement results from Sections 1 and 2



Figure 16. Accuracy verification of type-3 jig

highlighted the significant impact of LiDAR installation geometry on measurement accuracy. In Section 1, the LiDAR was installed at a low position, at a near-horizontal incidence angle relative to the excavation surface. This configuration caused two issues: (1) significant occlusion by obstacles due to the shallow angle and (2) frequent adjustment of the LiDAR angle with each excavation step. By contrast, the setup in Section 2, where the LiDARs were mounted facing downward on the struts, minimized occlusions and ensured a high incidence angle. The achieved accuracy of approximately ±20 mm in Section 2 satisfied the general tolerance for rough excavation, confirming the system's applicability to MG.

The primary challenge identified in Section 2 was the operational complexity of fine-tuning. This method required installing three or more reflective sheets on the excavation surface and performing labor-intensive calculations of the rigid-body transformation matrix, either manually or via programming, which disrupted construction activities. The type 3 jig addressed this issue by incorporating the bubble level and the fixed tribrach. By mechanically constraining the roll and pitch angles to known values (leveling), we simplified the positioning problem to determining only the translation and yaw angle. This improvement will significantly reduce installation times and eliminate the need to access the excavation floor, thereby enhancing safety and operational efficiency.

Static accuracy was verified while the system operated in a dynamic environment. The slight errors observed in Section 2 (Figures 13 and 14) may be attributed to sensor noise and strut vibration caused by heavy machinery and the minute displacement of the struts over time. Because the LiDAR was mounted on the struts, any shoring deformation directly affected the point cloud coordinates. Future work must distinguish between sensor registration errors and actual structural displacement to utilize this system for the safety monitoring of earth-retaining walls.

5 Conclusion

This study proposes a practical LiDAR-based MG system designed for open-cut tunnel excavation sites, where GNSS signals may be unavailable and geometric features are repetitive. By integrating multiple LiDAR sensors with a TS using a specialized registration jig, we established a workflow for accurately transforming point cloud data into the absolute site coordinate system.

Our field trials demonstrated the critical importance of sensor placement and registration method. The initial trial (Section 1) suffered from occlusion and a low incidence angle, but the improved setup in Section 2—with the LiDARs mounted vertically on earth-retaining struts—achieved a measurement accuracy of approximately ±20 mm. This accuracy is well within the tolerance required for excavation, validating the system's effectiveness for real-time MG. Furthermore, the development of the leveling-enabled (type 3) jig addresses the operational bottleneck of manual fine-tuning, offering a solution that significantly reduces installation time and labor costs.

In conclusion, the proposed system enables precise construction management in environments with limited features. Furthermore, it enhances safety and facilitates efficient site management by minimizing the need for site foremen to enter excavation areas. Future research will verify the efficiency of the type 3 jig in ongoing projects and investigate its potential application to other measurements, such as those involved in concrete placement and the structural displacement monitoring of retaining walls, contributing to comprehensive site safety management.

References

[1] MLIT. I-construction 2.0. Online: https://www.mlit.go.jp/tec/constplan/content/001738240.pdf, Accessed: 26/12/2025.

[2] R. Heikkila and M. Jaakkola. Automation of road construction - the state of the art in Europe. In Proceedings of the 23rd International Symposium on Automation and Robotics in Construction (ISARC2006), pages 7–10, Tokyo, Japan, 2006

[3] T. Hachijo, et al. 3D measurement system for soil loading by an autonomous backhoe using OPERA. In Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC2025), pages 1501–1506, Montreal, Canada, 2025

[4] KOMATSU. 3D machine guidance. Online: https://www.komatsu.com/en-us/technology/smart-construction/3d-machine-guidance, Accessed: 26/12/2025.

[5] Y. Zhang, L. Lu, X. Luo, and J. Pan. Global BIM-point cloud registration and association for construction progress monitoring. Automation in Construction, Vol. 168, Part A, 105796, 2024.

[6] Z. Xu, B. Schwab, Y. Yang, T. H. Kolbe, and C. Holst. L2M-Reg: Building-level uncertainty-aware registration of outdoor LiDAR point clouds and semantic 3D city models. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 233, pages 467–486, 2026.

[7] J. Kim, J. Kim, S. Paik, and H. Kim. Point cloud registration considering safety nets during scaffold installation using sensor fusion and deep learning. Automation in Construction, Vol. 159, 105277, 2024.

[8] Livox Technology Company Limited. LIVOX HAP. Online: https://www.livoxtech.com/jp/hap, Accessed: 26/12/2025.

[9] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communications of the ACM, Vol. 24, No. 6, pp. 381–395, 1981.

[10] J. Paul and N. D. McKay. A method for registration of 3-D shapes. IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239–256, 1992.

[11] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-9, No. 5, pages 698–700, 1987.

[12] SciPy. Online: https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.transform.Rotation.from_matrix.html#id3, Accessed: 24/3/2026.

[13] F. L. Markley. Unit quaternion from rotation matrix. Journal of Guidance, Control, and Dynamics, Vol. 31.2, pp. 440–442, 2008.

[14] KOLG Inc. KOLC+ BIM/CIM sharing cloud. Online: https://kolcx.com/, Accessed: 26/12/2025.

An Integrated Product-Process Representation for Deriving Robotic Task Structures using Graph Transformations

Maikel J.S. Brinkhoff[1,2], **Sebastian Esser**[1,2], **and André Borrmann**[1,2]

[1]Chair of Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2]TUM Georg Nemetschek Institute - AI for the Built World, Technical University of Munich, Germany

maikel.brinkhoff@tum.de

Abstract -

This paper introduces a graph-based method for deriving robot-executable task structures from Building Information Models (BIM). While BIM models are widely employed to coordinate design and planning, they are seldom used to facilitate on-site automation. A key challenge is that they typically describe building elements and properties, but not the stepwise procedures required for robotic construction. Existing BIM-to-robot pipelines therefore encode process logic in external artifacts (e.g., schedules, scripts, skill libraries) and connect this logic to BIM elements through ad-hoc mappings, which limits reuse across elements and projects. We propose a workflow that integrates BIM-derived product data with reusable construction method templates and performs systematic refinement into task structures. The BIM model is represented as a labeled property graph, construction methods are modeled as parametrized construction process templates, and graph transformations instantiate element-specific task structures. These graphs capture ordering, dependencies, and required parameters, and are mapped to Behavior Trees (BTs), resulting in an execution-level control structure for on-site robotic tasks. A case study on a multi-story building illustrates that the workflow can be applied at building scale, generating element-specific task structures and corresponding BTs from a single, integrated product-process representation.

Keywords -

BIM; Graph Modeling; Task Decomposition; Behavior Tree; Construction Robotics

1 Introduction

In construction projects, digital design intent must be translated into explicit tasks before it can be executed. Digital models support this transformation by coordinating information across disciplines and project stages. Building Information Models (BIM) have become the dominant representation for this purpose [1], informing planning and scheduling [2] as well as on-site progress monitoring during execution [3]. At the same time, construction work remains largely manual, even as robotic systems are beginning to demonstrate their potential for on-site tasks [4].

Despite its broad use, BIM remains sparsely used to enable on-site automation. These models contain geometric descriptions of building components and semantic information ranging from component properties (e.g., materials) to higher-level objects such as stories and zones. However, they often lack the procedural details needed to derive robot-executable tasks. According to the taxonomy of Everett and Slocum [5], construction processes are typically defined at the activity level, whereas robotic execution needs fine-grained tasks and elemental motions. This mismatch in abstraction leaves a gap between the information available in BIM and the task-level information required for robotic execution.

Recent research has sought to derive actionable information for robotics directly from BIM. Examples include extracting work locations and object context from Industry Foundation Classes (IFC) geometry [6], or combining BIM with onboard sensing to perceive workpiece geometry and update motion plans under changing conditions [7]. While these studies demonstrate the value of BIM-derived information, they remain narrowly scoped. They often address a single task or building element and rely on manually engineered mappings from BIM entities to robot actions. This limits their transfer beyond the specific task setting for which the mapping was designed.

The underlying challenge is representational. Product data, schedules, and robot actions are typically stored in separate artifacts, such as schedules in CSV files, loosely defined process graphs, and hard-coded task scripts. This fragmentation prevents systematic reuse and the consistent capture of process logic at the building scale. Addressing this challenge calls for a unified, higher-level representation that links construction processes and their task structures across building elements, while remaining compatible with robotic execution-level control structures. Graph modeling is identified as a promising approach to represent products, processes, dependencies, and their evolution over time in a single formalism [8].

This paper proposes an integrated product-process representation that links BIM-derived product data, construc-

tion methods, and graph-based process modeling to systematically derive element-specific task structures. These structures remain connected to their source BIM data and are mapped to execution-level representations for robotic systems. The remainder of the paper discusses related research on BIM for construction automation and process modeling with graphs. It then presents the proposed workflow and graph representation and illustrates their application in a case study. It concludes with a discussion of contributions, limitations, and directions for future research.

2 Background

2.1 Deriving Robotic Tasks from BIM Data

Existing BIM-to-robot pipelines mainly differ in where procedural knowledge is represented and at what granularity it becomes executable. A common pattern is to encode process logic in an external artifact and use BIM primarily to supply geometric and semantic parameters. Zhu et al. [9] represent construction intent in a CSV schedule that specifies task sequences per element. Execution parameters, such as IDs or locations, are retrieved from the BIM model. Terzer et al. [10] similarly derive mission-specific parameters from IFC data for wall painting, where high-level missions are decomposed into subtasks and executed using Behavior Trees (BTs) [11]. These approaches demonstrate the value of BIM as a contextual data source for robotic execution. However, process logic is generally defined outside the BIM model and linked to BIM entities through manual mappings. This makes the resulting task structures application-specific and limits their reuse.

A second pattern brings task planning closer to the BIM environment by embedding task definitions in scheduling workflows and linking actions to robotic skills. Oyediran et al. [12] integrate robotic task planning into 4D BIM by associating actions with skill identifiers stored in an external database. This database also specifies additional modeling inputs, such as pickup locations or installation reference frames, which have to be manually added to the BIM model. While this improves traceability and ties tasks more directly to building elements, it still depends on manual modeling inputs and does not provide a general, reusable representation of process logic.

Across these approaches, BIM serves as a repository of product data and execution parameters, while procedural knowledge is spread across heterogeneous artifacts (e.g., schedules, mission scripts, skill libraries, BTs). This split complicates the systematic refinement of product-level descriptions into reusable task structures. It therefore motivates representations that connect element context with explicit process logic in a unified form.

2.2 Graph Representations in BIM and Construction Robotics

Graph representations have increasingly been explored as a way to integrate heterogeneous information and to make relationships between products, processes, and resources explicit. Knowledge graphs are used to support semantic integration and reasoning by translating IFC-derived scene information into graph form and linking it to task and agent models. Chen et al. [13], for example, propose the iSTA (Scene-Task-Agent) model for inspection tasks. The model represents scene context extracted from IFC in a Resource Description Framework (RDF) format combined with robotics ontologies and a custom task ontology to support reasoning and task allocation. In a related fashion, Zhu et al. [9] use a Linked Building Data (LBD) approach to integrate building models and construction schedules by converting IFC and CSV schedule data into RDF graphs. The information stored in the LBD graph is queried by a Robot Construction Action Node (RCAN) system, which serves as an intermediate layer between BIM and robotic execution in ROS and supports task allocation and execution. These approaches highlight how graphs can unify and query contextual data, rather than generating task structures for robotic execution.

Beyond knowledge graphs, task graphs have been used to encode process dependencies more directly. Atanasova et al. [8] introduce a directed graph model for cooperative human-robot assembly. Unlike prior works focused on inspection or static task sequences, their approach captures real-time task dependencies, agent assignments, and evolving assembly states. This example demonstrates how graph models can capture dynamic process logic and dependencies relevant for execution.

Taken together, prior work shows that BIM information can parametrize robotic actions, while graph representations can make process relations and dependencies explicit. Yet these research directions remain insufficiently connected at the step where procedural knowledge is linked to element context. What remains unclear is how this connection can be established systematically and represented in a form that supports robotic execution.

2.3 Control Structures for Execution

Behavior Trees (BTs) organize robotic behavior in a modular hierarchy that separates control flow from execution nodes and provides runtime reactivity [11]. Complex missions are decomposed into executable actions, organized by control logic and supplied with the parameters needed for those actions. This makes BTs relevant not only as an execution formalism, but also as an indicator of what task representations must contain to support robotic execution.

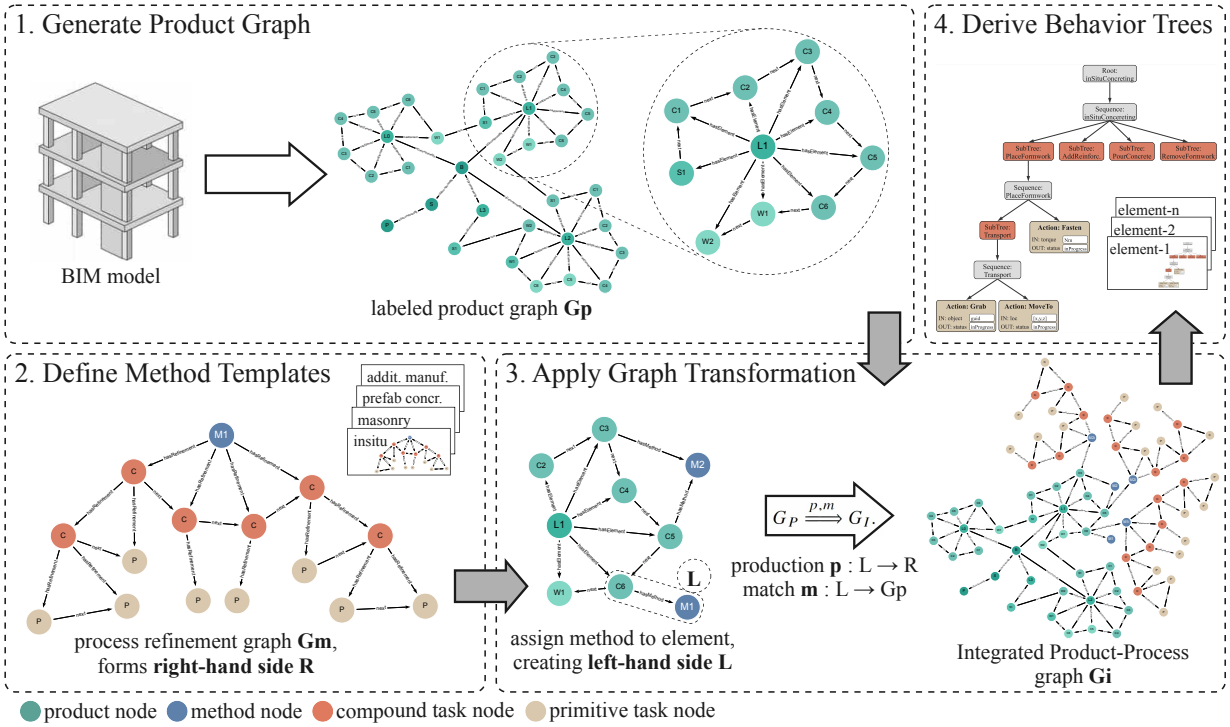


Figure 1. **Overview of the proposed method**: (1) represent a BIM model as a labeled product graph G_p; (2) decomposed construction processes as method templates G_m; (3) user-assigned methods link each product to its method and transform G_p into an integrated product-process graph G_I; and (4) derive Behavior Trees.

In construction robotics, BTs are used as execution structures for tasks parametrized from BIM. Terzer et al. [10] use BIM to provide semantic and geometric parameters for wall painting missions. Similarly, Tandon et al. [14] apply BTs to inspection missions, where BIM data provides target elements and inspection poses, and subsequently generate BTs for navigation and visual inspection. These examples show that BTs are a suitable target structure for describing construction processes. This motivates structuring the task graphs discussed in sec. 2.2 so that element context and process logic can be mapped to BTs, supporting a refinement workflow from BIM-derived product data to execution-level control.

3 An Integrated Product-Process Representation for Robotic Construction

This study proposes a unified workflow that uses graphs to link product data derived from BIM with procedural knowledge and refine this link into task structures suitable for robotic execution. As shown in Fig. 1, the workflow represents the BIM model as an LPG and models construction methods as parametrized templates. Binding these templates to elements triggers graph transformations that create task structures for each building element, yielding an integrated product-process graph. The resulting graph is then converted into a robot-executable format using Behavior Trees. The following sections detail these stages in turn, beginning with product representation, progressing to process structuring and generation, and finally deriving execution-level control structures.

3.1 Representing BIM Data as a Product Graph

We represent the building as a product graph, modeled as an LPG from semantic and geometric information contained in Industry Foundation Classes (IFC) models. This choice aligns with prior work showing that graphs can unify heterogeneous information within a single model [13, 8]. The LPG provides a transformable basis for deriving task structures that can later be converted into robot-executable control structures. Because we focus on construction methods for load-bearing elements, the resulting graph inherently represents the building's structural shell.

The graph follows the IFC spatial hierarchy, grouping building elements under their respective building stories, as illustrated in the upper-left section of Fig. 1. Nodes are labeled as either *SpatialElement* or *BuildingElement*, with their IFC subtype retained as a second label. Spatial elements include the site, building, and its stories, while

relevant building elements include, but are not limited to, beams, walls, columns, and slabs. The representation can be extended with *Resource* nodes (e.g., robots, equipment, or material supply points). The combined labels enable queries not only over spatial and containment relations, but also over specific element types.

Edges capture spatial decomposition and containment relationships from IFC, ensuring that each element remains embedded in its spatial and physical context. We additionally enrich the graph with adjacency relations between neighboring elements, and inter-element ordering relations that can be used to represent a construction sequence. When the graph is extended with process information, typed edges (e.g., *needsResource*) can connect tasks to required resources to support later allocation and feasibility checks. Together, the labeled nodes and edges form the product graph $\boldsymbol{G_p}$. This graph is fine-grained enough to associate construction methods and task structures with individual elements, yet abstract enough to be applicable to typical IFC models. Coarser representations, such as zones, do not expose individual elements as task targets. In contrast, finer decompositions, such as fasteners, are often absent in design-phase models and therefore cannot be assumed.

3.2 Capturing Construction Processes as Method Templates

A construction method describes how a building element is realized in practice, expressed as an ordered sequence of processes. In this work, we adopt this notion for four construction methods: *in-situ concreting, prefabricated concreting, masonry,* and *3D concrete printing*. Each method is modeled as a parametrizable template that organizes its process steps into a reusable structure. The template is element-agnostic at first, becoming element-specific task structures only when bound to a building element. This separation allows the same method to be reused across multiple elements while still adapting to their individual properties. However, construction processes are still too abstract to directly instruct robots [5]. We therefore refine them into tasks (e.g., Grab), which ultimately provide the lower-level units from which executable actions are derived.

3.2.1 Task Decomposition

We structure method templates using Hierarchical Task Network (HTN) planning [15] principles. HTN offers a hierarchical modeling approach that breaks high-level activities down into ordered subtasks. In HTN terminology, high-level activities are *compound* tasks, which can be refined into other compound tasks or into *primitive* tasks. Primitive tasks are directly executable actions such as *Grab, MoveTo,* or *Fasten*. In a graph representation, compound tasks map to internal nodes, and primitive tasks map to leaf nodes. Primitive task nodes specify the required operational properties, such as locations, tolerances, and element IDs. These properties are instantiated, that is, receive their values, when the method template is bound to a specific building element.

In the proposed representation, a construction method is formalized as a directed acyclic graph (DAG) $\boldsymbol{G_m}$, shown in the bottom-left quadrant of Fig. 1. The DAG consists of compound and primitive task nodes linked by refinement (parent-child) relations. For each compound task, the order of its child tasks defines the local execution sequence among siblings. This captures hierarchical decomposition and task ordering in a form that can later be converted into an executable control structure. A partial template example for in-situ concreting is shown below.

Partial in-situ template

```
InSituConcreting
├─ PlaceFormwork
│  ├─ Transport
│  │  ├─ Grab
│  │  └─ MoveTo
│  └─ Fasten
├─ AddReinforcement
├─ PourConcrete
└─ RemoveFormwork
```

3.3 Integrating Product and Process Data

Up to this point, product data and procedural knowledge have been represented separately. The transition from abstract method templates to actionable task structures requires associating these templates with the product data of each element. As shown in the bottom-right quadrant of Fig. 1, we achieve this in two steps: first, a method template is bound to a building element, and second, graph transformation rules instantiate task properties and transfer the template structure, yielding an element-specific task structure.

3.3.1 Method Binding

Assigning construction methods to elements is treated as a design decision and is therefore user-defined. Each assignment is stored in the product graph as a directed *hasMethod* edge from the element node to a *ConstructionMethod* node representing the chosen method. Elements that use the same method reference the same *ConstructionMethod* node, which reflects that the method is a reusable template rather than an element-specific object. Each element-method pair then forms the left-hand side pattern matched by the transformation rules.

3.3.2 Graph Transformation

After assigning a construction method to an element, we apply graph transformation rules [16] to generate element-specific task structures:

$$p : L \rightarrow R, \tag{1}$$

where $\boldsymbol{L}$ denotes the element-method pattern, and denotes $\boldsymbol{R}$ the right-hand side graph that introduces or enriches task nodes and edges according to the method template. We consider four kinds of rules: (i) rules that create task instances for all tasks in the method template and designate the root task for the bound element, (ii) rules that transfer refinement and sequencing relations from the template to the instances, (iii) rules that instantiate primitive task properties using product data, and (iv) rules that link instantiated task nodes to *Resource* nodes where resource requirements are specified. A match

$$m : L \rightarrow G_P, \tag{2}$$

selects an occurrence of the pattern $\boldsymbol{L}$ in $\boldsymbol{G_p}$ (created by the element-method assignment). Applying a rule $\boldsymbol{p}$ under a match $\boldsymbol{m}$ transforms the product graph $\boldsymbol{G_p}$ into an integrated product-process graph $\boldsymbol{G_I}$:

$$G_P \overset{p,m}{\Longrightarrow} G_I. \tag{3}$$

The transformation does not delete or rewrite the product side. Instead, it extends the graph with task structures anchored to building elements. The resulting integrated product-process graph $\boldsymbol{G_I}$ links spatial and physical building information with sequenced construction tasks and forms the basis for the execution model.

3.4 Deriving a Robot-Executable Control Structure

The integrated product-process graph specifies which tasks are required and how they are ordered, but its task structures still need to be expressed as an execution-level control structure. We therefore map the task structures to Behavior Trees (BTs) [11] that preserve the higher-level process logic captured in the graph. While typically the internal control program of a single robot, we use BTs as a shared representation of the construction process.

Compound tasks become *control* nodes (e.g., sequence or parallel) that organize their children according to the refinement and precedence relations in the DAG $\boldsymbol{G_m}$, preserving the hierarchical and ordered structure introduced in Sec. 3.2.1. Primitive tasks become *action* nodes that carry over the element-specific properties. To promote reuse and modularity, recurring compound tasks are represented as standalone *subtrees* that can be referenced from higher-level Behavior Trees (see Fig. 2). A project-level tree then composes the element-specific trees and orders them according to inter-element dependencies, providing a single entry point for execution. The resulting BT is generated per project from the integrated product-process graph, preserving traceability to elements and task data while providing a robot-executable control structure.

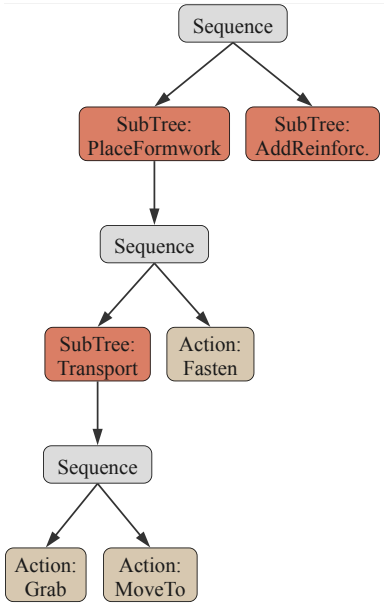


Figure 2. Part of an exemplary Behavior Tree for an in-situ concreting method, illustrating control nodes (sequence), action nodes, and reusable subtrees.

4 Case Study

To illustrate the proposed workflow, we apply it to a multi-story building model and demonstrate how BIM-derived product data, construction method templates, and graph transformations are combined to generate element-specific task structures and Behavior Trees.

4.1 Setup

The case study uses an IFC model of a three-story building, shown in Fig. 3. We restrict the scope to load-bearing elements and consider 51 beams, 36 columns, six walls, and four slabs. The product graph preserves the IFC spatial hierarchy and is enriched with adjacency relations by checking intersections between element axis-aligned bounding boxes. Four construction methods are considered: *in-situ concreting, prefabricated concreting, masonry,* and *3D concrete printing*. Their method templates follow the structure described in sec. 3.2. For the case study, method bindings are assigned manually to illustrate the workflow, for example, in-situ concreting for

slabs and beams, 3D concrete printing for columns and walls, and masonry for walls. The workflow was implemented in Python, using IfcOpenShell for IFC processing, Neo4j for graph storage and transformation, and a custom Python script to query the graph and write Behavior Tree XML files.



Figure 3. BIM model used for case study.

4.2 Instantiating the Product-Process Graph

The resulting product graph G_p contains seven spatial nodes with six spatial decomposition relationships, 97 building-element nodes and containment relationships, 223 adjacency relationships, and 96 precedence relationships between elements. We obtain the element precedence relationships through a simple centroid sorting along the Z, Y, and X axes. These relations are not intended to represent an accurate construction sequence, but provide an initial project-level order that can later be replaced by more explicit sequencing rules.

Method binding is realized using an interactive IFC viewer [17], where users select building elements and assign a construction method to each of them. Applying graph transformation rules yields the integrated product-process graph G_I, where method templates are instantiated and each element is linked to a rooted task structure. For the case study model, the transformations generate 1,138 task nodes, comprising 396 compound tasks and 742 primitive tasks, connected by 1,480 refinement edges and 792 precedence edges. On average, the in-situ concreting, prefabricated concreting, and masonry templates contribute 10, 7, and 5 primitive tasks per element, respectively, reflecting differences in method complexity. The resulting task structures are acyclic and anchored to their elements, indicating that the method templates can be reused consistently at the building scale.

4.3 Deriving Behavior Trees from Task Structures

The task structures in G_I provide the basis for deriving Behavior Trees that describe the construction process independently of a specific robot platform. They specify *what* must be done and in which order, while robot selection determines *how* tasks are executed. This separation is important because the considered methods may be realized using different platforms (e.g., in-situ concreting via a pump, prefabrication using lifting and placement equipment such as a crane, masonry with a bricklaying robot, and 3D concrete printing with a mobile printing system).

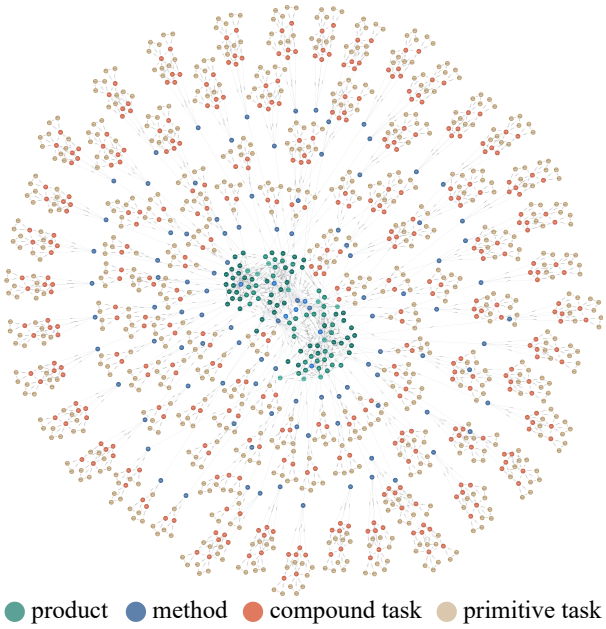


Figure 4. Product-Process Graph for a Simple Building with Multiple Methods.

For each element-method pair, the rooted task structure is mapped to a BT subtree. Compound tasks become control nodes, while primitive tasks become parametrized action nodes. This mapping is applied to all elements with assigned methods. Fig. 5 shows a representative BT for an element assigned to the in-situ concreting template. Its action (leaf) nodes carry the parametrized values of a particular element, for instance the element ID and target location used by the *GrabObject* and *MoveTo* actions.

Across the case study, the workflow generates 98 BTs: one BT per element in scope plus one project-level BT. These BTs preserve the method template structure and remain linked to their corresponding building elements. This demonstrates that the refinement of BIM-derived product data and procedural knowledge into task structures carries through to an execution-level control structure.

5 Discussion

The case study shows that construction method templates can be reused to generate element-specific task structures systematically. A small set of methods is applied across the load-bearing elements, and graph transformations instantiate rooted task subgraphs for each element from the corresponding method binding. This indicates that reusable method templates are applied consistently at

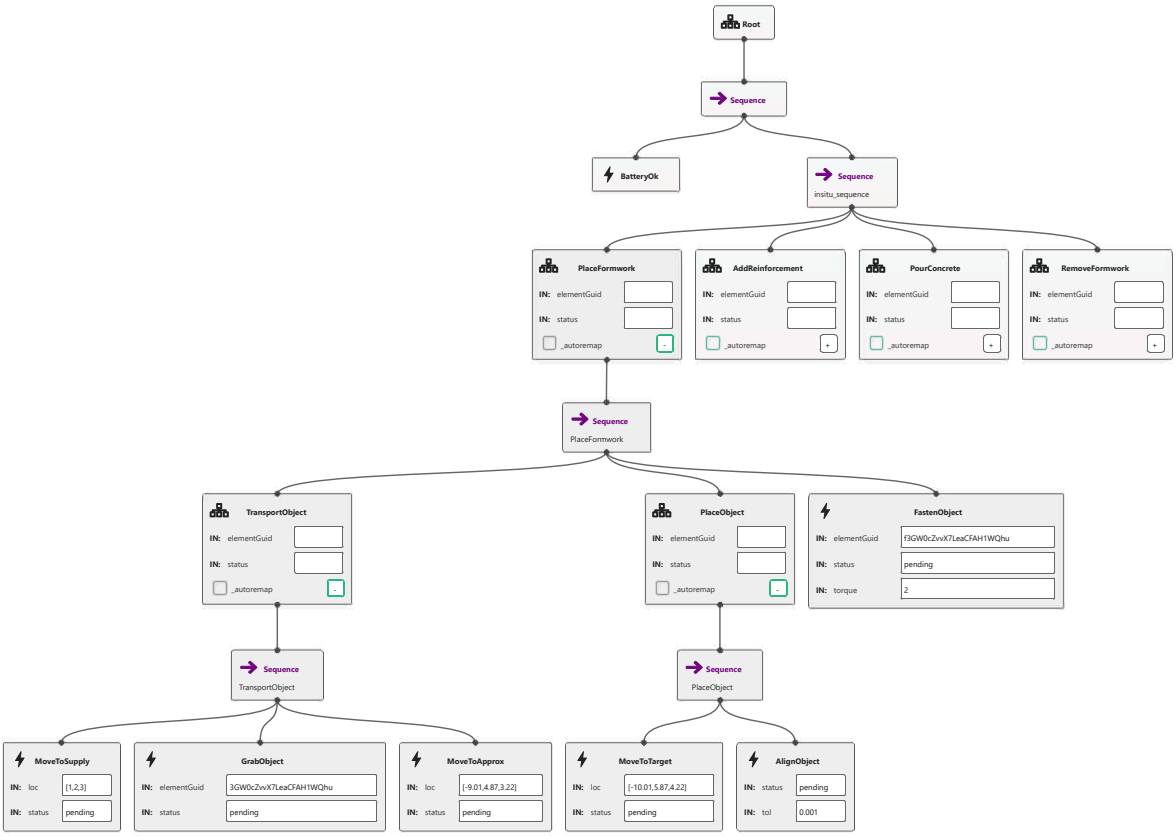


Figure 5. Behavior Tree derived from product-process graph.

the project level.

The proposed representation complements knowledge graph approaches in BIM [9, 13] by linking product information and process knowledge in a form that can be mapped to execution-level control structures. Instead of distributing process information across schedules, mission scripts, or skill libraries, the integrated product-process graph stores task decomposition and ordering together with element context. Queries can therefore access process semantics (e.g., predecessors, successors, refinement relations) alongside product attributes. The same graph structure can be extended with additional dimensions such as resources or temporal constraints. This creates a flexible layer for future functionalities such as robot-task allocation or discrete-event simulation.

Despite its merits, the current approach has limitations. LPGs are flexible and easy to extend, but their schemaless nature offers limited guarantees about validity and structure. A closer alignment with the IFC schema would not only provide more structure but also support more complex reasoning. For example, this would enable spatial reasoning, where queries could infer whether a robot can reach a target pose by passing through doors of sufficient width, or traverse terrain such as stairs. Such queries would extend the current lightweight representation beyond element-specific task generation toward a richer representation of construction context. Additionally, method templates are defined manually, which raises questions of modeling effort and scalability as more construction methods are added. Finally, the present work stops at Behavior Tree generation and does not yet validate runtime execution. Future work will therefore focus on more systematic template definition and on evaluating the generated structures in simulation and with physical robots.

6 Conclusion

This paper presents a refinement workflow that links BIM product data with procedural knowledge to generate element-specific task structures for robotic construction. Product data is represented as a labeled property graph (LPG), and construction methods as parametrized templates. Graph transformations instantiate element-specific task structures, which are then mapped to Behavior Trees (BTs) as execution-level control structures. We evaluate the approach on a multi-story building model. For the load-bearing elements, we build an integrated product-process graph, reuse a small set of method templates across

many elements, and generate one BT subtree for each element to match the intended process logic. This reduces the need for hand-crafted BTs and keeps each subtree explicitly linked to its source element. Because tasks and product context share a single graph, the model can be extended with resource and timing constraints, used for robot-task allocation, or updated during execution by integrating perception feedback.

In summary, the proposed workflow consolidates fragmented product, process, and execution artifacts into a single product-process model that yields robot-executable task structures. It supports systematic refinement and reuse across elements and provides a foundation for reuse beyond a single project, paving the way for more scalable BIM-to-robot workflows in construction robotics.

Acknowledgments

We gratefully acknowledge the support of the German Research Foundation (DFG) for funding the project under grant FOR 5672, subproject A "Generic framework for simulating and executing robotized construction", grant no. 517965147. Furthermore, we acknowledge the support of the TUM Innovation Network "Co-Construct", the DFG Transregio 277 "Additive Manufacturing for Construction" (grant 414265976), and the DFG project 510047432 as part of the program SPP 2187 "Adaptive modularized constructions made in flux" (grant 423969184).

References

[1] A. Borrmann, M. König, C. Koch, and J. Beetz. *Building Information Modeling: Technology Foundations and Industry Practice*, volume 1. Springer International Publishing AG, 2018. doi:10.1007/978-3-319-92862-3.

[2] P.F. Ahmadi and M. Arashpour. An analysis of 4d-bim construction planning: advantages, risks and challenges. In *Proc. of the 37th Int. Symp. on Automation and Robotics in Construction (ISARC 2020)*, pages 163–170, 2020. doi:10.22260/ISARC2020/0025.

[3] F. Pfitzner, S. Hu, A. Braun, A. Borrmann, and Y. Fang. Monitoring concrete pouring progress using knowledge graph-enhanced computer vision. *Automation in Construction*, 174, 2025. doi:10.1016/j.autcon.2025.106117.

[4] C.J. Turner, J. Oyekan, L. Stergioulas, and D. Griffin. Utilizing industry 4.0 on the construction site: Challenges and opportunities. *IEEE Transactions on Industrial Informatics*, 17(2), 2021. doi:10.1109/TII.2020.3002197.

[5] J.G. Everett and A.H. Slocum. Automation and robotics opportunities: Construction versus manufacturing. *Journal of Construction Engineering and Management*, 120, 1994. doi:10.1061/(ASCE)0733-9364(1994)120:2(443).

[6] S. Kim, M. Peavy, P. Huang, and K. Kim. Development of bim-integrated construction robot task planning and simulation system. *Automation in Construction*, 127, 2021. doi:10.1016/j.autcon.2021.103720.

[7] K.M. Lundeen, V.R. Kamat, C.C. Menassa, and W. McGee. Autonomous motion planning and task execution in geometrically adaptive robotized construction work. *Automation in Construction*, 100, 2019. doi:10.1016/j.autcon.2018.12.020.

[8] L. Atanasova, J.C. Kenny, E.V. Alexi, F. Chovghi, D. Mitterberger, and K. Dörfler. Building (with) human–robot teams: fabrication-aware design, planning, and coordination of cooperative assembly processes. *Construction Robotics*, 9(2), 2025. doi:10.1007/s41693-025-00158-w.

[9] A. Zhu, P. Pauwels, E. Torta, H. Zhang, and B. de Vries. Data linking and interaction between bim and robotic operating system (ros) for flexible construction planning. *Automation in Construction*, 163, 2024. doi:10.1016/j.autcon.2024.105426.

[10] M. Terzer, T. Flatscher, M. Magri, S. Garbin, J. Emig, and A. Giusti. A facilitated construction robot programming approach using building information modelling. In *10th Int. Conf. on Control, Decision and Information Technologies (CoDIT)*, pages 2656–2661, 2024. doi:10.1109/CoDIT62066.2024.10708285.

[11] M. Colledanchise and P. Ögren. *Behavior Trees in Robotics and AI: An Introduction*. CRC Press, Clermont, Florida, 2018. doi:10.1201/9780429489105.

[12] H. Oyediran, W. Turner, K. Kim, and M. Barrows. Integration of 4d bim and robot task planning: Creation and flow of construction-related information for action-level simulation of indoor wall frame installation. *Journal of Information Technology in Construction*, 30, 2025. doi:10.36680/j.itcon.2025.015.

[13] J. Chen, W. Lu, Y. Fu, and Z. Dong. Automated facility inspection using robotics and bim: A knowledge-driven approach. *Advanced Engineering Informatics*, 55, 2023. doi:10.1016/j.aei.2022.101838.

[14] A. Tandon, J. Stührenberg, and K. Smarsly. Automated building inspections coupling building information modeling and behavior trees. In *Proc. of the 2025 European Conference on Computing in Construction*, 2025. doi:10.35490/EC3.2025.360.

[15] K. Erol, J. Hendler, and D.S. Nau. HTN Planning: Complexity and Expressivity*. In *12th AAAI Conf. on Artificial Intelligence*, pages 1123–1128, 1994.

[16] A. Corradini, U. Montanari, F. Rossi, H. Ehrig, R. Heckel, and M. Löwe. *Handbook of Graph Grammars and Computing by Graph Transformation*, chapter Algebraic Approaches to Graph Transformation - Part I: Basic Concepts and Double Pushout Approach, pages 163–245. World Scientific, 1997. doi:10.1142/9789812384720_0003.

[17] T. Wrabel, S. Esser, and A. Borrmann. A framework architecture for the information backbone for robotic construction. In *Proc. of the 43rd Int. Symp. on Automation and Robotics in Construction (ISARC 2026)*, 2026.

Near-Field Laser-Line Pose Sensing for Absolute Localization of Mobile Robots and Handheld Tools in Construction

Pok Yin Victor Leung[1], **Alvario Cassinelli**[1] **and Miu Ling Lam**[1]

[1]City University of Hong Kong, Hong Kong S.A.R. (China)

victor.leung@cityu.edu.hk, acassine@cityu.edu.hk, miu.lam@cityu.edu.hk

Abstract -

This paper presents a low-cost, high-precision laser-line sensor designed for construction-robotics applications such as on-site tiling, leveling, and tool localization. Unlike similar projects built around custom electronics and hardware, our approach uses a off-the-shelf microcontroller with camera module and a semi-transparent projection screen to measure the vertical position of a laser line from distances of 1–6 m with sub-millimeter accuracy. We introduce a complete sensing pipeline, including controlled illumination, and subpixel peak extraction across multiple camera modes. A fully automated data-collection framework cycles through sensor resolutions, records timing jitter and peak-detection stability, and quantifies achievable measurement precision. Results show that a consumer-grade imaging system can provide consistent, high-frequency height measurements suitable for indoor construction tasks, offering an accessible alternative to specialized CCD-based receivers or high-end laser trackers.

Keywords -

Construction Robotics; Absolute Localization; Laser Plane Sensing; Digital Fabrication

1 Introduction

Precision alignment is essential in robotic construction tasks such as floor tiling, concrete smoothing, and wall rendering, where sub-millimeter tolerances are sometimes required. While many pose estimation methods exist, few can reliably achieve this level of precision, and those that can—such as laser trackers or robotic total stations—come at high cost and with significant deployment complexity. In practice, many robotic processes do not require high precision in full 6-DOF. Instead, they can be aligned relative to a fixed planar reference. Examples include controlling floor tile placement for horizontal level or guiding wall plastering tools to stay plumb.

This paper introduces a low-cost laser-plane height sensor that leverages the geometric consistency of construction-grade laser levels. The method uses a camera in combination with a near-field projection screen to detect where a laser plane intersects the screen surface. By analyzing the position of this projected line at subpixel resolution, the sensor estimates height deviation relative to the plane with high precision. This paper focuses on the design and characterization of the sensing system itself, while integration into a complete robotic workflow is left for future work.

The main contributions of this work are: (1) a reproducible open-hardware sensor comprising only a camera module, a 3D-printed housing, and a paper screen; (2) a lightweight processing pipeline including lens focusing, calibration, and real-time peak detection; and (3) an experimental evaluation of the sensor's performance across multiple resolutions and distances, demonstrating accuracy suitable for integration into on-site robotic workflows.

2 Related Work

Robotic construction applications often rely on external positioning systems to achieve high accuracy over large workspaces. Commercial systems such as robotic total stations and laser trackers offer sub-millimeter precision by optically tracking a retro-reflective target [1]. These systems have been used in mobile construction robotics to localize manipulators relative to a site-fixed reference. Similarly, ETH Zurich's Robotic Fabrication Laboratory has deployed a Nikon iGPS system to localize robots and objects in a shared workspace [2]. While these systems enable full 6-DOF pose tracking, they impose significant infrastructure demands: tripod-mounted cameras or lasers, calibrated room-scale volumes, and line-of-sight requirements. Motion capture systems provide similar accuracy in indoor settings but require multi-camera arrays mounted on trusses or frames and are less suitable for deployment in dynamic construction sites.

Marker-based vision systems have been developed as low-cost alternatives to motion capture or total station tracking. For example, a camera mounted on the robot's end-effector can detect fiducial markers placed in the construction environment [3]. However, vision-based localization methods require sufficient lighting, careful calibration of the optical system, and stable environmental geometry. Their accuracy is inherently dependent on resolution, lens distortion, and viewing angle—factors that

limit their effectiveness for long-range height sensing.

Planar Laser Sensing in Field Robotics More recently, several construction robots company have developed robots with laser line sensors, rely on laser planes as spatial references. Unfortunately, many of them are not published and their methods remain undisclosed. Published examples include a mobile robot that follows a laser line projected onto a raised-floor grid, achieving 2–3 mm positioning error using a custom-built sensor tuned to the laser profile [4]. While effective for gross alignment tasks, these sensors do not support sub-millimeter resolution or customizable processing. Only until recently, a custom laser receiver was developed for detecting the intersection of a construction-grade laser plane [5]. Using custom electronics and hardware, their spinning CCD sensor array is capable of sensing a laser plane height with accuracy of 25 µm(max) and 5.5 µm(average).

Other researchers have explored triangulation-based laser systems, where static beacons emit laser patterns and mobile receivers triangulate their position. However, such systems require multiple calibrated emitters and suffer from cumulative line-of-sight and angular ambiguity issues, particularly on cluttered sites. Camera-based sensing systems have also been used to monitor structured-light patterns or detect projected lines in manufacturing. These systems often require structured lighting conditions and are not intended for on-site construction tasks or over large distances.

3 Sensor Design and Implementation

3.1 Sensor Concept

The sensor we developed detects the vertical position of a laser plane by capturing the intersection of the laser beam with a near-field projection screen. The screen is positioned orthogonally to the incoming laser plane such that the projected line appears as a nearly straight segment across the visible field. As the sensor moves relative to the laser, the projected line shifts vertically along the screen surface. Since the sensor only measures the height of the laser line within its fixed field of view, the system effectively slides the laser intersection vertically over the screen. This converts the 3D spatial deviation relative to the plane into a 1D pixel displacement observable by the camera.

3.2 Hardware Architecture

The hardware consists of three components:

- OpenMV H7 Plus microcontroller with integrated OV5640 camera sensor, which supports resolutions up to 2592×1944. The board is officially priced at USD 95, though alternative sources offer it for approximately USD 40.The stock lens is replaced with a generic 5MP HD650 M12 lens (3.5 mm focal length, f/2.8) offering a 90° horizontal and 60° vertical field of view. The lens structure follows a 2G2P configuration and is marketed as "zero-distortion." The lens is readily available online and costs approximately USD 4.
- Mechanical enclosure and screen: The microcontroller is mounted to a custom 3D-printed PLA enclosure using two screws. The enclosure maintains a fixed 20 mm spacing between the lens and the screen. The case can be fabricated using an entry-level FDM 3D printer with a 0.2 mm layer height. No supports or post-processing are required.
- The screen is cut from standard 80 g/m^2 printing paper and adhered to the front of the housing using adhesive tape. The visible size of the screen is 30 × 40 mm

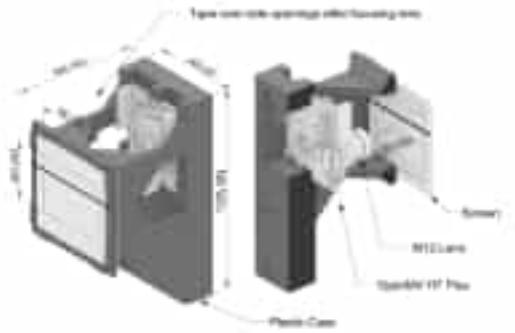


Figure 1. Dimensioned drawing (left) and cut-away (right) of the sensor hardware

After lens focusing is completed, the side openings are covered with black tape to prevent light from reaching the screen from the back side. The sensor geometry and dimensions are illustrated in Figure 1.

4 Software Architecture

The onboard firmware runs MicroPython scripts using the OpenMV IDE environment. An SD card is not required; all scripts execute from internal flash.

Lens Focusing Procedure To assist with focusing at the 20 mm working distance, a custom script is provided. The user replaces the white screen with a temporary pattern-printed paper (provided in the repository). The focus script evaluates local contrast in five image regions and computes

a focus score that is displayed in real time. The user adjusts the lens manually until the score is maximized, then locks the lens in place. After focusing, the pattern is replaced with a blank screen, and all side openings are taped over to block ambient light. The focusing aid is designed to be reproducible and requires no test targets beyond the printed pattern

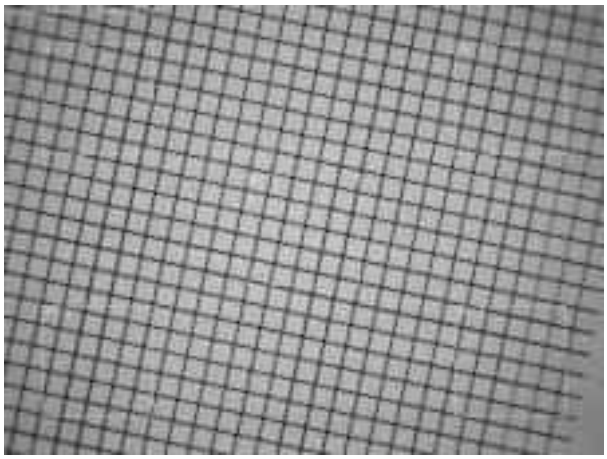


Figure 2. Frame displayed during lens-focusing. White square indicates multiple regions of interest.

Laser Line Detection Pipeline The laser detection script includes the following steps:

- Image acquisition: The sensor operates in grayscale mode at VGA resolution. Exposure and gain settings can be manually configured, though automatic exposure and gain are preferred.
- Brightness thresholding: Pixels with a brightness value below 80
- Row reduction: The captured image is collapsed into a 1-dimension vertical brightness profile. For each image row, a local mean brightness is computed using OpenMV's ROI statistics.
- Adaptive smoothing: A moving average filter is applied to the row profile. The window size is selected automatically using a statistical thresholding function (`auto_window`) that estimates the width of the bright region based on mean and variance. This step is essential for converting any flat-top area (caused by the thickness of the laser beam) into a brightness profile with well-defined peak.
- Peak detection: The smoothed profile is passed through a peak detector (`argmax`) to identify the brightest row, which corresponds to the laser line.
- Subpixel fitting: Two estimators are implemented:
  - Three-point fit: A quadratic curve is fit to the peak row and its neighbors using the standard parabola formula.
  - Five-point fit: A more robust version using five neighboring values and an approximation of a Savitzky–Golay filter, offering better stability on symmetric profiles.

The moving average smoothing helps the peak detection algorithm being insensitive to variation in image brightness. We demonstrate the robustness of the method by setting the brightness threshold to different values (from 20% to 80% at 10% increment) and confirmed that the detected peaks are identical (Fig.3). It is worth noting that the precision and repeatability produced by peak estimation does not require the laser line to have a symmetrical brightness.

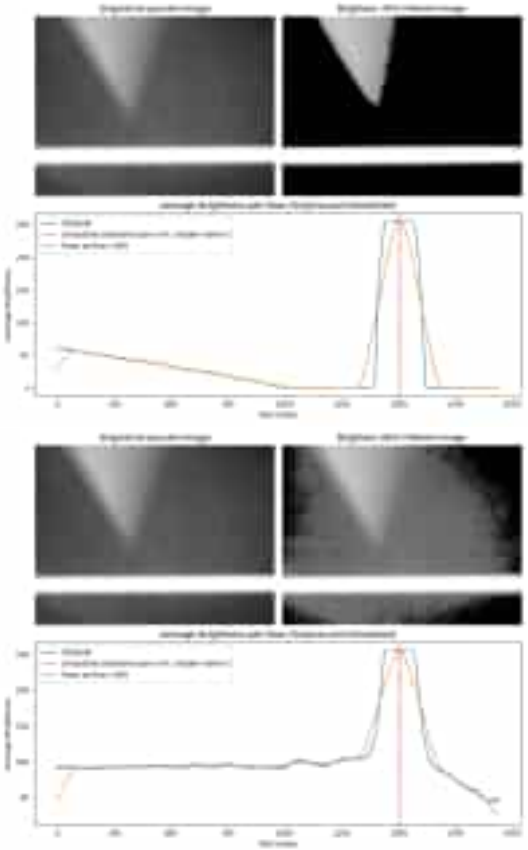


Figure 3. The project laser line (white horizontal stripe) as seen by the camera (top left) and after threshold (top right) at 20% and 80%. The brightness peak (dotted red) is identical after smoothing (orange) the raw brightness data (blue).



Figure 4. Photo illustrating Experiment Setup S1 (left), S2 and S3 (right)

Table 1. Overview of datasets used in evaluation, including varied parameters, spatial ranges, and emitter configurations. Emitter types are defined as S1–S3 in Section 5.1.

ID	Purpose	Resolution	z Range (mm)	d Range (mm)	Setup	Notes
T3	Precision, estimator comparison	7 res.	0–1.0	400	S1	Dense sweep (§5.2–5.3)
T4	Calibration linearity	7 res.	0–14.0	400	S1	Calibration (§5.1)
T6	Short-range emitter distance	VGA	5.0 (fixed)	60–860	S1	Emitter sweep (§5.4)
T7	Long-range (Vibration)	VGA	0–10.0	8000–13000	S2	Jitter (§5.5)
T8	Long-range (Vibration)	VGA	0–16.0	5000–8000	S3	Jitter (§5.5)
T9	Long-range (Concrete Floor)	VGA	5.0 (fixed)	1000–11000	S2	Jitter (§5.5)

5 Experimental setup

To evaluate the accuracy, stability, and geometric response of the sensor across a wide range of conditions, a set of structured experiments was conducted. These tests were designed to isolate variables such as image resolution, emitter distance, ambient lighting, and laser beam geometry, and to quantify their effect on the sensor's measurement performance.

5.1 Data Collection Apparatus

Experiments were conducted using combinations of a motorized linear size (for varying emitter distance, d), manual micrometer stage (sensor height, z), fixed or pendulum-mounted laser emitters, and varying site conditions. Three types of laser emitter setups were used (also refer to Figure 4): **S1**: Static red laser emitter on linear track: used in T3, T4, and T6 for controlled Z-sweeps and emitter-distance tests. **S2**: Pendulum green laser level (construction grade): used in T7, T9, and T10 for long-range jitter and linearity tests under passive self-leveling conditions. **S3**: Stationary green laser level: used in T8 for comparison against pendulum-based emitters on raised flooring. Ten datasets (T1 to T10) were collected during development (see Table 1); Note 1: T1, T2, and T5 were later discarded due to instability or redundancy. The remaining seven datasets (T3, T4, T6–T10) are used throughout the evaluation. Note 2: "7 res." indicates resolution sweep across QQVGA (160x120) to WQXGA2 (2592x1944).

6 Results and Discussion

6.1 Calibration linearity across resolutions

To verify the geometric behavior of the sensor, we evaluate whether the relationship between sensor height z and detected laser line position is linear across all resolutions. Dataset T4 was collected by sweeping z from 0 to 14 mm (in 15 steps) using a micrometer stage, while cycling through seven resolutions at each height. For each resolution, the mean subpixel peak position (using 3-point subpixel estimator) was fit with a linear regression against ground-truth z, and residuals (representing error) were computed per sample (in µm). Figure 5 shows the calibration lines and residuals for all resolutions. Results confirm consistent linearity across the entire range, with residuals bounded within ±20 µm.

6.2 Subpixel estimator comparison

We compare three peak detection methods: integer row index, 3-point quadratic fit, and 5-point fit. Using dataset T3 (dense 101-step Z sweep at VGA), each method's peak position is compared against the calibration curve from T4, and RMS residuals are computed. Figure 6 plots RMS error (µm) and normalized linearity error for each method and resolution. Integer peak performs worst (50–60 µm

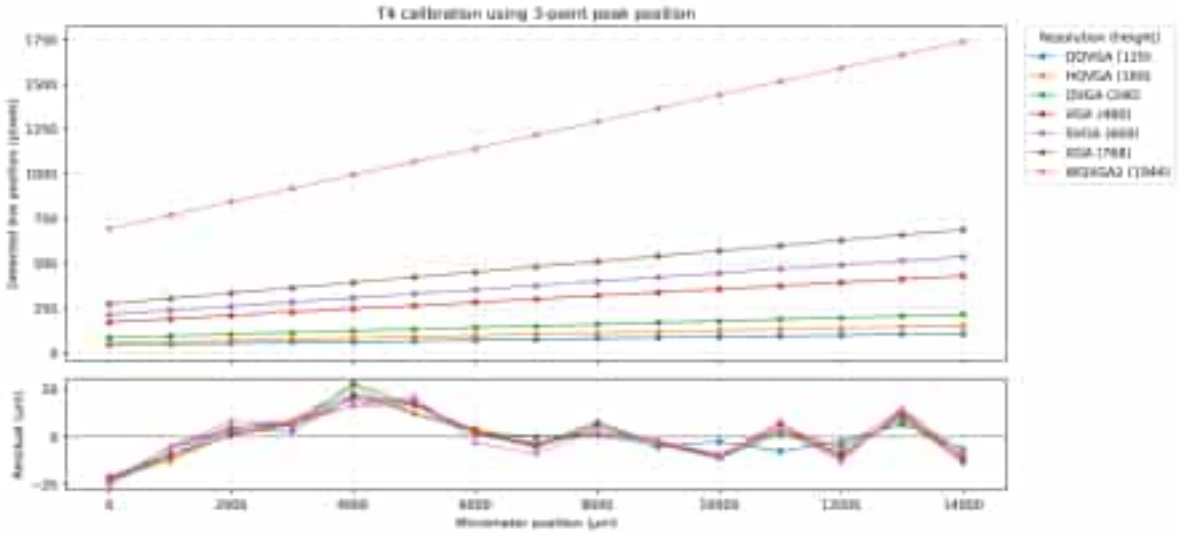


Figure 5. Calibration curves and residuals for all resolutions (dataset T4).

error), 3-point significantly improves accuracy (8–12 µm RMS), and 5-point yields no meaningful gain. The 3-point fit is therefore optimal.

6.3 Measurement error distribution at VGA

To characterize noise in a typical usage case, we analyze residual errors at VGA resolution using dataset T3. Measured heights are compared to the linear fit from T4, and residuals are converted to µm. Figure 7 shows a histogram of residuals along with symmetric percentile intervals. The median error is near zero; 90% of samples lie within ±38 µm and 95% within ±51 µm. The distribution is non-Gaussian with heavy tails but centered and repeatable. This defines the practical noise floor for VGA.

6.4 Sensitivity to emitter distance (short range)

Dataset T6 varies emitter distance from 60 mm to 860 mm while keeping sensor height fixed. At each distance, the laser line is detected and its height computed using calibration from T4. Figure 5.4 shows the mean deviation and frame-to-frame jitter versus emitter distance. A slow drift (200 µm) is observed over 800 mm, attributed to mechanical straightness of the linear track. Frame jitter remains low (¡10 µm), confirming that emitter distance has negligible impact on local precision within 1 m.

6.5 Long-range jitter on different floor types

To evaluate long-range performance under realistic conditions, we compare datasets T7–T9. The sensor is positioned at a fixed height (fixed micrometer) observing a laser emitter at large distances (up to 13m). At each testing condition, multiple observations were made (T7, T8: 10 samples; T9: 200 samples) to compute the jitter. Figure 9 shows jitter vs. distance on raised floor (T7–T8) and concrete floor (T9). The raised floor amplifies vibration and pendulum motion, leading to jitter of 50–300 µm at 8–13 m. On concrete, remains below 32 µm, an order of magnitude more stable than the test performed on raised floor. The jitter in T9 also grows linearly with distance, indicating that environmental stability, not sensor performance, dominates long-range error.

6.6 Long-range linearity

Dataset T10 tests linearity at 6 m emitter distance using 21 micrometer height steps (0–5 mm, 50 repeats each). Figure 10 shows the linear fit and residuals. The system maintains a consistent slope with RMS residuals of 14–21 µm and no curvature. Even at long distances, geometric linearity is preserved, and measurement accuracy remains stable.

7 Discussion and Conclusion

The sensor achieves sub-millimeter accuracy across a range of distances using only low-cost, readily available components. Its performance remains highly linear across all resolutions and working distances, with 95% confidence bounds of ±0.026 mm at 400 mm and ±0.050 mm at 6 m. While long-range jitter is observed, the reason is attributed to floor stability and emitter mounting instead of the sensor's internal precision. Note that, the present evaluation was conducted in indoor conditions and does not include direct sunlight or strongly mixed artificial lighting. Dust, debris, partial occlusion, and stronger ambient light can be further evaluated in future work. Such work

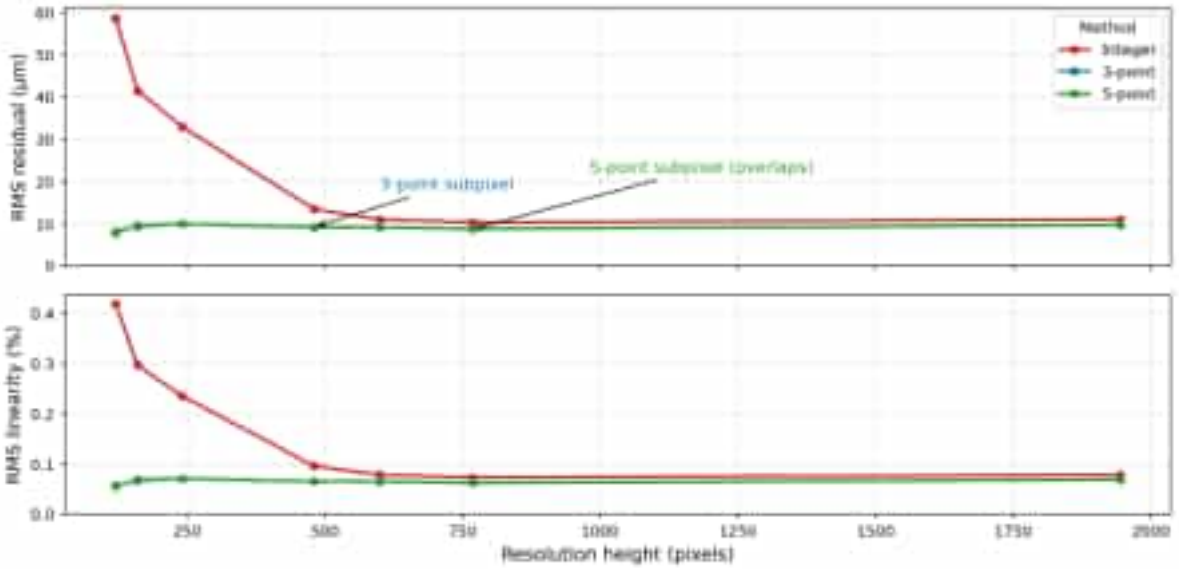


Figure 6. Comparison of integer, 3-point, and 5-point peak estimators (dataset T3). 3-point method achieves near-optimal precision with minimal complexity.

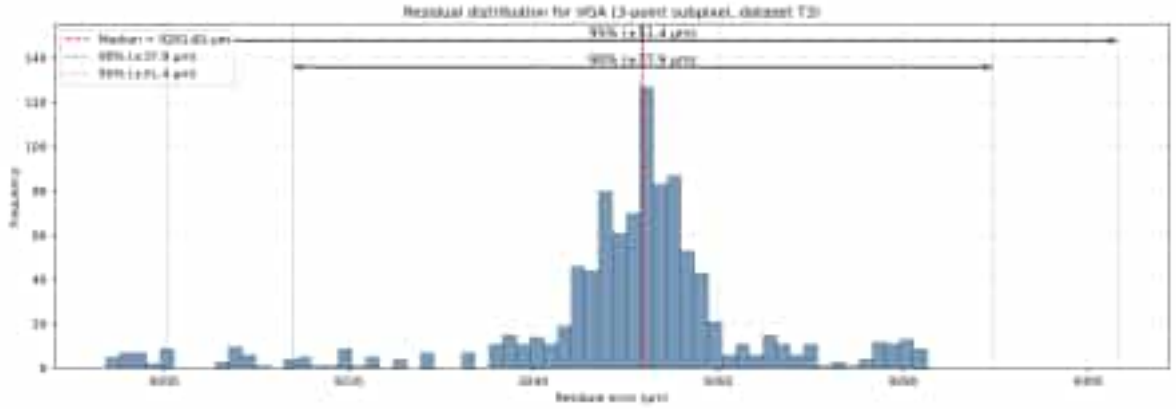


Figure 7. Histogram of residual errors at VGA (dataset T3). 90% and 95% intervals are marked with arrows; most samples lie within ±50 µm.

can try to expand towards a general hypothesis that if the human eye can discern a projected laser line in a specific environment, the camera should also be equally capable.

The measurable height span of a single sensor is bounded by the choice of screen dimensions, camera field of view, and calibration range. Thus the present prototype should be understood as one specific design while other design choices are likely possible. At longer emitter distances, usable performance is likely to be limited by laser visibility, signal-to-noise ratio and further exacerbated by environmental vibration. However, their effect is yet to be characterized.

Our method uses low-cost construction-grade laser planes as laser emitters, which lowers the financial barrier for system integration. For example, in robotic floor tiling, the sensor can be mounted on a tile gripper to regulate height relative to a site-fixed laser plane. Other sensing modules, such as accelerometers, can handle tile levelness. In addition, it also allows robots to share the same planar references used by human workers, potentially opening up new avenues for human-robot collaboration.

Our sensing method is reproducible, does not rely on optics calibration, and is intended to complement other sensors such as inclinometers or marker tracking [6]. The real-time subpixel pipeline runs entirely onboard the OpenMV camera module and requires no host computation, mak-

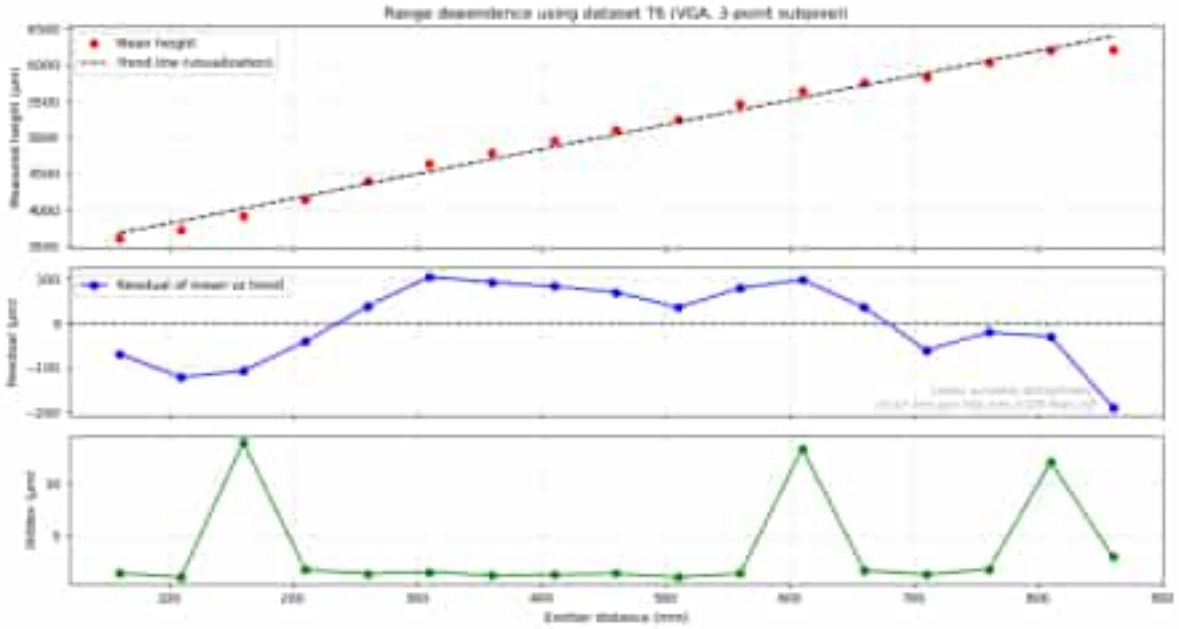


Figure 8. Measured height deviation and jitter vs. emitter distance (dataset T6). Drift remains ¡200 µm across 800 mm; jitter stays below 10 µm.

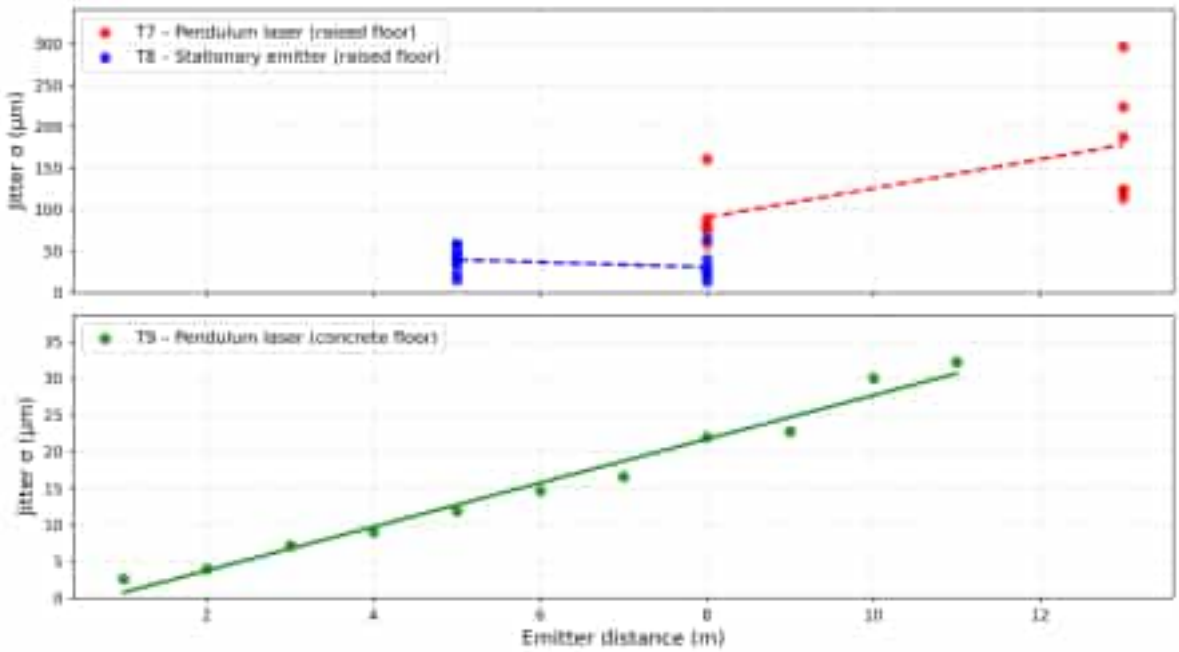


Figure 9. Long-range jitter vs. distance on different floor types (T7–T9). Tests on raised floor have large oscillations (max 300 µm); tests on concrete floor have substantially lower jitter (up to 32 µm at 11m).

ing it suitable for embedded applications. All hardware and software are released in an open-source repository `https://github.com/yck011522/laser_line_sensor/` to encourage adoption. The system avoids any custom circuit design, soldering, or firmware recompilation, and is intended to be assembled with only basic tooling and a consumer-grade 3D printer.

Future work can consider adapting the detection al-

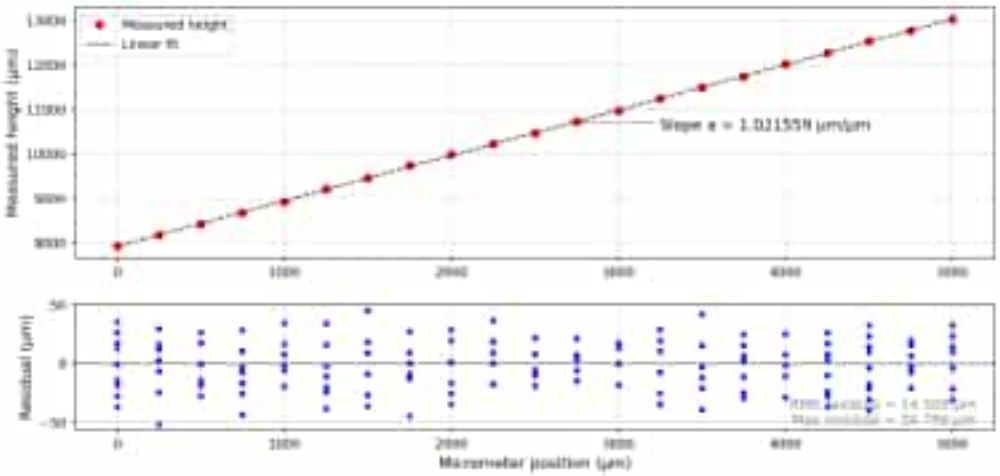


Figure 10. Linearity curve and residuals at 6 m distance (dataset T10). RMS error remains below 21 µm; no curvature or nonlinearity is observed.

gorithm to other embedded camera systems such as the ESP32-CAM or Seeed Studio XIAO ESP32S3. More broadly, the work illustrates how existing on-site tools can be reinterpreted—through near-field sensing and lightweight computation—to enable accessible precision feedback in construction robotics.

7.1 Acknowledgment

This research is supported by CityU Internal Grant 9610746.

References

[1] Selen Ercan, Sandro Meier, Fabio Gramazio, and Matthias Kohler. Automated Localization of a Mobile Construction Robot with an External Measurement Device. 2019 Proceedings of the 36th ISARC, Banff, Canada:929–936. ISSN 2413-5844. doi:10.22260/ISARC2019/0124.

[2] Manuel Lussi, Timothy Sandy, Kathrin Dörfler, Norman Hack, Fabio Gramazio, Matthias Kohler, and Jonas Buchli. Accurate and Adaptive in Situ Fabrication of an Undulated Wall Using an on-Board Visual Sensing System. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE Press. doi:10.1109/ICRA.2018.8460480.

[3] Kathrin Dörfler, Timothy Sandy, Markus Giftthaler, Fabio Gramazio, Matthias Kohler, and Jonas Buchli. Mobile Robotic Brickwork. In Dagmar Reinhardt, Rob Saunders, and Jane Burry, editors, *Robotic Fabrication in Architecture, Art and Design 2016*, pages 204–217. Springer International Publishing. ISBN 978-3-319-26378-6. doi:10.1007/978-3-319-26378-6_15.

[4] Takehiro Tsuruta, Kazuyuki Miura, and Mikita Miyaguchi. Mobile robot for marking free access floors at construction sites. 107:102912. ISSN 0926-5805. doi:10.1016/j.autcon.2019.102912.

[5] Siwei Chang, Haoqian Liu, Ching-wei Chang, Nan Wang, Haobo Liang, Haitao Wu, and Guangzhen Lin. High-resolution laser receiver for construction levelling robots using multi-linear charge-coupled device (CCD) arrays. 173:106077. ISSN 0926-5805. doi:10.1016/j.autcon.2025.106077.

[6] Julian Nubert, Shehryar Khattak, and Marco Hutter. Graph-based Multi-sensor Fusion for Consistent Localization of Autonomous Construction Robots. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 10048–10054. IEEE Press. doi:10.1109/ICRA46639.2022.9812386.

A Comparative Study of Mobile Device and Terrestrial Laser Scanning-Based Scan-to-BIM Workflows

Ahmet Bahaddin Ersoz[1,2] and Frédéric Bosché[2]

[1]Department of Civil Engineering, Middle East Technical University, Turkiye
[2]School of Engineering, University of Edinburgh, UK
abersoz@metu.edu.tr, f.bosche@ed.ac.uk

Abstract – Scan-to-BIM procedures using terrestrial laser scanning (TLS) are accurate but require significant time and financial investment. This study evaluates a faster alternative utilizing mobile device-based LiDAR and a mobile application built on the Apple RoomPlan API. The research first assesses geometric reliability through point-cloud-to-BIM confidence indices, specifically Coverage, Distribution, and Distance. Additionally, it employs BIM-to-BIM matching metrics, including 3D-IoU and 3D-Compactness, to compare the mobile output against a TLS baseline in residential and office buildings. Results demonstrate that the mobile pipeline reduces total scan-to-BIM time from 3.5 min/m² to 0.1 min/m². Although the mobile application provides reliable room layouts with space 3D-IoU scores reaching 0.93, it exhibits limitations regarding wall thickness accuracy and positional drift. Consequently, based on these initial findings, this workflow shows potential for facility management and initial retrofit assessments where rapid data acquisition is prioritized over high geometric precision.

Keywords – Scan-to-BIM, Mobile LiDAR, TLS, Mobile Device, Geometric Accuracy

1 Introduction

Scan-to-BIM procedures are key to generate digital models containing geometry and core semantic information that support a range of applications. These BIM models support retrofit design, energy modelling, and facility management (including digital-twins or digital-logbooks).

Most scan-to-BIM workflows start from reality capture data, in which point clouds are the dominant representation used to capture the existing conditions as a reference geometry for creating parametric BIM objects. They provide high-resolution geometric information, and point clouds are usually generated using terrestrial laser scanning (TLS) and mobile laser scanning (MLS). TLS captures dense and accurate geometry. However, it requires multiple stations with subsequent data registration, which can slow down the data collection procedure. By contrast, handheld systems can be faster to deploy, yet their accuracy depends on the stability of tracking on the scene structure, which can lead to drift and misalignment based on the long scanning paths.

Recently, lidar sensors have become available on mobile devices, particularly on iPhone Pro and iPad Pro models. These devices are low-cost compared with professional scanners and easy to deploy for quick capture tasks [1]. Many studies have assessed the feasibility of lidar-integrated mobile devices for scan-to-BIM workflows by comparing output point clouds against TLS reference data [1][2][3][4]. They show that, while, like MLS, they can acquire data more rapidly than TLS and can achieve promising local precision in indoor settings. Beyond standalone indoor scans, some studies merge mobile lidar with other sensing sources, such as TLS, UAV lidar or photogrammetry, principally to improve completeness [5][6].



Figure 1. Framework of the proposed assessment method

However, regardless of whether the scan comes from TLS, MLS, or mobile device lidar, most workflows require post-processing followed by geometric and semantic modelling, hereafter BIM modelling. In these studies, the BIM model is typically reconstructed through manual or semi-manual/automated modelling.

It must be noted that, besides lidar depth sensing, mobile devices also provide RGB camera data, another modality containing useful information about the scene. This data can also be leveraged to detect and segment

various elements in the scene, such as architectural elements (e.g., windows, doors) and furniture (e.g., tables, chairs) [7].

A key recent development in the use of mobile device lidar is Apple's RoomPlan [8]. Specifically, RoomPlan is an API that uses the device camera and lidar scanner to create parametric 3D room scans, including room dimensions and recognized furniture, producing a structured room representation rather than only raw points or a mesh. Several commercial apps use this type of pipeline, including Polycam [9], Magicplan [10] and Metaroom [11]. Even though many studies evaluate iOS lidar in terms of point cloud or mesh quality, few studies evaluate the full pipeline of RoomPlan using standardized and quantitative BIM-level metrics. Recently Mortazavi et al. [12] proposed an assessment system with criteria, a rating scheme, and utility value analysis to compare site-scanning methods for creating a BIM model, including Magicplan and Polycam against TLS and MLS. While this is valuable for tool selection, it mainly supports comparative scoring and does not report explicit geometric accuracy metrics for the produced models, such as deviation-to-ground-truth measures or systematic BIM-to-BIM overlap measures.

To address this gap, this study proposes an assessment of a smartphone-based scan-to-BIM application built using the RoomPlan API, focusing on the full pipeline from on-device capture to BIM model output. For evaluation, we adopt quantitative, geometry-focused metrics at two levels. First, we use point-cloud-to-BIM confidence indices to assess matching reliability by integrating Coverage, Distribution, and Distance measures [13]. Second, we use BIM-to-BIM matching metrics based on 3D-IoU and 3D-Compactness to compare the outputted BIM model against a reference BIM model, following the benchmark design of Liu et al. [14]. By combining these evaluation layers, the study aims to clarify the challenges and opportunities of fast scan-to-BIM for future research and practice, especially for reliable as-built modelling.

2 Method

This section outlines the method employed for the assessment of a smartphone-based scan-to-BIM workflow, as summarised in Figure 1.

2.1 TLS-based BIM Modelling

The traditional scan-to-BIM methodology utilises TLS to establish an accurate geometric baseline of the existing environment. As illustrated in Figure 1, the workflow follows a sequential, multi-stage process to transform physical space into a digital representation.

Reality capture is performed using a static, multi-station laser scanning approach. The scanner is strategically positioned at various locations throughout the building to ensure comprehensive coverage and minimise data gaps caused by occlusion shadows. Following the acquisition, the individual point clouds from each station are registered into a single coordinate system. This stage results in a high-density unified point cloud that serves as the precise geometric reference of the as-built conditions. Finally, the registered point cloud is used as a spatial reference for manual BIM modelling. During this phase, semantic BIM objects are created by interpreting the point cloud data to represent the building's geometry and components accurately. The resulting manually-created model serves as a validation reference (TLS-BIM) against which the automated outputs of the mobile workflow are quantitatively measured.

2.2 Mobile-app-based BIM Modelling

The mobile-app-based method focuses on the creation of automated BIM models using an app that deploys the RoomPlan API on a mobile device. Unlike the traditional workflow above, this procedure employs a continuous and integrated real-time data acquisition and processing process, which depends on the selected app. RoomPlan API was released by Apple in 2022 for generating a parametric 3D floor plan of an indoor environment using mobile scanning. The capture is performed using iPhone or iPad devices equipped with an RGB camera and a lidar scanner. The API simultaneously conducts real-time scanning and device pose tracking, also known as Simultaneous Localization and Mapping (SLAM). During scanning, RoomPlan combines RGB-D frames and associated camera poses to create semantic point clouds. Then, by using learning-based estimators, the framework creates walls, openings, doors, windows, and furniture objects.

Indoor environment reconstruction is composed of two parts: 3D room layout estimation and 3D object (e.g., furniture) detection. Room layout estimation detects walls, floors, ceilings, openings, doors, and windows in real time. Walls are first predicted as 2D lines in bird's-eye view and then extruded into 3D surfaces based on the distance between the detected floor and ceiling. All spaces bounded by walls, ceiling, and floor are also additionally defined. Doors and windows are detected using multimodal (RGB and lidar) data, and projected into 3D using projection information and camera position. The 3D object detection pipeline recognizes 14 object categories, in addition to windows and doors. These are furniture objects and include: tables, chairs, cabinets, fridges, etc. The object detection combines a local and global detector. Local detector is operating during scanning on a seen camera view, and a global detector is working after scanning on the full reconstructed scene to fuse results (which is particularly necessary for larger

items). All detected objects are then associated with the space within which they have been detected.

The output of the API is thus not only geometric, but also includes semantics, including object types and some object relationships (relationships of walls, ceilings, doors, and furniture with spaces). The results of indoor capturing can be exported in Universal Scene Description (USD) formats. Some apps also map the information into IFC. As previously mentioned, there are various apps in the market that leverage this API, including Polycam [9], Magicplan [10] and Metaroom [11].

To maintain consistency and data integrity during the acquisition process, the following best-practice procedures were followed during scanning:All doors and windows within the scan area were closed.Curtains and blinds remained open to ensure visibility of windows. Scanning was performed using a slow vertical oscillation (up and down movement) to capture ceilings, walls, and floor slabs. A maximum scanning distance of 5 meters from the device to the surface was maintained.Scanning was performed during maximum daylight conditions.The device camera remained open and active from start to end of the scanning.

2.3 Evaluation Criteria

To assess the proposed scan-to-BIM pipeline, an evaluation is conducted at two levels. The point cloud (PCD)-to-BIM evaluation measures how well the BIM elements created using the mobile app match the TLS PCD, and the BIM-to-BIM evaluation compares the mobile app-based BIM model against the BIM model manually created from the TLS PCD. The scope of this evaluation is limited to assessing the geometric fidelity and spatial placement of the core architectural elements

2.3.1 PCD-to-BIM Comparison

Before comparing the TLS PCD to the mobile-app BIM model, they must be aligned in the same coordinate system. We have done this using the 4-Plane Congruent Sets (4-PlCS) method [15] (see implementation at https://github.com/CyberbuildLab/4PlCS). The algorithm extracts planar patches from both the input point clouds and the BIM model. It randomly searches for bases of four planes where three are not pair-wise parallel and the fourth is not co-planar. To handle potential matching failures and reject outliers, the algorithm replaces standard RANSAC with a custom two-step support evaluation method. It first assesses centroid support by checking if the centroids of the point cloud patchs project inside the matched model patches. It then evaluates plane support by calculating the percentage of all remaining point cloud planar patches that support the transformation. Finally, it groups similar transformations and presents a ranked list of the most likely alignments for the user to quickly verify and select the correct one. While this semi-automated approach improves efficiency, manual alignment remains an alternative that does not affect subsequent analyses of this study

Then, for PCD-to-BIM comparison, we adopt the framework of confidence indices proposed by Ersöz and Bosché [13] (see implementation at https://github.com/CyberbuildLab/pcd-bim-matching-confidence).This framework evaluates matching reliability by computing an overall Confidence Index (I-confidence) per BIM element. This is achieved by integrating Coverage (I-coverage), Distribution (I-distribution), and Distance (I-distance) indices, which assess how well scan points align with visible BIM surfaces..Detailed mathematical definitions and algorithmic procedures for these indices are thoroughly described in [13].

2.3.2 BIM-to-BIM Comparison

To assess the geometric similarity between the MOBILE-BIM and TLS-BIM models, we employ the model-to-model evaluation metrics introduced by Liu et al. [14] (see implementation at https://github.com/CyberbuildLab/bim-bim-matching). This approach utilizes two complementary metrics: 3D-IoU and 3D-Compactness. 3D-IoU measures the spatial alignment and Boolean volume overlap between strictly matched semantic classes in the Ground Truth and estimated models. Meanwhile, 3D-Compactness penalizes many-to-many geometric mappings to evaluate whether the reconstructions align in a clean, one-to-one correspondence (highlighting over-segmentation or over-merging). Detailed definitions and the underlying computation logic for these volumetric metrics are provided in [14].

3 Experiments, Results, and Discussion

3.1 Test Environments

To evaluate the mobile device-based scan-to-BIM mobile-app pipeline, we conducted a comparative study across two distinct building typologies representing varying levels of geometric and semantic complexity. The residential case (RES) consists of a single-storey residential unit characterized by medium clutter with an area of approximately 50 m^2. This environment features standard furniture and opaque partition walls, representing a controlled baseline for occlusion. In contrast, the office case (OFF) involves a two-storey university facility characterised by high levels of clutter. The corresponding dataset is split into the first floor (OFF1) and the second floor (OFF2), both with an area of approximately 150 m^2.

3.2 TLS Workflow

To establish a GT baseline, we employed a Leica BLK360 terrestrial laser scanner. The scanner was positioned at multiple stations throughout each room and corridor to capture overlapping point clouds, which were then registered into a unified, high-density point cloud to serve as the precise reference of existing conditions. Using this registered point cloud as an underlay, BIM models were then manually created in Autodesk Revit (Figure 2) and then exported in IFC format. To ensure consistency and eliminate operator variability, the modelling was performed by a single BIM professional with over five years of specialized scan-to-BIM experience. The modelling scope was defined at a Level of Development (LOD) 300, focusing specifically on the elements (walls, floors, doors and windows) necessary for a direct comparative analysis with the mobile application's output.The TLS point cloud and resulting TLS BIM model jointly represent the reference GT.

3.3 Mobile device-based Scan-to-BIM Workflow

After the TLS data collection, the same environments were captured using a lidar-enabled iPad Pro 13-inch (M4) running iPADOS Version 18.1.1. Metaroom version 7.2.1 app, built on the RoomPlan API, was used for modelling [11]. The scanning protocol presented in Section 2.2 was strictly followed to ensure data consistency. While the residential unit was captured in a single continuous session, the complex two-storey office required a segmented approach. Initial attempts to scan both floors in a continuous sequence revealed a limitation in the device's vertical localization ability, where the algorithm struggled to maintain accurate floor-to-floor height consistency during the transition via the staircase, resulting in vertical misalignment (demonstrated in Figure 4a). Consequently, to ensure geometric integrity, we adopted a strategy where each floor is scanned and analysed separately.

Unlike the TLS workflow, no post-acquisition registration or modelling is required; the application performed real-time semantic segmentation, instantly generating and extruding 3D components (walls, doors, windows) into a parametric model immediately upon scan completion. Then, BIM models were exported from the app in IFC format.

3.4 PCD-to-BIM Geometric Confidence Analysis

We first computed the confidence indices of the generated BIM models against the TLS data using the PCD-to-BIM confidence indices. The confidence indices provide a granular view of how well the BIM geometries align with the reality capture data.

Following the method established by Ersöz and Bosché, wall elements were selected as the primary reference for evaluating the confidence indices due to their significant prevalence and geometric diversity within the datasets, and grouped as interior and exterior walls. The results are tabulated in Table 1 and Table 2. Note that the I-confidence reported in these tables is calculated by simply multiplying the I-coverage, I-distribution, and I-distance indices.

The main observation is that MOBILE-BIM models have lower values for all indices in all three cases (Table 1). This highlights the geometric deviation inherent in the mobile device tracking and reconstruction process, which spatial data patterns suggest is primarily driven by SLAM drift. As the device is moved, the real-time tracking over longer trajectories or within texture-poor scenes accumulates small errors, resulting in a cumulative SLAM drift (positional shift). Consequently, while individual wall segments are all detected successfully, their relative placement within the global coordinate system deviates from the GT. This in turn results in fewer TLS points being matched to the MOBILE-BIM elements (which is limited to 5 cm) which results in the observed lower index values.

Exterior Wall Challenges: Table 2 reveals significantly lower confidence scores for exterior walls across both TLS and MOBILE models (e.g., I-confidence ranging from 0.05 to 0.18). This is because scanning was performed indoors, so that the exterior faces of these walls were not scanned. Furthermore, a noteworthy observation in Table 2 is that the MOBILE-BIM appears to demonstrate higher confidence indices for exterior walls compared to the TLS-BIM. This is in fact misleading and is due to chance. Indeed, due to the global positional drift in the mobile workflow, the external faces of the MOBILE-BIM exterior walls were often positioned wrongly (Figure 3) in a way that they ended up residing close and were matched to the available point cloud of the interior wall surface to calculate the indices. The true confidence index values for those walls should have instead been lower than for the TLS-BIM.

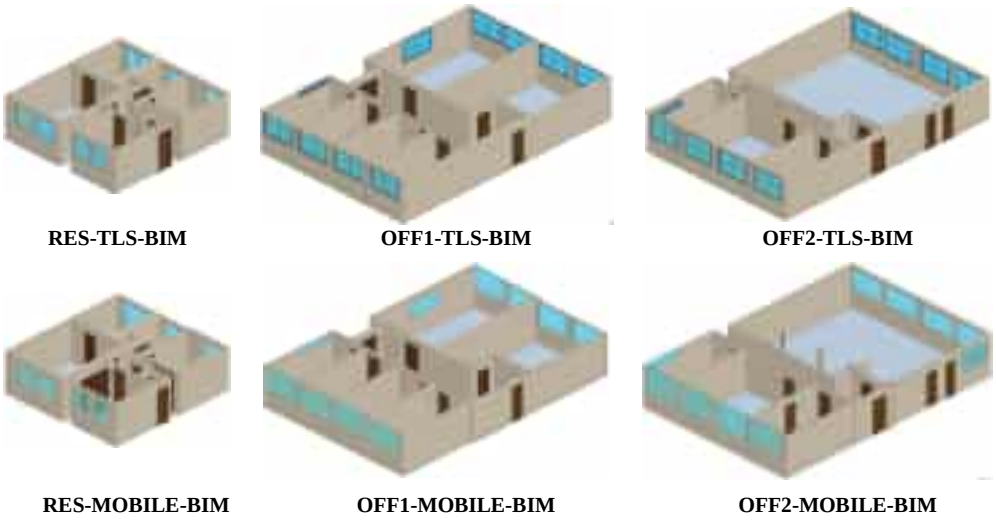


Figure 2. TLS, BIM models created via the TLS and mobile app-based workflow

3.5 BIM-to-BIM Geometric Comparison

We evaluate the BIM model's geometric fidelity using the 3D-IoU and 3D-Compactness indices. A comparative analysis of 3D-IoU and total area difference across the datasets reveals a critical distinction between volumetric precision and quantitative estimation.

Table 1. PCD-to-BIM confidence indices for interior walls

BIM Models	I-cov	I-distri	I-dista	I-confi
RES TLS	0.64	0.70	0.74	0.36
RES MOBILE	0.55	0.63	0.57	0.24
OFF1 TLS	0.71	0.94	0.64	0.44
OFF1 MOBILE	0.50	0.84	0.45	0.20
OFF2 TLS	0.71	0.91	0.67	0.45
OFF2 MOBILE	0.42	0.67	0.47	0.18

Table 2. PCD-to-BIM confidence indices for exterior walls

BIM Models	I-cov	I-distri	I-dista	I-confi
RES TLS	0.36	0.36	0.76	0.10
RES MOBILE	0.48	0.70	0.50	0.18
OFF1 TLS	0.29	0.48	0.55	0.08
OFF1 MOBILE	0.40	0.66	0.49	0.15
OFF2 TLS	0.24	0.42	0.49	0.05
OFF2 MOBILE	0.36	0.54	0.47	0.13

For both residential and office buildings the results were tabulated in Table 3 to Table 5. The total areas presented in the tables are calculated based on specific geometric projections. For vertical components, windows, doors, and walls, the areas are calculated using the projection of the 3D geometry onto their respective wall planes. The areas for space elements are determined by projecting the volumes onto the horizontal (Z) plane, which represents the net floor area.

Walls: Across all three datasets, for the wall elements, the mobile app achieved relatively low m3D-IoU scores, ranging from 0.20 (Office Floor 1) to 0.32 (Residential). Such low IoU values typically indicate issues with wall thickness and positional shift. However, the area difference remains remarkably low. For instance, in the Office Floor 1 dataset (Table 4), the predicted wall area (243.18 m^2) is nearly identical to the ground truth (242.87 m^2), with a difference of only 0.31 m^2. This highlights that the mobile application is highly effective at capturing the existence and length of walls (hence the accurate total area) but struggles with the precise placement and thickness (hence the low 3D-IoU).

Furthermore, as illustrated in Figure 4b, the app misinterprets square columns as short, isolated wall segments rather than distinct column elements, because the API does not have the capability to specifically detect columns. This error directly contributes to lower 3D-Compactness scores for walls, as the one-to-one correspondence between the TLS-BIM wall and the MOBILE-BIM geometry is broken by these false-positive wall segments.

Spaces: In contrast, the room spaces demonstrate better performance, with m3D-IoU scores reaching 0.90-0.93 and 1.00 for compactness in the office datasets. This confirms that the primary strength of the mobile app-based approach lies in room layout estimation. The app successfully generates watertight, logically correct room volumes and topology, making it highly reliable for space management where room boundaries and connectivity are more critical than the precise thickness of the partitioning walls. Top views of the space elements are shown in Figure 3.

Doors: In the residential dataset (Table 3), door elements exhibit a significant difference between directional compactness scores. The TLS-to-MOB score is high (0.88), indicating that the app successfully

detected most existing doors. However, the MOB-to-TLS score is notably lower (0.64), and the predicted door area (20.36 m^2) exceeds the reference TLS model (13.34 m^2). This indicates a high rate of false positives (redundancy). Indeed, the mobile app misinterpreted other wall features (e.g., storage) as doors, as shown in Figure 4c. This suggests that while the app has high recall (finds real doors), its precision is compromised by environmental clutter and class similarities.

Windows: Across all datasets, windows achieved very low 3D-IoU scores. This poor volumetric performance is a direct effect of the misplacement particularly observed for exterior walls. Since most windows are hosted on exterior walls, which suffer from positional shifts due to global drift, the generated windows inherit these spatial offsets. Consequently, even if a window is correctly sized, its 3D volume does not intersect correctly with the TLS-BIM, resulting in a low, even zero IoU score. Despite the positional failure, the area difference remains relatively stable, which implies that the detection logic (fusion of camera and lidar) successfully identifies the presence and dimensions of the openings. The failure is therefore not one of recognition, but of spatial dependency: the openings are correctly seen but placed in incorrectly modelled exterior walls (inadequate thickness).

3.6 Comparison of Workflow Efficiency

Table 6 illustrates the comparative analysis of workflow efficiency of TLS and mobile app workflows. The total time spent decreased from 20 hours for the TLS workflow to 32 minutes for the mobile app-based workflow. When normalized by floor area, this represents an efficiency improvement from 3.5 min/m² to 0.1 min/m².

All steps of the TLS workflow are time consuming. While acquisition required 4 hours (necessitated by the logistics of static setups, target placement, and repositioning to minimize occlusion), point cloud registration took 6 hours, and manual BIM modelling took 10 hours. This 10-hour duration reflects the efficiency of an experienced professional generating a tightly scoped LOD 300 architectural model. This baseline time could vary depending on operator proficiency or if higher LODs were required. In contrast, the mobile app conducts data acquisition, registration and modelling jointly in real-time. In particular, the mobility of the devices means that interior environments consisting of many rooms and occlusions can be visited very fast.

Table 3. Residential House BIM-to-BIM metrics

	m3D-IoU	3D-Comp (TLS-to-MOB)	3D-Comp (MOB-to-TLS)	m3D-Comp	TLS El. Count	MOB El. Count	TLS Total Area (m^2)	MOB Total Area (m^2)
Doors	0.34	0.88	0.64	0.76	8	11	13.34	20.36
Walls	0.32	0.60	0.61	0.61	29	31	143.19	145.86
Windows	0.05	0.25	0.29	0.27	8	7	9.14	8.57
Room spaces	0.84	1.00	1.00	1.00	8	8	49.48	47.04

Table 4. Office Building Floor 1 BIM-to-BIM metrics

	m3D-IoU	3D-Comp (TLS-to-MOB)	3D-Comp (MOB-to-TLS)	m3D-Comp	TLS El. Count	MOB El. Count	TLS Total Area (m^2)	MOB Total Area (m^2)
Doors	0.40	1.00	1.00	1.00	8	8	15.65	13.58
Walls	0.20	0.46	0.42	0.44	20	38	242.87	243.18
Windows	0.03	0.10	0.10	0.10	10	10	22.00	25.08
Room spaces	0.90	1.00	1.00	1.00	7	7	143.96	134.96

Table 5. Office Building Floor 2 BIM-to-BIM metrics

	m3D-IoU	3D-Comp (TLS-to-MOB)	3D-Comp (MOB-to-TLS)	m3D-Comp	TLS El. Count	MOB El. Count	TLS Total Area (m^2)	MOB Total Area (m^2)
Doors	0.32	0.88	0.78	0.83	8	9	13.34	13.95
Walls	0.26	0.57	0.48	0.53	22	29	209.36	209.97
Windows	0.00	0.00	0.00	0.00	9	10	21.40	26.32
Room spaces	0.93	1.00	1.00	1.00	5	5	148.16	140.40

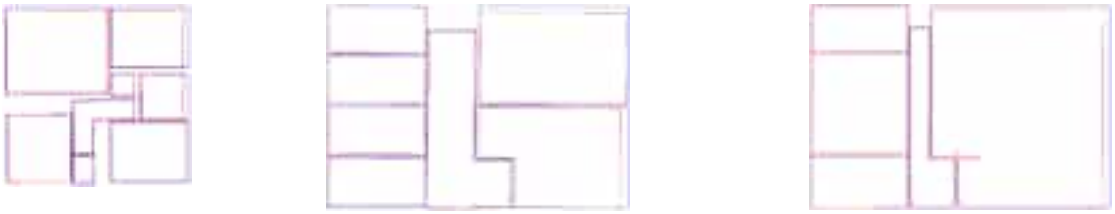


Figure 3. Room space comparison (Blue color: TLS-based BIM , Red color: Mobile-based BIM)

In conclusion, while the TLS workflow prioritises geometric precision at the cost of time, the Mobile device-based Scan-to-BIM prioritises immediacy and semantic readiness, suggesting it could be a suitable option for applications where accuracy is not so important (e.g., retrofit assessment, facility management).

Table 6. Total time required for Residential and Office Building

Workflow Stage	TLS-based	Mobile device-based
Data Acquisition	4 hr	
Data Processing	6 hr	32 min
Modelling	10 hr	
Total	20 hr (3.5 min/m^2)	32 min (0.1 min/m^2)

4 Conclusion

This study presented a quantitative assessment of a mobile app-based Scan-to-BIM workflow utilising a mobile app that uses lidar & RGB camera for data acquisition. The results were benchmarked against TLS scan data and BIM models generated from them. By applying PCD-to-BIM and BIM-to-BIM matching metrics across residential and office environments, we have assessed the capabilities and current limitations of this pipeline.

The major finding of this research is the improvement in workflow efficiency. The mobile app-based pipeline demonstrated a reduction in total production time of 97%.

Despite the observed issues with wall thickness and positioning precision, the high 3D-IoU scores for room space elements confirm that the generated watertight room volumes are accurate. Moreover, the minimal area differences observed for windows and walls indicate that the system is effective at identifying the existence and gross dimensions of building elements.

This makes the workflow uniquely suited for applications where room layouts and asset counts are prioritized over highly accurate geometric fidelity. The mobile app-based solution is also significantly cheaper.

However, this automation introduces geometric trade-offs that must be considered. A primary technical constraint identified in this study is the inaccuracy of wall thickness estimation. The system often struggles to determine actual wall thickness due to factors such as accumulated SLAM drift, frequently reverting to default thicknesses or misinterpreting the distance between finished surfaces. Future research utilizing independent, high-precision localization tracking (e.g., via ground control points) is recommended to quantitatively decouple sensor depth noise from algorithmic SLAM drift. As errors in wall thickness aggregate, the relative positioning of adjacent rooms shifts away from the ground-truth global coordinates. For practitioners, this implies that while the model is reliable for floor area calculations and interior finish schedules, it may require manual adjustment where exact wall-thicknesses are critical.

Another observed challenge is operational scalability. The current API operates on real-time processing with significant memory constraints, which typically limits accurate single-session captures to areas of approximately 15m x 15m; larger floor plans often force a scan completion to preserve data integrity. Furthermore, as evidenced by the vertical misalignment observed in the continuous scanning test, the lack of robust loop closure for vertical transitions yields significant elevation and rotational drifts between floors.

This study evaluates geometric fidelity by utilizing the mobile pipeline's semantic classifications to establish valid, class-strict correspondences between the ground truth and estimated models. Consequently, semantic errors inherently penalize the geometric evaluation; for instance, misidentifying columns as walls creates unmatched false positives that directly lower compactness scores. Building on this integrated approach, future research can incorporate dedicated semantic metrics, such as precision, recall, and confusion matrices, to quantify the system's ability to categorize complex indoor components alongside its geometric reconstruction.



(a) (b) (c)

Figure 4. Misalignments and misclassifications (a) Two-storey scanning test with the app yielded misalignments (b) Misclassified columns as walls in the OFF2 dataset (c) Storage detected as doors in the residential dataset

Acknowledgements

This work was partly supported by an International Postdoctoral Research Fellowship (2219) awarded by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Grant number: 1059B192302109). TÜBİTAK's support is gratefully acknowledged.

References

[1] C. Askar, and H. Sternberg. Use of Smartphone Lidar Technology for Low-Cost 3D Building Documentation with iPhone 13 Pro: A Comparative Analysis of Mobile Scanning Applications. *Geomatics*, 3(4):563–579, 2023.

[2] L. Díaz-Vilariño, H. Tran, E. Frías, J. Balado, and K. Khoshelham. 3d mapping of indoor and outdoor environments using apple smart devices. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2022:303–308, 2022.

[3] T.-A. Teo, and C.-C. Yang. Evaluating the accuracy and quality of an iPad Pro's built-in lidar for 3D indoor mapping. *Developments in the Built Environment*, 14:100169, 2023.

[4] Y. Zhang, R. Chang, W. Mao, J. Zuo, W. E. Zhang, and L. Liu. Low-Cost iOS-Based Automated Detection of Under-Construction Interior Drywalls: An Exploratory Study. *Journal of Construction Engineering and Management*, 2025.

[5] W. Blaszczak-Bak, A. Masiero, P. Bąk, and K. Kuderko. Integrating Data from Terrestrial Laser Scanning and Unmanned Aerial Vehicle with LiDAR for BIM Developing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1–2024:25–30, 2024.

[6] Harshit, P. Chaurasia, S. Zlatanova, and K. Jain. Low-Cost Data, High-Quality Models: A Semi-Automated Approach to LOD3 Creation. *ISPRS International Journal of Geo-Information*, 13(4):119, 2024.

[7] X. Su, D. Campos Zamora, and J. E. Froehlich. RAIS: Towards A Robotic Mapping and Assessment Tool for Indoor Accessibility Using Commodity Hardware. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–5, St. John's NL Canada, 2024.

[8] Apple Machine Learning Research. 3D Parametric Room Representation with RoomPlan. Apple Machine Learning Research. On-line: https://machinelearning.apple.com/research/roomplan, Accessed: 01/01/2026.

[9] Polycam Inc. 3D & 2D Floor Plan Creator - Generate Customizable Floor Plans | Polycam. On-line: https://poly.cam/floor-plans, Accessed: 02/01/2026.

[10] Sensopia Inc. Magicplan for Remodeling Professionals. magicplan. On-line: https://www.magicplan.app/industries/remodeling, Accessed: 02/01/2026.

[11] Amrax Inc. 3D Room Scanner App for iPhone and iPad | Metaroom Scan App. Metaroom by AMRAX. On-line: https://amrax.ai/3d-room-scan/, Accessed: 01/01/2026.

[12] M. Mortazavi, B. Wu, S. Steinbrenner, and S. Haghsheno. Evaluating Site Scanning Methods: An Assessment System for Quantitative Comparison and Selection. In *Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC)*, Lille, France, 2024.

[13] A. B. Ersöz, and F. Bosché. Evaluating confidence in geometric matching between 3D point clouds and BIM models by integrating coverage, distance, and distribution metrics. *Automation in Construction*, 181:106636, 2026.

[14] Y. Liu, H. Huang, G. Gao, Z. Ke, S. Li, and M. Gu. Dataset and benchmark for as-built BIM reconstruction from real-world point cloud. *Automation in Construction*, 173:106096, 2025.

[15] M. Bueno, F. Bosché, H. González-Jorge, J. Martínez-Sánchez, and P. Arias. 4-Plane congruent sets for automatic registration of as-is 3D point clouds with 3D BIM models. *Automation in Construction*, 89:120–134, 2018.

A BIM-IoT Based Digital Twin Framework for Real-Time Indoor Environmental Quality Optimization in Educational Buildings

Yashvi Rajendrakumar Unadkat[1], Varun Kumar Reja[1], Arnab Jana[1]

[1]Department of Civil Engineering, Indian Institute of Technology Bombay, India
yashviunadkat13786@gmail.com, varunreja@iitb.ac.in, arnab.jana@iitb.ac.in

Abstract –

Digital Twin (DT) is an emerging technology that can be used for Indoor Environmental Quality (IEQ) optimization. However, its application in educational buildings, especially in the Indian context, is limited. Students spend a significant amount of their total time indoors in educational buildings, making effective IEQ management critical for their health, comfort, performance, and productivity. This study proposes a bi-directional Digital Twin framework that integrates BIM models, IoT sensors, and actuators to enable real-time monitoring, analysis, and control of the indoor environment. The framework uses real-time data on temperature, humidity, air velocity, and CO_2, which are analyzed using Predicted Mean Vote (PMV) and Percentage People Dissatisfied (PPD), ensuring compliance with ASHRAE 55 and ASHRAE 62.1. To validate the proposed framework, it was implemented in a shared office space at IIT Bombay. Results show stable system performance, reliable real-time data synchronization, visualization, and standards-based IEQ assessment. The monitoring and analytical components were experimentally validated, while the actuation and control layer was established at the framework level for future deployment.

Keywords –

Indoor Environmental Quality, Thermal Comfort, Indoor Air Quality, Digital Twin, Building Information Modeling, Internet of Things, BIM-IoT Integration, IEQ Optimization

1 Introduction

IEQ plays a crucial role in ensuring the health, cognitive performance, and thermal comfort of occupants. People spend 80-90% of their time indoors, and students spend a large portion of this time in educational buildings. Hence, maintaining optimal IEQ in educational buildings plays a critical role in reducing long-term health risks and increasing quality performance [1]. IEQ consists of four parameters: Thermal comfort, Acoustic comfort, Indoor Air Quality (IAQ), and Lighting comfort [2]. Out of these, thermal comfort and IAQ are considered for this study. Currently, people use conventional methods for IEQ optimization that are often manual, inaccurate, inconsistent, and lack seamless information exchange between the physical and digital environments. Therefore, there is a need to enable accurate, efficient, and bi-directional data exchange, which is where DT serves the purpose.

DT is "a digital representation of a physical asset, updated with real-time data, collected using a network of IoT sensors" [2]. In the IEQ monitoring context, DT helps us set up a seamless exchange of real-time information between the physical and digital environments and control the facility. In this way, implementing DT in educational buildings will provide a better indoor environment for students, increasing their productivity, performance, and reducing the health risks.

To develop a DT, BIM-IoT integration plays a foundational role [5]. The primary role of BIM in developing a DT is to produce a virtual extension of a real-world item [6]. On the other hand, the IoT devices help sync the physical and virtual states by enabling real-time data measurement, which supports IEQ optimization in real-time [7], and forms the base of the DT. But the framework for the implementation of DT to optimize IEQ in developing countries is still limited, especially for educational buildings.

To tackle this limitation, this study proposes a framework for the deployment of a DT to monitor, analyze, and control IEQ in an educational building unit. DT enables enhanced dynamic spatial and temporal data visualization and monitoring, which is achieved using a DT software called Autodesk Tandem in this study. The real-time analysis of the sensor data helps the system to make decisions as per standard guidelines. The thermal comfort analysis utilizes two indices: Predicted Mean Vote (PMV) and Percentage People Dissatisfied (PPD), and the ASHRAE 55 standard. On the other hand, IAQ analysis is done as per ASHRAE 62.1 standard

compliance. The decisions made as per the analysis are used to control the physical environment using suggested actuators like the IR Blaster. The results of this study provide a unified framework for DT, serving as a bridge between monitoring, analysis, and control to achieve a healthier environment.

The paper is broadly divided into five sections: in Section 1, we introduce the problem statement, Section 2 reviews current approaches for IEQ monitoring, DT implementation, and identifies some research gaps. In Section 3, we will examine the proposed methodology and case study for deploying a DT for IEQ monitoring. In the last two sections, we will move on to the results, draw some conclusions, discuss future work, and summarize the study.

2 Literature Review

This section reviews existing studies on IEQ optimization techniques, their drawbacks, and how the application of DT, BIM, and IoT can serve the purpose. The review identifies technical research gaps and limitations that form the basis for the proposed framework.

2.1 IEQ Monitoring Techniques

IEQ has gained attention over the past few years, particularly in educational buildings, due to its direct impact on students' cognitive performance [8]. Especially in developing countries like India, the deteriorating outdoor environment quality significantly affects the indoor air quality, which is one of the 10 leading causes of illness [1]. Hence, monitoring and improving IEQ in buildings is a widely studied topic. Several studies have focused on assessing IAQ and thermal comfort in educational buildings, highlighting inadequate ventilation, high CO2 concentrations, and thermal discomfort, particularly in the Indian context [1]. IEQ monitoring studies generally follow two approaches: objective and subjective. Objective methods include field measurements, and subjective methods are based on personal judgment surveys [9]. Many of the studies rely on survey-based assessment and hence provide limited insight into the dynamic behaviour of the indoor environment.

Poor IAQ is reported as the second major health threat in India, after poor water quality and hygiene [1]. And thermal comfort directly affects user performance, which is why these two are included in this study. Thermal comfort depends on indoor environmental conditions as well as the subjective factors of occupants, like clothing and metabolic rate [2]. If we look into objective methods to measure IEQ, the technologies used to measure thermal comfort are directed towards data-driven models based on networks of connected sensors. [10]. Similarly, methods used to monitor IAQ include Wireless Sensor Networks (WSN) and IoT. But few parameters revealed that the performance of these systems is low compared to real-time and dynamic monitoring IoT systems [11]. Hence, integration of real-time IoT data acquisition with building models is important to monitor IEQ.

2.2 BIM, IoT, and Digital Twin for IEQ Optimization

BIM is an important component for information management, focusing on data interchange between different stakeholders, and has wide applications. BIM-based approaches improve spatial context but are generally disconnected from real-time data streams and control strategies. Therefore, IoT-based systems are being used to collect real-time data. IoT has made buildings go from ordinary to smart buildings by integrating everyday objects with digital controllers; such improvements have made it possible to create a controllable and integrated network supporting decision-making and human well-being [11]. However, most of these systems operate using one-way communication, where data flows only from the physical environment to digital dashboards for visualization, not for control. DT technology has emerged as a bridge between this gap by enabling bi-directional communication. Other industries have been using DT for many years, but it came to attention in the Architecture, Engineering, Construction, and Operations (AECO) industry with the concept of smart buildings and smart cities [4].

Recent studies have demonstrated the potential of DTs for monitoring thermal comfort, energy consumption, and IAQ in buildings. There are three stages of DT development:

1. Digital Model, which is a static representation of the physical environment, independent of sensor data, and cannot be changed [12].
2. Digital Shadow, which has one-way data flow, i.e., the model updates automatically when any change in the physical system is detected, but not the other way around [13].
3. Digital Twin, which is “a virtual replica of a physical asset that collects and sends real-time information”. It is a two-way communication loop, within which dynamic data flows and optimizes the physical environment [2]. DT effectively fills the communication gap between the physical and digital environments and creates a digital replica by integrating real-time IoT data with a BIM model. For example, a classroom with sensors sending real-time data and actuators like IR blasters controlling the devices based on the decisions made.

Data analysis serves as the core of the decision-making process, utilizing ASHRAE 55 and ASHRAE

62.1 guidelines. Nevertheless, most existing DT experimental setups are limited to residential or commercial buildings. DT application in educational buildings is very scarce, particularly in the Indian context, where educational buildings often operate under high occupancy and mixed-mode ventilation, making IEQ management more challenging under poor outdoor air quality conditions. Moreover, many studies have limited emphasis on automated actuation. Given these limitations, there is a strong need for a standardized and deployable framework to set up a DT for IEQ optimization in educational buildings. Therefore, this study aims to develop a DT by integrating a BIM model with real-time IoT data and actuators for a shared office space at IIT Bombay. The key contribution of this study includes:

1. Integration of BIM-IoT data to set up bi-directional communication.
2. Framework for the development of a deployable DT for IEQ Monitoring in educational buildings.
3. Mechanism for real-time optimization of IEQ.

3 Methodology

This section presents the proposed methodology for developing a BIM-IoT integrated Digital Twin framework for real-time IEQ monitoring and optimization. It describes the workflow, system architecture, data acquisition process, analysis approach, and actuation strategy. The methodology is further tested and demonstrated till the actuation part through a case study conducted in an educational building unit at IIT Bombay. The workflow of the proposed DT framework is shown in Figure 1. It illustrates the process from real-time data acquisition to BIM-based visualization, analysis, and actuation, forming a continuous loop.

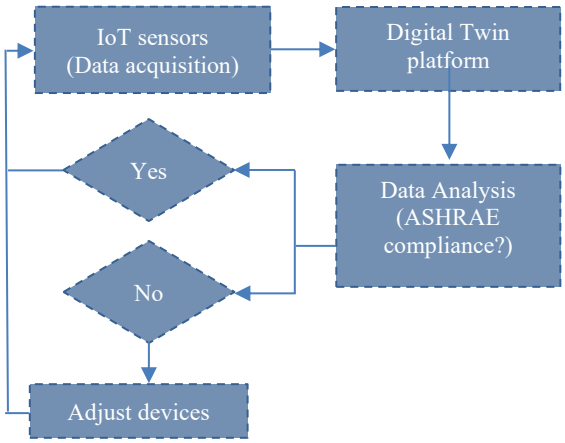


Figure 1. Proposed system workflow

3.1 DT Framework Architecture

The DT framework proposed in this study consists of five interconnected layers that collectively establish a bi-directional communication loop between the physical environment and its digital representation. These layers are described as follows:

1. Physical Layer: The physical layer comprises real-world assets like the building and IoT sensors. This layer is responsible for sensing, collecting, and providing real-time physical data for subsequent processing.
2. Data Transfer Layer: The data transfer layer enables the transmission of collected sensor data from the physical environment to the DT platform. This layer integrates real-time IoT sensor data with the BIM model on the digital twin platform, facilitating the creation of a synchronized DT.
3. Visualization Layer: The visualization layer supports real-time visualization of incoming data, allowing users to monitor system performance, observe alerts, and analyze historical trends through dashboards and charts.
4. Analysis Layer: The analysis layer plays a critical role by interpreting the received data to support informed decision-making. The outcomes of this layer form the basis for the actuation layer, as the results are translated into control commands for the actuation layer.
5. Actuation Layer: The actuation layer implements control actions based on analyzed results. Actuators, like IR Blasters, can be used to control devices like fans and ACs to maintain optimal IEQ. The integration of this layer is designed to enable closed-loop control of the system.

Together, these five layers establish a bi-directional DT system, forming a continuous feedback loop that spans real-time data acquisition, transfer, analysis, visualization, and control. The interrelation and sequence of action of all these five layers are as shown in Figure 2. These layers support some specific tasks, which are discussed below.

3.1.1 Data collection and transfer

The methodology adopts an integrated approach for developing a DT of an educational building unit to optimize Indoor Air Quality and Thermal Comfort. The framework integrates the above-mentioned five layers to enable a bi-directional DT. However, in the current stage of implementation, the focus is on real-time monitoring, visualization, and standards-based analysis, while the actuation logic is designed and planned to be deployed in upcoming stages. The data to be collected using IoT sensors must first be selected based on the requirements of the analysis. For assessing thermal comfort and IAQ,

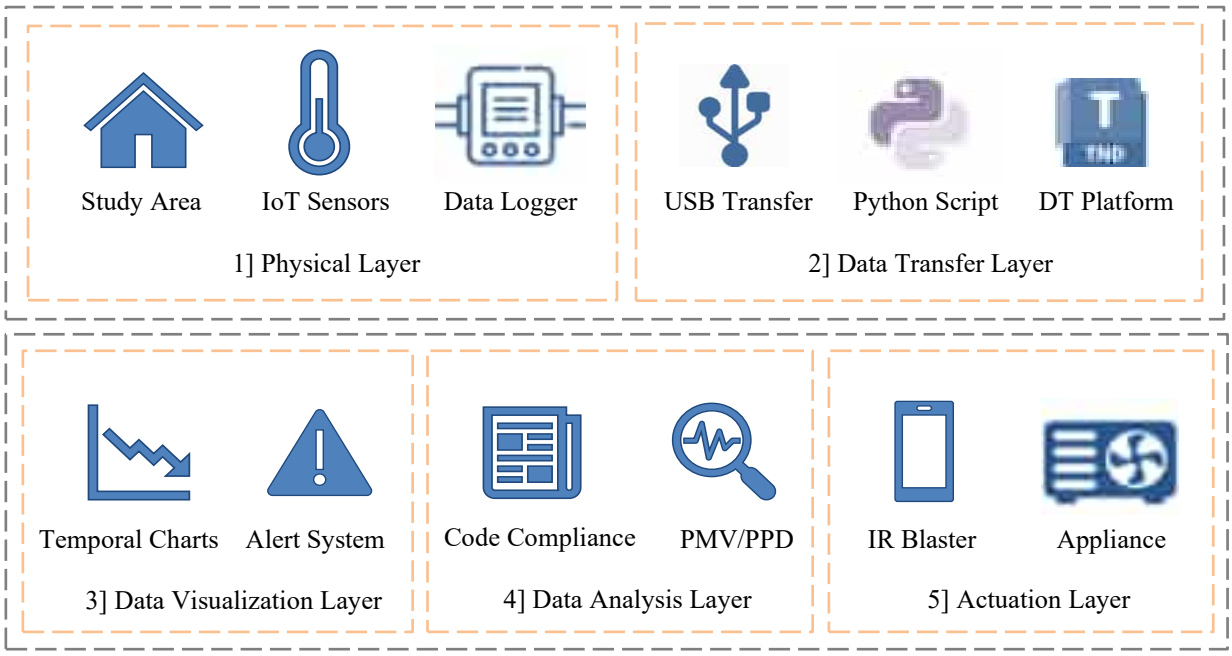


Figure 2. DT framework architecture

specific parameters should be chosen in accordance with ASHRAE standards and relevant research, as presented in Table 1 below.

Table 1. Selected parameters to assess thermal comfort and IAQ [14],[15], [16], [17]

Parameters	Description
Air Temperature (°C)	Directly impacts human perception of a hot/cold environment
Radiant Temperature (°C)	Radiative heat exchange from surfaces
Relative Humidity (% RH)	Affects Thermal comfort and sensation
Air speed (m/s)	Directly impacts the rate of convective and evaporative heat loss
CO2 (ppm)	Has an impact on cognitive performance

These parameters will be collected in real-time using IoT sensors, where thermal comfort parameters will be analyzed using indices PMV, PPD, along with checking compliance with ASHRAE 55. On the other hand, IAQ parameters will be analyzed by checking compliance with ASHRAE 62.1 standards. There are two more such parameters like: Particulate Matter (PM) and Volatile Organic Compounds (VOCs), which are responsible for IAQ, but this study utilizes only CO_2.

Real-time data acquisition is carried out using IoT sensors, where sensor selection and placement play a critical role in data accuracy and reliability. Selection criteria for IoT devices include measurement accuracy, availability, and cost-effectiveness. Sensor positioning follows ASHRAE guidelines with sensors placed near the occupant's breathing zone, working/sitting zones, and other representative points of the space. To develop a DT, the digital representation of the physical environment should be created first. A detailed BIM includes the architectural elements, furniture layout, and represents the physical environment, occupants' zones, and spatial layout. The BIM model is then to be integrated with real-time environment data collected from sensors deployed within occupant zones. These sensors continuously measure key IEQ parameters related to thermal comfort and IAQ, as mentioned in Table 1.

3.1.2 Data Visualization and Analysis

Real-time sensor data is transferred from the physical layer to the digital environment via the data transfer layer, enabling synchronization between the two systems. There are different software programs available to develop a DT, like Bentley iTwin, Autodesk Tandem, Ansys Twin Builder, etc. We can use any software that provides easy integration of BIM and Real-time IoT sensor data, and that offers multiple ways to visualize spatial and temporal data. For this study, we have worked on Autodesk Tandem because Tandem and Revit are released by the same software company, favoring the cross-platform compatibility and enabling Autodesk Tandem to read the Revit model correctly [6]. The

incoming data is mapped to corresponding zones, Zone A and Zone B, as shown in Figure 3, to allow users to visualize IEQ conditions using temporal charts, heatmaps, and customizable dashboards. These visualization tools help us study the trends for the selected parameters, which may help in future prediction as per historic trends. Additionally, the Autodesk Tandem provides a feature that triggers an alert whenever the parameter data exceeds the assigned permissible range as per ASHRAE 55 and ASHRAE 62.1.

Later, an analysis layer processes the sensor data to assess the compliance with thermal comfort and IAQ standards. Thermal comfort is evaluated using the PMV and PPD indices in accordance with ASHRAE 55, and IAQ parameters are assessed in accordance with ASHRAE 62.1. The permissible limits of all the selected parameters are shown in Table 2. The analysis layer then determines the corrective actions based on the predefined decision rules.

Table 2. Permissible range of thermal comfort and IAQ parameters as per ASHRAE, WHO guidelines [15],[16],[17]

Parameters	Permissible limit
Average Air Speed (m/s)	< 0.20 m/s
Air Temperature (°C)	22 °C - 26 °C
Mean Radiant Temperature (°C)	+2 °C of air temperature (Ta)
Relative Humidity (%RH)	30-60 %RH
CO_2 (ppm)	≤ 1000 ppm above outdoor (~1400 ppm max.)

3.1.3 Actuation Strategy

To enable bi-directional communication, the methodology incorporates an actuation layer that transforms the analytical decisions into control signals for the physical devices. These signals are generated to be transmitted to actuators like IR Blasters, which can regulate devices such as ACs and fans in a deployed setup. IR blaster devices work just like a normal remote, but based on the signals it receives from the analysis layer. This actuation strategy closes the loops between monitoring, analysis, and control, thereby distinguishing the proposed system from conventional one-way monitoring approaches.

It also provides the basis for the future evaluation of energy consumption impacts and occupant well-being improvements. In this study, the actuation workflow was defined and validated conceptually but not physically executed.

3.2 Case Study

This case study demonstrates the proposed framework to deploy a DT in an educational building unit. The selected unit is a shared office space at the Department of Civil Engineering at Indian Institute of Technology Bombay, which is a 50-year-old building with faculty offices, classrooms, conference rooms, and laboratories.

3.2.1 The study area

The study area shown in Figure 3 is a room occupied by research scholars for an average of 6 to 8 hours on weekdays. The space has an area of 25.6 m² (8m x 3.2m x 3m). A glass partition wall divides the room into two parts, each measuring 12.2 m² and 13.4 m², occupied by the professor and students, respectively. The walls are of brick masonry with plaster finish, while the floor is tiled, and the finished ceiling contains lighting fixtures and a fan. The partition wall with a door allows partial transparency and limited air circulation between the two spaces. Both zones are equipped with wooden desks with computer systems, chairs, and storage shelves.

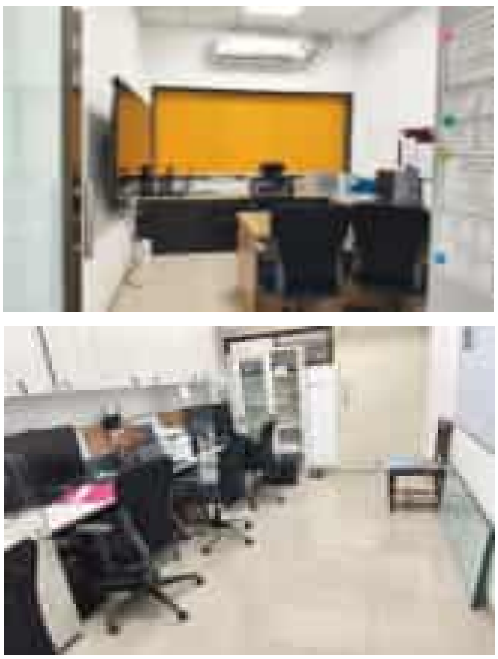


Figure 3. Selected study area

The Zone A (professors' seating area) has a split-type air conditioner unit, along with ceiling fans and a glass window on the wall. Zone B, occupied by research scholars, is equipped with a fan which is remote controlled. The BIM model (architecture and furniture layout) for the study area cabin was created using Autodesk Revit, as shown in Figure 4.

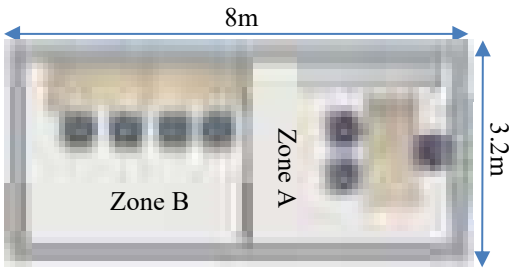


Figure 4. Study area model

3.2.2 Data Acquisition and Visualization

To capture real-time data, a Testo multi-sensor kit was used due to its high measurement accuracy, availability, and its compliance with ASHRAE standards. These sensors were deployed within the occupants' breathing zone using tripods set up as per ASHRAE 55 and ISO 7762 guidelines. The Testo multi-sensor kit, as shown in Figure 5, consists of some sensors and components as mentioned in the list below:

1. Globe Temperature sensor
2. CO_2 sensor
3. Air Velocity vein sensor
4. Air Temperature/ Humidity sensor
5. Testo 480 data display



Figure 5. Sensors used to collect data

The sensors recorded real-time data for temperature, humidity, air velocity, and CO_2 at predefined intervals of 30 seconds which were shown on Testo 480 display device. The data is then sent to the computer system on Easyclimate software via USB connection. The updated version of Easyclimate software allows users to send real-time data, but the available older version was not compatible with Autodesk Tandem. Therefore, to send real-time data, a pipeline using a Python script and Node-RED was created to transfer the average of collected data at an interval of 2.5 minutes; this frequency is selected as per ASHRAE standards. As the data arrives at the DT model, the visualization can be done in the form of heatmaps, temporal stream charts, and an alert system based on the set threshold values, as shown in Figure 5.

3.2.3 Data Analysis and Actuation

Next, the analysis layer will compute the thermal comfort indices (PMV and PPD) using ASHRAE 55, assuming a standard metabolic rate of 1.0 met (activity: sitting and typing) and clothing of 1.0 clo unit (can vary from case to case). IAQ parameters will be checked for compliance with ASHRAE 62.1 using a Python script. IAQ analysis is done based on CO_2 only. Other parameters like PM and VOCs can be used to get more accurate IAQ optimization. For thermal comfort analysis, if at any time thermal comfort indices (PMV and PPD) are going out of the permissible range, the parameters affecting thermal comfort indices will be checked, and whichever parameter is deviating will be adjusted by controlling the respective device. For example, if the temperature is going out of range, the AC temperature can be controlled. When any parameter deviates from the threshold values, the actuation layer is designed to be triggered. To enable automated control, an IR blaster has to be deployed within the study area. A Python script-generated signal can be sent to the IR Blaster via Wi-Fi network, thereby regulating thermal conditions and air quality in real-time. The actuation part is planned to be deployed. The IR blaster selected for use is the BroadLink RM4 Mini, due to its compatibility with Python and Node-RED, its widespread use in research, and its Wi-Fi-based design. This creates a bi-directional loop, as the adjusted data will be collected again through sensors on predefined intervals, and the entire procedure is designed to be repeated upon deployment.

3.2.4 Validation Strategy

Validation strategy to evaluate system performance under real-world conditions includes:

1. Assessing improvements in IEQ indicators before and after automated control.
2. Testing the responsiveness of the control system by measuring the time lag between the detection of parameter deviation and the actuation response.
3. Comparison of manual vs automated operation modes and reduction in electricity use.

The results will demonstrate the feasibility and effectiveness of deploying a bi-directional DT framework for IEQ monitoring and control in educational buildings within the Indian context.

4 Results and Discussion

This paper provides a framework for monitoring IAQ and thermal comfort in educational buildings using a DT. The results demonstrate the feasibility and effectiveness of integrating BIM and IoT systems to develop a DT for IEQ optimization. The proposed framework of DT successfully established continuous

data flow between the physical environment and the digital model. Real-time sensor data collected at a defined interval was found to be sufficient to capture meaningful variations in parameters. Despite software compatibility constraints, the integrated pipeline using Python and Node-RED enabled stable data synchronization. The observed latency between physical data acquisition and digital visualization remained within acceptable limits for facility management use cases, indicating quick system responsiveness. The DT platform enabled real-time visualization of IEQ parameters using temporal stream charts, color-coded heatmaps, and an alert-based system. The BIM-integrated visualization provided enhanced spatial awareness for facility managers, as shown in Figure 6.

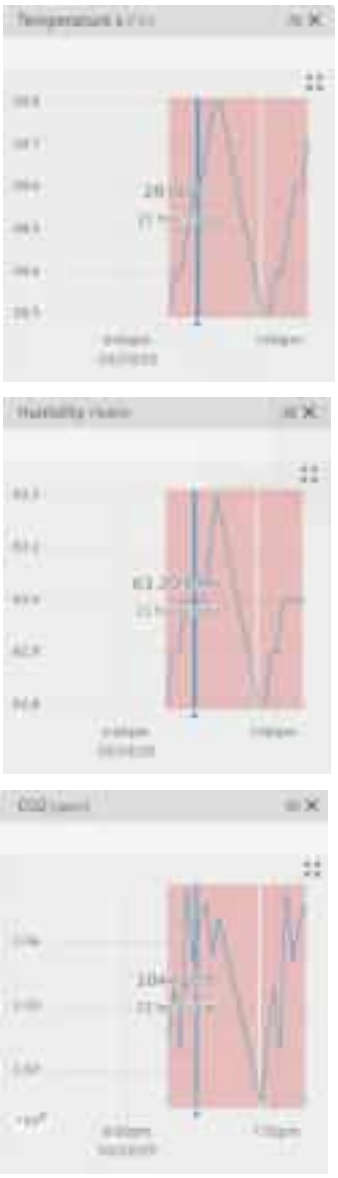




Figure 6. Visualization of real-time data on Autodesk Tandem

Despite the absence of the field-level actuation, the successful integration of monitoring, analysis, and control logic demonstrates framework readiness.

5 Conclusion and Future Work

This paper presents a bi-directional Digital Twin framework for monitoring and optimizing IEQ in educational buildings by integrating BIM, real-time IoT sensor data, analytical models, and actuation logic. The proposed approach addresses the limitations of conventional IEQ management systems by enabling continuous data exchange between the physical and digital environments and supporting standards-based decision-making in real time. The case study conducted at IIT Bombay demonstrates the feasibility of deploying the proposed framework in an educational setting. The DT enabled real-time visualization of thermal comfort and indoor air quality parameters through BIM-integrated dashboards and alert systems, with acceptable system responsiveness despite software compatibility constraints. The results confirm that real-time monitoring combined with ASHRAE 55 and ASHRAE 62.1 compliance assessment can effectively identify thermal comfort and air quality deviations. By incorporating an actuation strategy, the framework establishes the foundation for closed-loop control, distinguishing it from

traditional one-way monitoring approaches. This capability has the potential to improve occupant comfort and well-being while supporting more efficient building operations and energy use. Overall, the proposed Digital Twin framework offers a scalable and practical solution for intelligent IEQ management in educational buildings.

Although the present study validates the monitoring, visualization, and standards-based decision framework, physical actuator integration remains to be implemented and tested under real world conditions. Further extensions will include quantification of energy consumption impacts resulting from automated control actions and comparative analysis between manual and automated operation modes. Additionally, IAQ parameters, such as PM and VOCs, will be incorporated to enhance IAQ assessment. Occupant feedback surveys will be integrated to correlate objective IEQ indicators with subjective comfort responses.

References

[1] N. Vijay and J. D. Sarasamma. A critical review on the indoor air quality status of schools in India. Current World Environment, 19(3):1061–1076, 2024.

[2] S. Assaf, M. Assaf, X. Li, A. Bouferguene, and M. Al-Hussein. A digital twin-based system for indoor environmental quality monitoring in off-site construction facilities. ISARC Proceedings of the International Symposium on Automation and Robotics in Construction, 42(1):885–892, 2025.

[3] D. Accardo, S. Meschini, M. Boscariol, L. C. Tagliabue, and G. M. Di Giuda. Towards a digital twin for IAQ monitoring and control in educational facilities through asset management system platform. Proceedings of the European Conference on Computing in Construction (EC3), page 326, 2023. doi:10.36680/j.itcon.2023.326.

[4] Z. Zheng, J. D. Castaño Molina, E. T. Mengiste, S. A. Prieto, and B. García de Soto. Digital twins for healthier spaces: A scalable framework for monitoring indoor environmental quality. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-G-2025:1053–1060, 2025.

[5] C.-K. Lo. The application of digital twin for indoor air quality management: A case study in Hong Kong. Journal of Sustainable Built Environment, 2025. doi:10.70731/8sk2aa87.

[6] T. De Rubeis, A. Ciccozzi, A. Tata, D. Ambrosini, M. Ragnoli, V. Stornelli, and S. Brusaporci. A workflow for a building information modeling-based thermo-hygrometric digital twin: An experimentation in an existing building. Sustainability, 16(10):1–21, 2024.

[7] A. K. Saxena, V. K. Reja, and K. Varghese. IoT-enabled framework for real-time management of power tools at construction projects. Journal of Construction Engineering and Management, 1(1):1–12, 2019.

[8] A. A. Lawal, O. A. Dare-Abel, and O. O. Ajayi. A review of indoor environmental quality (IEQ) parameters and their impact on occupant satisfaction in educational buildings. Journal of Architecture and Built Environment, 1(1):1–12, 2021.

[9] L. Pistore, F. Cappelletti, P. Romagnoni, and A. Zonta. Assessment of IEQ in two high schools by means of monitoring, surveys, and dynamic simulation. Building and Environment, 1(1):1–15, 2019.

[10] E. T. Saligue, R. Y. Saligue, J. A. Olaybar, K. N. L. Sy, and P. D. Cerna. Smart home automation with IoT-driven IR blasters: Enhancing control of legacy appliances with predictive usage pattern using machine learning. Proceedings of the International Conference on Smart Electronics and Communication, 1(1):861–868, 2024.

[11] J. Saini, M. Dutta, and G. Marques. A comprehensive review on indoor air quality monitoring systems for enhanced public health. Sustainable Environment Research, 30(1):6–18, 2020.

[12] T. D. Nguyen. The role of BIM in integrating digital twin in building operations. Sustainability, 15(13):10462, 2023. doi:10.3390/su151310462.

[13] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn. Digital twin in manufacturing: A categorical literature review and classification. IFAC PapersOnLine, 51(11):1016–1022, 2018.

[14] P. T. B. S. Branco, S. I. V. Sousa, M. R. Dudzinska, D. G. Ruzgar, M. Mutlu, G. Panaras, G. Papadopoulos, J. Saffell, A. M. Scutaru, C. Struck, and A. Weersink. A review of relevant parameters for assessing indoor air quality in educational facilities. Environmental Research, 261:119713, 2024.

[15] ASHRAE. ANSI/ASHRAE Standard 55-2017: Thermal environmental conditions for human occupancy, Addendum d. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2020.

[16] ASHRAE. ANSI/ASHRAE Standard 62.1-2022: Ventilation for acceptable indoor air quality. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2022.

[17] World Health Organization. WHO guidelines for indoor air quality: Selected pollutants. World Health Organization, 2010.

An Integrated Unmanned Aerial Vehicle-Based Framework for Crack Detection and Classification of Bridge Infrastructure

Saleh Abu Dabous[1], Fatma Hosny[1], Bharadwaj R. K. Mantha[1,2] and Sainab Feroz[1]

[1]Department of Civil and Environmental Engineering, University of Sharjah, United Arab Emirates
[2]Department of Civil Engineering, Indian Institute of Technology Madras, Tamil Nadu, India
sabudabous@sharjah.ac.ae, fmohamed@sharjah.ac.ae, rmantha@sharjah.ac.ae, U19104579@sharjah.ac.ae

Abstract – Bridge infrastructure is highly susceptible to deterioration due to continuous exposure to harsh environmental conditions and increasing traffic volumes and loads, necessitating frequent inspections to ensure structural safety and serviceability. Traditional bridge inspection approaches are often time-consuming, labor-intensive, error-prone, and disruptive to traffic operations, highlighting the need for more efficient, reliable, and non-invasive inspection solutions. Although several studies have proposed sensor-based and unmanned aerial vehicle (UAV)-assisted inspection systems, limited research has focused on the systematic identification, classification, and quantitative assessment of surface crack severity using image-based approaches. To address this gap, this study develops an integrated UAV-based framework for automated crack detection and classification within a bridge monitoring system. The proposed framework advances existing approaches through the integration of crack identification using gray-scale pixel-based image analysis and crack classification based on geometry, orientation, and severity. The methodology comprises four main stages: UAV-based data acquisition, image preprocessing and crack detection, quantitative crack characterization and classification, and validation through laboratory testing and a real-world bridge case study. Preliminary case study results show that the proposed gray-scale pixel-based algorithm achieves minimum errors of 2.9% and 5.5% and maximum errors of 17.9% and 16.0% for crack length and width estimation, respectively. Severity classification was consistent for moderate and extended cracks (2 out of 4 segments), while fine cracks exhibited higher misclassification rates due to reduced contrast and image quality. Orientation profiling showed correct or partial agreement for most detected crack segments. Higher errors under real-world conditions highlight the need for improved image acquisition stability.

Keywords–Digital Inspection; Computer Vision; Structural Health Monitoring (SHM); Automated Quantification; Human Robot Interaction (HRI); Infrastructure Resilience Assessment.

1 Introduction

Bridge infrastructure is a critical component of highway transportation systems and plays a vital role in ensuring connectivity, safety, and economic development. Due to continuous exposure to harsh environmental conditions, heavy traffic, and aging, bridges are highly susceptible to deterioration and damage, necessitating regular inspection and maintenance [1,2]. However, transport authorities often operate under strict time and budget constraints, which limit the effectiveness and frequency of traditional inspection practices [1]. Conventional bridge inspection processes rely primarily on visual and manual inspection, making them labor-intensive, time-consuming, subject to personal judgment, and potentially dangerous for inspectors [3]. In addition, these inspections often require road closures and may fail to identify hard-to-reach or hidden defects, resulting in incomplete condition assessments [4]. To address these limitations, recent research has increasingly focused on non-contact and non-destructive testing (NDT) technologies for bridge inspection and monitoring, including photogrammetry, infrared thermography, and unmanned aerial vehicles (UAVs) [5,6]. Infrared thermography has also been used in combination with machine learning techniques to estimate crack properties such as crack width from temperature variations captured in thermal images [7]. Among these technologies, UAV-based inspection has received considerable attention due to its ability to safely access difficult-to-reach locations, reduce inspection time and cost, minimize traffic disruption, and provide high-resolution visual data for structural assessment [4,8]. UAV-mounted visual imaging systems have been widely applied for crack detection, surface damage

identification, and geometric measurements of bridge elements, with several studies reporting accuracy comparable to that of conventional field measurements [9]. UAV-based images further enable the reconstruction of three-dimensional (3D) models from overlapping images, supporting improved localization and quantitative assessment of surface defects [10]. Photogrammetry-based approaches have been successfully applied to generate textured 3D models of concrete elements and extract crack characteristics such as crack length, width, and deformation behavior from image data [11]. Previous studies have demonstrated the potential of UAV imagery for crack detection, deformation monitoring, corrosion assessment, and post-disaster damage evaluation of bridge infrastructure [12]. Despite these advances, UAV-based inspection approaches remain sensitive to imaging conditions such as lighting variability, image blur, and surface texture, which can affect the reliability of crack detection and measurement [13]. At the same time, significant efforts have been made to automate the crack detection using image processing and computer vision techniques. Traditional image-based methods rely on filtering, thresholding, edge detection, and morphological operations to extract crack features and estimate their dimensions [14]. More recently, artificial intelligence and deep learning approaches have been introduced to improve detection accuracy under changing environmental conditions [15]. However, such data-driven methods typically require large, well-labeled datasets and extensive training, which are often unavailable in practical bridge inspection scenarios.

Moreover, many existing crack detection approaches primarily focus on identifying the presence of cracks rather than systematically characterizing crack geometry and orientation [16]. In particular, limited attention has been given to extracting quantitative crack properties such as crack length, width, and orientation while simultaneously interpreting crack patterns relative to bridge components and infrastructure condition assessment. Such detailed crack characterization is essential for supporting maintenance planning and structural evaluation of bridge infrastructure.

1.1 Research Gaps

Despite significant progress in UAV-based bridge inspection, several research gaps remain in the current literature. First, many existing studies focus on individual techniques or standalone inspection methods and lack integrated, image-based frameworks that support systematic crack identification, profiling, and quantitative assessment within an automated and updatable bridge monitoring workflow [17]. Second, while crack detection has been widely investigated, limited attention has been given to extracting and utilizing gray-scale pixel-level information for robust crack identification and quantitative characterization of crack properties such as length, width, and orientation.

Finally, most existing UAV-based crack detection approaches emphasize automated detection but rarely incorporate systematic analysis of crack orientation patterns or their relationship with specific bridge components [13]. The geometric orientation of cracks, such as horizontal, vertical, and diagonal patterns, can provide valuable insight into potential structural behavior and deterioration mechanisms. However, this contextual interpretation is often missing in existing automated inspection frameworks. Therefore, there remains a need for integrated methodologies that combine crack identification, quantitative crack characterization, and orientation-based crack profiling to support practical bridge inspection and maintenance decision-making.

1.2 Objectives and Contributions

To address the identified research gaps, this study aims to develop a UAV-based bridge monitoring framework using image-based techniques for crack detection and assessment. A UAV-based image analysis approach is implemented for concrete highway bridges to support surface crack detection, characterization, and classification. Unlike many existing UAV inspection approaches that rely heavily on deep learning models requiring large labeled datasets, the proposed framework utilizes gray-scale pixel-based image analysis to detect crack features and extract their geometric properties. The methodology integrates crack detection with quantitative characterization of crack attributes and orientation-based profiling to enable a more comprehensive interpretation of crack patterns relevant to bridge infrastructure assessment. The specific objectives of this study are to 1) develop a gray-scale pixel-based approach for identifying surface cracks from UAV-acquired images; 2) extract quantitative crack characteristics, including length, width, and orientation, from detected crack features; and 3) profile cracks based on their geometric orientation (horizontal, vertical, and diagonal) to support systematic assessment of bridge infrastructure; 4) determine crack severity levels (e.g., low, moderate, and extended) by integrating gray-scale crack identification and geometric crack profiling results to support condition assessment of bridge components. The main contribution of this study lies in the development of an integrated UAV-based inspection framework that combines gray-scale image analysis with orientation-based crack profiling and quantitative crack characterization within a unified monitoring workflow. This approach enables systematic crack assessment without relying on extensive training datasets typically required by deep learning models. The methodology is verified through a real-world bridge case study conducted in the United Arab Emirates (UAE),

demonstrating its applicability to automated, non-invasive bridge inspection.

2 Materials and Methods

This study develops a UAV-based image analysis methodology for bridge condition assessment focused on surface crack detection and characterization. UAV-assisted inspection provides a non-contact and non-invasive means of acquiring high-resolution visual data, enabling access to hard-to-reach bridge components while reducing inspection time, cost, and traffic disruption. The proposed framework follows a sequential, image-driven workflow, as illustrated in Figure 1, consisting of six main actions: acquiring UAV images, detecting surface cracks, extracting crack geometry, profiling crack orientation, classifying crack severity, and supporting informed inspection and maintenance decision-making. First, high-resolution images of bridge components are acquired using a UAV under controlled environmental conditions to ensure adequate image quality. Second, the acquired images are directly processed using a gray-scale pixel-based crack detection algorithm to identify potential crack regions. Third, quantitative crack geometry, including crack length and width, is extracted from the detected crack features. Fourth, cracks are profiled based on their geometric orientation as horizontal, vertical, or diagonal. Fifth, crack severity levels are determined by integrating crack geometry and orientation information to distinguish between minor and more critical surface defects relevant to bridge inspection. Finally, the extracted crack characteristics are used to support systematic interpretation of bridge surface condition and inform inspection and maintenance considerations.

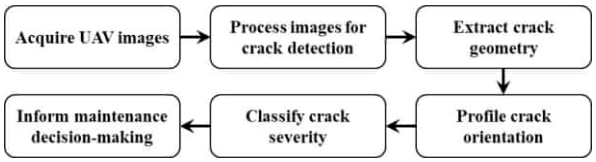


Figure 1. Research methodology framework

2.1 Acquire UAV Images

The data acquisition stage focuses on collecting sufficient high-resolution visual information to support reliable image-based crack detection and quantitative crack characterization of bridge surface defects. UAVs are employed to capture high-resolution, overlapping images of bridge components from multiple viewpoints, ensuring adequate coverage and redundancy for image processing and crack analysis [10,18]. Careful flight planning is undertaken to achieve appropriate image overlap, viewing angles, and spatial resolution required for reliable crack identification and measurement. Data acquisition planning considers environmental conditions, including lighting variability, wind speed, and operational constraints, to minimize image blur and distortion during flight. Particular attention is given to capturing close-range images of critical structural components to support gray-scale, pixel-based crack identification and the subsequent extraction of crack characteristics, such as length, width, and orientation.

2.2 Process Images for Crack Detection

The acquired UAV images are directly processed using image processing techniques without relying on 3D surface reconstruction. Images are converted to gray-scale format and analyzed based on pixel intensity variations to enhance crack features relative to the surrounding surface texture. Preprocessing operations, including noise suppression and contrast enhancement, are applied to reduce the influence of illumination variability and background irregularities. This processing stage establishes a consistent basis for crack detection by improving the separability of crack pixels from non-crack regions. The processed images are then prepared for segmentation, enabling reliable identification of crack features across different bridge components and imaging conditions.

2.3 Extract Crack Geometry

Following crack detection, quantitative crack geometry is extracted from the segmented crack regions. Crack length is computed based on the spatial extent of detected crack features, while crack width is estimated using pixel-based geometric analysis techniques. These measurements provide objective and repeatable metrics for characterizing crack extent and severity. The extracted geometric properties form the quantitative foundation for crack profiling and severity classification. By relying on gray-scale pixel information and geometric measurements, the methodology enables consistent crack characterization without the need for prior training data or complex learning-based models.

2.4 Profile Crack Orientation

Detected cracks are subsequently profiled according to their geometric orientation and classified as horizontal, vertical, or diagonal. Orientation profiling is performed by analyzing the spatial alignment of crack features relative to the image coordinate system. This classification supports systematic comparison of crack patterns across different bridge components and aids in interpreting crack behavior with respect to structural configuration, as different crack orientations may be associated with different damage mechanisms or

structural responses. In addition to orientation profiling, crack severity is classified into three discrete levels, namely fine, moderate, and extended, based on quantitative crack width measurements. These severity categories are defined using width thresholds commonly adopted in bridge inspection practice, enabling consistent differentiation between less critical and more significant surface defects. The combined use of crack orientation and severity classification provides contextual information essential for systematic condition assessment of bridge infrastructure.

2.5 Classify Crack Severity

Crack severity is classified by integrating quantitative crack geometry and orientation information. Severity is defined using three discrete categories, namely fine, moderate, and extended, which are determined primarily based on crack width thresholds commonly adopted in bridge inspection practice [14], enabling differentiation between less critical and more significant surface defects. The integration of crack geometry and orientation supports a structured and interpretable severity assessment framework. This classification approach facilitates consistent evaluation of crack conditions across bridge elements and supports prioritization of defects that may require closer inspection or maintenance intervention.

2.6 Inform Maintenance Decision-Making

The combined outputs of crack detection, geometric extraction, orientation profiling, and severity classification are used to inform the interpretation of bridge surface condition. By providing quantitative and contextual crack information, the proposed methodology supports inspection reporting and maintenance planning, enabling engineers and asset managers to better understand defect characteristics and their potential implications. Although the framework does not replace engineering judgment, it provides a transparent and systematic basis for supporting inspection decisions and identifying areas that may require further investigation or monitoring.

3 Methodology Implementation

This section presents the implementation of the proposed methodology described in Section 2. The methodology is applied to a real-world bridge case study to demonstrate its practicality and effectiveness for bridge condition assessment. The implementation includes site reconnaissance, UAV-based data acquisition, execution of the image-based crack detection and quantification process, and validation through laboratory testing and field application.

3.1 Site Reconnaissance

A prestressed concrete highway bridge located in Ghalilah, Ras Al Khaimah, UAE, was selected to implement and evaluate the proposed UAV-based bridge inspection methodology. The bridge is located along the E11 highway and comprises four traffic lanes supported by six simply supported spans, each 11.5 m long, resulting in a total length of 71 m and a vertical clearance of approximately 9.1 m. The superstructure consists of a precast reinforced concrete deck supported by cast-in-place concrete piers and closed abutments with wingwalls, with expansion joints, parapets, and safety barrier railings provided along the deck edges. The bridge was selected due to the availability of historical inspection reports, which supported validation of the proposed methodology, as well as its representative lighting conditions and limited surrounding obstructions. The inspection records were reviewed to obtain information on structural characteristics, component geometry, and previously reported defects, aiding UAV flight planning and identification of damage-prone areas. As shown in Figure 2, reported defects included concrete efflorescence, spalling, and vertical and horizontal cracking on the bridge deck. A site reconnaissance survey was conducted to identify potential hazards to UAV operations, including traffic flow and nearby obstructions, and to determine suitable take-off and landing locations. A preliminary visual inspection was also performed to identify bridge elements requiring close-range UAV imaging. Required approvals for UAV operation were obtained from the relevant authorities, and pre-flight checks were conducted in accordance with regulatory and manufacturer guidelines to ensure safe UAV operation during data acquisition.



Figure 2. Condition of the bridge structure extracted from inspection reports

3.2 Acquire UAV Images

A UAV-based data acquisition strategy was adopted to capture high-resolution images suitable for image-based crack analysis. Data collection was conducted under favorable environmental conditions, including

sufficient lighting, low wind speeds, and moderate ambient temperatures. A DJI Inspire 2 quadcopter vertical take-off and landing (VTOL) UAV equipped with a Zenmuse X7 camera featuring a 24-megapixel, 35-mm sensor was used for image acquisition. Live video streaming provided a first-person view of the UAV trajectory and enabled real-time identification of areas requiring additional inspection. Due to proximity to structural elements and the presence of GPS-denied environments beneath the bridge deck, the UAV was operated in manual flight mode with approximately 80% horizontal overlap. Manual operation allowed adaptive flight-path adjustment and improved control during close-range imaging. The UAV was flown at an average distance of approximately 2m from the bridge surface with an inclination angle of about 15º. An initial flight focused on imaging the underside of the bridge deck, followed by two additional flights targeting cracked and spalled regions identified during site reconnaissance. In total, 64 high-resolution images and two video sequences were acquired for subsequent analysis.

3.3 Process Images for Crack Detection

The acquired UAV images were processed using MATLAB®-implemented image processing algorithms. Images were directly analyzed without three-dimensional reconstruction. Each image was loaded into the processing interface, where resizing was applied as required to reduce computational demand. Image preprocessing included smoothing, Gaussian filtering, and adaptive histogram equalization to suppress noise, reduce image disturbances, and enhance contrast while limiting amplification of background noise.

Primary segmentation was performed using gray-scale thresholding to convert the image into a binary representation, highlighting potential crack regions. A global gray-scale threshold was applied to separate dark crack pixels from the surrounding concrete surface based on pixel intensity differences following preprocessing operations. The threshold value was determined empirically during algorithm calibration to enhance crack visibility while suppressing background noise. In the context of the conducted experiments, the fine-tuned gray-scale threshold value was set to 90, based on empirical calibration across representative sample images, where threshold values between 70 and 120 were evaluated to achieve optimal separation between crack pixels and background intensity levels. Secondary segmentation employed an active contour region-growing approach to refine crack boundaries and separate crack features from the background.

3.4 Extract Crack Geometry

Post-processing operations were applied to reduce false detections, including area-based filtering to remove small, isolated segments and mean-intensity-based dark-blob filtering to eliminate regions insufficiently distinct from the background. Following post-processing, quantitative crack geometry was extracted from detected crack segments. Crack length was determined using skeletonization, which extracts crack centerlines while preserving topological continuity. Crack width was estimated using a distance transform approach, whereby distances from crack boundaries to the centerline were averaged and converted to physical dimensions. These geometric measurements provided objective metrics for crack characterization. Based on the laboratory validation results, the smallest crack width successfully detected during laboratory validation was approximately 0.72 mm under controlled imaging conditions. Under field conditions, the effective detectable crack width may increase due to image resolution limitations, UAV vibration, lighting variability, and perspective distortions during image acquisition.

3.5 Profile Crack Orientation and Classify Crack Severity

Detected cracks were profiled based on geometric orientation and classified as horizontal, vertical, or diagonal according to predefined angular thresholds. Crack segments with orientation angles less than 30° or greater than 150º were classified as horizontal cracks. Segments with angles between 30º and 70º or between 110º and 150º were classified as diagonal cracks, while remaining orientations were classified as vertical cracks.

Crack severity was subsequently classified by integrating crack width measurements with predefined severity categories. Crack width thresholds were defined as fine cracks (0.1-1 mm), moderate cracks (1-5 mm), and extended cracks (>5 mm), consistent with commonly adopted bridge inspection practices [19].

3.6 Inform Maintenance Decision-Making and Validation

The extracted crack geometry, orientation profiles, and severity classifications were used to inform the interpretation of bridge surface condition. By providing quantitative and contextual crack information, the methodology supports inspection reporting and identification of areas requiring closer monitoring or intervention. Validation was performed at both laboratory and field scales. Laboratory validation used four concrete beam specimens (50×15×15cm) with vertical cracks induced through four-point flexural testing. Crack dimensions were measured using an electronic handheld digital microscope with a maximum magnification of 1600x, and corresponding images were captured under controlled conditions (Figure 3).

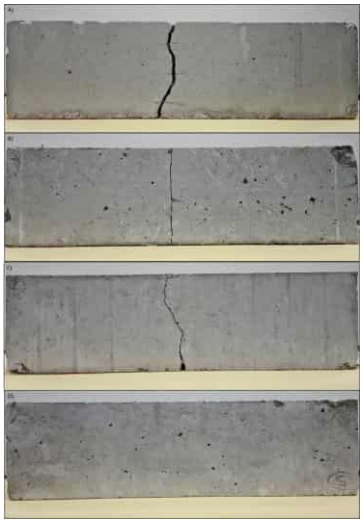


Figure 3. Images of cracked beam: A) specimen 1, B) specimen 2, C) specimen 3, and D) specimen 4

The crack delineation results obtained using the MATLAB®-based algorithm are illustrated in Figure 4. In all specimens, the algorithm successfully identified crack segments. Minor background noise, including surface discoloration, markings, and small spalled regions, was occasionally misclassified as cracks. However, these artifacts were excluded through manual refinement. Manual intervention was primarily required to remove false-positive regions and recover crack segments that were partially undetected due to low image contrast or segmentation discontinuities. In the laboratory validation stage, manual refinement was limited to excluding a small number of non-crack segments, representing less than approximately 10% of detected regions. This indicates that the majority of crack detection and geometric characterization was performed automatically by the algorithm, while manual intervention was limited to correcting segmentation artifacts and refining crack boundaries. The algorithm correctly identified crack orientation in all laboratory specimens. Quantitative validation results indicate that crack length and width estimation errors were less than 17.9% and 16.0%, respectively. Specimen 3 exhibited the highest accuracy, with errors of 2.9% for crack length and 5.5% for crack width. In contrast, specimen 4 exhibited the most significant errors, highlighting the increased difficulty in identifying very fine cracks that exhibit limited contrast relative to background noise. These results indicate that uncertainty in crack detection arises from several factors, including image resolution, crack visibility relative to background texture, segmentation threshold sensitivity, and UAV image acquisition stability. In particular, motion-induced image blur and non-perpendicular viewing angles may reduce the detectability of fine cracks and introduce measurement uncertainty in crack geometry estimation.

A comparison between measured and estimated crack dimensions is provided in Table 1.

Table 1. Comparison of measured and estimated crack dimensions

Specimen #		Original	Estimated	Percent Error (%)
1		16.100	16.710	**3.789**
2	Length	16.700	17.395	**4.162**
3	(cm)	18.200	18.742	**2.978**
4		6.300	5.171	**17.921**
1		0.280	0.251	**10.357**
2	Width	0.100	0.092	**8.000**
3	(cm)	0.072	0.076	**5.556**
4		0.025	0.029	**16.000**

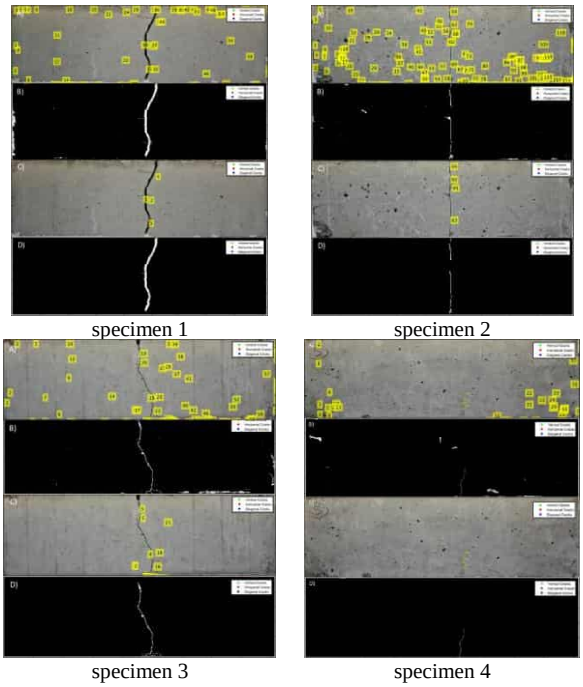


Figure 4. Overlay and binarized crack detection results for laboratory specimens before and after manual enhancement.

Following laboratory validation, the crack detection algorithm was applied to UAV-acquired images of the bridge deck. Visual inspection identified multiple cracks within an approximately 80×100cm region of the deck surface. As shown in Figure 5, diagonal fine cracks were observed in segments 1 and 2, a vertical extended crack in segment 3, and a horizontal moderate crack in segment 4. This image region is presented as a representative example to illustrate the implementation of the proposed crack detection and classification methodology. The algorithm was applied to several UAV-acquired images collected during the inspection campaign; however, the selected region was used to demonstrate the step-by-step crack detection process and the interpretation of crack geometry, orientation, and severity classification results.

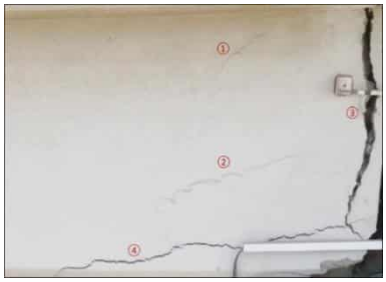


Figure 5. Cracks identified on an image of the bridge deck extracted using a UAV

Automated crack detection results are presented in Figure 6. Initial segmentation results revealed incomplete crack delineation and misclassification of certain background elements as cracks, particularly in regions containing fine cracks. Manual refinement was therefore performed by marking undetected crack regions and excluding background elements such as cables and spalled areas. Manual adjustments were required primarily for two of the four crack segments detected in the case study, particularly for fine cracks where contrast between crack features and surrounding concrete surfaces was limited. The initial identification and geometric extraction of crack features were performed automatically by the algorithm, while manual refinement was applied only to correct segmentation artifacts and improve crack continuity in challenging regions. Despite these adjustments, some crack segments remained partially disconnected due to image blur and non-perpendicular viewing angles during image acquisition.

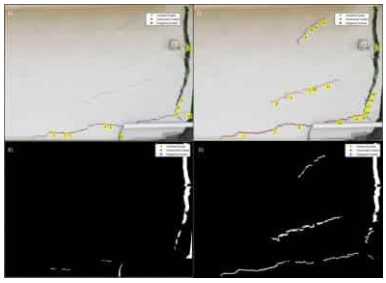


Figure 6. Images obtained from crack detection analysis, showing automatic segmentation and corresponding overlay results before and after manual refinement.

The dimensional statistics reported by the algorithm are summarized in Table 2. Crack segments 3 and 4 were correctly classified in terms of both orientation and severity, while segments 1 and 2 were misclassified as moderate cracks and partially misidentified as horizontal cracks. These discrepancies are attributed primarily to UAV vibration, image blur, and perspective distortion, which reduced crack visibility and contrast in fine and moderate crack regions.

Table 2. Comparison of crack dimensions, orientation, and severity

Segment #	Width (mm)	Height (mm)	Dimension		Orientation	
			Original	Classified	Original	Classified
1	3.87	223.08	Fine	Moderate	Diagonal	Horizontal / Diagonal
2	4.93	403.94	Fine	Moderate	Diagonal	Horizontal
3	7.32	646.02	Extended	Extended	Vertical	Vertical
4	3.87	750.17	Moderate	Moderate	Horizontal	Horizontal

4 Conclusions and Future Work

This study presented a proof-of-concept UAV-based bridge inspection framework for crack assessment that integrates gray-scale pixel-based crack detection, geometric characterization, and severity classification. The framework was implemented and evaluated through preliminary laboratory experiments and a real-world bridge case study in the UAE to demonstrate its feasibility and practical applicability under controlled and field conditions.

The preliminary results indicate that the proposed image-based approach is capable of identifying crack geometry, orientation, and severity trends, with acceptable accuracy under laboratory conditions and increased uncertainty under real-world imaging constraints. These findings highlight both the potential and current limitations of UAV-acquired imagery for detailed crack assessment, particularly when image clarity and acquisition stability are affected. Sources of uncertainty include variations in lighting conditions, image blur caused by UAV vibration, surface texture variability, and perspective distortion during image acquisition. These factors may affect the reliability of crack segmentation and introduce uncertainty in crack geometry estimation, particularly for very fine cracks. Based on the laboratory validation results, the smallest crack width successfully detected using the proposed methodology was approximately 0.72 mm under controlled imaging conditions. However, under field conditions, the effective detectable crack width may increase due to image resolution limitations and environmental factors affecting image clarity.

By combining crack geometry and orientation profiling with severity classification, the framework supports context-aware interpretation of surface cracks relevant to bridge components and inspection reporting. The methodology, therefore, provides a structured and transparent basis for identifying crack patterns and prioritizing areas that may require closer inspection or maintenance intervention.

Overall, the proposed methodology demonstrates the feasibility of a non-invasive, image-driven approach for UAV-based bridge inspection and provides a foundation for more comprehensive and robust implementations. Although the presented results focus on a representative

example image region to illustrate the methodology, further validation using larger annotated image datasets would enable more comprehensive statistical evaluation of detection and classification performance. Future research will therefore focus on expanding the dataset and incorporating quantitative performance metrics such as precision, recall, and confusion matrices to provide a more rigorous assessment of algorithm performance.

Acknowledgment

The authors acknowledge the support of industry professionals and technical experts whose insights contributed to the data collection, analysis, and validation of this research. Individual names are withheld for confidentiality.

References

[1] S. Abu Dabous and S. Alkass, Managing bridge infrastructure under budget constraints: A decision support methodology, *Can. J. Civ. Eng.*, vol. 38, no. 11, pp. 1227–1237, 2011, doi: 10.1139/l11-082.

[2] W. Cook, P. J. Barr, and M. W. Halling, Bridge Failure Rate, *J. Perform. Constr. Facil.*, vol. 29, no. 3, p. 04014080, Dec. 2015, doi: 10.1061/(asce)cf.1943-5509.0000571.

[3] D. Agdas, J. A. Rice, J. R. Martinez, and I. R. Lasa, Comparison of Visual Inspection and Structural-Health Monitoring As Bridge Condition Assessment Methods, *J. Perform. Constr. Facil.*, vol. 30, no. 3, p. 04015049, Jun. 2016, doi: 10.1061/(asce)cf.1943-5509.0000802.

[4] J. Seo, L. Duque, and J. P. Wacker, Field Application of UAS-Based Bridge Inspection, *Transp. Res. Rec.*, vol. 2672, no. 12, pp. 72–81, Dec. 2018, doi: 10.1177/0361198118780825.

[5] M. Rashidi, M. Mohammadi, S. S. Kivi, M. M. Abdolvand, L. Truong-Hong, and B. Samali, A decade of modern bridge monitoring using terrestrial laser scanning: Review and future directions, *Remote Sens.*, vol. 12, no. 22, pp. 1–34, 2020, doi: 10.3390/rs12223796.

[6] S. Feroz and S. A. Dabous, Uav-based remote sensing applications for bridge condition assessment, *Remote Sens.*, vol. 13, no. 9, 2021, doi: 10.3390/rs13091809.

[7] A. Hamdan *et al.*, Crack Width Detection for Concrete Structures Using Thermal Images and Artificial Neural Network, *Proc. - Int. Conf. Dev. eSystems Eng. DeSE*, pp. 19–24, 2024, doi: 10.1109/DeSE63988.2024.10911945.

[8] F. Hosny, S. Barakat, and S. A. Dabous, Statistical Analysis of Operational Carbon for Different Building Types in the UAE, *Jordan J. Civ. Eng.*, vol. 18, no. 2, pp. 346–359, 2024, doi: 10.14525/JJCE.v18i2.15.

[9] L. Duque, J. Seo, and J. Wacker, Bridge Deterioration Quantification Protocol Using UAV, *J. Bridg. Eng.*, vol. 23, no. 10, p. 04018080, Oct. 2018, doi: 10.1061/(asce)be.1943-5592.0001289.

[10] Y. F. Liu, X. Nie, J. S. Fan, and X. G. Liu, Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction, *Comput. Civ. Infrastruct. Eng.*, vol. 35, no. 5, pp. 511–529, May 2020, doi: 10.1111/mice.12501.

[11] R. Al-Ruzouq, S. Abu Dabous, M. T. Junaid, and F. Hosny, Nondestructive deformation measurements and crack assessment of concrete structure using close-range photogrammetry, *Results Eng.*, vol. 18, no. September 2022, p. 101058, 2023, doi: 10.1016/j.rineng.2023.101058.

[12] F. Jalinoos, M. Amjadian, A. K. Agrawal, C. Brooks, and D. Banach, Experimental Evaluation of Unmanned Aerial System for Measuring Bridge Movement, *J. Bridg. Eng.*, vol. 25, no. 1, p. 04019132, Oct. 2020, doi: 10.1061/(asce)be.1943-5592.0001508.

[13] S. Chen, D. F. Laefer, E. Mangina, S. M. I. Zolanvari, and J. Byrne, UAV Bridge Inspection through Evaluated 3D Reconstructions, *J. Bridg. Eng.*, vol. 24, no. 4, p. 05019001, Apr. 2019, doi: 10.1061/(asce)be.1943-5592.0001343.

[14] S. Dorafshan, R. J. Thomas, and M. Maguire, Benchmarking image processing algorithms for unmanned aerial system-assisted crack detection in concrete structures, *Infrastructures*, vol. 4, no. 2, Apr. 2019, doi: 10.3390/infrastructures4020019.

[15] S. Yoon, G. H. Gwon, J. H. Lee, and H. J. Jung, Three-dimensional image coordinate-based missing region of interest area detection and damage localization for bridge visual inspection using unmanned aerial vehicles, *Struct. Heal. Monit.*, vol. 20, no. 4, pp. 1462–1475, Jul. 2021, doi: 10.1177/1475921720918675.

[16] S. Abu Dabous and S. Feroz, Condition monitoring of bridges with non-contact testing technologies, *Autom. Constr.*, vol. 116, no. April, p. 103224, 2020, doi: 10.1016/j.autcon.2020.103224.

[17] G. Morgenthal *et al.*, Framework for automated UAS-based structural condition assessment of bridges, *Autom. Constr.*, vol. 97, pp. 77–95, Jan. 2019, doi: 10.1016/j.autcon.2018.10.006.

[18] F. Nex and F. Remondino, UAV for 3D mapping applications: A review, Nov. 08, 2014, *Springer*. doi: 10.1007/s12518-013-0120-x.

[19] Federal Highway Administration's (FHWA) Specifications, Bridge Inspector' s Reference Manual (BIRM), 2023.

Assessing Network Level Reconnaissance Risks in ROS2 Enabled Service Robots for Smart Infrastructure Operations

Abdul Shakir[1], Hiba Jomaa[1], and Bharadwaj R. K. Mantha*[1,2]

[1]College of Engineering, University of Sharjah, Sharjah, United Arab Emirates.
[2]Department of Civil Engineering, Indian Institute of Technology, Madras, India.
shakirabdul31@gmail.com, u22105324@sharjah.ac.ae, rmantha@sharjah.ac.ae

Abstract –

Service robots have been increasingly deployed across several infrastructure applications. ROS2 has become the widely used operating system for service robots due to its performance and compatibility. However, its default configuration lacks authentication, encryption, or access control. While prior studies have highlighted general ROS2 security weaknesses, the feasibility of reconnaissance attacks in realistic service robot deployments remains underexplored. A controlled setup using ROS2 Humble was developed, in which a custom publisher node continuously transmitted robot positional data (i.e., real time location) over data distribution service (DDS) topics. Fast DDS Spy was used to discover domain participants, topics, data writers, while wireshark captured and analyzed underlying real-time publish–subscribe (RTPS) traffic. The experiments were repeated across ten independent simulation runs to ensure consistent observations. The results demonstrate that in all ten simulation runs, an unauthorized entity on the same network could access eight types of sensitive operational data (e.g., robot's internet protocol (IP)) with consistent observation across trials. These findings highlight integrity and confidentiality risks in default ROS2 deployments and emphasize the need for stronger security controls to safeguard service robots particularly in collaborative environments. These observations can serve as valuable guidance for researchers and practitioners designing and utilizing autonomous service robots. In addition, the study's findings provide valuable insights and significantly complement the smart city initiatives and the UAE vision of a resilient, secure, sustainable, and intelligent infrastructure ecosystem.

Keywords – On-demand Cobots; Cybersecurity; Vulnerabilities; Construction automation; AEC industry; Connected sites; Futuristic infrastructure.

1 Introduction

Service robots are increasingly deployed to perform task-oriented functions across a wide range of real-world environments, including infrastructure inspection, facility management, hospitality, healthcare, and public services. These robots are designed to execute specific operational tasks such as monitoring, inspection, navigation, or human interaction, with development efforts primarily focused on achieving reliable application-level functionalities. As a result, less attention is often given to the underlying operating system and security architecture that support these capabilities, particularly in early-stage development and deployment [1]. In many cases, service robots operate while connected to shared or public wireless networks, remaining continuously exposed to the internet as they roam through open environments, which significantly expands their attack surface. In architecture engineering construction and operations (AECO) industry, robots and drones are used for site inspection, progress monitoring, and hazard detection. In indoor built environments such as commercial buildings and facilities, service robots support environmental monitoring, security patrols, and asset inspection [2][3]. Comparable task driven deployments are observed in hospitality and healthcare sectors, as well as in emerging autonomous transportation systems. The practical importance of these applications is reflected in city-level initiatives; for example, Dubai aims to transition 25% of its transportation systems to autonomous modes by 2030, signaling the growing reliance on robotic systems within critical urban infrastructure.

To enable communication and coordination among robotic components, the Robot Operating System (ROS) is widely adopted as a modular software framework for robotic development. Since its initial open-source release in 2009, ROS1 facilitated rapid prototyping and system integration by providing reusable libraries and

standardized interfaces across heterogeneous hardware and software platforms. However, ROS1 was not designed with security or real-time guarantees as primary considerations, and its key limitations are discussed in detail in Section 2.2 [4]. To address these drawbacks, ROS2 was introduced as a redesigned framework targeting modern service robot requirements. ROS2 improves real-time performance, supports multi-robot and distributed deployments, and adopts middleware technologies such as the Data Distribution Service (DDS) to enable scalable communication [4] [5]. DDS is a middleware standard that enables real-time, reliable, and scalable data exchange between distributed systems using a data centric publish subscribe communication model.

Despite these architectural improvements, many ROS2 deployments continue to operate using default middleware configurations that do not enforce authentication, encryption, or fine-grained access control. As service robots largely operate over shared wireless networks or public Wi-Fi, these default configurations expose communication channels to passive and active adversaries. Vulnerabilities at the operating system or middleware level can compromise confidentiality, integrity, and availability of robotic data and control flows, potentially affecting mission execution and human safety [6]. Therefore, systematic analysis of network-level exposure, protocol behavior, and security weaknesses in ROS2-based systems is essential. This applies to all service robot deployments, since continuous connectivity and human interaction can increase risks, potentially affecting the confidentiality, integrity, and availability across all layers of robotic operation. This study addresses this gap by examining the type of security relevant information that can be identified using currently available tools. Unlike prior studies that limited their analysis to traffic observation and protocol identification, this study uniquely combines Fast DDS Spy at the middleware layer with Wireshark at the network layer to systematically enumerate eight categories of sensitive operational data extractable through passive reconnaissance, without authentication or direct system interaction. It also serves to enhance the awareness among business stakeholders by demonstrating how the under prioritization of these risks can lead to significant financial and operational consequences incase of potential cybersecurity incidents involving service robotic operation.

2 Literature Review

The reviewed literature indicates that although ROS2 introduces significant improvements over its predecessor such as enhanced modularity and the adoption of DDS based communication, its security architecture is still not fully mature [7]. Multiple studies such as [5],[6],[8] report persistent weaknesses, including 1) insufficient authentication mechanisms between nodes, 2) limited or optional encryption configurations, and the 3) absence of built-in intrusion detection capabilities. As a result, ROS based systems remain vulnerable to a range of cyber threats. These threats can affect the performance of any service robot, regardless of its application, by putting the robot's primary tasks at risk. These threats can lead to several potential impacts on robotic systems, including 1) disruption of service continuity, 2) degradation of system performance, and 3) compromise of one or more elements of the Confidentiality, Integrity, and Availability (CIA) triad [9]. Furthermore, the literature highlights that the security posture of a ROS2 system is strongly influenced by the underlying DDS middleware implementation.

For the safe operation of service robots, multiple aspects were examined from prior studies to identify key findings and research gaps. The review begins at the build environment level, where Section 2.1 discusses operating environments, followed by Section 2.2, which analyzes the prominent DDS in robot communication. The discussion then extends to the larger system level, where Section 2.3 focuses on network-level communication and analysis based on relevant studies. This motivated the division of the literature review into three sections: 2.1 Analysis of different types of environments used to assess vulnerabilities of service robots in ROS2, 2.2 Analysis of DDS suppliers, i.e., the middleware entities, to examine the state-of-the-art literature in this domain, and 2.3 Review of studies related to understanding the network operability of service robots. Additionally, the review aimed to critically analyze the current literature to identify gaps in ROS2 security research, highlighting areas where cybersecurity risks and vulnerabilities have been underexplored. Together, these sections offer a cohesive understanding of the environment, middleware characteristics, and network interactions, highlighting opportunities to expand on existing research

2.1 Analysis of Environment Type

A total of 25 recent highly relevant studies were reviewed and discussed in detail to examine the experimental environments used in prior ROS2 and service robot research. The main objective was to examine the types of experimental environments used in previous studies such as simulations, hybrid or live deployments. Table 1 summarizes the distribution of experimental environments across the reviewed studies.

Simulation based experiments were the most adopted approach, with 13 out of 25 studies conducted entirely in simulation. Tools such as Gazebo and Webots, alongside virtual machines running Ubuntu, were employed for

fully virtual testing and system modeling. The prevalence of simulation reflects its practicality due to lower costs, reduced logistical complexity, and the ability to safely conduct experiments without disrupting operational service robots. Simulation also allows the construction of diverse scenarios, including potential attack vectors and operational conditions, enabling systematic testing that closely mirrors real-world interactions [10]. This controlled approach ensures repeatable and consistent results, providing confidence in the findings and supporting their relevance to operational service robots. 5 out of 25 studies adopted a hybrid approach, combining simulation with validation on mobile robots such as TurtleBot3 and MiR100. This method balances the safety and flexibility of simulation with the realism of physical testing, addressing some limitations of purely simulated environments. 7 studies performed experiments entirely on physical robots, including UR5 robotic arms and Unitree G1 humanoids, or conducted real-world usability evaluations and security audits. While real-world testing provides high-fidelity insights, it is often limited by high costs, complex integrations, and the potential for disrupting operational service robots [9].

Table 1 : Frequency (out of 25) of the most recently reviewed literature based on environment type.

Environment Type	Frequency
Simulation	13
Hybrid	5
Physical robot	7

Given that simulation is the most widely adopted and practical experimental environment, this study focuses on simulation-based experiments to assess ROS2 vulnerabilities using reconnaissance i.e. active or passive information gathering. By focusing on simulation, this study can construct multiple scenarios yet independent, systematically test ROS2-based systems in a controlled, repeatable, realistic, and safe manner.

2.2 Analysis of DDS suppliers

This section examines the most widely used DDS middleware in ROS2 deployments as elaborated in the latest most relevant literature. Figure 1 summarizes the middleware distribution across the reviewed literature. The analysis focuses on identifying the DDS middleware used in each study, as middleware selection directly influences communication behavior and security characteristics in ROS-based systems. Out of the 25 studies, 9 explicitly stated the use of the default DDS middleware provided by ROS2. These studies primarily focused on system performance and communication behavior without modifying the default middleware configuration. Since the default middleware is widely adopted and explicitly mentioned in previous studies, minimal additional setup is required, making it a practical choice for this study. Consequently, we proceeded with the default Fast DDS implementation. In 13 of the 25 studies, the DDS middleware was either not treated as a primary variable or was not explicitly mentioned, often marked as not applicable (N/A). This highlights a gap in literature, suggesting that middleware selection is frequently overlooked and security is not the primary focus, as most studies concentrate on application-level aspects of the robot. Only one study specifically analyzed Cyclone DDS, indicating limited attention to alternative DDS implementations, while two studies were based on legacy ROS1, which does not natively support DDS.

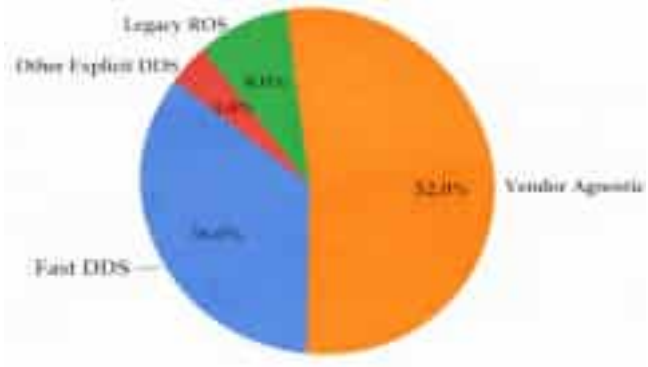


Figure 1: Frequency (in % out of 25) of the most recently reviewed literature based on Middleware distribution.

Fast DDS not only requires minimal configuration but also provides a standardized and widely supported environment ensuring consistency with existing studies. According to [11], it facilitates a focus on security analysis and scenario development without altering underlying communication mechanisms. Overall, there is a clear research gap in the explicit analysis and comparison of DDS middleware implementations in ROS2, particularly regarding security and network communication properties. Therefore, this study leverages the use of Fast DDS, a widely adopted and well documented middleware, while focusing on addressing security concerns that have been largely overlooked.

2.3 Relevant Studies related to Network Analysis

In several studies [12], [13], and [14], network analysis tools such as Wireshark were employed to capture and analyse data packets, including Address Resolution Protocol (ARP), Dynamic Host Configuration Protocol (DHCP), Domain Name System (DNS), Multicast DNS (mDNS), as well as traffic related to the Real-Time Publish–Subscribe (RTPS) protocol. These works demonstrated the effectiveness of network-level monitoring in exposing communication behaviors within ROS-based systems and highlighted

interoperability considerations at the middleware layer. However, their analysis largely remained at the level of traffic observation and protocol identification, without extending to an evaluation of how such observable communications could be exploited or how they might affect robots during active operation. For instance, the potential impact of network-level vulnerabilities on robots performing active tasks such as sensing, navigation, or service delivery was not investigated. In particular, the implications of reconnaissance and passive observation attacks on robots operating over shared or public networks remain unexplored.

Furthermore, a common pattern emerges i.e. regardless of the robot's specific application, whether collecting environmental data in a corridor or delivering goods within a neighborhood the underlying system architecture remains predominantly consistent. This architectural similarity implies comparable attack surfaces, especially at the network and middleware layers. Nevertheless, prior research has only partially examined these shared vulnerabilities, stopping short of evaluating their significance in operational service robot deployments. Accordingly, a representative network-based reconnaissance scenario involving a service robot system is included in the methodology section to contextualize the discussion. Notably, no prior study has systematically quantified the specific types of sensitive operational data that can be passively extracted during an active reconnaissance phase targeting a ROS2-based service robot.

3 Methodology

The main goal of this study is to evaluate network level vulnerabilities in ROS2 based service robots by combining middleware level monitoring with operational use-case validation. Figure 2 illustrates the overall methodology, designed to address the critical gaps identified in sections 2.1, 2.2 and 2.3. The methodology is structured into four main steps as described below. Step 3.1 describes the experimental environment and the tools required to conduct the study. Step 3.2 focuses on standard reconnaissance scenario, detailing the tools used and the validation process to determine the type of observational data that can be collected. Building on this, Step 3.3 applies the approach to a specific use case i.e. robot position monitoring in ROS2 using Fast DDS Spy to assess how the observations translate into practical insights. Finally, Step 3.4 synthesizes and analyzes the additional operational data collected across the experiments, highlighting the implications for system security and exposing potential attack surfaces. Together, these steps form a cohesive framework that combines controlled experimentation, scenario-based evaluation, and operational analysis, enabling a comprehensive assessment of network-level behaviors in ROS2 service robots.

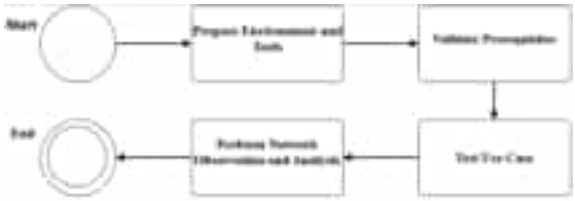


Figure 2: Methodology Diagram

3.1 Prepare Environment and Tools

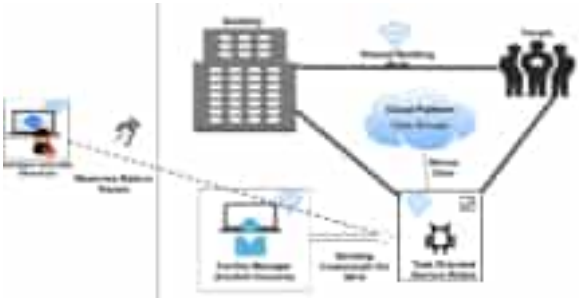
The experimental setup used Ubuntu 24.04 LTS as the host operating system, ROS2 humble as the robotic framework. To analyze communication at the middleware and network layers, multiple diagnostic tools were employed. Fast DDS spy was used on the Ubuntu system to monitor and inspect Data Distribution Service (DDS) communication between ROS2 nodes, enabling visibility into topic discovery and messages exchange at the middleware level. To capture those packets, a software tool called wireshark was executed on a separate windows-based host to capture and analyze network traffic across different machines, supporting cross-host packet-level inspection. For simulation-based evaluation and validation, the Webots 3D robot simulator was used to provide a physics-based virtual environment for modeling, programming, and testing robotic systems. In addition, Turtlebot3 platform was utilized to asses navigation, sensing, and control behaviors under realistic operational conditions. Together, these tools provided a practical setup for simulation and analysis within ROS2 systems.

3.2 Validate Prerequisites

Figure 3 depicts a task-oriented service robot operating within an open wireless network (e.g., Wi-Fi), where it communicates bidirectionally with human operators and smart building infrastructure to support routine operational functions. Such communication may include status updates, sensor data exchange, and control or coordination messages required for normal robot operation. The figure also introduces a potential intruder (left) who gains physical access to the building premises and comes into proximity of the service robot. Without interacting directly with the robot or its control systems, the intruder passively monitors the network by running standard diagnostic tools, such as Wireshark, on the same wireless network. Through this passive observation, the intruder can capture and inspect network traffic associated with the robot's ongoing operations, thereby gaining visibility into various forms of operational and communication data. For the purposes of this study, the attacker is assumed to be an external, unprivileged individual with no prior knowledge of the target system's

architecture or credentials. The attacker is assumed to have physical proximity to the shared wireless network and is equipped only with standard, publicly available diagnostic tools. No active interaction with the robot or its control systems is assumed, confining the threat model to passive network observation consistent with a reconnaissance-phase adversary.

Figure 3: Overview of the network-based reconnaissance



targeting of a task oriented service robot operating in built environment

The scenario depicted in Figure 3 is validated using a controlled experimental environment with a limited set of tools. The validation setup consisted of an Ubuntu 24.04 LTS system running ROS2 with the default DDS middleware, Fast DDS (Fast RTPS), complemented by a Windows host system running Wireshark for capturing and analyzing network traffic. This setup enables observation of operational communications in a controlled manner while reflecting the reconnaissance scenario introduced earlier. Three terminal sessions were opened within the Ubuntu environment. The first session was used to source the ROS2 environment, while the second and third sessions were used to execute ROS2 publisher and subscriber nodes, respectively. ROS2 communication was generated. The execution of these nodes resulted in continuous RTPS communication over user datagram protocol (UDP) using the Fast DDS middleware. Simultaneously, Wireshark was used to capture network traffic on the active network interface, where RTPS packets corresponding to ROS2 communication were successfully observed and identified. This validation confirms that ROS2 DDS based communication can be passively captured and analyzed at the network level using standard packet inspection tools. That is, RTPS traffic generated by ROS2 nodes was captured using Wireshark on the host system by monitoring the shared network interface of the virtualized Ubuntu environment.

Figure 4 illustrates a Wireshark capture of RTPS traffic generated by a ROS2 system using the Fast DDS middleware. The capture shows multiple RTPS packets exchanged over UDP between a ROS2 participant and a multicast destination address. The packets are identified as RTPS data messages, including INFO_TS and DATA sub messages, which are characteristic of DDS based communication. For example, this RTPS packet exchange corresponds to a service robot operating in built environment publishing laser scan data on /scan (similarly odometry and velocity commands).



Figure 4: Captured RTPS traffic from a ROS2 system observed using Wireshark

The source IP address (192.168.1.38) corresponds to a ROS2 node within the system, while the destination address (239.255.0.1) indicates multicast communication used for DDS participant discovery and data dissemination. The transport layer is UDP, with the destination port set to 7400, which is the default port range used by DDS/RTPS. The packet lengths range between approximately 500 to 600 bytes, reflecting periodic data exchange between ROS2 nodes. For instance, DDS multicast communication compromises the robot's real-time position and sensor state (e.g., door access). The consistent timing and repetition of packets demonstrate continuous publisher–subscriber communication, confirming active ROS2 node interaction over the network.

From the captured RTPS traffic, several key reconnaissance insights were obtained through active network monitoring. The analysis revealed critical communication details, including the source and destination IP addresses, as well as the corresponding source and destination ports used by ROS2 nodes. These observations expose the internal network structure and confirm the use of multicast-based communication for data distribution [15]. Furthermore, the protocol and packet characteristics allows the identification of the DDS middleware in use, which in this setup was eProsima Fast DDS. In addition, application layer information embedded withing the RTPS traffic discloses the username of the source system account from which the ros2 nodes were executed. Interestingly, these findings also demonstrate that, even without active interaction, an external observer can extract sensitive system and operational details during the reconnaissance phase, highlighting the exposure present when DDS communication is left unencrypted. Building upon the observed and validated ability to passively capture RTPS traffic, the subsequent phase of this study examines what

additional operational information can be obtained during reconnaissance usingaxz standard tools and configurations.

3.3 Test Use Case

Robot position monitoring in ROS2 via Fast DDS Spy is presented as a foundational and representative monitoring scenario for evaluating data exposure in DDS-based robotic systems. Access to a robot's positional data is critical because position information underpins core robotic functions such as navigation, task execution, collision avoidance, and interaction with the surrounding environment.

The base platform used was ROS2 Humble, with a custom python publisher node (publisher.py) that generates random robot positions at one second interval. The published topic was rt/robot_position with advanced features. The tools used in this experiment include: 1) ROS2 humble, serving as the foundational operating system 2) Fast DDS Spy with version 1.4.0 , a command line introspection tool developed by eProsima for discovering DDS entities such as participants, topics, writers, and readers, 3) Wireshark, a network traffic analyser utilized to capture RTPS packets, 4) Webots Simulator, and 5) TurtleBot3, both serving as robot simulators for generating motion data. The procedure for conducting this experiment is as follows: 1) Publisher Setup: A Python based ROS2 node was developed within the robot_position_publisher package to publish three dimensional coordinates (x, y, z) at a fixed frequency of 1 Hz on the rt/robot_position topic. The node generated randomized positional data to simulate real time robot movement, enabling subsequent monitoring and analysis. 2) ROS2 Workspace Setup: The ROS2 workspace was sourced and built using colcon build, ensuring that the robot_position_publisher package and its associated nodes were correctly recognized and accessible within the ROS2 environment. This setup facilitated seamless execution and integration of the publisher node into the DDS based communication network.

Continuous access to this data provides deep insight into a robot's movement patterns or operational behaviour. By observing positional data over time, an external entity can track where a robot is moving, infer routines, predict future locations, or identify areas of operational importance. This capability is particularly relevant for autonomous service robots operating across multiple locations, including smart buildings, hospitals, or industrial facilities, where positional awareness directly correlates with sensitive operational workflows. For example, during building audits or patrol operations, knowledge of robot trajectories may inadvertently reveal restricted zones, high-value assets, or confidential activities [5][16]. If such information is exposed without appropriate protection, it can enable targeted security, privacy, or safety-related attacks. In this study, robot position data is deliberately selected as a baseline DDS data stream to demonstrate non-intrusive monitoring within a ROS2 environment. The experiment focuses on capturing and analysing position updates published on DDS topics to verify real-time data flow, timing behaviour, or observability without interfering with system operation. While the implementation monitors three-dimensional coordinates (x, y, z) as a proof of concept, the same monitoring methodology is directly applicable to other operational DDS topics, including sensor readings, task states, navigation goals, or interactions with smart building infrastructure. As such, this scenario establishes a reusable framework for analyzing DDS-level data exposure beyond positional information alone.

3.4 Perform Network Observation and Analysis

The following observations were recorded, and the results are illustrated in the Figure 5 below and are explained in detail as follows.



Figure 5: Participants recorded by Fast DDS Spy

1) Participants: In the context of DDS and ROS2, a participant refers to a robot or software application actively engaged in network communication. Each participant is associated with a Globally Unique Identifier (GUID), which serves as a digital identifier to distinguish individual nodes or applications. For instance, a participant may have an ID such as 01.0f.ae.c2.03.75.b4.90.00.00.00.00, uniquely identifying that specific robot or program within the DDS domain.

2) Topics: Each line represents a topic discovered on the DDS network. rt/robot_position is the topic to which the Python publisher was publishing simulated robot positions, with type Point (from geometry_msgs). (1|0) indicates one DataWriter publishing while no DataReaders are subscribed. [0.780132 Hz] indicates the publishing frequency. Topics such as rt/parameter_events and rt/rosout are standard ROS2 topics for parameter updates and log messages.As shown in Figure 6, Fast DDS Spy monitored the DDS network as a passive introspection tool, discovering active

participants, topics, data writers, and data readers without impacting running nodes. This confirmed that the rt/robot_position topic was broadcasting successfully at 0.78 Hz.Analysis of the topics and participants highlights potential cybersecurity implications. For example, an inspection robot continuously collecting position and environmental data in a corridor could be passively observed by an external party, revealing the robot's location, movement patterns, planned paths, and sensor feeds without interfering with its normal operation. This study focuses on read-only access to evaluate how much operational detail can be inferred; data modification and write access are reserved for future research. This setup demonstrates the capability of Fast DDS Spy to audit ROS2 systems at the DDS middleware level, emphasizing the importance of securing communication infrastructure

3.4.1 Key Inferences

While the observability of unencrypted DDS traffic is broadly acknowledged in ROS2 documentation, this study advances beyond this general observation by systematically identifying and validating eight specific categories of sensitive operational data extractable through passive reconnaissance, demonstrating the practical extent of exposure in a realistic service robot deployment scenario. Using basic reconnaissance tools, it was possible to successfully observe and capture 8 forms of sensitive operational data from the ROS2 system. This included source and destination IP addresses, source and destination ports, user identity associated with the source system, the underlying DDS middleware in use, as well as DDS participants, ROS2 nodes, and active publisher nodes. These findings demonstrate that even limited, non-intrusive tools can reveal critical system and communication details without requiring authentication or direct interaction with the target robot. Such exposure highlights a significant attack surface, where reconnaissance alone can serve as an effective entry point for subsequent attack phases, including command-and-control (C2) based attacks.



4 Conclusions, Future work and Limitations

This study demonstrates that effective reconnaissance of a ROS2-based robotic system is feasible using a minimal set of passive, non-intrusive tools when the attacker is positioned within the same network domain. By deploying Fast DDS Spy and Wireshark on a shared network, it was possible to discover DDS participants, enumerate active topics, and observe real-time robot telemetry without modifying the target system or requiring authentication. Through this process, eight distinct types of operational data were identified, including positional information transmitted over the rt/robot_position topic. The conclusions are strictly confined to the reconnaissance phase; references to subsequent attack phases such as C2, spoofing, and man-in-the-middle attacks are intended solely to contextualize reconnaissance as a precursor and do not form part of the claims of this study. Future work will extend beyond reconnaissance to evaluate C2 feasibility, focusing on correlating velocity and positional data to reconstruct full motion trajectories in real time, enabling a deeper assessment of integrity and availability risks [18][19]. Further research will explore active attack techniques including man-in-the-middle attacks, message replay, and topic spoofing, as well as mitigation strategies such as DDS-Security, network segmentation, and secure ROS2 configurations. Practitioners are advised to enable DDS-Security plugins, restrict DDS domain IDs, and isolate robotic systems from shared networks as immediate best practices. While Fast DDS Spy proved effective for passive discovery and monitoring, the experiment has several limitations. First, the tool operates only with DDS implementations compatible with eProsima Fast DDS and relies on default ROS2 middleware configurations. The experiments were conducted exclusively using Fast DDS, which was selected based on its status as the default ROS2 middleware and its prevalence in the reviewed literature. Whether alternative DDS implementations such as Cyclone DDS exhibit different reconnaissance behaviors under similar conditions remains an open question and is identified as a direction for future investigation. Furthermore, the experimental scope was intentionally limited to default ROS2 configurations, reflecting the most commonly encountered real-world deployment conditions. A comparative evaluation between default and security-enabled DDS configurations is recognized as an important direction and is explicitly identified as part of the planned future work.Systems using alternative DDS vendors, customized discovery settings, or secured DDS (DDS Security) profiles may not be fully observable using this approach[20]. Second, Fast DDS Spy is primarily an introspection tool and does not natively support message manipulation or injection, limiting its use to reconnaissance rather than active exploitation. The experiment also focused on positional data only and did not evaluate the impact of QoS policies, encryption, authentication, or access control mechanisms

that may be present in hardened deployments. Finally, the experimental setup was conducted in a controlled environment, and real-world conditions such as packet loss, network segmentation, or intrusion detection systems were not considered. Performance-related metrics such as data rates and traffic patterns fall outside the scope of this reconnaissance-focused study and are reserved for future work.

5 References

[1] Mayoral-Vilches, V., Makris, A., & Finisterre, K. (2025). Cybersecurity AI: Humanoid Robots as Attack Vectors, https://doi.org/10.48550/arXiv.2509.14139.

[2] Mantha, B. R. K., Menassa, C. C., & Kamat, V. R. (2016). Ambient data collection in indoor building environments using mobile robots. *33rd ISARC* (pp. 428–436). https://doi.org/10.22260/isarc2016/0052

[3] Mantha, B. R., Garcia de Soto., B., Menassa C. C., and Kamat V. R., (2020) "Robots in Indoor and Outdoor Environments", *Taylor and Francis*, DOI: https://doi.org/10.1201/9780429398100.

[4] Kim, J., Smereka, J. M., Cheung, C., Nepal, S., & Grobler, M. (2018). Security and Performance Considerations in ROS 2: A Balancing Act https://doi.org/10.48550/arXiv.1809.09566.

[5] DiLuoffo, V., Michalson, W. R., & Sunar, B. (2019). Credential Masquerading and OpenSSL Spy: Exploring ROS 2 using DDS security. https://doi.org/10.48550/arXiv.1904.09179.

[6] Aartsen, M., Banga, K., Talko, K., Touw, D., Wisman, B., Meinsma, D., & Bjorkqvist, M. (2022). Analyzing Interoperability and Security Overhead of ROS2 DDS Middleware. 30th MED, 976–981. https://doi.org/10.1109/MED54222.2022.9837282.

[7] Rivera, S., & State, R. (2021). Securing Robots: An Integrated Approach for Security Challenges and Monitoring for the Robotic Operating System (ROS).

[8] Bistrom, D., Westerlund, M., Duncan, B., & Jaatun, M. G. (2022). Privacy and security challenges for autonomous agents: A study of two social humanoid service robots. IEEE CloudCom, 230–37, https://doi.org/10.1109/CloudCom55334.2022.00040.

[9] Botta, A., Rotbei, S., Zinno, S., & Ventre, G. (2023). Cyber security of robots: A comprehensive survey. Intelligent Systems with Applications, 18, 200237. https://doi.org/10.1016/j.iswa.2023.200237

[10] Deng, G., Xu, G., Zhou, Y., Zhang, T., & Liu, Y. (2022). On the (In)Security of Secure ROS2. ACM SIGSAC, 739–753. https://doi.org/10.1145/3548606.3560681.

[11] Patel, Y., & Rughani, P. H. (2023). qcROS2: An Architecture to Secure Network Communication of ROS2-based Environment. 8th ICRAE, 253–258. https://doi.org/10.1109/ICRAE59816.2023.10458571.

[12] Patil, U., Gunasekaran, A., Bobba, R., & Abbas, H. (2024). ROS2-Based Simulation Framework for Cyberphysical Security Analysis of UAVs. https://doi.org/10.48550/arXiv.2410.03971

[13] Goerke, N., Timmermann, D., & Baumgart, I. (2021). Who Controls Your Robot? An Evaluation of ROS Security Mechanisms. 7th ICARA, https://doi.org/10.1109/ICARA51699.2021.9376468.

[14] Kim, J., Smereka, J. M., Cheung, C., Nepal, S., & Grobler, M. (2018). Security and Performance Considerations in ROS 2: A Balancing Act. https://doi.org/10.48550/arXiv.1809.09566.

[15] Fernandez, I. A., Neupane, S., Chakraborty, T., Mitra, S., Mittal, S., Pillai, N., Chen, J., & Rahimi, S. (2024). A Survey on Privacy Attacks Against Digital Twin Systems in AI-Robotics. IEEE CIC, https://doi.org/10.1109/CIC62241.2024.00019.

[16] Mayoral-Vilches, V., White, R., Caiazza, G., & Arguedas, M. (2022). SROS2: Usable Cyber Security Tools for ROS 2. 2022 IEEE/RSJ IROS, https://doi.org/10.1109/IROS47612.2022.9982129.

[17] Oruma, S. O., Sánchez-Gordón, M., Colomo-Palacios, R., Gkioulos, V., & Hansen, J. K. (2022). A Systematic Review on Social Robots in Public Spaces: Threat Landscape and Attack Surface. Computers, 11(12), 181. https://doi.org/10.3390/computers1112018. 1

[18] Patel, Y., Rughani, P. H., & Desai, D. (2022). Network Forensic Investigation of Collaborative Robots: A Case Study. 2022 7th ICMERR, 51–54. https://doi.org/10.1109/ICMERR56497.2022.10097787

[19] Rivera, S., Lagraa, S., Iannillo, A. K., & State, R. (2019). Auto-Encoding Robot State Against Sensor Spoofing Attacks. 2019 IEEE ISSREW, 252–257. https://doi.org/10.1109/ISSREW.2019.00080.

[20] Rivera, S., Lagraa, S., Nita-Rotaru, C., Becker, S., & State, R. (2019). ROS-Defender: SDN-Based Security Policy Enforcement for Robotic Applications. 2019 IEEE SPW, 114–119. https://doi.org/10.1109/SPW.2019.00030.

Field Evaluation of a Wearable Framework for Worker Motion Recognition and Zone-Level Localization*

Minjun Chang[1], Sai Machiraju[2], and Yong K. Cho[3,†]

[1]School of Civil and Environmental Engineering, Georgia Institute of Technology, United States of America
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, United States of America
[3]Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, United States of America

minjunchang@gatech.edu, saimachi@gatech.edu,yong.cho@ce.gatech.edu

Abstract -

Unobserved worker motions on active construction jobsites contribute to injuries, near-misses, and productivity loss; however, continuous monitoring remains difficult under real deployment constraints. This study presents a real-time worker activity monitoring framework that integrates lightweight Temporal Convolutional Networks (TCN) with wearable sensing and zone-level localization to enable continuous and deployable behavior monitoring. The proposed system employs a three-IMU configuration mounted on the wrist, neck, and back to capture complementary upper-body motion while maintaining compatibility with personal protective equipment. Each sensor streams six-axis inertial data at 50 Hz and performs sliding-window inference using a TCN architecture with dilated causal one-dimensional convolutions and residual connections, supporting low-latency and memory-efficient processing suitable for edge deployment. Task-relevant construction activities are recognized using anatomically grounded motion patterns without relying on vision-based monitoring. To incorporate spatial context, Bluetooth Low Energy beacons provide RSSI-based zone localization that is temporally smoothed and synchronized with activity predictions to generate unified "what-and-where" behavioral labels. The framework is validated through a real factory deployment, demonstrating robust performance under occlusion, signal interference, and natural task variability. Results show that the system achieves reliable real-time inference with average accuracies of 76.28% at the motion level and 78.62% at the activity level, supporting scalable and privacy-preserving activity analytics in real construction environments. The field deployment further identifies key operational challenges, including sensor placement tradeoffs, signal occlusion effects, and user comfort constraints—and provides practical implementation guidelines for wearable monitoring systems in construction environments

Keywords -

Wearable Sensing; Worker Activity Recognition; BLE Localization; Edge AI; Construction Productivity

*This work was supported in part by the National Science Foundation (NSF) under Award No. 2515786 and by the Georgia Research Alliance (GRA). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or GRA.

1 Introduction

Construction jobsites continue to experience injuries, near-misses, and productivity losses caused by unsafe or inefficient worker actions that often remain unobserved. Supervisors are typically responsible for monitoring multiple crews across dynamic work zones, making continuous observation difficult and limiting the ability to provide timely interventions. Although automated monitoring systems have been widely explored, many remain challenging to deploy in real construction environments due to occlusion, privacy concerns, infrastructure requirements, and operational disruptions.

Wearable sensing has emerged as a promising alternative for continuous worker monitoring because it is less sensitive to environmental visibility constraints and better supports privacy-preserving data collection. However, practical field deployment requires balancing recognition performance with usability, latency, and robustness. Many existing wearable monitoring approaches rely on dense sensor configurations or non-causal inference pipelines, which increase worker burden, computational requirements, and system complexity, thereby limiting adoption on active jobsites. To address these challenges, this study adopts a lightweight wearable configuration consisting of three sensing devices and applies a causal Temporal Convolutional Network (TCN) to enable low-latency motion recognition suitable for real-time edge deployment.

While motion recognition provides valuable behavioral information, it offers limited situational awareness without spatial context. Understanding where worker actions occur is critical for safety monitoring, task compliance evaluation, and productivity assessment. To provide this contextual information, the proposed framework integrates motion recognition with zone-level localization, enabling simultaneous interpretation of worker activities and spatial positioning.

This paper presents the first field-evaluated wearable

framework that tightly integrates causal TCN-based motion recognition with BLE zone-level localization into a unified real-time behavioral monitoring pipeline, validated on an active construction site with professional workers performing real installation tasks. The system is evaluated through laboratory testing and a real-world field case study to assess deployment feasibility and operational performance under realistic construction conditions. Rather than focusing on algorithmic development, this study emphasizes deployment-oriented evaluation through field implementation. Specifically, this case study investigates: (1) the minimum wearable sensor configuration that maintains monitoring performance while remaining acceptable to workers, (2) the reliability of activity recognition under uncontrolled field variability, including task transitions and signal noise, (3) deployment barriers that emerge during continuous real-time monitoring, and (4) stakeholder perceptions regarding the usefulness and usability of integrated activity and localization monitoring systems.

2 Literature Review

2.1 IMU based Human Motion Recognition

Wearable IMU-based motion recognition has been widely studied due to its robustness to visual occlusion and independence from environmental lighting conditions [1, 2]. High recognition accuracy has been reported in controlled laboratory settings, often using multiple IMUs distributed across the body [3, 4]. Kim and Cho [5] demonstrated automatic recognition of highway construction worker motions using wearable inertial sensors and Long Short-Term Memory (LSTM) networks, highlighting the effectiveness of deep sequence models for construction activity recognition.

Despite these advances, dense sensor configurations introduce challenges for field deployment, including increased setup time, user discomfort, interference with personal protective equipment, and maintenance overhead. To reduce sensing burden, prior studies have explored sparse sensor configurations or upper-body placements [3]. While these approaches improve usability, they often exhibit reduced robustness under natural task variability, motion transitions, and inter-subject differences. Most wearable sensing research remains laboratory validated, with limited reporting of long-duration field deployment outcomes.

In addition, many deep learning-based motion recognition methods rely on long temporal windows, non-causal inference, or offline post-processing [6]. These assumptions limit applicability for real-time field deployment, where bounded latency and causal operation are required. These gaps motivate the development of motion recognition systems designed explicitly for continuous on-site operation using minimal wearable instrumentation.

2.2 BLE-based Localization

Bluetooth Low Energy (BLE) beacon systems have been widely adopted for indoor localization due to their low infrastructure cost and deployment flexibility [7]. In construction and industrial environments, BLE-based approaches offer advantages where fixed camera systems or dense infrastructure are impractical. However, RSSI-based distance estimation is sensitive to multipath effects, body shadowing, and environmental interference [8], which are common on active jobsites.

As a result, many studies favor zone-level localization rather than high-precision tracking, using calibration and temporal filtering to improve robustness [7]. Zone-level localization aligns well with safety and productivity applications, where presence within a designated work area is often sufficient. Despite this, BLE localization is frequently evaluated independently of activity recognition systems, limiting its ability to provide actionable behavioral context. Integrating localization with motion recognition remains an open challenge for field-ready worker monitoring.

2.3 Construction Productivity

Construction productivity and safety research increasingly emphasizes the need for continuous and scalable monitoring systems [9]. Vision-based approaches provide rich contextual data but face persistent deployment barriers related to occlusion, privacy concerns, and maintenance in dynamic environments [10]. Wearable-based systems address these challenges but often focus on either motion recognition or localization rather than their integration.

Recent work highlights that actionable field monitoring requires joint understanding of worker actions and spatial context [11]. Knowing what motion is being performed and where it occurs enables zone-specific risk assessment, task compliance monitoring, and productivity analysis. However, many existing systems remain confined to laboratory validation or depend on complex sensing and computation that are difficult to sustain in the field. There remains a gap for field-ready frameworks that integrate lightweight motion recognition, practical localization, and real-time visualization while operating under realistic deployment constraints.

3 Methodology

The system is designed as two components: (1) IMU-based Motion Recognition, (2) BLE-based Localization. Each Personal Protection Unit (PPU) embeds one IMU and a Bluetooth signal receiver within a single microcontroller-based device, and workers are equipped with three PPUs to enable distributed motion and localization sensing.



Figure 1. TCN Network Inference Pipeline

3.1 Motion Recognition

As illustrated in Figure 1, the motion recognition module performs real-time inference using a three-device wearable configuration. Three PPUs are mounted on the wrist, rear neck, and lower back to capture complementary distal and proximal body movements while remaining compatible with personal protective equipment. Each PPU integrates a 9-axis IMU consisting of an accelerometer, gyroscope, and magnetometer. Prior to deployment, the accelerometer and magnetometer are calibrated to correct hard-iron and soft-iron distortions by rotating the IMU along a spherical trajectory and fitting the sampled data to a sphere, yielding offset parameters and a correction matrix. The gyroscope is calibrated by averaging 500 stationary samples and subtracting the resulting bias from subsequent measurements. Although each IMU provides nine-axis data, only six-axis inertial measurements from the accelerometer and gyroscope are streamed at 50 Hz for inference, while magnetometer channels are excluded due to magnetic interference commonly encountered near steel structures and powered equipment on construction sites.

IMU streams are segmented into short sliding windows to bound latency and memory usage. Each window is processed independently using channel-wise z-normalization to reduce inter-subject variation and sensitivity to mounting differences. This design avoids reliance on long-term state, which can drift during extended field deployment, and preserves responsiveness during motion transitions.

A Temporal Convolutional Network (TCN) is used as the core recognition model. The network consists of three dilated residual blocks with 64, 64, and 128 channels, a kernel size of 5, and dilation rates of 1, 2, and 4. Input windows span 100 samples (approximately 2s at 50 Hz). All convolutions are causal, and predictions depend only on past samples within each window. No future frames, bidirectional context, or offline smoothing are used. Causality is enforced through right padding and truncation, ensuring consistent alignment between input windows and output predictions. This structure enables parallel computation and predictable inference time suitable for edge deployment.

The TCN outputs per-window motion probabilities, and a lightweight temporal smoothing step is applied over recent outputs to stabilize predictions during short transitions. This smoothing operates only on past predictions and introduces no look-ahead latency. The resulting pipeline achieves reliable real-time motion recognition while balancing accuracy, responsiveness, and deployability for field use.

3.2 Training Configuration

Model training and evaluation follow the protocol established in prior work to balance recognition performance and deployment practicality. IMU data are segmented into short sliding windows and processed using identical preprocessing across all experiments. Two training strategies are evaluated: a fixed single-split configuration with an 80:20 train/test split, and a leave-one-subject-out (LOSO) configuration for cross-subject generalization analysis. All models use the same Temporal Convolutional Network architecture, windowing parameters, and optimization settings so that observed differences reflect only sensor combinations and training strategies.

Seven sensor combinations are evaluated, as summarized in Table 1, including single-IMU, dual-IMU, and tri-IMU configurations. In-lab results indicate that multi-sensor configurations consistently outperform single-sensor setups. The wrist provides strong discriminative cues for upper-limb motion, while the neck and back contribute complementary proximal information. Among all tested configurations, the Neck+Back+Wrist (NBW) combination achieves the most stable performance across accuracy and F1 metrics in repeated in-lab inference trials. Although certain dual-sensor configurations achieve comparable peak accuracy, the NBW configuration demonstrates greater robustness during motion transitions and sustained operation.

Based on these results, the single-split (8:2) model trained with the NBW sensor combination is selected for the real-world case study described in Section 4. The single-split model is chosen over the LOSO model because

Table 1. Combinations and Lab Test Results

ID	Selected IMUs	Acc.	ID	Selected IMUs	Acc.
1	Neck+Back+Wrist (NBW)	0.8326	5	Neck (N)	0.7720
2	Neck+Back (NB)	0.7785	6	Back (B)	0.6878
3	Neck+Wrist (NW)	0.8353	7	Wrist (W)	0.7363
4	Back+Wrist (BW)	0.7926			

the case study evaluates continuous system performance under a fixed deployment setup rather than cross-subject generalization. This selection prioritizes temporal stability and real-time inference reliability, which are critical for evaluating system behavior during live monitoring in a factory environment.

3.3 Zone-Level Localization

Zone-level localization is implemented using commercial BC04P Bluetooth Low Energy (BLE) beacons deployed across the worksite. All beacons operate with identical advertising frequency and transmission intervals to maintain consistent signal behavior across zones. Transmit power is adjusted based on beacon placement height to balance coverage and signal stability. Beacons placed near floor level are configured at low transmit power (-16 dBm) to limit signal overlap and reduce multipath effects. Beacons mounted on ceilings higher than 10 ft are configured at higher transmit power, up to +4 dBm, to ensure adequate coverage across open work areas. The selected transmission settings also ensure long operational stability, with beacon battery life exceeding six months under the configured test conditions and extending beyond one year when operating at minimum transmit power.

Each wearable PPU continuously scans for BLE advertisements and records received signal strength indicator (RSSI) values from nearby beacons. Relative distances to beacons are estimated using a calibrated RSSI-to-distance model established during site configuration. To improve distance estimation accuracy, the PPUs apply an environment-specific signal propagation constant in the log-distance path loss model, which characterizes indoor attenuation conditions. This constant is determined experimentally by measuring RSSI values at multiple known distances from each beacon prior to deployment and typically ranges between 2 and 3 for indoor environments. The calibrated distance estimates are combined with known beacon coordinates defined on a preconfigured worksite map, and the PPU computes its instantaneous planar position (x_i, y_i) at each sampling step. Localization is performed at the zone level rather than high-precision tracking to reduce sensitivity to short-term RSSI fluctuations while remaining sufficient for safety and productivity analysis.

To improve robustness against signal noise, body shadowing, and transient interference, raw position estimates are temporally smoothed on the server side using a weighted averaging scheme. Let (x_i, y_i) denote the estimated planar position at time step i within a sliding window of length $N = 10$. Each estimate is assigned a confidence weight w_i derived from the corresponding BLE signal quality, such as normalized RSSI magnitude or beacon visibility count. The displayed position $(\bar{x}, \bar{y})$ is computed as:

$$\bar{x} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \qquad \bar{y} = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i} \tag{1}$$

This weighted averaging reduces the influence of low-confidence measurements while preserving responsiveness suitable for real-time visualization.

The resulting localization output is used to support higher-level site analytics. First, worker positions are compared against predefined work zone boundaries to determine whether a worker remains within assigned task areas or enters unauthorized zones. This enables detection of potential safety violations and unintended access. Second, temporal changes in estimated position are used to quantify travel distance and dwell time. By correlating movement distance with recognized motion labels, the system distinguishes between walking-intensive tasks and stationary work patterns. This joint interpretation of spatial displacement and motion recognition provides additional context for assessing task compliance and productivity without requiring fine-grained positioning accuracy.

4 Case Study

A case study was conducted at a warehouse construction site for window and door installation in Atlanta, Georgia, USA.

4.1 Data Collection for Feasibility Study

The case study was conducted at an indoor showroom construction site. At the time of deployment, major structural elements such as drywall partitions had already been installed, while interior construction activities were ongoing. Compared to large-scale building projects, the site covered a relatively small area with multiple workers operating in close proximity and dense wall structures, creating constrained movement paths and frequent line-of-sight interruptions. On the test date, workers were engaged in glass wall and door installation. Other tasks such as painting and silicone moulding were present, but the door installation team was selected because their work involved more dynamic and varied movements.

A floor map of the site was manually generated using laser-based distance measurements to define work zones

and beacon locations. The environment included reinforced concrete walls and large metallic components that were frequently repositioned during installation, introducing potential signal obstruction and multipath interference. This setup provided a realistic and moderately challenging indoor environment for evaluating zone-level localization performance.



Figure 2. Workers wearing PPUs during work

4.2 Case Study Description

Three workers participated in the study, all of whom were professional contractors employed by Stellar W&D. Two participants served as test subjects for the case study evaluation, while data from the third participant were collected to support further refinement of the training model. All participants were right-handed and performed collaborative tasks related to glass wall and door installation. Each worker was equipped with three Personal Protection Units (PPUs) mounted on the rear neck, lower back, and right wrist, as illustrated in Figure 2. Each PPU integrates an ICM-20948 IMU. The wrist unit measures approximately 80mm in diameter and weighs under 100g including the battery, with a continuous operating life exceeding 4 hours at the configured 50 Hz streaming rate. Seven beacons were selected based on site geometry and empirical calibration to achieve full coverage of the installation work zone while minimizing signal overlap between adjacent zones.

The total study duration was 2 hours and 13 minutes. Seven Bluetooth Low Energy (BLE) beacons were deployed across the test site at floor level with transmit power 16dBm and advertising interval of 100ms, as shown in Figure 3, where blue markers indicate beacon locations. Four fixed cameras were installed to capture the full work process. These recordings were used solely for offline annotation and ground truth labeling of worker motions to support accuracy analysis and evaluation of the recognition model.

5 Result

5.1 Deployment Feasibility

Qualitative feedback was collected through interviews with a workers who wore the PPUs during the case study to assess wearability and potential interference with daily tasks. The participant reported that the PPUs mounted on the back and neck using clip attachments did not interfere with work activities. However, the wrist-mounted PPU was perceived as slightly bulky during tasks involving frequent wrist motion, such as hammering. The participant suggested that reducing the device size to approximately 50mm in diameter could further improve comfort and usability.

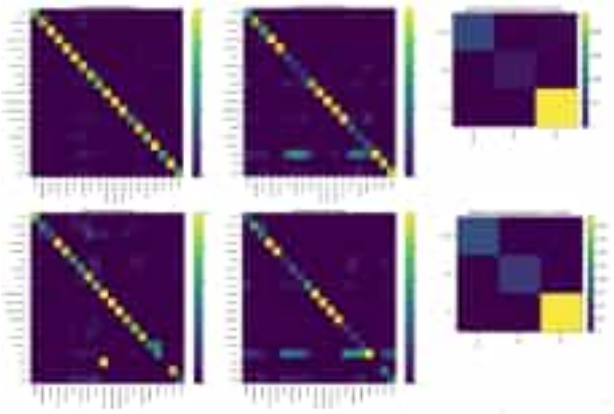


Figure 3. Worker 1 and 2 Confusion Matrices

In addition, the real-time worker localization and motion recognition interface was demonstrated to the site business owner to obtain feedback on practical value. The owner indicated that the combined presentation of worker location and motion provided meaningful insight for managing productivity and manpower allocation. In particular, the ability to observe where workers are located and what they are performing was identified as useful for task assignment and future planning. The owner also expressed interest in incorporating walking distance and task-specific activity distribution as additional indicators for more comprehensive productivity monitoring.

5.2 Feature Accuracy

Motion-level performance analysis was conducted using recognition data obtained from the pretrained Temporal Convolutional Network (TCN) model developed through in-lab data acquisition. The model was trained and validated under scripted laboratory scenarios, where in-lab inference tests achieved an average accuracy of 83.26%. This pretrained model was then deployed without modification during the case study to evaluate its feasibility and reliability under real field conditions.

Figure 3 presents confusion matrices for the case study results, with the top row corresponding to Worker 1 and the bottom row corresponding to Worker 2. The left column shows row-normalized confusion matrices, which reflect recall-oriented behavior by indicating how true motion classes are confused by the model. The middle column shows column-normalized confusion matrices, which emphasize precision by indicating how often predicted classes are correct. The right column presents joint-level accuracy, where shoulder- and elbow-related motions are concatenated to evaluate whether the model correctly identifies the active joint rather than the exact motion subtype. This analysis was performed to assess robustness under uncontrolled site conditions, which differ substantially from structured laboratory settings. The resulting motion-level accuracy shown in Table 2 was 75.86% for Worker 1 and 76.70% for Worker 2, demonstrating stable performance and supporting the feasibility of deploying the model on active construction sites. The lower F1 observed for Worker 1 is primarily attributed to the lab-to-field domain gap and a task composition dominated by overhead hammering sequences, which involve coupled elbow and shoulder motions that are historically the most confused class pair in training evaluation. Worker 2's higher F1 reflects a task distribution more consistent with the training set. Class-weighted training and domain adaptation are identified as direct future remedies.

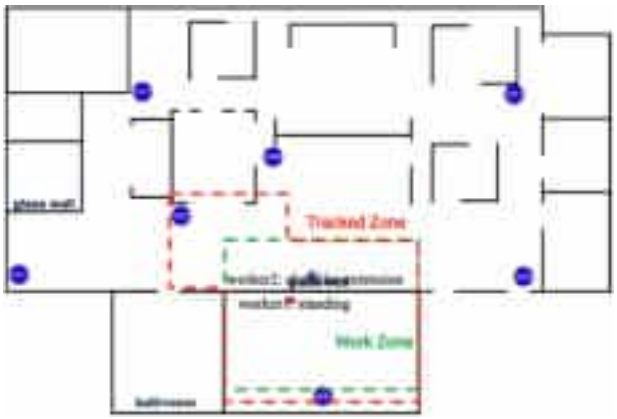


Figure 4. Work Zone and Tracked Zone Coverage

Activity-level accuracy was evaluated to examine the viability of the proposed method for assessing worker productivity in a more realistic setting. Activities performed during the test period were manually annotated, and each activity was defined as a sequence of constituent motion primitives, as illustrated in Figure 4. A recognized motion was considered correct if it corresponded to any motion primitive associated with the annotated activity at the given time frame. The resulting activity-level accuracy was 80.07% for Worker 1 and 77.16% for Worker 2, indicating that individual motion patterns influenced recognition performance while maintaining sufficiently high accuracy for practical productivity assessment.

Table 2. Motion/Activity Recognition Performance

Metric	Worker 1	Worker 2
Motion Level Accuracy	75.86%	76.70%
Motion F1-Score	0.5510	0.6401
Activity Level Accuracy	80.07%	77.16%



Figure 5. Task to Motion Breakdown

Localization performance was evaluated using area coverage as the primary metric. The designated work zone for the installation task covered approximately 774.8ft^2, while the area in which workers were visualized on the web-based interface covered approximately 563.5ft^2, as shown in Figure5 and Table3. Quantified localization error was not computed in this study as the primary objective was zone-level presence detection. The tracked area fully overlapped the defined work zone, resulting in 100% work zone coverage. However, approximately 27.27% of the tracked area corresponded to spatial estimates outside the intended zone. These inaccuracies are primarily attributed to variations in line-of-sight conditions between the wearable PPUs and nearby BLE beacons. For example, when a worker's body orientation obstructed direct signal paths, multiple PPUs occasionally received stronger RSSI measurements from an adjacent beacon, leading to temporary location estimates biased toward that beacon.

Table 3. Area Inspection Summary

Area Inspection	
Work Zone	$\approx$ 774.8 ft^2
Tracked Zone	$\approx$ 563.5 ft^2
Work Zone (WZ) Coverage	100%
WZ Portion	72.73%

6 Discussion

The case study evaluated the feasibility of deploying the proposed system under real field conditions and assessed its performance in real-time motion recognition and zone-level localization. Results indicate that the system can operate continuously in an active work environment using

a minimal wearable configuration and lightweight localization infrastructure. Motion-level predictions remained stable throughout the deployment, and zone-level localization provided consistent spatial context suitable for safety monitoring and preliminary productivity assessment.

Activity-level analysis was conducted through manual annotation of motion sequences rather than automated inference. Individual activities were identified by aggregating consecutive motion-level predictions based on visual inspection and task context. While this approach enabled initial evaluation of task execution under real field conditions, it does not scale and introduces subjectivity; developing automated activity aggregation from motion primitives is therefore a direct priority for future work. Furthermore, productivity assessment requires distinguishing work-related actions from idle or unrelated behaviors, which is not explicitly addressed at the motion level. Developing automated activity aggregation and work-behavior classification methods is therefore essential for scalable productivity analytics.

Several deployment lessons emerged from the field study. The three-PPU configuration successfully captured overall body motion patterns with acceptable accuracy; however, detecting fine-grained arm movements without a dedicated forearm sensor remains challenging. In such cases, upper-limb activity must be inferred indirectly from torso inclination and whole-body kinematics, which may reduce sensitivity for localized hand tasks. Worker compliance also affects the interpretation of productivity metrics. Natural behaviors such as standing, conversing, or checking mobile devices occur intermittently during work hours and must be separated from task-related actions to avoid overestimation of productive time. In addition, the case study was limited to three workers within a single task team, and activity-level analysis was manually annotated. Although feasibility was demonstrated, scalability claims refer to the lightweight system architecture rather than the current validation scope. Future work will extend deployment to multiple task types, larger worker groups, and varied site conditions to substantiate broader applicability. The current evaluation is limited to two workers, which precludes formal statistical comparison across worker profiles or task types; future work will expand the cohort size.

Based on observations from the field deployment, several practical recommendations are identified to support future implementation of the wearable IoT system in construction environments:

- Sensor Placement: Wrist-mounted IMUs provide strong discriminative information for upper-limb motion but may introduce minor comfort concerns during repetitive hand-intensive tasks. Neck and lower-back placements offer stable proximal motion signals with minimal interference to worker movement and personal protective equipment. Multi-location sensing improves robustness under natural task variability and motion transitions.

- Beacon Deployment: Zone-level localization can be achieved using low-power BLE beacons strategically positioned to reduce signal overlap. Floor-level placement improves zone discrimination in partitioned indoor environments but may introduce location bias due to line-of-sight obstruction. Elevated placement increases coverage in open spaces but may require additional installation effort. Beacon spacing and transmit power should be adapted to site geometry and obstruction conditions to maintain localization reliability.

- Calibration and Field Deployment: RSSI calibration and IMU calibration should be conducted during initial system setup to capture site-specific signal propagation characteristics and improve data accuracy. Periodic verification of signal quality is recommended to account for environmental changes, material interference, and body-shadowing effects that may arise during active construction. In addition, standardized worker onboarding and device fitting procedures should be established to ensure consistent sensor placement, maintain data reliability, and reduce variability caused by improper attachment.

7 Conclusion

This paper presented a field-evaluated wearable framework that integrates real-time motion recognition and zone-level localization to support continuous worker monitoring in construction environments. By combining a lightweight Temporal Convolutional Network with a reduced IMU configuration and BLE-based localization, the system demonstrated stable performance under real deployment conditions without relying on cameras or dense sensing infrastructure. Results from a factory case study showed reliable motion-level and activity-level recognition, as well as consistent spatial coverage suitable for safety and productivity analysis. While activity-level inference and productivity assessment were conducted manually in this study, the findings highlight the feasibility of the proposed framework and motivate future work on automated activity aggregation and work-related behavior classification. Overall, the results suggest that the proposed approach provides a practical foundation for scalable, privacy-preserving worker monitoring on active construction sites.

References

[1] Francisco J. Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[2] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiaoli Li, and Shonali Krishnaswamy. Physical activity recognition with mobile phones: Challenges, methods, and applications. *IEEE Signal Processing Magazine*, 32(4):52–59, 2015.

[3] Kinam Kim and Yong K. Cho. Effective inertial sensor quantity and locations on a body for deep learning-based worker motion recognition. *Automation in Construction*, 113:103126, 2020.

[4] Minsoo Park, Seongwoo Son, Yuntae Jeon, Dongyoung Ko, Mingeon Cho, and Seunghee Park. Scaffolding worker imu time-series dataset for deep learning-based construction site behavior recognition. *Advanced Engineering Informatics*, 65:103232, 2025.

[5] Kinam Kim and Yong K. Cho. Automatic recognition of workers' motions in highway construction by using motion sensors and long short-term memory networks. *Journal of Construction Engineering and Management*, 147, 2020.

[6] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[7] Ramsey Faragher and Robert Harle. Location fingerprinting with bluetooth low energy beacons. *IEEE Journal on Selected Areas in Communications*, 33 (11):2418–2428, 2015.

[8] Faheem Zafari, Athanasios Gkelias, and Kin K. Leung. A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, 21(3):2568–2599, 2019.

[9] Jochen Teizer, Brian S. Allread, Craig E. Fullerton, and Jimmie Hinze. Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. *Automation in Construction*, 19(5):630–640, 2010.

[10] Jian Gong and Carlos H. Caldas. Computer vision-based video interpretation model for automated productivity analysis of construction operations. *Journal of Computing in Civil Engineering*, 23(3):151–160, 2009.

[11] Arun Singh Chatterjee, Bivas Mitra, and Sandip Chakraborty. Accoustate: Auto-annotation of imu-generated activity signatures under smart infrastructure. *arXiv preprint arXiv:2112.06651*, 2021.

From satellite images to urban-scale air temperatures: A Multi-Layer Perceptron approach

Geunchan Song[1], Azarakhsh Rafiee[1] and Martín Mosteiro-Romero[1,2]

[1]Faculty of Architecture and the Built Environment, Delft University of Technology, Netherlands
[2]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Netherlands

M.A.MosteiroRomero@tudelft.nl

Abstract -

A detailed understanding of the Urban Heat Island is crucial for energy efficiency, renewable integration, building system automation and demand response in district energy systems. Satellite images provide high resolution information about land surface temperatures (LST) at urban scale, but do not provide information about air temperature (AT), which affect energy performance and thermal comfort. This paper presents a methodology to estimate AT at urban scale for the case of Seoul, South Korea. A Multi-Layer Perceptron (MLP) neural network was developed to convert satellite-derived LST to AT using 12 years of satellite imagery and air temperature measurements from local automatic weather stations (AWS). The method is then independently tested by comparing the predicted temperatures to the measurements from 1000+ distributed "Seoul Data of Things" (S-DoT) sensors in the city. The model was found to achieve an R^2 value of 0.968 during validation using the AWS data from the period 2013–2024. The mean difference during the independent testing was 0.98°C, with an R^2 of 0.807, confirming the model successfully predicts actual air temperatures rather than sensor-specific values.

Keywords -

Urban Heat Island; Air Temperature; Remote Sensing; Multi-Layer Perceptrons

1 Introduction

Rapid urbanization and industrialization have intensified the challenges posed by global warming. As population in cities increases, urban areas have become significant contributors to greenhouse gas emissions while simultaneously experiencing the adverse effects of climate change. Among these effects, the increased air temperature of a city compared to the surrounding rural area, known as the Urban Heat Island (UHI) [1], exacerbates global warming by amplifying energy demands and deteriorating air quality.

A detailed understanding of the UHI in urban areas is crucial for urban planning decisions to support both energy efficiency [2] and renewable integration [3] As urban temperatures increase, space cooling demand increases, with implications on the design and operation of buildings and energy systems. Effective building automation and demand response in district energy systems require accurate load forecasting, where a detailed understanding of the expected air temperatures during operation is crucial. In spite of the considerable effects of the urban environment on local air temperatures, such data is rarely available at sufficient spatial resolution, creating a gap between climate analysis and practical urban energy applications.

1.1 Measuring the UHI

The most direct way to measure the UHI is through weather stations that record air temperature at standard heights within the urban canopy, from ground level to the tops of trees and buildings [4]. However, weather stations have limited spatial coverage, while the effects of the urban climate can be highly localized based on street layout and building materials. Another solution is adding low-cost weather sensors. For example, Schatz and Kucharik [5] installed 150 temperature and humidity sensors at 3.5 meters height on streetlights and utility poles in Madison, USA, to capture fine-scale temperature variations.

One of the most widely used methods for obtaining precise temperature data in UHI studies is using satellite imagery. Satellite imagery can provide a very high spatial coverage of urban land surface temperatures by measuring thermal radiation from surfaces compared to weather stations. Thermal remote sensing has been used for decades to assess the UHI, to perform land cover classifications and in models of urban surface atmosphere exchange [4].

Ground-based weather stations, while providing accurate point measurements, typically offer sparse spatial coverage in urban areas, often one station per several square kilometers. This spatial limitation makes it challenging to capture the fine-scale temperature variations essential for understanding urban heat island effects and their relationships with UFEs at the building or neighborhood scale. Satellite thermal imagery offers complete spatial coverage at resolutions of 30-100m, providing temperature information for every pixel across entire cities. However, satellites

measure surface temperature rather than air temperature, necessitating conversion methods to obtain the air temperature values required for building energy analysis and human comfort assessment. Combining these data sources can offer the spatial completeness of satellite data while maintaining relevance to energy consumption patterns and urban planning applications.

1.2 Converting Land Surface Temperature to Air Temperature

The most fundamental approach to obtain urban-scale air temperature from satellite data involves simple linear regression between LST and air temperature [6]. Multiple linear regression (MLR) incorporates additional predictor variables beyond LST to improve accuracy [7]. Chen et al. [8] developed an enhanced empirical regression method that incorporated multiple predictors and achieved an average error within ±2.5K for 98% of weather stations. Another option is Regression Kriging (RK), which applies regression to establish relationships between temperature and auxiliary variables, and kriging to spatially interpolate based on their spatial structure. RK has been successfully implemented to generate temperature maps at high resolution [9]. However, it assumes linear relationships in trend, while temperatures have complicated non-linear relationships, making it infeasible to analyze detailed relationships [10].

Among machine learning methods, Random Forests (RF) have emerged as a powerful method for air temperature estimation [11]. The RF algorithm has a great ability to handle non-linear relationships and multiple variables, making it particularly suitable for complex urban environments. However, RF can create artificial spatial patterns due to its tree-based structure, which may be problematic for continuous temperature mapping. Also, RF significantly underestimates temperature extremes and struggles to recover finer scale spatial resolution.

Neural networks have also been successfully implemented. Choi et al. [12] used Artificial Neural Networks (ANN) and achieved high accuracy, though they noted that model performance is highly dependent on input variables and network architecture. Multi-Layer Perceptron (MLP) neural networks have shown superior performance for air temperature estimation, particularly in challenging conditions [13]. MLP can effectively combine LST with auxiliary variables like NDVI, elevation and solar angles to create comprehensive air temperature maps [7].

While more advanced algorithms and neural network architectures could be explored, MLP was selected in this work due to its robustness, lower data requirements and demonstrated performance in air temperature predictions. MLP produces smooth, continuous temperature surfaces by learning the underlying physical relationships between variables. MLP can capture the non-linear relationships between LST and air temperature that vary with surface types, time of day and weather conditions. The ability to integrate multiple input variables simultaneously is another key strength. MLPs work well with standard table-formatted data and do not need special data structures to perform effectively [14].

1.3 Scope of this paper

This paper presents an MLP-based methodology for the conversion of LSTs from satellite images to air temperatures at urban scale for energy planning and system operation. While many studies have examined the UHI effect and building energy consumption, several research gaps remain in current approaches. Most research analyzes these phenomena over short periods, missing long-term urban transformations and their cumulative impacts [15]. While Raj and Yun [16] analyzed surface UHI effect in Seoul over 20 years using remote sensing, they specifically focused on LST rather than air temperatures. Furthermore, few studies validate their models using independent, spatially dense sensor networks.

These limitations highlight the need for an integrated, long-term approach that can handle complex non-linear relationships while providing practical insights for urban planning. Long-term analysis capturing actual urban transformations, rather than static snapshots, is essential for understanding how development patterns influence environmental outcomes. In this context, this research aims to contribute to these gaps through: (1) the integration of long-term satellite and ground-based datasets over a 12-year period; (2) the use of independent high-density IoT sensor networks (S-DoT) for external validation; and (3) the demonstration of spatially continuous air temperature mapping suitable for urban energy applications. Thus, this research provides a first step towards a more detailed understanding of the UHI in Seoul and its effects on urban energy systems.

2 Methodology

The methodology employs a two-stage approach, as shown in Figure 1. Stage 1 focuses on training the MLP using 39 AWSs data as ground truth. Air temperature measurements at 11:12 AM from 2013 to 2024 are matched with the Landsat 8 overpass time. The MLP model learns the complex non-linear relationships between 4 input variables: Land Surface Temperature (LST), Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM) and Solar Zenith Angle (SZA), and the measured air temperature from AWSs. Stage 2 tests the trained model using independent data from over 1,100 S-DoT sensors across Seoul since 2021. This independent test-

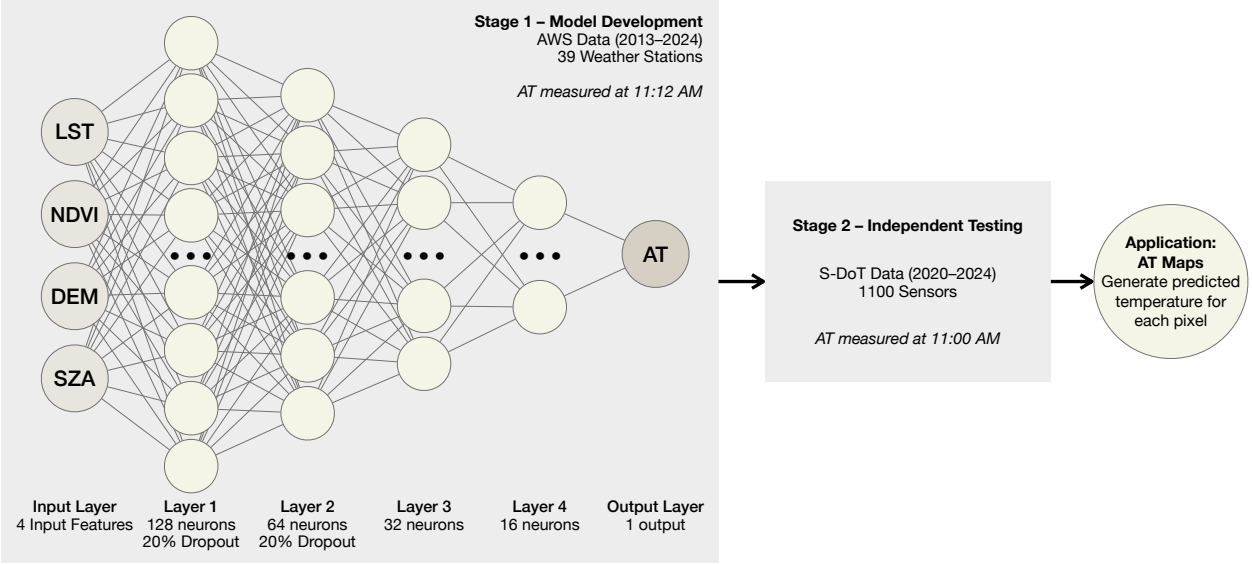


Figure 1. Model development and testing methodology showing the five-layer MLP architecture with dropouts

ing confirms the model's performance on a spatially dense network not used during training.

2.1 Data collection

2.1.1 Satellite data

This research uses Landsat 8 due to its extensive temporal coverage of over 10 years and high spatial resolution of 30 by 30 meters. From Landsat 8, the digital elevation model (DEM) and solar zenith angle (SZA) can be acquired, while LST and normalized difference vegetation index (NDVI) can be calculated using three spectral bands: Band 4 (red), Band 5 (near-infrared, NIR) and Band 10 (Thermal Infrared) [17]. This research uses Google Earth Engine to collect these bands and calculate the LST and NDVI of the research area.

Landsat 8 orbits the Earth every 16 days, which means that the satellite only captures the area 1 or 2 times per month. Seoul is captured at around 11:12 in the morning, which is close to the peak temperature in a day. This temporal limitation means that the model outputs consist of single monthly values derived from these satellite observations at their specific capture times rather than continuous temporal data.

Satellite images from Landsat 8 were collected from March 2013 (the first available month) to December 2024. During this period, 118 images were collected, however not all satellite images from Landsat are useful for temperature calculations. Satellite imagery is very vulnerable to cloud cover, which can obscure most of the land area. Thus, only images where the cloud coverage was less than 30% sky and the research area was fully visible were selected. When no suitable images were available within this threshold for specific time periods, the cloud coverage limit was incrementally increased if acceptable quality images could still be obtained. As a result, we were able to collect 7 valid images per year, which corresponded to 7 valid images per month over the 12-year period of study.

The input features for the MLP model are shown in Fig. 2. Land Surface Temperature (LST) serves as the primary input, providing thermal information measured by Landsat 8 bands at 11:12 AM. The Normalized Difference Vegetation Index (NDVI) represents vegetation greenness, which significantly affects the LST and air temperature relationship through the evapotranspiration process. NDVI values range from -1 to 1, with higher values indicating denser vegetation. The Digital Elevation Model (DEM) represents terrain elevation, affecting air temperature through the environmental lapse rate where temperature decreases approximately 6.5°C per 1000 m elevation gain. Elevation influences wind exposure and cold air pooling, making it an essential predictor for temperature estimation. Finally, the Solar Zenith Angle (SZA) represents the angle between the sun and the vertical direction, which determines the intensity of solar radiation. It influences the amount of solar radiation reaching the surface, affecting both LST and air temperature.

2.1.2 Weather data

The Korean Meteorological Administration (KMA) operates 554 automatic weather stations (AWSs) throughout the country that record air temperatures, wind speed, pressure and other climatic factors at subhourly scale [18]. Out of these, 28 AWSs are installed in Seoul, shown in Fig. 3. Although the spatial resolution of the AWS is low com-

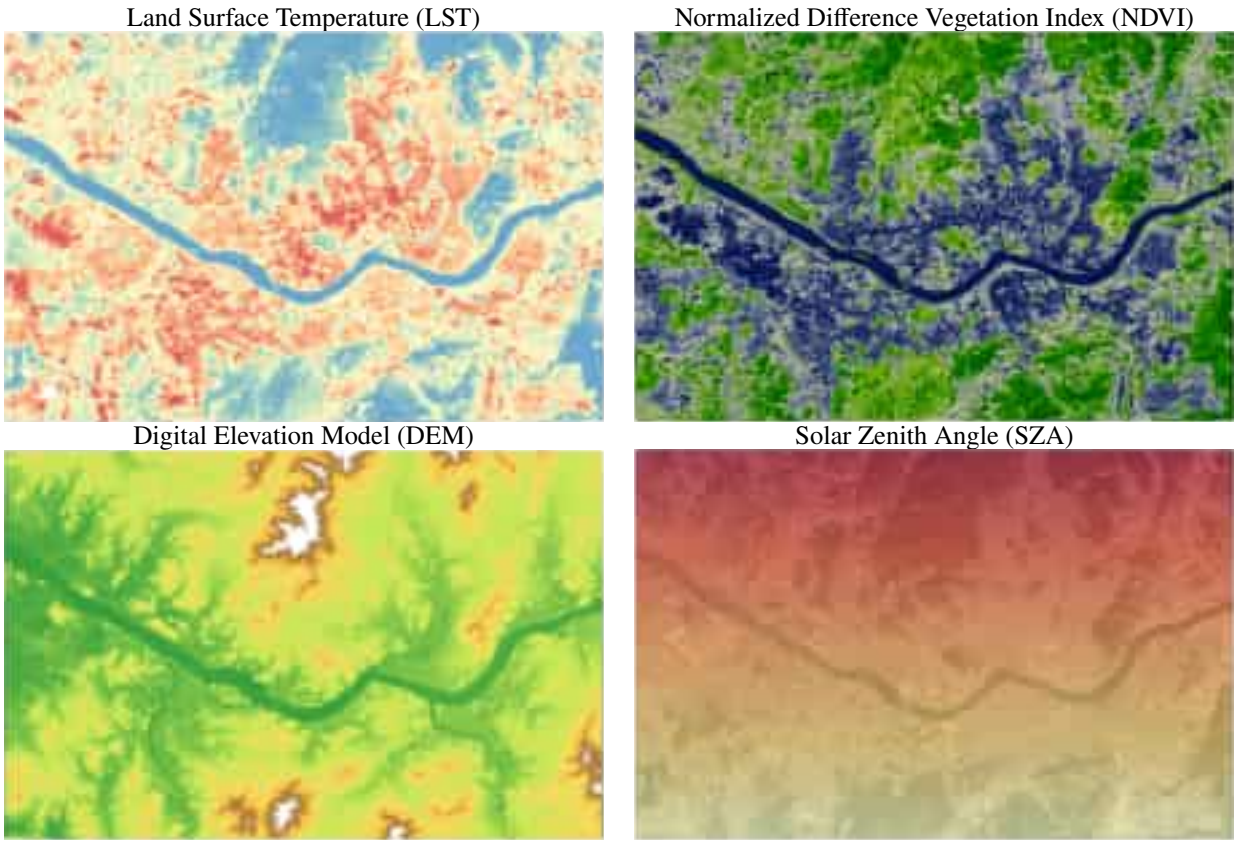


Figure 2. Examples of the four input features in the MLP model

pared to the satellite imagery (28 AWSs in Seoul, which has an area 605.25 km), it has a very high temporal resolution. This research additionally considers 9 AWSs near the Seoul border to increase spatial resolution.

In addition to these weather stations, the Korean government installed weather sensors in 2020 as part of an initiative called Seoul Data of Things (S-DoT). S-DoT is a city data sensor that collects various urban phenomena data, such as fine dust level, temperature, relative humidity, noise, illumination level, ultraviolet radiation, and other parameters at hourly scale [19]. As of 2024, 1159 sensors have been installed, as shown in Fig. 3.

2.2 Stage 1: MLP development

The MLP employs a four hidden layer architecture with an input layer of 4 features and an output layer that progressively reduces the number of neurons from input features to air temperature prediction. The choice of 128 neurons in the first hidden layer follows the common heuristic of having the first hidden layer size as a power of 2 between 2-5 times the number of inputs [20]. This architecture size has proven effective in similar atmospheric and environmental prediction tasks where the relationship between inputs and outputs involves multiple non-linear interactions [21].

Dropout is one of the most popular regularization methods for preventing neural network models from overfitting in the training phase [22]. 20% dropout is applied to the first 2 hidden layers. The Rectified Linear Unit (ReLU) activation function is used throughout the hidden layers. ReLU provides computational efficiency and addresses the vanishing gradient problem that affects traditional activation functions in deep networks [23]. The output layer employs linear activation, appropriate for continuous temperature prediction.

The MLP model follows the common 70-15-15 split for training, validation and testing sets [21]. The model's weights are updated during training using Adam optimizer, which adapts the learning rate for each weight individually and incorporates momentum, which helps the optimization process continue moving in productive directions and avoid getting stuck in local minima [24]. A learning rate of 0.001 is used in weight updates, which is a widely tested starting point that provides stable convergence for neural networks across various applications [25]. The model uses a batch size of 32, as it provides a sufficient gradient averaging to smooth out the randomness inherent in individual samples [26]. The model trains for a maximum of 200

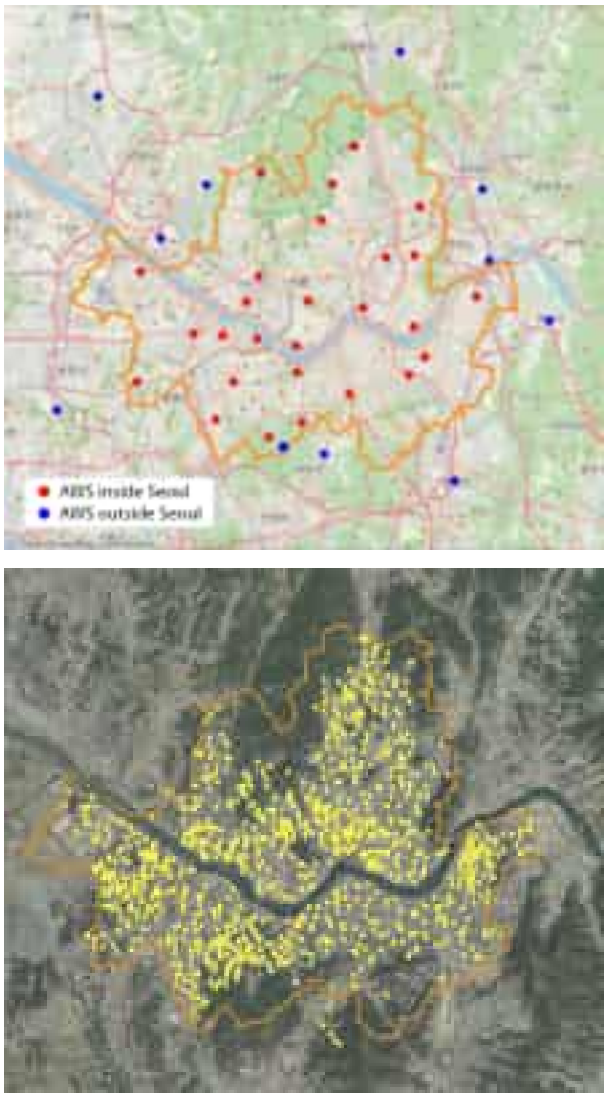


Figure 3. Location of AWS weather stations (top) and S-DoT sensors (bottom)

epochs, with an early stopping patience of 20 epochs to avoid overfitting.

2.3 Stage 2: Independent Testing with S-DoT Sensors

The independent testing begins after the MLP model has completed training using the AWS data. The trained model generates air temperature predictions at each S-DoT sensor location, and these are then directly compared with the actual air temperature measurements from the S-DoT sensors at 11:00 AM. The S-DoT sensors are more densely distributed across Seoul, which provides a spatial testing that the sparse AWS cannot provide. The dense network captures microclimate variations within the urban environment, testing the model's ability to predict air temperatures at locations far from any training stations.

3 Results

The MLP model for converting LST to air temperature was developed, validated and tested using AWS data (2013–2024) as ground truth and further tested using the S-DoT sensor data (2020–2024).

3.1 MLP model performance

Figure 4 shows the model training loss, showing how well the model fits the training data with each epoch, as well as the validation loss, which shows the model's performance on the held-out validation set. Both curves show successful learning behavior, as they decrease rapidly in early epochs as the model learns patterns, then gradually stabilize around the 25th epoch. The convergence of both training and validation curves at similar loss values indicates successful training without overfitting. The early stopping was activated at epoch 71 when validation loss showed no improvement for 20 consecutive epochs.

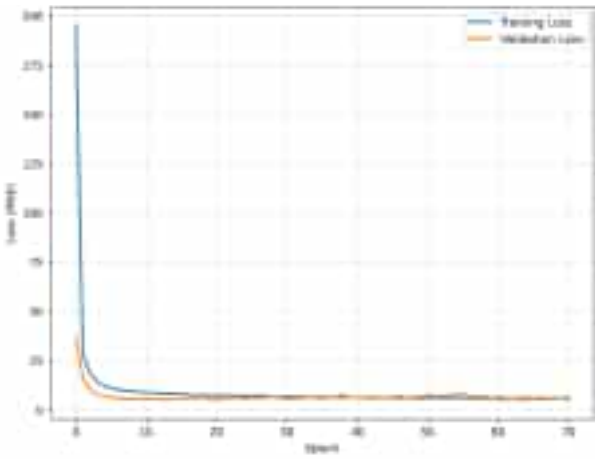


Figure 4. Training and validation loss curves during MLP development

The performance of the model is quantified in Table 1. The coefficient of determination (R^2) value of 0.968 indicates a well-captured nonlinear relationship between LST and air temperature. The performance of this MLP aligns with recent studies in urban temperature estimation, where Salih et al. [13]'s MLP model obtained an R^2 of 0.91. The Root Mean Square Error (RMSE) of 2.1°C is also in a similar range compared to other papers, where Choi et al. [12] got RMSE of 2.19°C using an artificial neural network (ANN). Finally, the Mean Absolute Error (MAE) of 1.62°C provides the average absolute difference between predicted and actual air temperatures.

Table 1. MLP performance metrics for each of the two stages

Data	Period	R^2	RMSE	MAE
AWS	2013–2024	0.968	2.10°C	1.62°C
S-DoT	2020–2024	0.807	4.73°C	3.87°C

3.2 Independent S-DoT testing results

The trained MLP model is further tested with independent S-DoT sensors for the period 2020–2024. Figure 5 shows the predicted versus observed air temperatures at S-DoT sensor locations, with points clustering along the 1:1 line. The mean difference is 0.98°C, with a standard deviation of 4.63°C. The performance of the model during independent testing (Table 1) is lower than for the validation test set, though the R^2 value of 0.807 still indicates a reasonable predictive capability. This is in spite of the fact that S-DoT sensors, unlike AWS, are not installed according to meteorological standards, adding to the uncertainty in their measurements.

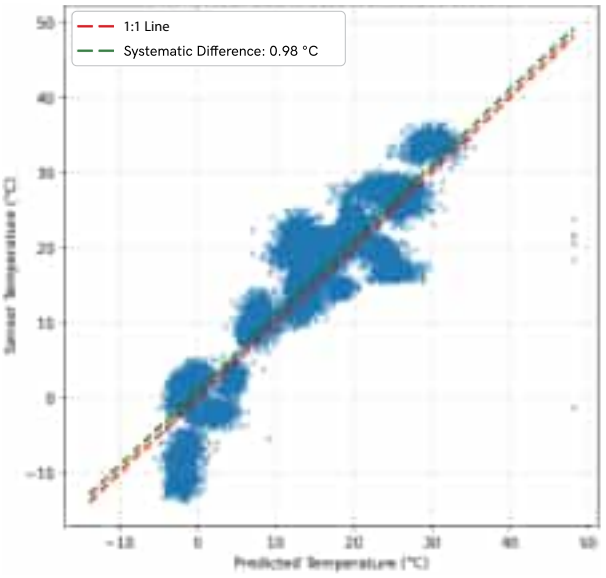


Figure 5. Predicted air temperature testing with S-DoT sensors (2020-2024)

3.3 Model application: Air temperature maps

Figure 6 shows a scatter plot of the measured LST from the satellite images and the predicted air temperature (AT) from the MLP model during the 12-year analysis period. The results show that the difference between LST and AT increases as the measured LST increases. This aligns with Naserikia et al. [27]'s findings, which observed greater temperature differences between LST and air temperature in built environments during warm days.

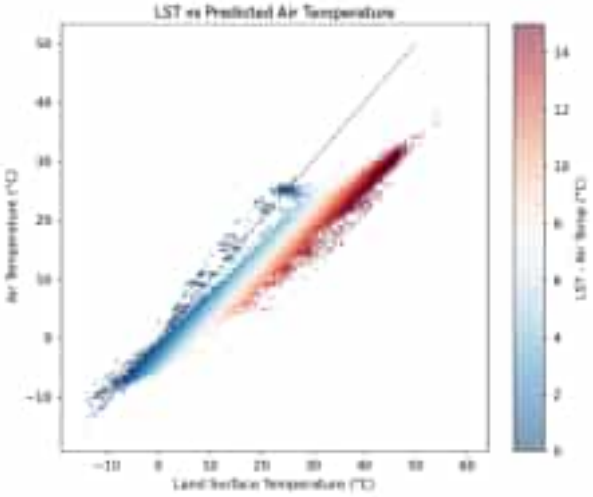


Figure 6. Scatter plot of land surface temperature and final predicted air temperature for the period 2013–2024

As shown in Table 2, the LST tends to be much larger than AT in the summer, with mean differences ranging from 3.94°C in winter to 12.24°C in summer. The larger summer difference reflects intense solar heating of urban surfaces during peak radiation hours, while winter's lower sun angle and reduced radiation result in smaller differences. This consistent seasonal pattern validates the model's physical realism and demonstrates that the MLP successfully captures the varying thermal relationships throughout the year. The finalized model is used to generate air temperature distributions based on LST, as shown in the maps in Fig. 7.

Table 2. Difference between land surface temperature and final predicted air temperature

Metric	Value
Mean Temperature Difference	7.78°C
Spring	9.40°C
Summer	12.24°C
Fall	7.51°C
Winter	3.94°C
Median Temperature Difference	7.58°C
Standard Deviation	3.88°C

4 Conclusions

The research successfully developed a Multi-Layer Perceptron model to convert satellite-based Land Surface Temperature (LST) to air temperature using Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM) and Solar Zenith Angle (SZA) us-

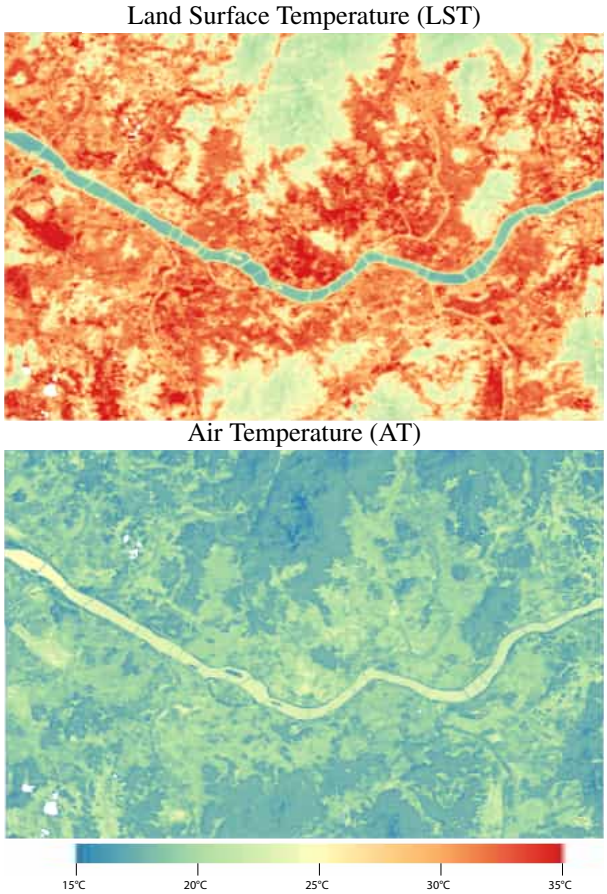


Figure 7. Land surface temperature and predicted air temperature on 25th May 2022 11:10:19

ing air temperature data from Automatic Weather Stations (AWSs) as ground truth data. The MLP model achieved an R^2 of 0.968 and tested with independent S-DoT sensor data. The independent testing was conducted using data from 1000+ S-DoT weather sensors for the period 2020–2024, achieving a mean temperature difference was 0.98°C and an R^2 of 0.807, confirming that satellite imagery provides a reliable method for urban temperature monitoring when ground-based stations are limited.

However, it is important to acknowledge that the analysis uses single hourly temperature observations, since Landsat provides only 1-2 images per month at a fixed time of 11:12 AM. Future research should focus on the development of methods to model diurnal variations in temperature distributions. The use of more frequent satellite observations, if available, could better capture temporal scales to demonstrate the cascading effects from urban form through temperature to energy consumption. Furthermore, more advanced machine learning methods or hybrid approaches combining satellite images with physical models to capture diurnal patters should be explored.

Despite these limitations, the decade-long analysis provides valuable insights into the spatial patterns and long-term trends that inform urban planning strategies. While many researchers have tried to use satellite imagery for urban temperature analysis, few have validated their models with an independent data source. The testing using S-DoT sensors, based on over 23,000 data points from a dense sensor network, shows that satellite-based approaches can accurately convert data to air temperature. Unlike most urban studies that analyze data over a short period, this research examines a full decade of data from 2013 to 2024.

The findings from this research provide specific, quantitative guidance for urban planning policies aimed at mitigating air temperature and managing energy consumption in dense urban districts.

References

[1] J.W. Oh, J. Ngarambe, P.N. Duhirwe, G.Y. Yun, and M. Santamouris. Using deep-learning to forecast the magnitude and characteristics of urban heat island in Seoul Korea. *Scientific Reports*, 10:3559, 2020. doi:10.1038/s41598-020-60632-z.

[2] D. Maiullari. *Urban form influence on microclimate and building cooling demand: An analytical framework and its application in the Rotterdam case*. PhD thesis, Delft University of Technology, 2023.

[3] M. Mosteiro-Romero and Schlueter A. Effects of occupants and local air temperatures as sources of stochastic uncertainty in district energy system modeling. *Energies*, 14:2295, 2021. doi:10.3390/en14082295.

[4] J. Voogt and T. Oke. Thermal remote sensing of urban climates. *Remote Sensing of Environment*, 86:370– 384, 2003. doi:10.1016/s0034-4257(03)00079-8.

[5] J. Schatz and C. J. Kucharik. Urban climate effects on extreme temperatures in Madison, Wisconsin, USA. *Environmental Research Letters*, 10:094024, 2015. doi:10.1088/1748-9326/10/9/094024.

[6] H. Meyer, M. Katurji, T. Appelhans, M. Müller, T. Nauss, P. Roudier, and P. Zawar-Reza. Mapping daily air temperature for Antarctica based on MODIS LST. *Remote Sensing*, 8:732, 2016. doi:10.3390/rs8090732.

[7] P. Noi, J. Degener, and M. Kappas. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. *Remote Sensing*, 9:398, 2017. doi:10.3390/rs9050398.

[8] Y. Chen, J. Quan, W. Zhan, and Z. Guo. Enhanced statistical estimation of air temperature incorporating nighttime light data. *Remote Sensing*, 8:656, 2016. doi:10.3390/rs8080656.

[9] M. Ignatius, J. Lim, B. Gottkehaskamp, K. Fujiwara, C. Miller, and F. Biljecki. Digital twin and wearables unveiling pedestrian comfort dynamics and walkability in cities. *ISPRS Ann. Photogramm. Remote Sens. and Spatial Inf. Sci.*, X-4/W5-2024:195–202, June 2024. doi:10.5194/isprs-annals-X-4-W5-2024-195-2024.

[10] X. Ding, Y. Zhao, Y. Fan, Y. Li, and J. Ge. Machine learning-assisted mapping of city-scale air temperature: Using sparse meteorological data for urban climate modeling and adaptation. *Building and Environment*, 234:110211, 2023. doi:10.1016/j.buildenv.2023.110211.

[11] K. Tang, H. Zhu, and P. Ni. Spatial downscaling of land surface temperature over heterogeneous regions using random forest regression considering spatial features. *Remote Sensing*, 13:3645, 2021. doi:10.3390/rs13183645.

[12] S. Choi, D. Jin, N. Seong, D. Jung, S. Sim, J. Woo, U. Jeon, Y. Byeon, and K. Han. Near-surface air temperature retrieval using a deep neural network from satellite observations over South Korea. *Remote Sensing*, 13:4334, 2021. doi:10.3390/rs13214334.

[13] W. Salih, E. M. E. Khalki, V. Ongoma, R. Lguensat, B. Aithssaine, H. Ouatiki, F. Driouech, B. Sebbar, S. Achli, and A. Chehbouni. TEMLI: A high-resolution air temperature estimation using machine learning and land surface data across Morocco. *Earth Systems and Environment*, 9:1311–1331, 2025. doi:10.1007/s41748-025-00629-8.

[14] T. Kotsiopoulos, P. Sarigiannidis, D. Ioannidis, and D. Tzovaras. Machine learning and deep learning in smart manufacturing: The smart grid paradigm. *Computer Science Review*, 40:100341, 2021. doi:10.1016/j.cosrev.2020.100341.

[15] W. Liao, T. Hong, and Y. Heo. The effect of spatial heterogeneity in urban morphology on surface urban heat islands. *Energy and Buildings*, 244:111027, 2021. doi:10.1016/j.enbuild.2021.111027.

[16] S. Raj and G. Y. Yun. Exploring the role of strategic urban planning and greening in decreasing surface urban heat island intensity. *Journal of Asian Architecture and Building Engineering*, 24:866–879, 2024. doi:10.1080/13467581.2024.2308592.

[17] N. Sahani. Assessment of spatio-temporal changes of land surface temperature (LST) in Kanchenjunga Biosphere Reserve (KBR), India using Landsat satellite image and single channel algorithm. *Remote Sensing Applications: Society and Environment*, 24: 100659, 2021. doi:10.1016/j.rsase.2021.100659.

[18] Korea Meteorological Administration. KMA data service. `https://data.kma.go.kr/cmmn/main.do`, 2025. Accessed: 11/09/2025.

[19] M. Park and K. Baek. Quality management system for an IoT meteorological sensor network–Application to Smart Seoul Data of Things (S-DoT). *Sensors*, 23:2384, 2023. doi:10.3390/s23052384.

[20] J. Heaton. *Introduction to Neural Networks with Java*. Heaton Research, USA, 2008.

[21] S. Rasp, M. S. Pritchard, and P. Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115:9684–9689, 2018. doi:10.1073/pnas.1810286115.

[22] I. Salehin and D. Kang. A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics*, 12:3106, 2023. doi:10.3390/electronics12143106.

[23] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

[24] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference for Learning Representations*, 2015. doi:10.48550/arxiv.1412.6980.

[25] Y. Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade*, pages 437–478. Springer, Berlin, Heidelberg, 2012. doi:10.1007/978-3-642-35289-8_26.

[26] D. Masters and C. Luschi. Revisiting small batch training for deep neural networks. *arXiv preprint*, 1804.07612, 2018. doi:10.48550/arxiv.1804.07612.

[27] M. Naserikia, M. A. Hart, N. Nazarian, B. Bechtel, M. Lipson, and K. A. Nice. Land surface and air temperature dynamics: The role of urban form and seasonality. *The Science of the Total Environment*, 905: 167306, 2023. doi:10.1016/j.scitotenv.2023.167306.

A Survey-Controlled Photogrammetric Framework for Consistent Multi-Floor Geometry Estimation in Robotic Facade Installation

Danya Liu[1] **Niki Kentroti**[1] **and Kepa Iturralde**[1]

[1]Chair of Digital Transformation in Construction, Institute of Construction Management, Faculty of Civil and Environmental Engineering, University of Stuttgart, Stuttgart, Germany

danya.liu@ibl.uni-stuttgart.de, niki.kentroti@ibl.uni-stuttgart.de kepa.iturralde@ibl.uni-stuttgart.de

Abstract -

Robotic facade installation across multiple floors necessitates millimeter-level geometric accuracy to ensure reliable anchor alignment and panel placement. The present paper proposes a survey-controlled sensing and modeling framework that integrates a construction tool with high-resolution close-range imaging to establish a consistent multi-floor reference frame. The utilization of facade anchors is facilitated through the employment of survey markers, which serve to localize these anchors. The registration of dense, image-based geometry to the survey frame enables the execution of precise height transfer and inter-floor alignment. The resulting survey-referenced geometry is aligned with an existing design Building Information Modeling (BIM)/Industry Foundation Classes (IFC) model and evaluated against survey ground truth to verify multi-floor consistency, error propagation, and anchor-level suitability for robotic facade installation. The proposed framework is predicted on the explicit addressing of multi-floor alignment and error propagation, thereby providing a reliable geometric foundation for robotic facade installation and reproducible accuracy evaluation across floors.

Keywords -

Construction, visual, localization, BIM, Measurement, installation, robotics.

1 Introduction

The installation of robotic facades across multiple floors necessitates geometric accuracy at the millimeter to centimeter level. Errors in inter-floor alignment, encompassing anchor positions, floor-to-floor height, and coordinate transfer, have the potential to accumulate and directly impact panel fit-up.

Recent advancements in fiducial-marker localization, high-precision surveying, and photogrammetric reconstruction have the potential to enhance on-site construction geometry. Specifically, markers facilitate reliable camera pose estimation, total stations facilitate the establishment of multi-floor control networks, and high-resolution imaging supports detailed facade and edge reconstruction. However, extant research has largely treated these components in isolation or focuses on documentation, with limited attention to survey-controlled multi-floor alignment and its direct integration into robotic facade installation.

The proposed framework integrates a construction tool with high-resolution imaging into a unified sensing and modeling pipeline with iterative robotic feedback. By anchoring measurements to a survey-defined reference frame, it ensures consistent multi-floor alignment and accurate execution-time synchronization between the physical construction and the design BIM.

2 Related Work

This part is devoted to a critical review of the extant literature concerning marker-based localization, high-precision surveying, photogrammetric building modeling, and BIM–robotics integration. It identifies the research gaps that this paper aims to address.

2.1 Marker-based localization

Visual fiducial markers are widely used for camera pose estimation and calibration in robotics and computer vision. AprilTag [1] provides reliable 6-DOF pose estimation from a single image and is robust to lighting changes and partial occlusion, with later versions improving detection accuracy and efficiency [2][3]. ArUco markers [4] offer a similar approach with built-in error correction.

Fiducial markers are also commonly used for sensor calibration, including camera–laser calibration and ChArUco boards that combine checkerboard features with marker IDs [5][6]. Prior studies have further evaluated detection robustness and pose ambiguity under challenging conditions [7]. Beyond pose estimation, recent work [8] shows that dense AprilTag grids can be used to reconstruct survey-grade facade geometry from calibrated images via triangulation and plane fitting, extending fiducial markers from local localization to large-scale geometric referencing in the built environment. In construction applications, markers are typically used for localization, while

workflows that integrate surveyed fiducials across multiple floors with BIM and geospatial datasets remain largely unexplored.

2.2 High-precision surveying and Scan-to-BIM in construction

High-precision surveying is a cornerstone of construction quality control. Recent studies have evaluated scan-to-BIM workflows based on terrestrial laser scanning (TLS), with a focus on the impact of scanner accuracy, registration strategy, and modeling choices on the quality of the resulting BIM model. Skrzypczak et al. [9][10] assess the accuracy of scan-to-BIM for a multi-story office building, while others demonstrate the application of scan-to-BIM in the context of historic churches and complex interior spaces. The study [11][12][13] demonstrates that achieving centimeter- or millimeter-level agreement with control measurements is contingent upon the meticulous design of the survey network. Related work on online data acquisition and building modeling for building renovation highlights the role of timely and built-in data integration in renovation workflows [14], which will be further improved in future research.

The majority of Scan-to-BIM studies concentrate on documentation of entire spaces and neglect to address the anchor-level, cross-floor alignment necessary for the installation of facade panels. Furthermore, there is a paucity of integration between total-station control networks and high-resolution close-range imaging in the context of survey-grade ground truth generation and rigorous validation of multi-floor alignment accuracy. In contrast to Scan-to-BIM workflows, which prioritize post-construction documentation or renovation, the present study focuses on execution-time geometric control in new building construction, where a design BIM is already available.

2.3 Image-based photogrammetric reconstruction

Image-based photogrammetry has been the subject of extensive research as a method for extracting geometric information from perspective imagery. Classical methods rely on multi-view geometry, camera calibration, and feature correspondence to recover three-dimensional structure from overlapping images, typically via structure-from-motion and projective reconstruction pipelines. These techniques facilitate the recovery of facade planes, edges, and relative spatial relationships directly from images without direct contact measurements.

Recent advancements in this field have extended classical photogrammetry by incorporating learning-based components and architectural priors to enhance its resilience under conditions of limited viewpoints, occlusions, and complex urban scenes. The viability of such approaches has been demonstrated in the reconstruction of facade geometry and metric relationships from sparse or street-level imagery at large scale, primarily for applications in mapping, visualization, and urban analysis [15][16][17]. Other studies infer heights and facade geometry from single images or panoramas using deep learning and geometric reasoning [18][19][20], and extend these methods to 3D city modeling and Scan-to-BIM pipelines [21][22].

Nevertheless, image-only photogrammetric methods are susceptible to camera pose uncertainty, scale ambiguity, and environmental conditions, which restrict their applicability for precision-critical construction tasks. Consequently, the present paper adopts photogrammetry as a complementary measurement modality within a survey-controlled framework. In this framework, image-derived geometric features are anchored to a stable reference system to support the accurate installation of robotic facades.

2.4 BIM–robotics integration

The integration of BIM and robotics is a pivotal subject in the domain of construction automation. BIM functions as a centralized information repository for robotic assembly, additive manufacturing, inspection, and autonomous construction systems, as evidenced by [23]. Bibliometric studies [24] indicate that the field of construction robotics is undergoing rapid growth and diversification, particularly in the domains of automated fabrication and mobile on-site platforms.

Research on BIM–GIS integration shows that combining spatial and semantic models enables robot navigation and planning in complex construction environments [25]. Recent work further extends BIM by incorporating laser scanning and photogrammetry to support digital twins and automated quality verification [21], while also providing essential geometric and semantic constraints for safe robotic operation [26].

The majority of extant studies presuppose a precise design BIM and concentrate on the manner in which robots consume BIM information. In contrast, this project focuses on generating as-built BIM/IFC models from survey control and image-based measurements, and on using the resulting geometry to support robotic multi-floor facade installation.

3 Problem Definition

3.1 Multi-floor alignment challenges

The installation of robotic facades across multiple floors necessitates the maintenance of consistent geometric alignment between anchors, panels, and reference frames over vertical distances. In practice, minor inaccuracies in floor-to-floor height transfer, control point propagation, or coordinate transformations can accumulate across

levels, resulting in misalignment between panels installed on different floors. These effects are of particular significance when facade elements span multiple storeys or must satisfy tight tolerance constraints at panel joints and anchors.

3.2 Challenges of maintaining consistent reference frames

In the context of multi-floor construction, geometric errors frequently emerge not from individual measurements but from inconsistencies in reference frames, transformation chains, and floor-to-floor transfer. Despite the fact that individual sensing modalities can offer reasonable local accuracy, minor inconsistencies in alignment, referencing, or propagation across floors can accumulate and result in misalignment at panel joints and anchors. The central challenge in robotic facade installation is therefore to maintain a consistent, traceable, and verifiable geometric reference across multiple floors throughout the planning and execution phases.

3.3 Requirements for precision in robotic installation

In new building construction, a detailed design BIM is typically available prior to facade installation. However, ensuring that robotic installation across multiple floors conforms to the design geometry remains challenging due to cumulative inter-floor alignment errors, reference frame inconsistencies, and execution-time uncertainties.

3.4 Need for BIM-constrained sensing and execution

Addressing these challenges necessitates a framework that integrates high-precision sensing with an existing design BIM to support execution-time alignment during robotic facade installation. In order to ensure consistent cross-floor alignment with verifiable precision, a unified pipeline combining survey control, image-based reconstruction, and BIM/IFC representation is necessary.

4 System Architecture

Fig.1 illustrates the execution-time workflow. The establishment of a global reference frame across multiple floors is the primary function of survey control. Based on this reference frame, images are acquired and utilized for photogrammetric localization, and the resulting measurements are registered to the survey frame to obtain survey-anchored geometry. This geometry is employed to generate robot commands for facade installation. Subsequent to the execution, re-localization and verification are carried out to effect an update to the geometric state. Subsequently, verified states and deviations are integrated with the existing design BIM/IFC model as as-built information. And Fig.2 offers a comprehensive overview of the proposed system architecture and data flow for robotic facade installation across multiple floors.

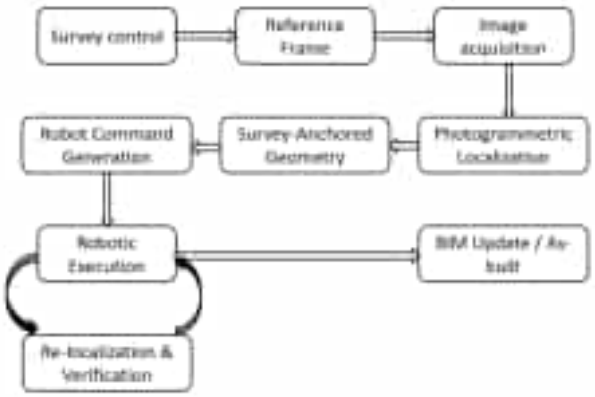


Figure 1. Survey-controlled workflow for multi-floor robotic facade installation

4.1 Survey-controlled reference framework

A Leica iCON iCS50 construction tool establishes a global reference frame that spans multiple floors and serves as the geometric backbone of the system. Each facade anchor is equipped with a Leica survey marker, enabling repeatable and survey-grade localization. By observing these markers across storeys, consistent floor-to-floor height transfer and inter-floor alignment are ensured. This survey-controlled reference frame provides stable geometric ground truth for subsequent sensing, fusion, and robotic operations.

4.2 High-resolution sensing and geometric reconstruction

High-resolution close-range images are captured using a Hasselblad camera to obtain dense geometric information of facade surfaces, anchors, and panel interfaces. Image observations are linked to the survey reference frame through fiducial markers, enabling reconstructed geometry to be expressed in global coordinates. This combination of sparse, high-accuracy survey measurements and dense image-based reconstruction supports both local installation verification and building-scale geometric analysis.

4.3 Robotic installation context

Robotic platforms operating on different floors are referenced to the same global coordinate system defined by the survey framework. This shared reference enables coherent spatial alignment during multi-floor facade installation

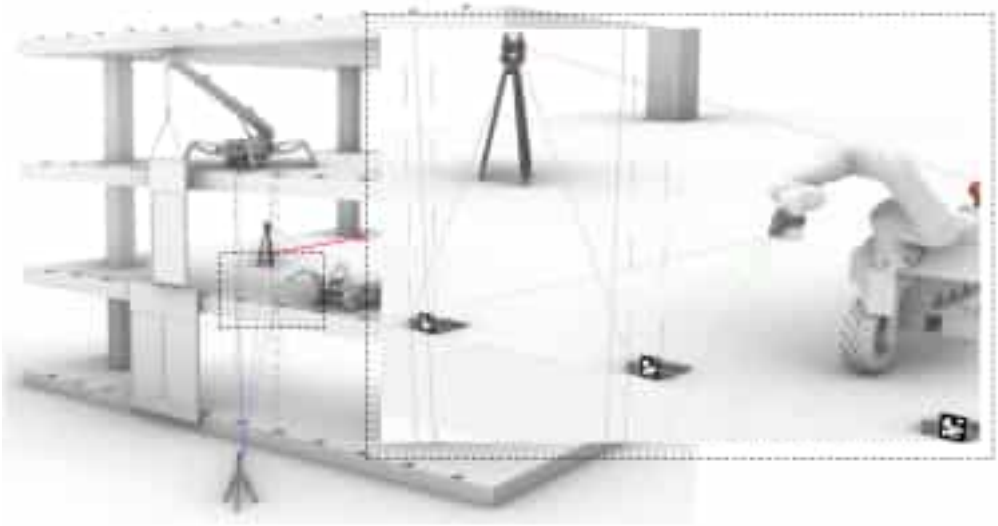


Figure 2. Conceptual illustration of a multi-floor robotic facade installation workflow

and reduces cumulative geometric errors. The architecture is designed to support installation planning, execution, and verification without assuming perfect design geometry, by grounding alignment and verification in survey control.

4.4 BIM / IFC integration

Survey- and image-derived measurements are integrated with an existing design BIM/IFC model to provide execution-time verification and structured documentation of as-built deviations during robotic facade installation. Anchor positions, facade geometry, and inter-floor relationships are stored within a unified BIM model, enabling comparison with design BIM and survey-observed as-built measurements. This integration supports automated quality documentation, traceability of geometric deviations, and downstream analysis throughout the building lifecycle.

5 Methodology

A methodology has been proposed for the systematic evaluation of millimeter-level accuracy in robotic facade installation across multiple floors. This methodology integrates controlled experiments, survey- and image-based sensing, data fusion, and evaluates the results against survey ground truth. Each methodological component addresses a specific aspect of the problem definition outlined in ***Problem Definition***, and collectively, they form a coherent evaluation pipeline.

5.1 Controlled and multi-floor experiments

Controlled experiments are conducted on buildings with multiple floors to evaluate whether robotic facade installation can satisfy millimetre-level geometric accuracy requirements. Anchors equipped with Leica survey markers are distributed across storeys and measured using a Leica iCON iCS50 construction tool, forming a vertically connected survey control network.

Survey-measured anchor coordinates are treated as ground truth. Image-based and fused measurements are evaluated relative to this reference in order to distinguish local single-floor errors from cumulative inter-floor alignment effects.

For the f-th floor, the mean vertical residual is defined as

$$r_f = \frac{1}{|\mathcal{A}_f|} \sum_{a \in \mathcal{A}_f} \left(z_a^M - z_a^S \right), \tag{1}$$

where $\mathcal{A}_f$ denotes the set of anchors on floor f, and z_a^S and z_a^M are the survey-measured and model-derived heights of anchor a, respectively.

The cumulative vertical drift up to floor F is expressed as

$$D(F) = \sum_{f=1}^{F} r_f. \tag{2}$$

The installation accuracy requirement is defined as

$$|D(F)| \leq \tau, \qquad \tau \in [2, 8] \text{ mm}, \tag{3}$$

reflecting typical tolerance limits for robotic facade installation.

Fig.3 illustrates how anchor-level vertical residuals are averaged at each floor and accumulate across multiple

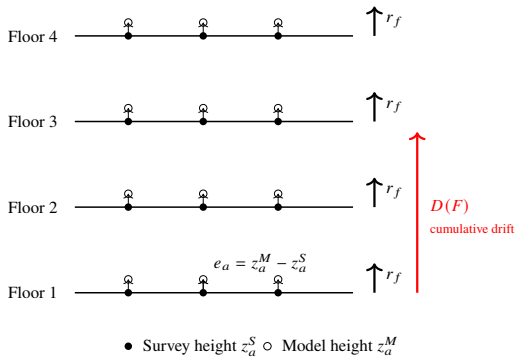


Figure 3. Illustration of equations(1)(2)(3)

floors into a cumulative vertical drift that affects multi-floor robotic facade installation accuracy.

5.2 Camera calibration and photogrammetry

The Hasselblad camera is calibrated using a standard photogrammetric procedure to estimate intrinsic parameters and lens distortion. Based on solid theory basic [27][28][29], Camera poses are initialized via fiducial marker detection and refined using bundle adjustment by minimizing reprojection error.

Given image observations $\mathbf{x}_{ij}$ of 3D points $\mathbf{X}_j$ in camera i, the optimization problem is formulated as

$$\min_{\{\mathbf{P}_i\},\{\mathbf{X}_j\}} \sum_{i,j} \left\| \mathbf{x}_{ij} - \pi(\mathbf{P}_i, \mathbf{X}_j) \right\|^2, \tag{4}$$

where $\pi(\cdot)$ denotes the camera projection model.

To enforce metric consistency with survey control, anchor points measured by the Leica iCS50 are incorporated as constraints:

$$\min \sum_{i,j} \left\| \mathbf{x}_{ij} - \pi(\mathbf{P}_i, \mathbf{X}_j) \right\|^2 + \lambda \sum_{k \in C} \left\| \mathbf{X}_k - \mathbf{X}_k^S \right\|^2, \tag{5}$$

where C denotes the set of survey-observed anchors and $\mathbf{X}_k^S$ their survey coordinates.

5.3 Leica and camera data fusion

Survey measurements and image-derived geometry are fused within the global coordinate system defined by the Leica iCON iCS50. Anchor-mounted survey markers serve as control points linking image space and survey space.

A rigid transformation $(\mathbf{R}, \mathbf{t})$ is estimated using weighted least squares:

$$\min_{\mathbf{R},\mathbf{t}} \sum_k w_k \left\| \mathbf{p}_k^S - (\mathbf{R}\mathbf{p}_k^I + \mathbf{t}) \right\|^2, \tag{6}$$

where $\mathbf{p}_k^S$ and $\mathbf{p}_k^I$ denote survey and image-derived coordinates of anchor k, and w_k represents measurement confidence [30].

To improve robustness against occlusions and misdetections, a robust cost function $\rho(\cdot)$ is applied:

$$\min_{\mathbf{R},\mathbf{t}} \sum_k \rho\left(\left\| \mathbf{p}_k^S - (\mathbf{R}\mathbf{p}_k^I + \mathbf{t}) \right\| \right). \tag{7}$$

The residual covariance after fusion is estimated as

$$\mathbf{\Sigma}_{\text{reg}} = \frac{1}{N-1} \sum_{k=1}^{N} \mathbf{e}_k \mathbf{e}_k^\top, \tag{8}$$

where $\mathbf{e}_k$ denotes the post-registration residual of anchor k.

5.4 Accuracy evaluation against survey ground truth

Geometric quantities derived from the fused survey–image model are evaluated against survey control measurements provided by the Leica iCON iCS50. Survey-measured anchor coordinates are treated as ground truth, enabling a direct assessment of anchor-level consistency and multi-floor error propagation.

For each anchor a, the vertical residual is defined as

$$\Delta z_a = z_a^M - z_a^S, \tag{9}$$

where z_a^S denotes the survey-measured vertical coordinate and z_a^M the corresponding model-derived value.

To assess suitability for robotic facade installation, a pass-rate metric is introduced:

$$\eta(\tau) = \frac{1}{M} \sum_{m=1}^{M} I\left(|\Delta z_m| \le \tau \right), \quad \tau \in [2, 8] \text{ mm}, \tag{10}$$

where M is the number of evaluated anchors and $I(\cdot)$ denotes the indicator function.

This evaluation focuses on anchor-level geometric consistency, inter-floor error propagation, and compliance with tolerance requirements relevant to precision-critical robotic installation.

To further characterize error distribution, standard statistics of the vertical residual Δz_a are evaluated:

$$\text{RMSE}_z = \sqrt{\frac{1}{M} \sum_{a=1}^{M} (\Delta z_a)^2}, \quad e_z^{\max} = \max_a |\Delta z_a| \tag{11}$$

In addition, full three-dimensional anchor deviation is defined as:

$$e_a^{3D} = \|\mathbf{p}_a^M - \mathbf{p}_a^S\|_2 \tag{12}$$

To assess cross-floor consistency, the inter-floor transfer error between adjacent floors is introduced:

$$E_{\text{transfer}}(f, f+1) = \left| (z_{f+1}^M - z_f^M) - (z_{f+1}^S - z_f^S) \right| \tag{13}$$

These metrics complement the pass-rate definition by capturing error magnitude and multi-floor consistency.

5.5 IFC data mapping

Survey- and image-derived as-built geometry [31][32] is mapped into IFC representations to enable structured storage, comparison, and documentation.

For each anchor a, the geometric deviation is expressed as

$$\Delta \mathbf{p}_a = \mathbf{p}^{\text{IFC}}_{a,\text{as-built}} - \mathbf{p}^{\text{IFC}}_{a,\text{design}}, \tag{14}$$

with the vertical component Δz_a explicitly recorded. These deviations are stored in IFC property sets and associated with anchors, panels, and floors.

6 Experimental Plan

The experimental plan represents the most critical component of the proposed future work, as it provides the quantitative evidence required to validate multi-floor alignment accuracy, error propagation behavior, and robustness under realistic construction conditions. Although experimental data will be acquired in later project phases, measurement layouts, marker placement rules, and observation strategies are defined in advance to ensure reproducibility and comparability.

6.1 Benchmark scenarios

Three benchmark scenarios are defined to progressively evaluate geometric accuracy from local to building-scale conditions:

- Single-floor scenario: This scenario evaluates local geometric accuracy within a single storey, isolating sensing, calibration, and fusion errors without inter-floor propagation. It establishes the baseline performance of the Leica-camera system.
- Cross-floor scenario: This scenario evaluates two adjacent floors jointly, focusing on floor-to-floor height transfer and coordinate consistency. It represents the minimum requirement for facade panels spanning multiple storeys.
- Full-height scenario: This scenario evaluates anchors distributed across the full building height. It exposes cumulative vertical drift and long-range error propagation, representing the most demanding case for robotic facade installation.

6.2 Marker placement strategy

Leica survey markers are placed on the facade following a regular grid layout to provide stable geometric control in both horizontal and vertical directions. The placement strategy defined to balance coverage, redundancy, and practical deployment constraints is as follows:

Table 1. Planned Leica marker layouts for benchmark scenarios.

Scenario	Floors	Spacing (m)	Grid (C×R)	Markers
Single-floor	1	2.0	3–4 × 2	6–8
Cross-floor	2	2.0	3–4 × 2	12–16
Full-height	3–5	2.0	3–4 × 2	18–40

- Horizontal and vertical spacing: 2.0 m.
- Markers are mounted at or near facade anchor locations where possible.
- At least one vertically aligned marker column is maintained across floors to support height transfer.
- All markers are observable from at least one construction tool setup.

6.3 Planned marker layout summary

The planned marker configurations for the three benchmark scenarios are summarised in Table 1.

6.4 Marker visibility and construction tool setup plan

To ensure reliable multi-floor alignment and redundancy in survey observations, a structured construction tool setup plan is defined for future experiments using the Leica iCON iCS50.

- Each marker is observed from at least two distinct station setups.
- Adjacent floors share common visible markers to form a continuous vertical control network.
- Station locations are selected to maximize facade coverage while avoiding shallow incidence angles.

6.5 Deviation measurement

Geometric deviations are quantified relative to survey-grade ground truth provided by the Leica iCON iCS50. Deviations are evaluated at anchor level, panel level, and across floors, with a focus on millimetre-level tolerances required for robotic installation.

Deviation metrics include local anchor residuals, inter-floor height differences, and cumulative vertical drift. All measurements are evaluated against predefined tolerance thresholds to assess installation feasibility.

6.6 Reproducibility and robustness testing

To assess reproducibility, experiments are repeated under varying acquisition conditions, including different

camera viewpoints, lighting conditions, and station configurations. Robustness is evaluated by introducing controlled perturbations such as partial marker occlusions and reduced visibility.

Statistical consistency across repeated trials is analysed using variance, confidence intervals, and failure rates relative to millimetre-level thresholds.

7 Conclusions

This paper presented a survey-controlled sensing and execution framework for robotic facade installation in new building construction, where a design BIM is available. By integrating a Leica iCON iCS50 construction tool with high-resolution close-range imaging and BIM-based representation, the proposed pipeline establishes a consistent multi-floor reference frame for precision-critical installation.

The framework defines a unified process that links survey control, dense image-based geometry, and BIM/IFC mapping, enabling reliable robotic facade installation and quantitative validation of multi-floor accuracy, error propagation, and installation tolerances against survey ground truth. The proposed pipeline is designed to address inter-floor alignment and error propagation within a survey-controlled framework, thereby providing a coherent and reproducible solution for precision-critical robotic facade installation across multiple floors.

The proposed framework contributes to the **AMALTEA** project by providing a survey-controlled execution pipeline for robotic facade installation in new building construction.

Acknowledgement

The research presented in this paper was originally carried out in the context of the Horizon project **AMALTEA**. **AMALTEA** has received funding from the EU's Horizon Europe program under Grant Agreement No. **101189678**.

References

[1] Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*, pages 3400–3407. IEEE, 2011.

[2] John Wang and Edwin Olson. Apriltag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4193–4198. IEEE, 2016.

[3] Joshua Springer and Marcel Kyas. Evaluation of orientation ambiguity and detection rate in april tag and whycode. In *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*, pages 281–286. IEEE, 2022.

[4] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014. ISSN 0031-3203.

[5] Dengqing Tang, Tianjiang Hu, Lincheng Shen, Zhaowei Ma, and Congyu Pan. Apriltag array-aided extrinsic calibration of camera–laser multi-sensor system. *Robotics and biomimetics*, 3(1):13, 2016.

[6] André Brunn, Frank Michaux, Osmund Tedjasukmana, Christian Willert, and Stephan Kallweit. Camera calibration using fiducial markers: A novel tool for automated piv systems. 2025.

[7] Víctor Manuel Mondéjar-Guerra, Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Manuel J. Marín-Jiménez, and Rafael Medina Carnicer. Robust identification of fiducial markers in challenging conditions. *Expert Syst. Appl.*, 93:336–345, 2018.

[8] Aravind Srinivasaragavan, Danya Liu, Kepa Iturralde, and Christoph Holst. A robust pipeline for facade geometry estimation using apriltags and digital imaging. In *Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC)*, 2025.

[9] Izabela Skrzypczak, Grzegorz Oleniacz, Agnieszka Leśniak, Krzysztof Zima, Maria Mrówczyńska, and Jan K Kazak. Scan-to-bim method in construction: assessment of the 3d buildings model accuracy in terms inventory measurements. *Building Research & Information*, 50(8):859–880, 2022.

[10] Agnieszka Leśniak, Izabela Skrzypczak, Leopold Kruszka, and Justyna Jaskowska-Lemańska. Scan-to-bim method and analysis of measurement accuracy on the example of a historic church. *Inżynieria Bezpieczeństwa Obiektów Antropogenicznych*, 2023.

[11] Juan Moyano, Ángel Justo-Estebaranz, Juan E. Nieto-Julián, Alfonso Ojeda Barrera, and María Fernández-Alconchel. Evaluation of records using terrestrial laser scanner in architectural heritage for information modeling in hbim construction: The case study of the la anunciación church (seville). *Journal of Building Engineering*, 62:105190, 2022. ISSN 2352-7102.

[12] Enrique Valero, Frédéric Bosché, and Martin Bueno. Laser scanning for bim. *Journal of Information Technology in Construction (ITcon)*, 27(23):486–495, 2022.

[13] MH Zakaria, KM Idris, Z Majid, MFM Ariff, N Darwin, MA Abbas, K Zainuddin, RA Shukor, and MA Aziz. Practical terrestrial laser scanning field procedure and point cloud processing for bim applications–tnb control and relay room 132/22kv. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:705–710, 2019.

[14] Danya Liu, Kepa Iturralde, and Christoph Holst. Online data acquisition and building modelling for building renovation. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 42, pages 949–955. IAARC Publications, 2025.

[15] Hao Li, Zhendong Yuan, Gabriel Dax, Gefei Kong, Hongchao Fan, Alexander Zipf, and Martin Werner. Semi-supervised learning from street-view images and openstreetmap for automatic building height estimation. *arXiv preprint arXiv:2307.02574*, 2023.

[16] Kaixuan Fan, Anqi Lin, Hao Wu, and Zhenci Xu. Pano2geo: An efficient and robust building height estimation model using street-view panoramas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215:177–191, 2024. ISSN 0924-2716.

[17] Xiucheng Liang, Jinheng Xie, Tianhong Zhao, Rudi Stouffs, and Filip Biljecki. Openfacades: An open framework for architectural caption and attribute data enrichment via street view imagery. *arXiv preprint arXiv:2504.02866*, 2025.

[18] Pratik Agarwal, Wolfram Burgard, and Luciano Spinello. Metric localization using google street view. 03 2015.

[19] Shibo Ge, Jiping Liu, Xianghong Che, Yong Wang, and Haosheng Huang. A novel method for estimating building height from baidu panoramic street view images. *ISPRS International Journal of Geo-Information*, 14(8), 2025. ISSN 2220-9964.

[20] Yizhen Yan and Bo Huang. Estimation of building height using a single street view image via deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:83–98, 2022.

[21] Andrzej Szymon Borkowski and Alicja Kubrat. Integration of laser scanning, digital photogrammetry and bim technology: A review and case studies. *Eng*, 5(4):2395–2409, 2024. ISSN 2673-4117.

[22] Theresa Meyer. Change detection and geometric bim verification for indoor construction site monitoring based on 3d point clouds and uncertainties.

[23] Hongzhe Yue, Qian Wang, Zixuan Zhao, Sha Lai, and Guanying Huang. Interactions between bim and robotics: Towards intelligent construction engineering and management. *Computers in Industry*, 169: 104299, 2025. ISSN 0166-3615.

[24] Manpreet Singh and Jeevanjot Chagger. Review study: Robotics and automation in construction. *International Research Journal of Innovations in Engineering and Technology*, 8:260–264, 11 2024.

[25] Sina Karimi and Ivanka Iordanova. Integration of bim and gis for construction automation, a systematic literature review (slr) combining bibliometric and qualitative analysis. *Archives of Computational Methods in Engineering*, 28(7), 2021.

[26] Obiora Odugu, Fatemeh Ghafari, Ehsan Shourangiz, Muhammad Tahir Khan, and Chao Wang. Building information model (bim) and robotic systems integration for construction: A comprehensive workflow analysis and future perspectives. In *International Conference on Human-Computer Interaction*, pages 272–282. Springer, 2024.

[27] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2002.

[28] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999.

[29] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[30] K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700, 1987.

[31] Thomas Liebich. Ifc4—the new buildingsmart standard. In *IC Meeting*, pages 1–25. bSI Publications Helsinki, Finland, 2013.

[32] Nuno Abreu, Andry Pinto, Aníbal Matos, and Miguel Pires. Procedural point cloud modelling in scan-to-bim and scan-vs-bim applications: A review. *ISPRS International Journal of Geo-Information*, 12 (7):260, 2023.

Real-Time Temporal Convolution based Worker Motion Recognition with Three IMU Sensor Combinations*

Minjun Chang[1], Sai Machiraju[2], Francis Baek[3] and Yong K. Cho[†]

[1]School of Civil and Environmental Engineering, Georgia Institute of Technology, United States of America
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, United States of America
[3]Assistant Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, United States of America
[†]Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, United States of America

minjunchang@gatech.edu, saimachi@gatech.edu,francis.baek@ce.gatech.edu,yong.cho@ce.gatech.edu

Abstract -

On active jobsites, unobserved worker motions drive injuries and near-misses, unsafe habits evade periodic audits, so reliable activity monitoring is essential to prevent delays and costly rework. In recent years, Inertial-Measurement-Unit(IMU) based motion recognition using Deep Learning approach is used widely in many domains. While the method has been proven to effectively recognize human motion using 5 to 21 IMUs, it is impractical to attach that many IMUs to construction workers. Minimal-sensing approaches using a single sensor are also proposed for limited motion patterns. To address these limitations, this paper presents a Temporal Convolutional Network (TCN) based real-time human motion recognition system that uses only three IMU sensors. TCN was chosen because it replaces recurrent states with small 1-D convolutions and dilations, keeping parameters and memory modest while inference runs in parallel on short windows, making it suitable for on-board deployment. Three IMUs are attached to the neck, back, and right wrist (NBW) that deliver 6-axis data at 50Hz, which are windowed for per-window z-normalized inference using a lightweight TCN with dilated, causal 1D convolutions and residual blocks. The model predicts 19 basic motion classes, classified by anatomical terms for body movements. This paper evaluated a lab-made 15-participant dataset, achieving the highest average prediction accuracy of 83.26% on the NBW combination in a single-split configuration, demonstrating the capability to achieve reasonable accuracy for field deployment. It also compares training results across diverse combinations of the three sensors to demonstrate the possibility of reducing sensor counts.

Keywords -

Inertial-Measurement-Unit, Worker Motion Recognition, 1D-CNN, Construction Safety Management

*This work was supported in part by the National Science Foundation (NSF) under Award No. 2515786 and by the Georgia Research Alliance (GRA). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or GRA

1 Introduction

Construction sites face frequent injuries, near-misses, and costly rework. Unsafe motions often occur between periodic audits, when supervisors cannot observe early warning signs[1]. Manual observation cannot scale across multiple crews or shifts. Paper based reports lag behind events, and fixed surveillance cameras struggle in dynamic construction environments[2]. Current practice relies on periodic visual audits, manual incident reports, and fixed cameras that lack continuous coverage across multiple crews. Unlike attendance-based check-in systems, the proposed approach requires no manual input from workers; the PPU devices automatically timestamp and stream sensor data continuously, enabling passive and unobtrusive monitoring. Real-time IMU-based motion recognition fills this gap by flagging hazardous postures before injuries occur, supporting ergonomic risk assessment through cumulative high-strain motion tracking, and quantifying task productivity without dependence on line-of-sight or observer presence.

Field-deployable human-motion recognition for construction workers remains constrained by sensor burden and latency, making many systems impractical on active jobsites due to donning time, discomfort, cost, battery swaps, RF congestion, and synchronization drift. Commercial motion-capture suits typically mount a large number of IMUs across the body[3] to achieve stable pose tracking, trading usability for coverage and drift robustness. Academic "sparse-IMU" efforts reduce sensor count but often depend on non-causal or offline temporal models that leverage future frames[4], additional sensing modalities[5], or heavyweight optimization pipelines compromising strict real-time operation on embedded hardware. The resulting gap is a method that delivers reliable, real-time recognition of a rich motion set with minimal sensor attachment and short, causal windows suitable for noisy and time-sensitive construction environments.

We addressed this gap by targeting three IMUs placed on the rear neck, lower back, and right wrist and en-

forcing strict causality with short windows (≈ 2 s at 50 Hz). Instead of relying on look-ahead or offline smoothing, our model is a lightweight Temporal Convolutional Network (TCN)[3] with dilated, causal 1D convolutions and residual blocks, providing a large effective receptive field without future context. The streaming pipeline performs per-window z-normalization and runs inference under low-memory conditions, targeting wearability and deployability on constrained edge hardware. Our working hypothesis is that the temporal structure of human motion can be captured by dilated causal convolutions over short windows when the three IMUs are chosen to provide complementary distal and proximal information. We compare several sensor configurations: neck+back+wrist (NBW), neck+back (NB), back+wrist (BW), neck+wrist (NW), and single sensor setups under matched computational budgets. We evaluate two training methods that reflect practical deployment: leave-one-subject-out (LOSO)[6] for cross-worker generalization and a single fixed split for rapid development. We measure accuracy, macro-F1, latency, and throughput, and perform ablation studies to characterize the trade-offs among different sensor sets and training methods. The selected sensor–training configuration is determined through a multi-objective criterion that jointly accounts for recognition reliability and field deployability, including performance, inference latency, model footprint, and sensor donning burden. Overall, our contributions are (i) a causal, edge-ready TCN pipeline for real-time worker motion recognition with only three IMUs, and (ii) a systematic ablation across seven sensor combinations and two training techniques that offer accuracy–practicality trade-offs for job site deployment.

2 Background

2.1 IMU based Human Motion Recognition

Wearable IMUs are widely used for human activity and pose estimation because they are inexpensive and robust to lighting or occlusion. Early and influential human activity recognition(HAR) systems paired multiple body-worn IMUs with deep neural networks and achieved high accuracy on curated datasets, but typically at the cost of many sensors and offline processing. Ordoñez et al. [7] proposed DeepConvLSTM, which demonstrated strong performance on multimodal IMU streams using a convolutional front-end and a recurrent back-end, establishing a deep-learning baseline for wearable HAR. Sparse-IMU pose estimators such as Deep Inertial Poser (DIP)[4] reconstruct full-body kinematics from six IMUs but rely on bidirectional temporal models and sliding windows with look-ahead, which limit strictly causal, real-time use on embedded devices. In manufacturing and construction-adjacent settings, multimodal systems often combine an IMU armband with video to improve accuracy, again trading off deployment simplicity for additional sensors and compute[8]. Prior construction-focused studies have shown that decomposing construction tasks into basic motion primitives enables detailed safety and productivity analysis, but these studies are often conducted in controlled settings, highlighting the gap between laboratory validation and field deployment[9]. Reviews and case studies report growing interest in using wearables for worker activity monitoring and safety, but note persistent barriers, such as the burden of multiple sensors, annotation costs, and synchronization challenges in dynamic construction environments. These factors point toward methods that minimize sensor count, remain causal, and operate with tight latency budgets while retaining recognition fidelity[10]. Recent construction focused IMU work by Park et al.[11] introduces the Scaffolding Worker IMU Time-series(SWIT) dataset for deep-learning recognition of hazardous and routine behaviors on scaffolds. The paper motivates IMU-based monitoring to overcome occlusion limits in camera systems and addresses the shortage of public, construction-specific datasets by releasing labeled 100 Hz time-series across ten behavior categories. State-of-the-art wearable HAR often assumes dense IMU coverage, which inflates cost and increases donning time and maintenance overhead[12]. On active construction sites many body-worn units hinder movement and PPE use and degrade reliability through frequent battery swaps, RF congestion, and synchronization drift. The gap is a method that uses few IMUs, runs causally in real time, and remains practical for noisy construction sites.

2.2 Deep Learning Methods in Motion Recognition

Given the constraints of existing wearable systems, deep learning has been widely explored to extract meaningful motion patterns from IMU signals. Architectures for IMU time series generally fall into three families. CNN–RNN hybrids[7] extract local temporal features with 1-D convolutions and model longer dependencies with recurrent units, but recurrent state adds memory overhead and can complicate low-latency deployment. Pure convolutional approaches, including Temporal Convolutional Networks (TCNs), replace recurrence with dilated causal convolutions and residual connections as they parallelize efficiently, support long receptive fields with short windows, and are well-suited to streaming inference on edge hardware[13]. Temporal convolutional stacks deliver strong accuracy with efficient inference on short windows[14] for action segmentation. TCNs have shown competitive or superior accuracy to recurrent models while preserving strict causality, motivating their use for real-time HAR from IMUs. Some approaches use multimodal fusion, pairing IMUs with cameras or audio to

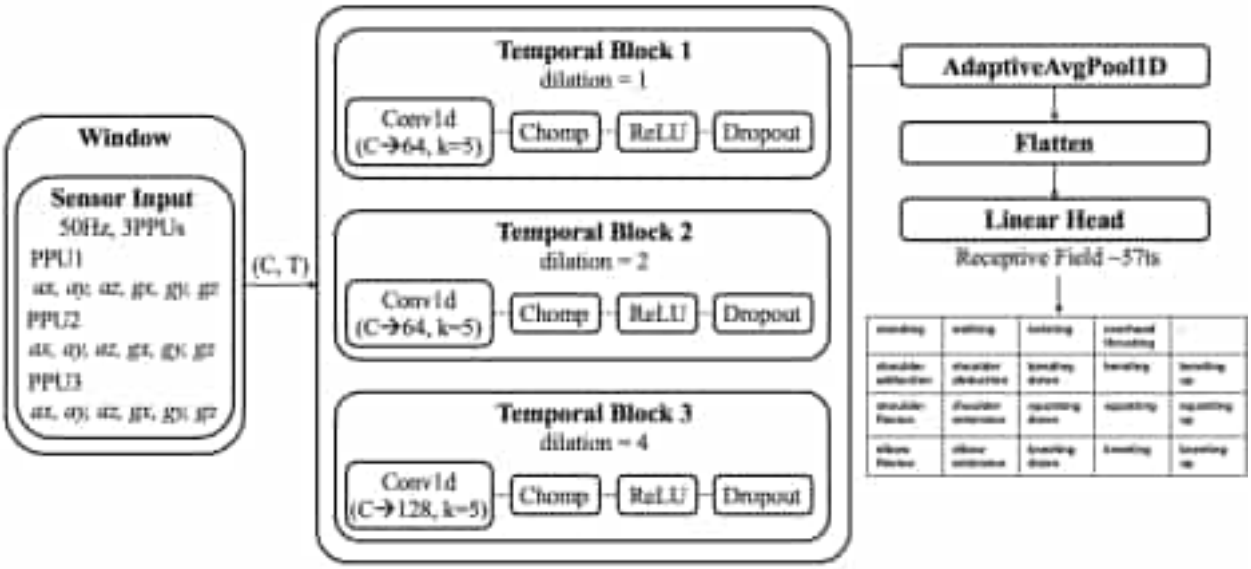


Figure 1. TCN Network Structure

improve recognition in controlled environments[6]. Another method explores optimization-based pose estimation from sparse IMUs, achieving high fidelity [9], but depends on non-causal inference or heavy priors that are hard to sustain on embedded systems. This leads to delayed monitoring time and extended computational cost. As a result, many existing approaches trade deployment practicality for accuracy or model complexity, assuming long temporal contexts or computationally intensive inference pipelines. These limitations underscore a gap in prior work for causal, lightweight temporal models that operate on short windows and support real-time inference with minimal sensor burden.

3 Methodology

3.1 Network Architecture

Our objective is to design and validate a light, edge-deployable system that recognizes 19 anatomically defined motion classes in real-time using three IMUs, while systematically identifying the most effective sensor combination and training protocol for downstream device design. We adopted TCN because it is causal, low-latency, and efficient on short windows typical of embedded streaming. Unlike bidirectional RNNs or Transformers, which require future context, a causal TCN operates on past data only and avoids recurrent state, making it easier to deploy in nonstationary jobsite conditions.

Our model described in Fig. 1 ingests a (C, T) IMU window (T=100≈2s at 50 Hz; C depends on the sensor combination). The backbone has three dilated residual blocks (channels 64/64/128, kernel 5) with shared dilations 1→2→4, ReLU+dropout, and 1×1 skips when needed. An AdaptiveAvgPool1d and a linear head produce logits for 19 motions. Right-padding and truncation ensure strict causality. This topology yields an effective receptive field of 57 steps (1.14 s), large enough to capture brief transients and sub-second pose evolution, helping separate look-alike actions like bend-down and kneel-down. We apply per-window, channel-wise z-normalization and a majority-vote smoother over recent outputs. Majority-vote smoother is a simple method of temporal post-processing where instead of displaying each prediction from time window independently, the smoother covers a short sequence(≈2s) of neighboring window predictions to select the most frequent label as output.

Computation scales modestly with kernel and channel counts, enabling real-time throughput on edge processors as memory is bounded by the fixed window. This balance of dilated temporal coverage, causal inference, and compactness suits field deployment with just three IMUs and tight latency under power budgets.

3.2 Sensors

Each ICM-20948 IMU is integrated into a wearable Personal Protection Unit (PPU) enclosure. We deploy three PPUs at the neck, lower back, and right wrist (Fig. 2). Upper-body placements such as the wrist, neck, and lower back have been shown to remain compatible with protective gear and routine work activities while maintaining user comfort[11, 15]. Prior work by Kim and Cho [12] systematically evaluated IMU quantity and placement on the body for deep learning-based worker motion recognition, identifying upper-body nodes as providing the best balance

ID	Motion	ID	Motion	ID	Motion	ID	Motion	ID	Motion
1	standing	2	walking	3	twisting	4	overhead thrusting	5	shoulder adduction
6	shoulder abduction	7	bending down	8	bending	9	bending up	10	shoulder flexion
11	shoulder extension	12	squatting down	13	squatting	14	squatting up	15	elbow flexion
16	elbow extension	17	kneeling down	18	kneeling	19	kneeling up		

Table 1. Anatomical terms for movement classification

of complementary motion coverage and minimal donning burden. Alternative locations such as the thigh and ankle were excluded because they would interfere with kneeling activities, footwear, and lower-body PPE common on construction sites. The neck unit sits at the base of the neck on the dorsal side to capture head-torso coupling while avoiding helmet interference. The back unit is centered over L5 with a belt clip to track trunk orientation and large sagittal motions. The wrist unit is strapped proximal to the ulnar styloid to capture hand activity with minimal glove obstruction. All mounts use rigid, anti-slip interfaces with consistent arrow markings to control sensor orientation.

After mounting, a brief functional check confirms sensor placement and synchronization. Subjects then perform a standardized calibration motion consisting of slow trunk flexion, an arm raise, and a return to neutral. This sequence verifies orientation consistency across the neck, back, and wrist channels and helps detect any strap slippage. All PPUs time-stamp packets on-device, and a lightweight offset correction is applied at the receiver to keep inter-PPU clock skew within 100 ms. To support deployment on remote sites with limited or absent network connectivity, each PPU stores raw sensor data locally on an onboard SD card.



Figure 2. Neck, Back, Wrist PPUs Placement

3.3 Data Collection

Motion data was collected from 15 subjects in a height range of 153 cm to 185 cm and weight range of 44kg to 83kg. Subjects were directed to perform 13 classified motions(19 including -up, -down breakdowns) in a lab environment. Sensors streamed nine-axis IMU signals at 50 Hz to a local server; all devices were timestamp-synchronized with time tolerance of 100ms. Trials were executed in pre-defined and separate orders of motions for corresponding ground truth labels.

To address motions, construction related activities such as hammering, lifting, placing, and carrying were broken down into basic motions as described in Table 1. Motions were defined using standard anatomical terms (e.g., adduction/abduction and flexion/extension), and we split them into direction-specific phases (e.g., . . . -down and . . . -up) to disambiguate symmetric transitions where applicable. The specific classification of the motions are displayed on Table 1.

3.4 Training Configuration

We trained exclusively on six-axis IMU features (three-axis accelerometer and three-axis gyroscope) and excluded magnetometer channels, as magnetic interference from steel members and powered equipment can destabilize headings on active jobsites. Two evaluation regimes were used: (i) a single split with 80% train and 20% test (equivalently, 8:2 train:test) fixed across all experiments for ablation consistency, and (ii) leave-one-subject-out(LOSO), which withholds each participant in turn to assess subject-level generalization. Windowing and inference settings matched deployment (short, causal windows), and all models were trained with identical optimization hyperparameters so differences reflect only sensor combinations and split protocols.

Table 2. Sensor Combinations used for Training

combination	selected IMUs	combination	selected IMUs
1	Neck+Back+Wrist(NBW)	5	Neck(N)
2	Neck+Back(NB)	6	Back(B)
3	Neck+Wrist(NW)	7	Wrist(W)
4	Back+Wrist(BW)		

3.5 Model Inference

3.5.1 Combination of Sensors

To identify the minimal instrumentation that preserves recognition accuracy, we trained the TCN on seven sensor configurations as described in Table 2: three single-IMU baselines (wrist, neck, back), three dual-IMU pairs (wrist+neck, wrist+back, neck+back), and one tri-IMU set (wrist+neck+back). This factorial ablation isolates the marginal value of each body location and their complementarities for dynamic and quasi-static motions. For comparability, all runs used identical windowing, optimization settings, and preprocessing. Only the input channel count varied with the chosen combination (6 channels per IMU, excluding the magnetometer, as specified in 3.4). The tri-IMU model serves as an upper-bound reference, while dual and single configurations quantify the trade-off between wearability and performance needed for field deployment.

Table 3. Training results across combinations

Method	Comb.	Acc.	Bal. Acc.	Macro F1	Wtd. F1
single split	NBW	0.9670	0.9396	0.9402	0.9671
single split	NB	0.9526	0.9343	0.9249	0.9531
single split	**NW**	**0.9748**	**0.9451**	**0.9485**	**0.9743**
single split	BW	0.9644	0.9176	0.9253	0.9636
single split	N	0.9570	0.9276	0.9246	0.9570
single split	B	0.8874	0.8227	0.8227	0.8893
single split	W	0.9470	0.8783	0.8893	0.9448
LOSO	**NBW**	**0.9607**	**0.9315**	**0.9302**	**0.9608**
LOSO	NB	0.9266	0.8828	0.8845	0.9263
LOSO	NW	0.9522	0.9137	0.9155	0.9517
LOSO	BW	0.9494	0.9013	0.9032	0.9491
LOSO	N	0.9160	0.8653	0.8673	0.9159
LOSO	B	0.8714	0.8029	0.8042	0.8702
LOSO	W	0.9263	0.8565	0.8593	0.9247

3.5.2 Training Output

Table 3 summarizes single-split performance across sensor combinations. The results are evaluated in terms of Accuracy and Balanced Accuracy (weighted by the number of samples per class), along with F1 scores. The F1-score as defined in Equation (1) is a main machine learning evaluation metric that balances precision and recall into a single score to represent the harmonic mean. We selected the Macro F1-score, which is the unweighted average of the F1-scores computed independently for each class, and the Weighted F1-score which averages the F1-scores per class, weighted by the number of samples per class.

$$\mathrm{F1}_c = \frac{2 \cdot \mathrm{Precision}_c \cdot \mathrm{Recall}_c}{\mathrm{Precision}_c + \mathrm{Recall}_c} \tag{1}$$

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^{C} \mathrm{F1}_c \tag{2}$$

Table 4. Activity–motion lists used in the evaluation.

No	Activity	Motion
1	walking	walking squatting-down
2	brick lifting	squatting squatting-up elbow-flexion
3	carrying	walking
4	brick placing	bending-down bending-up
5	grab hammer in closet	shoulder-abduction thrusting shoulder-abduction
6	hammering	elbow-flexion elbow-extension shoulder-flexion shoulder-extension
7	grab heavy toolbox	kneeling-down
8	place toolbox on top	kneeling-down thrusting
9	stretching	twisting shoulder-extension shoulder-flexion

$$\text{Weighted-F1} = \frac{\sum_{c=1}^{C} N_c \cdot \mathrm{F1}_c}{\sum_{c=1}^{C} N_c} \tag{3}$$

Neck–wrist achieves the highest accuracy 0.9748 and the highest balanced accuracy of 0.9451. It also delivers the best Macro-F1 of 0.9485 and weighted F1 of 0.9743. The tri-sensor NBW configuration achieves a competitive accuracy of 0.9670 and a balanced accuracy of 0.9396. Neck–back and back–wrist trail by small margins across all metrics. Single-sensor settings degrade, with back only at 0.8874 accuracy and wrist at 0.9470. These results indicate that the wrist provides strong discriminative cues and that the neck complements it well. The limited gains of NBW over neck–wrist suggest a two-sensor design is sufficient for field deployment

3.5.3 Real-time Inference

In-lab performance test was conducted with three participants who did not participate in training data collection process. Performance test consists of a series of activities commonly observed in construction fields, and each activity is a combination of 1 to 4 basic motions defined for data collection. The test sequence of activities and their motion composition are described in Table 4.

4 Result

4.1 Result of Single Split

We outline an 8:2 ratio single-split TCN training using 7 sensor combinations to verify its accuracy on the inference test. Results in Table 5. show that the model achieves an average accuracy of 77.64%. The optimal combination is NBW, achieving an average accuracy of 83.26%,

Table 5. Single-split inference results

Combination	Run#1	Run#2	Run#3	Run#4	Run#5	Run#6	Run#7	Run#8	Run#9	AVG
NBW	**0.8182**	**0.8608**	**0.8356**	**0.8033**	**0.8194**	**0.8209**	**0.8361**	**0.8358**	0.8636	**0.8326**
NB	0.7848	0.7595	0.7848	0.7237	0.7722	0.7975	0.7910	0.8058	0.7879	0.7785
NW	0.8228	0.8354	0.8481	0.8033	0.8472	0.8060	**0.8710**	0.8507	0.8333	0.8353
BW	0.7975	0.8101	0.8101	0.7105	0.7595	0.7848	0.8217	0.8358	0.8030	0.7926
N	0.7595	0.7470	0.7397	0.7377	0.7778	0.7848	0.8230	0.8060	0.7727	0.7720
B	0.6835	0.6962	0.7089	0.6579	0.6330	0.6962	0.7164	0.7015	0.6970	0.6878
W	0.7342	0.7215	0.7260	0.6842	0.7342	0.7468	0.7761	0.7463	0.7576	0.7363

Table 6. LOSO inference results

Combination	Run#1	Run#2	Run#3	Run#4	Run#5	Run#6	Run#7	Run#8	Run#9	AVG
NBW	**0.7468**	**0.8075**	**0.7733**	**0.8226**	**0.8194**	**0.7971**	**0.7903**	**0.8060**	**0.8182**	**0.7979**
NB	0.6582	0.6076	0.6027	0.5672	0.5890	0.6076	0.6000	0.6119	0.5909	0.6039
NW	0.7179	0.7722	0.7534	0.7143	0.7361	0.7612	0.7869	0.7463	0.7576	0.7495
BW	0.7215	0.6582	0.6579	0.6212	0.6489	0.6418	0.6515	0.6716	0.6061	0.6532
N	0.6962	0.6709	0.6962	0.6557	0.6667	0.7206	0.6825	0.6866	0.7273	0.6892
B	0.5897	0.5949	0.5890	0.5385	0.5833	0.6119	0.6615	0.6567	0.6218	0.6053
W	0.6456	0.6582	0.6712	0.6190	0.6667	0.6418	0.6769	0.7313	0.6970	0.6675

confirming that three sensors provide robust performance. BW and NB form a middle tier with averages of 79.26% and 77.85%, indicating moderate loss when one sensor is removed. Among single-sensor setups, Neck performs best at 77.20%, while Wrist reaches 73.63% and Back remains lowest at 68.78%. Including the wrist consistently improves performance across all multi-sensor configurations. These results suggest that single-split training better captures richer temporal patterns, and that the NW configuration offers the best balance between accuracy and sensor reduction.

4.2 Result of LOSO

We present accuracy results for 7 sensor combinations and 19 motions for the TCN model trained using the LOSO method. Results show the model predicts with an average of 68.09% accuracy total. The optimal combination is NBW, with an average accuracy of 79.79%. Numerical results of accuracy are shown in Table 6. NW is the next strongest at 74.95%, which indicates a viable two-sensor option. Removing the wrist most adversely affects accuracy, with NB averaging 60.39% and BW dropping to 65.32%. For single sensors, the ranking is Neck at 68.92%, Wrist at 66.75%, and Back at 60.53%. Adding the wrist to the neck raises accuracy by 6.03 points, while adding the wrist to the back raises it by 4.79 points.

4.3 Sensor Combination x Training Method

Both single-split and LOSO training methods achieve the highest accuracy with the NBW combination, indicating that using all three sensor points is superior to leaving out one or two sensors during training. Single split training performed better across all 7 combinations than LOSO training in the current dataset, achieving an average accuracy of 83.26% in the NBW combination.

5 Discussion

The causal TCN is benchmarked across seven sensor configurations under simple single-split and leave-one-subject-out regimes. The LOSO method was applied to enhance person-to-person generalization. Although the LOSO method was expected to perform better due to its generalization to different human samples, the results indicate better inference performance on the simple 8:2 split. This result may have occurred due to a small number of training samples that were not large enough to ensure cross-person generalization with the LOSO method.

The results show that Neck+Back+Wrist(NBW) combination of sensors is optimal for worker motion monitoring with an average of 83.26%(8:2 split) and 79.79%(LOSO) accuracy. This indicates adding sensors to major nodes of body movement will increment motion prediction accuracy but not significantly. However, the hazardous nature of the construction field makes it difficult to add more sensors especially on body nodes since devices on movement nodes can interfere with natural movements. Excluding the NBW combination, Neck+Wrist(NW) combination also displayed reasonably high accuracy compared to the other 5 combinations, showing the possibility of reducing sensors while maintaining high prediction accuracy.

The results in both training methods reveal similar error sources: (i) predicting a single motion when the activity contains multiple motions, (ii) using a two-second window that can blur short transitional states, and (iii) misclassifying coupled motions such as elbow extension occurring with shoulder extension. The first and second error sources

share similar ground which means solving one is likely to solve the other together. The first issue can be addressed by scaling up the prediction to the activity level, where an activity is a combination of multiple consecutive motions. This will still predict motions, but inaccurate predictions may occur due to the order of predictions, or transition motions may be ignored on a larger scale. The third issue should be solved by weighting wrist gyro and acceleration data during training, as most of the confusion arises from elbow and shoulder actions that rely heavily on wrist PPU motion data.

To address the suggested problems while maintaining the project's real-time prediction concept, we plan to scale from motion to activity to align with practical field display. This is to define an activity as a combination of multiple basic motions; for example, brick lifting as a combination of squatting down, elbow flexion, and squatting up. Our data collection and inference were conducted in a controlled laboratory setting with scripted motion sequences and fixed timing. This environment is less dynamic and more predictable than active construction sites; thus, real-world variability, interruptions, and tool interactions may reduce performance. We will validate on live construction sites in future work. Our next step is to extend the TCN to predict real-world activities as time-ordered combinations of motions, and to embed a pre-trained model on a microcontroller for live construction-site deployment. On the hardware side, we are actively working to improve the wearability of the PPU system. Current efforts focus on reducing the wrist unit to approximately 2 inches in diameter through improved circuit design and power management, bringing it closer to the form factor of a standard wristband. This size reduction is a stepping stone toward embedding the sensing units into existing PPE such as hard hat liners and safety vests, which would eliminate separate donning steps and lower the adoption barrier on active jobsites.

6 Conclusion

This work presented a causal TCN for real-time worker motion recognition using three IMUs. The system uses short windows and per-window normalization, and it runs at 50 Hz with modest compute. Across seven sensor configurations, the NBW setup achieved the best accuracy. Single-split averaged 83.26% for NBW, and LOSO averaged 79.79%, which shows good cross-person generalization. NW was the strongest two-sensor choice and suggests a path to lower burden.

We will expand subjects and environments, add latency and power benchmarks, and deploy a compact model on a PPU. We will also study placement variants and class-weighted training to reduce upper-limb confusion. Overall, the results show that a small, causal model with three IMUs can support practical jobsite monitoring.

References

[1] Jochen Teizer, Brian S. Allread, Craig E. Fullerton, and Jimmie Hinze. Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. *Automation in Construction*, 19(5):630–640, 2010.

[2] J. Gong and C.H. Caldas. Computer vision-based video interpretation model for automated productivity analysis of construction operations. *Journal of Computing in Civil Engineering*, 23(3):151–160, 2009.

[3] S. Bai, J. Z. Kolter, and V. Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[4] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll. Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics*, 2018. SIGGRAPH Asia.

[5] Y. Ren et al. Lidar-aid inertial poser: Large-scale human motion capture by sparse inertial and lidar sensors. *arXiv preprint arXiv:2205.15410*, 2022.

[6] Davoud Gholamiangonabadi, Nikita Kiselov, and Katarina Grolinger. Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *IEEE Access*, 8:133982–133994, 2020.

[7] F. Ordoñez and D. Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 2016.

[8] W. Tao, M. C. Leu, and Z. Yin. Multi-modal recognition of worker activity for human-centered intelligent manufacturing. *Engineering Applications of Artificial Intelligence*, 2019.

[9] Kinam Kim and Yong Cho. Automatic recognition of workers' motions in highway construction by using motion sensors and long short-term memory (lstm) networks. *Journal of Construction Engineering and Management*, 147, 10 2020.

[10] S. Chatterjee et al. Accoustate: Auto-annotation of imu-generated activity signatures under smart infrastructure. *arXiv preprint arXiv:2112.06651*, 2021.

[11] Minsoo Park, Seongwoo Son, Yuntae Jeon, Dongyoung Ko, Mingeon Cho, and Seunghee Park. Scaffolding worker imu time-series dataset for deep learning-based construction site behavior recognition. *Advanced Engineering Informatics*, 65:103232, 2025.

[12] Kinam Kim and Yong K. Cho. Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition. *Automation in Construction*, 113:103126, 2020.

[13] Jinwoo Kim, Kyeongsuk Lee, and JungHo Jeon. Systematic literature review of wearable devices and data analytics for construction safety and health. *Expert Systems with Applications*, 257:125038, 2024.

[14] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. Temporal convolutional networks for action segmentation and detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

[15] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiaoli Li, and Shonali Krishnaswamy. Physical activity recognition with mobile phones: challenges, methods, and applications. *IEEE Signal Processing Magazine*, 32(4):52–59, 2015.

UAV Surveying for Emergency Facility Construction: Recent Advances in Data Collection and Processing

Congzhen Yang, Gengrong Zhang, Ke Chen

[1]School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China
yczhen2022@163.com, 405827337@qq.com, chenkecm@hust.edu.cn

Abstract –

This study provides a review of the literature regarding unmanned aerial vehicles (UAV) surveying for emergency facility construction, with a focus on data collection and processing methods. Synthesizing publications between 2016 and 2025, this study categorizes the applications of visible-light cameras, thermal infrared cameras, light detection and ranging (LiDAR), and multi-source sensing configurations. Discussion of data processing highlights key methodologies in image analysis, point clouds, and multimodal fusion, emphasizing semantic segmentation, three-dimensional (3D) reconstruction, and simultaneous localization and mapping (SLAM) as pivotal for transforming raw data into geospatial information. Reported applications focus on natural disaster scenarios and support tasks such as survivor search, route planning, and emergency facility site selection and assessment. These findings provide a basis for the future development of UAV-integrated emergency response systems.

Keywords –

UAV; Emergency Facility Construction; Data Collection; Data Processing

1 Introduction

The global increase in the frequency and severity of disasters, particularly large-scale flooding, earthquakes, and extreme weather events, has placed critical infrastructure and human lives at escalating risk. In the immediate aftermath of such events, priorities shift toward humanitarian relief and the rapid restoration of essential services. Central to these efforts is the timely construction of emergency facilities such as temporary shelters, logistics hubs, and provisional access routes [1]. Effective planning for these facilities requires high-accuracy geospatial information. Yet, conventional ground-based surveying techniques are infeasible in disaster contexts, where access is restricted, terrain is unstable, and time is of the essence [2]. Insufficient terrain mapping can lead to suboptimal or even hazardous facility placement, increasing exposure to secondary risks (e.g., renewed flooding, landslides), and delaying aid delivery.

Unmanned aerial vehicles (UAV) have been adopted for rapid disaster response [3]. Their deployment in Hurricane Wilma [4] and the Lushan earthquake in China [5] demonstrated the value of aerial assessments for identifying damaged areas, guiding rescue teams, and supporting situational awareness. Existing reviews have documented UAV contributions across disaster management, search and rescue, damage detection [6, 7], and civil infrastructure monitoring[8]. However, these applications emphasize visual inspection and reconnaissance rather than the acquisition of high-precision geospatial information. The transition from general aerial awareness to reliable mapping introduces challenges, including stringent accuracy requirements, robust sensor calibration, complex data processing, and the need to operate under hazardous environmental conditions [9].

Recent technological advances are beginning to mitigate these limitations. UAV equipped with high-resolution sensors now facilitate the rapid generation of Digital Elevation Models (DEMs), orthophotos, and other high-fidelity geospatial products [10]. Nevertheless, the literature on UAV-based mapping specifically tailored to emergency facility construction remains fragmented, with limited systematic synthesis of technologies, methodologies, and validated practices

This study addresses this gap by synthesizing research from the past decade, guided by two core research questions:

(1) What recent advances in UAV sensing techniques improve data collection accuracy and speed in challenging environments?

(2) What advanced data processing methods ensure the geometric accuracy required for emergency facility construction?

The remainder of this paper is structured as follows: Section 2 details the methodology for the literature selection and analysis. Sections 3 and 4 present recent advances in UAV-based data collection and processing, respectively. Section 5 summarizes the paper and

highlights research gaps for future work.

2 Literature Search and Bibliometric Analysis

This study employs bibliometric analysis to evaluate English-language literature concerning UAV-based mapping for emergency facility construction published between 2016 and 2025. The search targeted core databases across engineering and technology, geoinformation science, and disaster management. A multi-dimensional search strategy was implemented, integrating UAV-related terms (e.g., unmanned aerial vehicle, UAV, drone); emergency facility-related terms (e.g., emergency infrastructure, disaster infrastructure, temporary facility, relief camp, emergency facility); and disaster scenario-related terms (e.g., disaster relief, post-disaster, post-disaster construction). To ensure a comprehensive dataset, subject terms, keywords, and abstracts were cross-referenced, initially yielding over 100 publications. Following a rigorous screening process to exclude papers with weak relevance to "core UAV mapping technologies" or "practical applications," 93 publications were retained. These were analyzed using CiteSpace for keyword co-occurrence and clustering.

Figure 1 presents the keyword co-occurrence network. High-frequency terms such as “disaster management” and “disaster response” delineate the operational context of emergency facility construction. The network also indicates that deep learning and computational modeling are prominent technical approaches. Because emergency facility construction relies on high-precision geospatial information, deep learning is used for image recognition, target segmentation, and disaster-feature extraction, while computational modeling, mainly referring to three-dimensional (3D) reconstruction, supports terrain simulation and facility layout optimization.

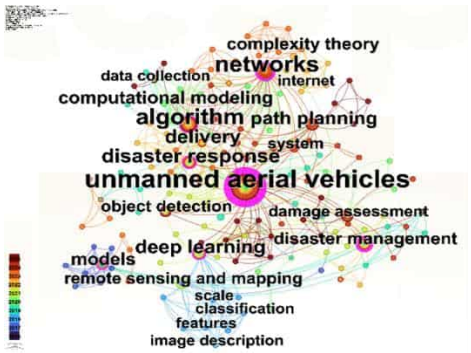


Figure 1. Keyword co-occurrence network

Using the log-likelihood ratio (LLR) algorithm in CiteSpace, seven clusters were identified (see Figure 2). The clustering results yield Q = 0.7455 (> 0.3) and S = 0.9473 (> 0.7), suggesting good separation among clusters and consistency within clusters. The data sensing cluster covers remote data collection and air–ground hybrid data collection, reflecting developments in data acquisition and sensor integration. The data processing cluster includes remote-sensing semantic segmentation and vision-based methods. The practical application cluster includes disaster rescue, security applications, and hazard-mitigation planning, reflecting the use of UAV in emergency facility construction.



Figure 2. Keyword clustering map

3 Data Collection techniques

Emergency facility construction requires survey data with high accuracy and short acquisition cycles, and the ability to accommodate changing disaster conditions. To meet these needs, UAV-based sensing techniques are used to collect different types of data.

3.1 Optical Sensing Technique

Optical sensors capture high-resolution spectral data essential for topographic mapping, feature identification, and damage assessment in emergency contexts.

3.1.1 Visible-light Camera

Visible-light cameras are widely used optical sensors. Eltner et al. [11] used a single UAV-mounted camera to capture optical imagery of small-scale topographic features, including slope micro-relief and debris-flow surface textures. Prokešová et al. [12] extended this approach to larger disaster settings by collecting imagery of landslide-prone areas. Galve et al. [13] investigated debris-flow and landslide sites along Spain’s A-7 highway using an UAV equipped with an ultra-high-resolution camera to collect multi-temporal imagery. To address challenges in pavement deformation measurement in landslide terrain, Cardenal et al. [14] used a mirrorless interchangeable-lens camera and

increased pixel density to support precise deformation measurements. However, visible-light imaging is sensitive to environmental interference. Atmospheric disturbances such as rain and fog can severely degrade image clarity and texture detail, or even obstruct imaging entirely. Furthermore, because visible-light cameras primarily capture planar data, they may not able to provide precise 3D spatial coordinates in occluded disaster environments.

3.1.2 Thermal Infrared Camera

Thermal infrared cameras detect the infrared radiation emitted by objects, translating temperature differentials into thermal signatures. This capability allows them to function in low-visibility conditions, including total darkness or light fog. Hoshino et al. [15] used thermal infrared imaging to exploit radiance differences between human bodies and surrounding backgrounds for victim detection and localization. This mitigates limitations of visible-light imaging in low light or smoke. Similarly, Dong et al. [16] mounted thermal cameras on UAV to collect data on victim postures and occlusions, which can improve survivor detection in earthquake rubble. Tavasoli et al. [17] demonstrated that thermal signatures can be used to map the spatial distribution of victims to prioritize emergency response. However, thermal infrared cameras typically feature lower spatial resolution than visible-light cameras.

3.2 Light Detection and Ranging (LiDAR)

Unlike optical sensors, LiDAR acquires dense 3D point clouds by emitting laser pulses and measuring returned signals. This provides 3D geometric detail along with spatial coordinates and distance measurements. LiDAR performance is largely independent of ambient lighting, making it suitable for nocturnal operations. Additionally, its multi-echo capabilities allow it to penetrate partial vegetation or debris cover, maintaining high ranging accuracy where optical sensors fail. Pellicani et al. [18] used LiDAR to collect high-density point clouds in landslide areas. Zhou et al.[19] used pre- and post-hurricane LiDAR point clouds to quantify damage to residential buildings, enabling assessment without additional contact with unstable structures during emergency operations. Despite these advantages, LiDAR systems remain high-cost, and their ranging precision can still be attenuated by extreme weather conditions.

3.3 Multi-Source Sensor Fusion

Table 1 summarizes the strengths and limitations of visible-light cameras, thermal infrared cameras, and LiDAR. In applications, these sensors are often combined to address the limitations of any single sensor.

Table1 Functions, limitations and factors of UAV-borne sensor types

Sensing type		Functions	Limitations	Factors
Optical sensor	Visible light camera	2D images	Poor adaptability to low-light and harsh weather; Only 2D information	Sunlight; Rainfall; Fog
	Thermal infrared camera	2D thermal radiation images	Low spatial resolution; No texture information	Heavy fog; Overcast and rainy; Severe dust
LiDAR sensor	LiDAR	3D spatial information	High cost; No spectral and thermal information	Heavy fog; Overcast and rainy
Multi-source sensor	Visible light camera+ Thermal infrared camera	2D texture and thermal radiation information	Dominated by 2D information; Limited data accuracy	Vegetation occlusion; Heavy fog
	Visible light camera+ LiDAR	3D spatial and 2D texture information	High cost; Poor adaptability to harsh weather	Heavy fog; Heavy rain
	Thermal infrared camera+ LiDAR	3D spatial and thermal radiation information	High cost; No texture information	Heavy fog; Overcast and rainy
	Visible light camera + Thermal infrared camera +LiDAR	3D spatial, 2D texture and thermal radiation information	Highest cost; Complex data fusion process	Extreme harsh weather (e.g., severe rainstorm and heavy fog)

Sensor fusion approaches combine sensors according to the data requirements of emergency surveys, disaster monitoring, and related tasks during raw data acquisition. Mineo et al. [20] used visible-light imagery

to reconstruct the 3D geometry and surface texture of landslides and integrated thermal infrared data to identify thermal anomalies in rock and soil. This combination supports characterization of landslide morphology while capturing subsurface- or surface-related thermal signals. Yu et al. [21] improved post-earthquake facility safety assessment by combining visible-light imagery of surface damage with LiDAR-derived 3D point clouds at millimeter-level accuracy. Cella et al. [22] used thermal infrared imaging to locate survivors and identify high-temperature hazards, while using LiDAR to generate 3D models to support spatial context in rescue operations. Shen et al. [23] integrated visible-light data for texture-based damage detection, thermal infrared data for thermal anomaly detection, and LiDAR for 3D geometric reconstruction, covering multiple information needs in post-disaster surveys. Overall, multi-source sensor fusion provides task-specific raw datasets that support subsequent data processing.

4 Data Processing Methods

Raw datasets acquired in disaster zones are typically voluminous, heterogeneous, and redundant. Data processing in this field focuses primarily on image and point cloud data.

4.1 Image Data Processing Methods

Processing of optical imagery emphasizes texture information and semantic attributes, and commonly uses deep learning to convert raw imagery into structured outputs. This section focuses on semantic segmentation and 3D reconstruction.

4.1.1 Deep Learning-based Image Semantic Segmentation

To improve semantic segmentation accuracy for UAV imagery, Zhou et al. [24] proposed a lightweight UAV frequency-aware and attention-enhanced network. Sui et al. [25] proposed the typical disaster bearing objects segmentation network semantic segmentation model, which can supported disaster target recognition in flood-related UAV imagery. To further enhance the specificity and precision of semantic segmentation, Stache et al. [26] combined a deep learning–based accuracy model with a UAV adaptive path-planning algorithm to link segmentation performance with data collection strategy, supporting decisions in emergency facility planning. However, the efficacy of these models remains contingent upon image quality, and the high computational overhead often hinders the real-time performance critical for immediate response.

4.1.2 3D Reconstruction using Structure from Motion (SfM)/Multiview Stereo (MVS)

With advances in computer vision, integrated SfM/MVS pipelines are widely used for 3D reconstruction from UAV imagery. These methods extract image feature points, estimate camera poses, generate sparse point clouds, and then perform dense reconstruction [27]. To mitigate computational latency, Zhang et al. [28] proposed an end-to-end latency optimization strategy that reduced modeling delay by about 40% while maintaining reconstruction accuracy. Wang et al. [29] used region-based parallel reconstruction and plane-detection optimization to improve efficiency and model completeness without requiring high-performance hardware. Despite these gains, SfM/MVS remains largely an offline, batch-processing approach, making it less suitable for incremental mapping in dynamic environments.

4.1.3 3D Reconstruction using Visual Simultaneous Localization and Mapping (SLAM)

Compared with SfM/MVS, visual SLAM supports real-time localization and incremental mapping, which aligns with the time constraints of emergency data processing [30]. To address accuracy and timeliness in post-disaster UAV surveying and mapping, Bi et al. [31] developed a UAV visual SLAM processing system based on ORB-SLAM2, improving both accuracy and real-time performance. Huo et al. [32] extended the workflow to include image correction and stitching, as well as obstacle detection. They further applied map fusion and path-planning algorithms to connect While visual SLAM aligns well with the time-sensitive nature of emergency response, it frequently struggles with feature-point matching in highly dynamic environments or areas with dense obstacles.

4.2 Point Cloud Data Processing Methods

Point clouds are the primary 3D data product collected by LiDAR, and their processing affects the accuracy of emergency surveying and mapping. This section reviews point cloud semantic segmentation and LiDAR-SLAM–based 3D reconstruction.

4.2.1 Deep Learning-based Point Cloud Semantic Segmentation

Deep learning methods for LiDAR point cloud semantic segmentation use neural networks to extract point-level features and assign category labels, providing semantic information for tasks such as 3D scene understanding and emergency mapping. Atik et al. [33] proposed the SegUNet3D ensemble approach, which converts point clouds into distance images and extracts

geometric features. Li et al. [34] addressed the trade-off between data complexity and processing efficiency in large-scale segmentation by proposing the slight filter learning network. However, a persistent challenge in this area is the heavy reliance on large, annotated datasets, which are often unavailable in unique disaster contexts.

4.2.2 LiDAR-SLAM–based 3D reconstruction

Compared with image-based reconstruction, LiDAR-SLAM does not require inferring 3D structure from 2D pixels. Instead, it uses the inherent spatial coordinates of point clouds to support geometric reconstruction and mapping [35]. Kaul [36] developed a compact quadcopter-based reconstruction method that generates dense point cloud maps using continuous-time SLAM. Real-time performance remains a constraint in many deployments. Mojtaba et al. [37] proposed a low-latency LiDAR-SLAM framework for incremental 3D reconstruction using lightweight point cloud segmentation and parallel matching, targeting high-dynamic, small-scale scenarios. Ćwian et al. [38] used geometric feature representations and factor-graph optimization to improve accuracy and efficiency in large-scale reconstruction. Together, these studies illustrate how LiDAR-SLAM pipelines can be adapted to different operational settings. However, positioning accuracy is often compromised in feature-sparse environments, as the scarcity of structural landmarks limits the algorithm's ability to maintain a stable spatial reference.

4.3 Processing Method for Fused Data

Image–point cloud fusion aims to combine complementary information: images provide texture and semantic cues, while point clouds provide 3D geometry. This section focuses on semantic fusion segmentation and SLAM-based 3D reconstruction using visual/LiDAR fusion.

4.3.1 Image–point cloud semantic fusion segmentation

Fusion-based segmentation methods vary by scenario and by the constraints of annotation, scale, and runtime. Dong et al. [39] proposed a parallel dual-stream (PD) network for semi-supervised semantic segmentation to reduce dependence on extensive labeled data in bimodal image–point cloud segmentation. Zhang et al. [40] proposed a deep learning-based image-point cloud fusion proposed an image–point cloud fusion scheme for large-scale 3D emergency scenarios to address segmentation error and missing information. Bultmann et al. [41] used a lightweight CNN with an embedded inference accelerator to integrate multimodal semantic cues into point cloud and image segmentation masks, producing global semantic maps suitable for time-constrained data processing. All these advances significantly improve the spatial resolution and reliability of semantic segmentation.

4.3.2 Visual/LiDAR SLAM–based 3D Reconstruction

By combining modalities, researchers can overcome the field-of-view limitations of LiDAR and the illumination sensitivity of vision [42]. Tong et al. [43] proposed a multi-sensor fusion framework that extends LiDAR-SLAM–based reconstruction to support emergency infrastructure surveying under complex conditions. Lou et al. [44] combined LiDAR with monocular vision to perform SLAM with 3D semantic reconstruction, addressing the limited semantic content of point clouds and supporting applications such as 3D surveying and post-disaster damage assessment. Consequently, this technology is well-suited for emergency infrastructure surveying and the assessment of building damage following seismic or fire events. However, the sophisticated fusion logic inherent in this technology imposes great demands on both algorithmic efficiency and hardware architectures.

5 Conclusion

This study applies bibliometric analysis to review research on UAV surveying and mapping for emergency facility construction published between 2016 and 2025. Using CiteSpace, keyword co-occurrence and clustering analyses were performed on 93 papers. The results indicate concentrations of research on sensor-based raw data acquisition and data processing methods based on deep learning and 3D reconstruction. For data acquisition, the review covers optical sensors (visible-light and thermal infrared cameras), LiDAR, and multi-sensor sensing/fusion configurations. For data processing, the literature centers on image data, point clouds, and multimodal fusion, with emphasis on semantic segmentation, 3D reconstruction, and SLAM-based approaches that convert raw data into geospatial information suitable for operational use.

The presented review focuses on sensors deployed on UAV platforms, related data processing methods, and their applications in emergency facility construction. Future work can assess endurance limitations and mitigation strategies to support sustained surveying and mapping in extended-duration operations, including landslide-prone and flood-affected areas.

6 Reference

[1] S.M.S. Mohd Daud, M.Y.P. M. Yusof, C. C. Heo, L.S. Khoo, M. K. C. Singh, M.S. Mahmood, and H. Nawawi. Applications of drone in disaster management: A scoping review. *Science & Justice*,

62: 30-42,2022

[2] W. G. William, P. L. Jerome, and D. Zekkos, Applications of UAVs in civil infrastructure. *Journal of Infrastructure Systems*, 25 :04019002, 2019.

[3] Y. Chen, X. Liu, B. Zhu, D. Zhu, X. Zuo, and Q. Li. UAV image-based 3D reconstruction technology in landslide disasters: A review. *Remote Sensing*, 17: 3117-3117, 2025.

[4] W. Miller, and D. L. Zhang. Tracing extreme updrafts in Hurricane Wilma (2005) to their boundary layer roots using trajectories: A thermodynamic and structural analysis. *Monthly Weather Review*, 148: 3605-3630, 2022.

[5] Y. S. Zhang, S. W. Dong, C. T. Hou, C. G. Guo, X. Yao, B. Li, J. J. Du, and J. G. Zhang. Geohazards Induced by the Lushan Ms7.0 Earthquake in Sichuan Province, Southwest China: Typical Examples, Types and Distributional Characteristics. *Acta Geologica Sinica (English Edition)*, 87: 646-657, 2013.

[6] N. Kerle, F. Nex, M. Gerke, D. Duarte, and A. Vetrivel. UAV-based structural damage mapping: A review. *ISPRS International Journal of Geo-information*, 9(1): 14, 2019.

[7] S. A. Shafian, and D. Hu. Integrating machine learning and remote sensing in disaster management: A decadal review of post-disaster building damage assessment. *Buildings*, 14: 2344-2344, 2024.

[8] A. Villarino, H. Valenzuela, N. Antón, M. Domínguez, and X. C. M. Cubillos. UAV applications for monitoring and management of civil infrastructures. *Infrastructures*, 10: 106-106, 2025.

[9] C.T. Recchiuto, and A. Sgorbissa. Post‐disaster assessment with unmanned aerial vehicles: A survey on practical implementations and research approaches. *Journal of Field Robotics*, 35: 459-490, 2018.

[10] B. Szypuła. Accuracy of UAV-based DEMs without ground control points. *GeoInformatica*, 28 :1-28, 2024.

[11] A. Eltner, A. Kaiser, C. Castillo, G. Rock, F. Neugirg, and A. Abellán. Image-based surface reconstruction in geomorphometry‐merits, limits and developments. *Earth Surface Dynamics*, 4: 359-389, 2016.

[12] R. Prokešová, M. Kardoš, and A. Medveďová. Landslide dynamics from high-resolution aerial photographs: A case study from the Western Carpathians, Slovakia. *Geomorphology*, 115: 90-101, 2010.

[13] J. P. Galve, J. L. Pérez-García, P. Ruano, J. M. G. López, C. R. Carmona, M. M. Sánchez, P.S. J. Longres, M. Ghadimi, A. Barra, R. M. Mateos, O. Monserrat, and J. M. Azañón. Applications of UAV digital photogrammetry in landslide emergency response and recovery activities: The case study of a slope failure in the A-7 highway (S Spain). *Landslides*, 22: 1383-1396, 2025.

[14] J. Cardenal, T. Fernández, J. L. P. García, and J. M. G. López. Measurement of road surface deformation using images captured from UAVs. *Remote Sensing*, 1507, 2019.

[15] W. Hoshino, J. Seo, and Y. Yamazaki. A study for detecting disaster victims using multi-copter drone with a thermographic camera and image object recognition by SSD. *2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, Pages.162-167, Delft, Netherlands, 2021.

[16] J. Dong, K. Ota, and M. Dong. UAV-based real-time survivor detection system in post-disaster search and rescue operations. *IEEE Journal on Miniaturization for Air and Space Systems*. 2(4): 209-219, 2021.

[17] S. Tavasoli, S. Poorghasem, X. Pan, T. Y. Yang, and Y. Bao. Autonomous post‐disaster indoor navigation and survivor detection using low‐cost micro aerial vehicles. *Computer-Aided Civil and Infrastructure Engineering*, 40: 130-144, 2024.

[18] R. Pellicani, I. Argentiero, P. Manzari, G. Spilotro, C. Marzo, R. Ermini, and C. Apollonio. UAV and airborne lidar data for interpreting kinematic evolution of landslide movements: The case study of the Montescaglioso Landslide (Southern Italy). *Geosciences*, 248, 2019.

[19] Z. Zhou, J. Gong, and X. Hu. Community-scale multi-level post-hurricane damage assessment of residential buildings using multi-temporal airborne LiDAR data. *Automation in Construction*, 98: 30-45, 2019.

[20] S. Mineo, D. Caliò, M. Intelisano, and G. Pappalardo. Landslide studying and monitoring by combining digital models from aerial visible and infrared photogrammetry. *Landslides*, 22: 1789-1804, 2025.

[21] R. Yu, P. Li, C. Liu, G. Zhai, and J. Shan. Post‐earthquake structural damage identification and safety evaluation using point clouds. *Structural Concrete*, 26: 5335-5353, 2024.

[22] M. Cella, J. Biber, P. Bournez, J. Salzinger, V. Widhalm, F. Vultaggio, F. d' Apolito, F. Bruckmüller, C. Sulzbachner, and P. F. Jende. ResQDrone: Autonomous indoor quadcopter for multi-layered mapping in crisis and disaster management scenarios. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 47-54, 2025.

[23] S. C. Y. Shen, Z. J. Li, S. Q. Zhou, X. L. Luo, and

Y. Zhang. Designing an efficient emergency response airborne mapping system with multiple sensors. *International Journal of Optics*, 2021, 2021.

[24] D. Zhou, H. Wang, S. Pang, D. Chen, H. Yao, and J. Yu. UAV-FAENet: Frequency-aware and attention-enhanced network for remote sensing semantic segmentation of UAV imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18: 1-17, 2025.

[25] Y. W. Sui, W. F. Shan, M. F. Wang, Y. Yang, and M. Y. Wang. TDBOSegNet: A semantic segmentation model for typical disaster-bearing objects in UAV images of flood disasters. *Canadian Journal of Remote Sensing*, 51: 19, 2025.

[26] F. Stache, J. Westheider, F. Magistri, M. Popović, and C. Stachniss. Adaptive path planning for UAV-based multi-resolution semantic segmentation. *Robotics and Autonomous Systems*, 159: 10488, 2023.

[27] M. Kholil, I. Ismanto, and M. N. Fu'ad, 3D reconstruction using Structure from Motion (SfM) algorithm and Multi View Stereo (MVS) based on computer vision. *IOP Conference Series: Materials Science and Engineering*, 1073(1): 012066, 2021.

[28] X. Zhang, M. Li, A. Hilton, A. Pal, S. Dey, and S. Debroy. End-to-End latency optimization of multi-view 3D reconstruction for disaster response. *2022 10th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, pages. 17-24, San Francisco, CA, USA, 2022.

[29] J.A. Wang, H.T. Ma, C.M. Wang, and Y.J. He. Fast 3D reconstruction method based on UAV photography. *ETRI Journal*. 40:788-793, 2018.

[30] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J.J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32: 1309-1332, 2016.

[31] Z. Wong, M. Zhang, and W. Bi. A UAV visual SLAM system for post-disaster emergency mapping. *2023 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2023)*, pages. 499-512, Springer Nature Singapore, Singapore, 2024.

[32] J. Huo, S.L. Zenkevich, A.V. Nazarova, and M. Zhai. Path planning based on map matching in UAV/UGV collaboration system. *International Journal of Intelligent Unmanned Systems*, 9: 81-95, 2019.

[33] M.E. Atik, and Z. Duran. An Efficient ensemble deep learning approach for semantic point cloud segmentation based on 3D geometric features and range images. *Sensors*, 22: 6210, 2022.

[34] X. Li, Z. Zhang, Y. Li, M. Huang, and J. Zhang. SFL-NET: Slight filter learning network for point cloud semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*. 61: 1-14, 2023.

[35] D. Droeschel, and S. Behnke. Efficient continuous-time SLAM for 3D Lidar-based online mapping. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages. 5000-5007, Brisbane, QLD, Australia, 2018.

[36] L. Kaul, R. Zlot, and M. Bosse. Continuous‐time three‐dimensional mapping for micro aerial vehicles with a passively actuated rotating laser scanner. *Journal of Field Robotics*. 33(1): 103-132, 2016

[37] K. Mojtaba, O. Martin, S. Oliver, B. Edwin, and S. Eckehard. LoLa-SLAM: Low-Latency LiDAR SLAM using continuous scan slicing. *IEEE Robotics and Automation Letters*. 6: 2248-2255, 2021.

[38] K. Ćwian, M.R. Nowicki, J. Wietrzykowski, and P. Skrzypczyński. Large-scale LiDAR SLAM with factor graph optimization on high-level geometric features. *Sensors*, 21(10): 3445, 2021.

[39] Q. Dong, J. Li, and S. Deng. Towards semi-supervised dual-modal semantic segmentation. *IEEE Transactions on Multimedia*. 27: 8212-8224, 2025.

[40] R. Zhang, G. Li, M. Li, and L. Wang. Fusion of images and point clouds for the semantic segmentation of large-scale 3D scenes based on deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 143: 85-96, 2018.

[41] S. Bultmann, J. Quenzel, and S. Behnke. Real-time multi-modal semantic fusion on unmanned aerial vehicles with label propagation for cross-domain adaptation. *Robotics and Autonomous Systems*. 159:104286, 2023.

[42] T. Shan, B. Englot, C. Ratti, and D. Rus. LVI-SAM: Tightly-coupled Lidar-visual-inertial odometry via smoothing and mapping. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages. 5692-5698, Xi'an, China, 2021.

[43] B. Tong, F. Jiang, B. Lu, Z. Gu, Y. Li, and S. Wang, Three-dimensional environment mapping with a rotary-driven Lidar in real time. *Sensors*, 25(15): 4870, 2025.

[44] L. Lou, Y. Li, Q. Zhang, H. Wei, SLAM and 3D semantic reconstruction based on the fusion of lidar and monocular vision, *Sensors*, 23(3):1502, 2023.

AI-Based Internal Structural Modeling Using Muon Tomography and Few-shot Parameter Calibration

Mudan Wang[1], Erika Parn[2], Sihan Liu[1,3], Borja García de Soto[2], Soheila Kookalani[1], Sander Sein[4], Tengiz Pataraia[4], Ioannis Brilakis[1]

[1]Department of Engineering, University of Cambridge, 7a JJ Thomson Avenue, Cambridge, CB3 0FA, United Kingdom
[2]S.M.A.R.T. Construction Research Group, New York University Abu Dhabi, PO 129188, Saadiyat Island, United Arab Emirates
[3]College of Civil Engineering, Tongji University, 1239 Siping Road, Shanghai, 200092, China
[4]GScan Ltd, The Bradfield Centre 184 Cambridge Science Park, Milton Road, Cambridgeshire CB4 0GA, United Kingdom

mw923@cam.ac.uk, eap9920@nyu.edu, sl2282@cam.ac.uk, garcia.de.soto@nyu.edu, sk2268@cam.ac.uk, sander.sein@gscan.eu, tengiz.pataraia@gscan.eu, ib340@cam.ac.uk

Abstract -

Reconstructing internal structural elements remains a significant challenge in the construction industry due to limited access to concealed components. Muon tomography provides a non-destructive approach for capturing information about internal structural conditions, yet its application in construction remains limited, especially in 3D model reconstruction. This paper proposes a few-shot parameter calibration–based framework for automated segmentation and reconstruction of internal building elements from muon tomography data, with a focus on reinforcement bars. The proposed method learns geometric and spatial patterns from a limited number of annotated datasets and generalizes to unseen configurations, reducing reliance on large labeled datasets. Experimental results demonstrate improved accuracy in rebar identification compared to existing segmentation approaches. The reconstructed rebar elements can be converted into BIM-ready representations, enabling applications such as defect detection, structural condition assessment, and digital twin development.

Keywords -

Few-shot geometric calibration; Muon tomography; Point cloud segmentation; Internal reconstruction; Scan-to-BIM

1 Introduction

Reconstruction of the internal condition of concrete structures is essential for ensuring structural safety, assessing durability, and supporting maintenance and retrofit decisions throughout the lifecycle of built assets [1, 2]. Internal features such as reinforcement layout, voids, and concealed damage directly influence structural performance, yet they are often undocumented or difficult to access in existing buildings and infrastructure [3]. As-built documentation may be incomplete, outdated, or unavailable, particularly for ageing structures, further increasing uncertainty in structural assessment and decision-making [4].

To address these challenges, internal concrete structure reconstruction has become an important research topic in construction informatics and structural health monitoring, aiming to infer concealed structural components from indirect sensing data [2]. A range of non-destructive testing techniques has been explored, including ultrasonic testing, electrical resistance tomography, and electromagnetic sensing, with demonstrated potential for detecting internal anomalies and reinforcement [3, 5]. However, many existing approaches are limited in penetration depth, sensitivity to material heterogeneity, or scalability to large and densely reinforced structures [4]. In addition, the resulting outputs are often image-based or qualitative, which restricts their direct use in engineering-oriented digital workflows.

With the growing adoption of digital twins and as-built Building Information Modeling (BIM) in the construction industry, there is an increasing demand for internal sensing methods that can provide volumetric information at structural scale and support downstream modeling and analysis tasks [6]. This has motivated interest in alternative imaging techniques capable of overcoming the physical and practical limitations of conventional non-destructive testing methods [2]. Currently, muon tomography has emerged as a promising solution, offering a fundamentally different inspection mechanism based on naturally occurring cosmic-ray muons [7]. Due to the high penetration capability of muons, this technique enables internal inspection of large and dense concrete structures with millimeter-level accuracy that are difficult to assess using traditional approaches [8, 9].

When applied to construction assets, muon tomography produces volumetric datasets that encode internal material distributions at structural scale [9]. However, these datasets are inherently complex, characterized by large spatial dimensions, indirect measurements, and uncertainty introduced during tomographic reconstruction [10]. Although recent studies have demonstrated the feasibility of applying muon tomography to reinforced concrete structures and other civil infrastructure [6], the interpretation of such data remains largely limited to qualitative visualization or density-based assessment [11]. System-

atic methods for transforming muon tomography outputs into structured, engineering-compatible representations, such as those required for BIM models, have received limited attention, highlighting an important gap between sensing capability and practical construction informatics applications [12].

This paper proposes a few-shot geometric calibration framework for processing muon tomography data to enable automated segmentation and reconstruction of internal building elements, with a focus on reinforcement bars. The method learns geometric priors from a limited number of annotated samples and generalizes to unseen configurations, reducing reliance on large labeled datasets. It is evaluated against existing instance segmentation approaches and demonstrates superior accuracy and robustness in rebar identification and reconstruction. The segmented elements are further converted into structured geometric representations suitable for BIM integration, supporting downstream applications such as structural assessment, defect detection, and digital twin development. The proposed framework advances muon tomography–based construction informatics and provides a practical pathway for as-built internal structure modeling.

2 Literature review

2.1 Internal Concrete structure reconstruction

Reconstructing the internal condition of concrete elements, including reinforcement layout and hidden defects, is fundamental to structural assessment and maintenance. Prior work has primarily relied on non-destructive testing and imaging techniques to infer internal features from indirect measurements. Among these, ultrasonic-based methods are widely adopted due to their practicality for field inspection. Early studies demonstrated the feasibility of ultrasonic testing and acoustic emission for characterizing damage in concrete elements [13] and summarized ultrasonic techniques for reinforced concrete evaluation [3]. To obtain volumetric representations, ultrasonic tomography has been developed and applied at structural scale, including full-scale reinforced concrete inspection [1] and broader reviews of nondestructive acoustic tomography methods for concrete [2].

A major challenge in ultrasonic tomography lies in the ill-posed inverse problem, which has motivated research into improved reconstruction algorithms and regularization. Iterative reconstruction approaches were explored early for ultrasonic imaging in concrete [14]. Subsequent work adopted algebraic reconstruction techniques and iterative solvers to enhance stability and image quality for reinforced concrete inspection [15]. More recently, network-based strategies have been investigated to improve ultrasonic tomography performance [16], while group sparsity has been introduced to reconstruct velocity field images of internal bridge structures [17]. In parallel, methodological advances have been proposed to refine the reconstruction mesh and improve ultrasonic tomography fidelity [18]. A systematic review further confirms the active development of reconstruction strategies and highlights persistent limitations in resolution, robustness, and deployment in complex concrete environments [19].

Beyond ultrasound, complementary modalities have been explored for internal concrete feature detection. Electrical resistance tomography has been used to locate reinforcement and visualize damage-related conductivity changes [4, 20]. Electromagnetic approaches such as microwave imaging have also been investigated for reinforced concrete inspection [5]. For high-resolution material characterization, X-ray and neutron tomography have been applied to reconstruct concrete microstructures and internal features [21], while CT-based 3D reconstruction has been used to analyze pore structures in concrete-related materials [22]. These approaches demonstrate strong imaging potential, but may be constrained by equipment requirements, data complexity, or applicability to field-scale construction scenarios.

Recent studies have also introduced learning-based components to improve internal condition detection, including deep learning applied to ultrasound arrays for two-step detection of internal concrete conditions [23]. However, across sensing modalities, existing work typically outputs voxel-level images or qualitative visualizations rather than explicit, engineering-ready representations of structural components. This creates a gap between internal imaging and downstream construction applications, particularly for workflows requiring component-level reconstruction and integration into digital models.

2.2 Muon tomography technique and construction applications

Muon tomography, also referred to as muography, is a passive non-destructive imaging technique that exploits naturally occurring cosmic-ray muons to investigate the internal structure of large and dense objects [7, 10]. Due to their high energy and strong penetration capability, muons can traverse thick materials such as concrete, rock, and steel, making them suitable for imaging structures that are inaccessible to conventional sensing methods [7, 9]. As muons pass through matter, their trajectories are altered through attenuation and multiple Coulomb scattering, with the degree of deviation dependent on the density and atomic composition of the traversed material [10]. By measuring the incoming and outgoing muon tracks using dedicated detector systems, three-dimensional density distributions of the inspected object can be reconstructed via tomographic inversion techniques [9, 7]. These characteristics have established muon tomography as a robust imag-

ing modality in domains such as nuclear material monitoring, geological exploration, and underground structure inspection, providing a solid technical foundation for its extension to large-scale construction and civil infrastructure applications [7, 10].

In the context of construction and civil infrastructure, muon tomography has recently emerged as a promising technique for inspecting reinforced concrete structures that are difficult to assess using conventional non-destructive testing methods [8]. Early proof-of-concept studies demonstrated the feasibility of imaging the interior of reinforced concrete elements using muon tomography, showing that reinforcement layouts and internal features could be detected with a quality comparable to, and in some cases exceeding, that of ultrasonic and radar-based methods [8, 11]. These studies highlight muon tomography's potential to overcome the penetration depth limitations of traditional techniques when applied to large or densely reinforced concrete components [24].

More targeted investigations have focused on the detection and localization of reinforcement bars within concrete using muon scattering tomography. Simulation-based studies have shown that muon tomography can detect rebars of practical dimensions and distinguish between single and multiple reinforcement layers embedded at different depths [24]. Subsequent experimental and methodological developments further validated the capability of muon scattering tomography to identify reinforcement meshes within large-scale concrete objects and to resolve their relative spatial positions [11, 12]. These works demonstrate the viability of muon tomography for reinforcement detection and highlight its relevance for assessing ageing infrastructure where as-built documentation may be incomplete or unavailable [12].

Beyond reinforced concrete inspection, muon tomography has also been explored for broader infrastructure monitoring applications, including the structural health assessment of large civil assets such as dams. Field deployments have shown that muon-based density imaging can reveal internal deterioration and material loss within massive concrete structures, supporting its potential role as a complementary tool for long-term structural health monitoring [6, 7]. While these studies primarily focus on sensing feasibility and density imaging, they collectively establish muon tomography as a powerful technique for construction-scale internal inspection [9].

Overall, existing research confirms the potential of muon tomography for construction applications, particularly for imaging internal conditions of large reinforced concrete structures. However, most studies emphasize sensing principles, detector development, and qualitative density reconstruction [10]. The extraction of explicit structural components and their integration into engineering-oriented workflows, such as as-built modeling and digital twins, remains largely unexplored, motivating further research on advanced interpretation and reconstruction methods for muon tomography data [12, 6].

2.3 Research gaps

Existing non-destructive inspection methods for reinforced concrete (e.g., ultrasonic and electrical resistance tomography) support internal inspection, but their outputs remain image- or voxel-level and are difficult to convert into component-based models for scan-to-BIM and digital twin workflows [1, 25]. Muography offers high penetration for imaging dense reinforced concrete, yet current construction-oriented studies mainly focus on sensing feasibility and qualitative density reconstruction rather than automated extraction of discrete reinforcement elements for BIM-ready modeling [8, 12].

Few-shot segmentation enables recognition under limited labeled data, typically via meta-learning, but such paradigms rely on large pretraining datasets and are sensitive to domain shifts, making them less applicable to sparse and geometry-driven muon tomography data. Therefore, this study adopts a few-shot geometric calibration strategy, using limited annotated rebar exemplars to extract geometric priors and calibrate segmentation parameters for data-efficient and interpretable reinforcement extraction.

Meanwhile, conventional clustering or model-fitting methods may fragment or merge rebars in dense layouts, while fully supervised learning remains constrained by scarce labeled muon datasets. Thus, approaches that leverage geometric priors and limited supervision are needed to robustly extract reinforcement components and support BIM-oriented internal structural modeling.

3 Proposed method

This section presents an AI-enabled framework for internal structure reconstruction from muon tomography data, designed to enable robust segmentation and BIM-ready representation of internal structural elements, with a focus on reinforcement bars. Figure 1. provides an overview of the proposed framework workflow.

The proposed few-shot parameter calibration method identifies reinforcement bars from a large point cloud by leveraging a small number of labeled rebar examples as geometric references rather than relying on extensive supervised training. Each few-shot sample is first analyzed to estimate its principal axis, radius, length, and straightness using PCA-based geometric analysis. When the number of labeled points is insufficient, the sample is automatically augmented by retrieving geometrically consistent points from the full point cloud. These augmented few-shot examples are then used to compute representative



Figure 1. Enhanced workflow diagram of the AI-enabled framework

statistics, including typical rebar radius, axial variation, local orientation deviation, and radial noise. The proposed few-shot geometric calibration method identifies rebars in large point clouds using a small number of labeled rebar exemplars as geometric references rather than extensive supervised training. Each exemplar is analyzed via PCA-based geometric fitting to estimate axis, radius, length, and straightness, and is augmented with geometrically consistent points when necessary. Representative geometric statistics are then computed to derive calibration parameters controlling neighborhood size, angular consistency, radial deviation, and minimum rebar length. Using these calibrated parameters, a region-growing strategy is applied to segment rebar clusters, with seed selection based on local linearity and iterative growth under spatial and geometric constraints. Cluster geometry is periodically updated to refine axis and radius estimates. A subsequent fill-in step recovers missed but consistent points, and clusters are validated by size, straightness, and length criteria before extracting geometric descriptors for BIM-compatible output.

4 Implementation details

4.1 Dataset preparation

In this study, a simulated approach was first employed to generate volumetric data representing the internal structure, which were stored in the NIfTI (.nii.gz) format. The simulated data were subsequently processed using software provided by GSCAN, through which structural components, including reinforcement and concrete, were identified and encoded in the volumetric representation. The resulting .nii.gz files contain three-dimensional voxel-based data that describe material distributions within the structure.

The NIfTI data are characterized by high dimensionality and volumetric continuity, making them well suited for representing internal structural conditions but less convenient for direct geometric analysis. To facilitate subsequent processing and reconstruction tasks, the volumetric data were therefore converted into point cloud representations, enabling more efficient handling and geometric-based analysis in later stages of the workflow.

The experimental dataset consists of point cloud data representing reinforced concrete elements with irregular rebar layouts. The point clouds were acquired through simulated scanning generation, capturing variations in rebar diameter, curvature, spacing, and overlap. Ground truth geometric information was used only for evaluation purposes and not for training. Figure 2 presents representative reconstruction results obtained using the proposed framework..

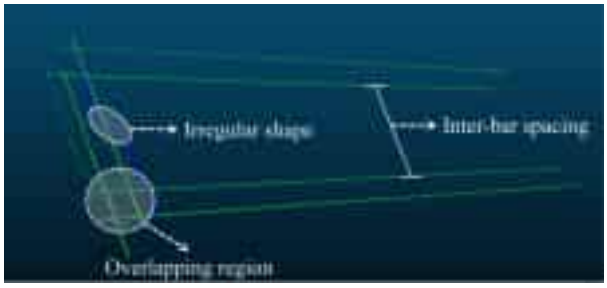


Figure 2. Labeled rebar data for training

4.2 Results of the proposed methods

This subsection presents the results of rebar segmentation from raw point cloud data. The proposed method effectively isolates rebar points from surrounding concrete and noise, even in cases of complex intersections and curved geometries. Quantitative and qualitative results demonstrate stable segmentation performance across different rebar configurations. Figure 3 shows representative examples of challenging rebar layouts, including dense arrangement, local overlap, and irregular spacing.

It is obvious that the segmentation results produced by the proposed AI-enabled few-shot calibrated method. Each reinforcement bar is identified as an individual cluster and visualized using a distinct color, demonstrating the method's ability to properly segment closely spaced rebars within a dense internal reinforcement layout.

As shown in the figure, the proposed approach successfully separates individual rebars even in regions with strong spatial proximity and complex intersections.

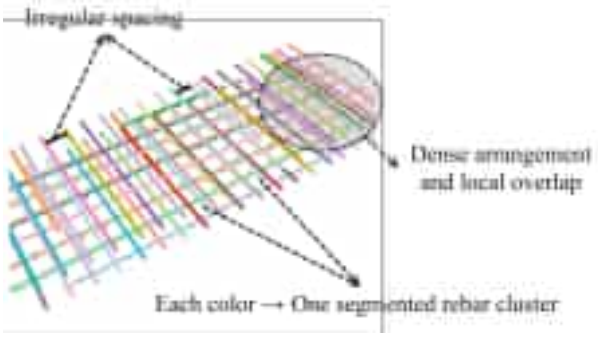


Figure 3. Segmentation results of using the proposed AI-enabled method

Importantly, rebars with irregular geometries and nonlinear connections are also correctly identified as continuous structural elements rather than being fragmented or merged with neighboring bars. This indicates that the geometry-calibrated region growing process effectively preserves rebar continuity while maintaining clear separation between adjacent elements.

Overall, the results confirm that the proposed method can (1) accurately segment individual rebars from complex internal point cloud data and (2) reliably identify connected rebars with irregular shapes, providing structured outputs suitable for downstream applications such as BIM reconstruction, defect detection, and structural condition assessment.

4.3 Comparison with other segmentation methods

This research further compares the proposed method with commonly used instance segmentation algorithms for individual rebar identification, including DBSCAN, cylinder-model fitting, and region growing segmentation.

As shown in Figure 4, DBSCAN detects a large number of rebar instances, with 220 rebars identified. However, the results exhibit severe over-segmentation. Individual rebars are frequently divided into multiple clusters, and the algorithm fails to identify connected rebars with complex or irregular geometries. This behavior is mainly due to DBSCAN's sensitivity to point density and distance thresholds, which limits its robustness in dense reinforcement layouts.

Figure 5 presents the results obtained using cylinder-model fitting, where only 36 rebars are detected. Many rebars are missed due to the strict geometric assumptions of the cylinder model, and connected rebars cannot be correctly identified. This approach is particularly sensitive to incomplete or noisy data and struggles to represent irregular rebar shapes.

As illustrated in Figure 6, region growing segmentation detects 56 rebars but suffers from significant fragmentation, especially for vertical rebars. While region growing partially captures linear structures, it fails to maintain rebar continuity and cannot reliably identify connected rebars, leading to incomplete and fragmented segmentation results.

Overall, the comparison demonstrates that existing instance segmentation methods either over-segment rebars or fail to capture their connectivity and irregular geometry. In contrast, the proposed few-shot calibrated method achieves accurate individual rebar segmentation while preserving connected and irregular rebar structures, highlighting its robustness and suitability for construction-oriented applications.



Figure 4. Segmentation results of using DBSCAN

5 Results and discussion

The experimental results demonstrate clear performance differences between the proposed method and commonly used instance segmentation algorithms for rebar identification, highlighting the challenges of density-based, model-based, and purely geometric approaches when applied to complex internal reinforcement layouts. DBSCAN detects a large number of rebar instances but suffers from severe over-segmentation. Individual rebars are frequently fragmented into multiple clusters, particularly under varying point density or close spatial proximity, preventing reliable identification of connected rebars and irregular geometries. Cylinder-model fitting exhibits the opposite behavior, detecting only a small subset of rebars. The method relies on strong geometric assumptions and is sensitive to incomplete data and deviations

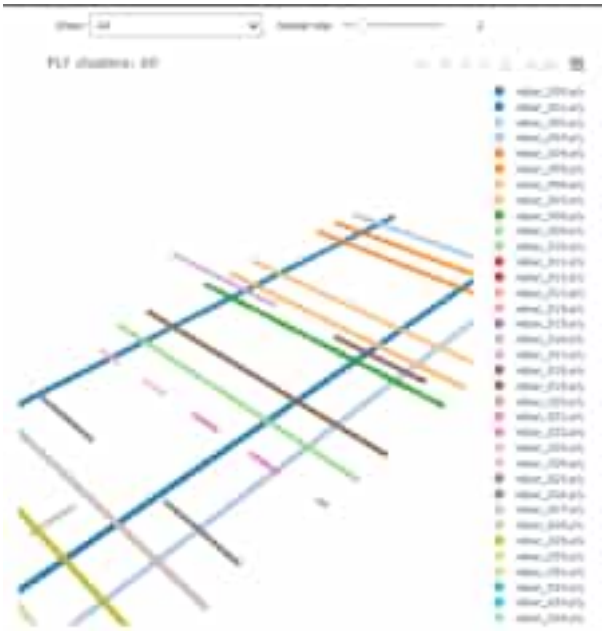


Figure 5. Segmentation results of using cylinder-model fitting

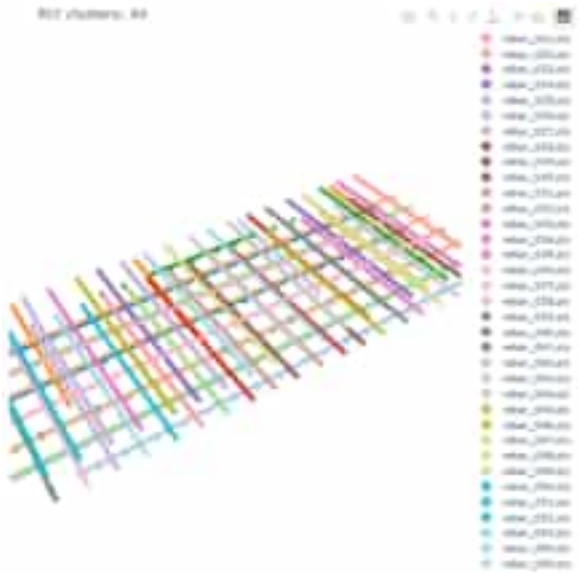


Figure 6. Segmentation results of using the region growing segmentation

from ideal cylindrical shapes, leading to missed rebars, under-segmentation, and incomplete internal structure reconstruction. Region growing achieves moderate detection performance but produces fragmented results, especially for vertical rebars. Although linear features are partially captured, the absence of calibrated geometric constraints prevents reliable identification of continuous and connected rebars. In contrast, the proposed few-shot calibrated method achieves accurate individual rebar segmentation while preserving continuity and connectivity, including irregular rebars embedded in structural components. This capability supports internal reinforcement modeling and downstream applications such as structural condition assessment, defect detection, retrofitting planning, as-built BIM generation, and digital twin development.

Several limitations of this study should be noted. First, standard quantitative instance segmentation metrics (e.g., IoU, precision, recall, and F1-score) were not reported, as no benchmark dataset was available for this evaluation. The current assessment is therefore based on manual observation and qualitative analysis. The current datasets are generated through simulation-based muon tomography, and robustness under real acquisition noise and reconstruction artifacts remains to be validated. Experiments were conducted on a limited range of reinforcement layouts and reconstruction settings; broader tests on varying densities and resolutions will be explored. Finally, while scan-to-BIM conversion is demonstrated, detailed evaluation of geometric parameter extraction and IFC/BIM generation accuracy will be investigated in future work.

6 Conclusions

Reconstructing internal structural elements remains a significant challenge in the construction industry due to the limited accessibility of concealed components. While muon tomography offers a promising non-destructive means of capturing internal structural information, its practical use in construction has been constrained by the complexity of the data and the lack of effective processing methods. Addressing this gap, this paper proposed a few-shot parameter calibration–based framework for the automated segmentation and reconstruction of internal building elements from muon tomography data, with a particular focus on reinforcement bars.

By leveraging a small number of labeled examples to calibrate geometry-driven segmentation, the proposed method generalizes effectively to unseen rebar configurations and reduces reliance on large annotated datasets. Comparative experiments demonstrate that the proposed approach achieves improved accuracy and continuity in rebar identification compared to commonly used segmentation methods. Furthermore, the reconstructed rebars are directly convertible into BIM-ready representations, enabling practical applications such as defect detection, structural condition assessment, retrofitting planning, and digital twin development for existing structures.

7 Acknowledgments

This research is part of the "Muon Tomography for Assessing Bridge Structures" (M-TABS) project, funded by Innovate UK (Project Ref: 10132366). The authors would like to thank their project partner GScan Ltd for their collaboration and support during this study. This research is partially supported by by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013.

References

[1] Hajin Choi and John S. Popovics. Nde application of ultrasonic tomography to a full-scale concrete structure. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 62(6):1076–1085, JUN 2015. ISSN 0885-3010. doi:10.1109/TUFFC.2014.006962.

[2] A. V. Kozlov and V. N. Kozlov. The development and current state of methods for the nondestructive testing and acoustic tomography of concrete. *Russian Journal of Nondestructive Testing*, 51(6):329–337, JUN 2015. ISSN 1061-8309. doi:10.1134/S1061830915060054.

[3] M. Schickert and M. Krause. Ultrasonic techniques for evaluation of reinforced concrete structures. In C Maierhofer, HW Reinhardt, and G Dobmann, editors, *Non-Destructive Evaluation of Reinforced Concrete Structures, Volume 2: Non-Destructive Testing Methods*, Woodhead Publishing in Materials, pages 490–530. 2010. ISBN 978-1-84569-960-4; 978-1-84569-950-5. doi:10.1533/9781845699604.2.490.

[4] K. Karhunen, A. Seppanen, A. Lehikoinen, J. P. Kaipio, and P. J. M. Monteiro. Locating reinforcing bars in concrete with electrical resistance tomography. In MG Alexander, HD Beushausen, F Dehn, and P Moyo, editors, *Concrete Repair, Rehabilitation and Retrofitting II*, pages 229+, 2009. ISBN 978-0-415-46850-3. 2nd International Conference on Concrete Repair, Rehabilitation and Retrofitting, Cape Town, SOUTH AFRICA, NOV 24-26, 2008.

[5] Shengxiang Qi, Hui Xu, and Jian Ren. Microwave imaging of reinforced concrete and design of a broadband antenna. In *Proceedings of 2017 International Conference on Video and Image Processing (Icvip 2017)*, pages 223–227, 2017. ISBN 978-1-4503-5383-0. doi:10.1145/3177404.3177424. International Conference on Video and Image Processing (ICVIP), Singapore, SINGAPORE, DEC 27-29, 2017.

[6] Laszlo Olah, Hiroyuki K. M. Tanaka, Toshio Mori, Yukihiko Sakatani, and Dezso Varga. Structural health monitoring of sabo check dams with cosmic-ray muography. *iScience*, 26(10), OCT 20 2023. doi:10.1016/j.isci.2023.108019.

[7] H. Gomez. Muon tomography using micromegas detectors: From archaeology to nuclear safety applications. *Nuclear Instruments & Methods in Physics Research Section A-Accelerators Spectrometers Detectors and Associated Equipment*, 936:14–17, AUG 21 2019. ISSN 0168-9002. doi:10.1016/j.nima.2018.10.011.

[8] Ernst Niederleithinger, Simon Gardner, Thomas Kind, Ralf Kaiser, Marcel Grunwald, Guangliang Yang, Bernhard Redmer, Anja Waske, Frank Mielentz, Ute Effner, Christian Koepp, Anthony Clarkson, Francis Thomson, Matthew Ryan, and David Mahon. Muon tomography of the interior of a reinforced concrete block: First experimental proof of concept. *Journal of Nondestructive Evaluation*, 40(3), SEP 2021. ISSN 0195-9298. doi:10.1007/s10921-021-00797-3.

[9] Ignacio Lazaro Roche. A compact muon tracker for dynamic tomography of density based on a thin time projection chamber with micromegas readout. *Particles*, 4(3):333–342, SEP 2021. doi:10.3390/particles4030028.

[10] H. Gomez, D. Attie, D. Calvet, X. Coppolani, B. Gallois, M. Kebbiri, M. Lehuraux, P. Magnier, I. Mandjavidze, S. Procureur, and M. Vandenbroucke. Muon tomography with micromegas: Archaeology, nuclear safety and new developments for geotechnics. In *9th International Symposium on Large TPCs for Low-Energy Rare Event Detection*, volume 1312 of *Journal of Physics Conference Series*, 2019. doi:10.1088/1742-6596/1312/1/012013. 9th International Symposium on Large TPCs for Low-Energy Rare Event Detection, Paris Diderot Univ, Inst Astroparticle Phys Campus, Paris, FRANCE, DEC 12-14, 2018.

[11] Magdalena Dobrowolska, Jaap Velthuis, Anna Kopp, Chiara De Sio, Ruaridh Milne, and Philip Pearson. Development of muon scattering tomography for a detection of reinforcement in concrete. *Engineering Research Express*, 3(3), SEP 1 2021. ISSN 2631-8695. doi:10.1088/2631-8695/ac21a8.

[12] M. Dobrowolska, J. J. Velthuis, C. De Sio, A. Kopp, R. Milne, and P. Pearson. Development of muon scattering tomography for detection of reinforcement

in concrete. *Journal of Instrumentation*, 17(9), SEP 2022. doi:10.1088/1748-0221/17/09/C09011.

[13] Dimitrios G. Aggelis, Tomoki Shiotani, Shouhei Momoki, and Akinobu Hirama. Acoustic emission and ultrasound for damage characterization of concrete elements. *ACI Materials Journal*, 106(6):509–514, NOV-DEC 2009. ISSN 0889-325X.

[14] Xi Fang and Zhuo-qiu Li. Study on iterative regularization method in ultrasonic imaging reconstruction for concrete's testing. In *2009 International Conference on Information Technology and Computer Science, Vol 1, Proceedings*, pages 213–216, 2009. ISBN 978-0-7695-3688-0. doi:10.1109/ITCS.2009.49. International Conference on Information Technology and Computer Science, Kiev, UKRAINE, JUL 25-26, 2009.

[15] Kevin Soetomo, Talitha F. Rahma, Endang Juliastuti, and Deddy Kurniadi. Ultrasonic tomography for reinforced concrete inspection using algebraic reconstruction technique with iterative kaczmarz method. In *Proceedings of the 2016 International Conference on Instrumentation, Control, and Automation (ICA)*, International Conference on Instrumentation Control and Automation, pages 16–21. Inst Teknologi Bandung, Fac Ind Technol, Instrumentat & Control Res Grp; IEEE Indonesia Sect, 2016. ISBN 978-1-5090-1335-7. International Conference on Instrumentation, Control and Automation (ICA), Bandung, INDONESIA, AUG 29-31, 2016.

[16] Lourenco Panosso Perlin and Roberto Caldas de Andrade Pinto. Use of network theory to improve the ultrasonic tomography in concrete. *Ultrasonics*, 96:185–195, JUL 2019. ISSN 0041-624X. doi:10.1016/j.ultras.2019.01.007.

[17] Jian Li, Jin Li, Chenli Guo, Hongtao Wu, Chuankun Li, Rui Liu, and Lujun Wei. Method for reconstructing velocity field images of the internal structures of bridges based on group sparsity. *Electronics*, 13(22), NOV 2024. doi:10.3390/electronics13224574.

[18] Lara Guizi Anoni and Vladimir Guilherme Haach. Mesh increment methodology for improving concrete ultrasonic tomography. *Journal of Computing in Civil Engineering*, 40(1), JAN 1 2026. ISSN 0887-3801. doi:10.1061/JCCEE5.CPENG-6872.

[19] Lara Guizi Anoni, Vladimir Guilherme Haach, and Lev Khazanovich. Image reconstruction in concrete ultrasound tomography: A systematic review. *Construction and Building Materials*, 441, AUG 30 2024. ISSN 0950-0618. doi:10.1016/j.conbuildmat.2024.137472.

[20] Chenhui Su, Wenchao Zhang, Jianyu Zhao, Hongjuan Yang, and Chengdong Li. Research on damage visualization of concrete structures based on electrical resistance tomography. *Frontiers In Physics*, 10, SEP 5 2022. doi:10.3389/fphy.2022.993260.

[21] R. D. F. Tiana Razakamandimby, Hyeung-Tae Kim, and Kyoungsoo Park. Characterization of 3d microstructure of concrete using reconstructed combined x- ray and neutron tomography image. In R Ghai, LM Chang, R Sharma, and AK Chandrappa, editors, *Sustainable Design and Eco Technologies For Infrastructure, Cecar 9 2022*, volume 441 of *Lecture Notes in Civil Engineering*, pages 433–438, 2024. ISBN 978-981-99-8467-1; 978-981-99-8465-7; 978-981-99-8464-0. doi:10.1007/978-981-99-8465-7_30. 9th Civil Engineering Conference in the Asian Region, INDIA, SEP 21-23, 2022.

[22] Sizhe Du, Yu Zhang, Jie Zhang, Nina Selyutina, Ivan Smirnov, Gang Ma, Xiang Zhang, Beibei Li, Yanchun Miao, Yuanzhen Liu, and Wenjing Wang. Study on pore characteristics of recycled aggregate concrete mixed with glazed hollow beads at high temperatures based on 3-d reconstruction of computed tomography images. *Construction and Building Materials*, 323, MAR 14 2022. ISSN 0950-0618. doi:10.1016/j.conbuildmat.2022.126564.

[23] Yanqi Wu, Yongping Wang, Dan Li, and Jian Zhang. Two-step detection of concrete internal condition using array ultrasound and deep learning. *NDT E International*, 139, OCT 2023. ISSN 0963-8695. doi:10.1016/j.ndteint.2023.102945.

[24] Magdalena Dobrowolska, Jaap Velthuis, Anna Kopp, Marcus Perry, and Philip Pearson. Towards an application of muon scattering tomography as a technique for detecting rebars in concrete. *Smart Materials and Structures*, 29(5), MAY 2020. ISSN 0964-1726. doi:10.1088/1361-665X/ab7a3f.

[25] Kimmo Karhunen, Aku Seppänen, Anssi Lehikoinen, Paulo J.M. Monteiro, and Jari P. Kaipio. Electrical resistance tomography imaging of concrete. *Cement and Concrete Research*, 40(1):137–145, 2010. ISSN 0008-8846. doi:10.1016/j.cemconres.2009.08.023.

Vision-Enabled Supervisory Feedback Loop for Extrusion Failure Mitigation in Robotic Additive Manufacturing

Hendrik Benz[1], The Vinh Nguyen Trong[1], Kunaljit Chadha[2], Massimo Visonà[3], and Katharina Klemt-Albert[1]

[1]Institute for Construction Management, Digital Engineering and Robotics in Construction, RWTH Aachen University, Aachen, Germany
[2]Gramazio Kohler Research, ETH Zürich, Zürich, Switzerland
[3]Erratic, Valdagno, Italy

benz@icom.rwth-aachen.de

Abstract -

Extrusion-based robotic additive manufacturing for construction-scale fabrication is highly sensitive to process faults such as unstable flow, clogging, and fluctuations in feed supply. Without continuous monitoring, these failures can cause prolonged non-deposition and geometric error. To address this limitation, this work presents a vision-guided supervisory framework that couples nozzle-centric process perception with controller-level intervention on an industrial robot. Visual process feedback is obtained from an eye-in-hand RGB-D camera, whose observations are processed online by an instance-segmentation model to classify extrusion behavior. In parallel, motion and process data are retrieved from the robot controller, establishing a common temporal basis between sensed process state and executed robot motion. Upon detection of persistent extrusion failure, the supervisory layer adjusts a bounded program override to reduce or halt robot motion, thereby constraining air-printing duration and facilitating prompt intervention. A proof-of-concept demonstration on an industrial robotic extrusion platform compares nominal reference prints with an induced pump-off failure event in which extrusion loss is detected online and followed by controller action. The results show a practical route toward process-aware supervision without invasive controller modifications and provide synchronized logs for future learning-based methods.

Keywords -

Additive manufacturing; Computer vision; In-process Monitoring; Feedback-Loop; KUKAVarProxy (KVP)

1 Introduction

Additive manufacturing (AM) is increasingly explored for large-scale construction, yet robust and repeatable printing remains challenging due to the interplay of multiple time-coupled sub-processes [1]. These challenges become more pronounced for rheology-sensitive and non-cementitious feedstocks, where process stability depends strongly on material consistency and environmental conditions [2]. Motivated by reducing embodied emissions of cement-intensive construction, earth-based materials are receiving growing attention as printable alternatives, commonly discussed under the emerging framework of 3D Printed Earth (3DPE) [3, 4]. However, printable mixtures in 3DPE can be highly sensitive to soil composition and moisture content, which affects fresh-state extrudability and build-up performance [5]. Consequently, intermittent flow, clogging, and bonding-related defects can occur during long-duration prints and may remain unnoticed without continuous monitoring [6]. For construction-scale extrusion AM, prior work further highlights that quality assessment should be performed in-process and in-line across multiple scales rather than relying on a single indicator [7].

To overcome the limited availability of practical supervision strategies for extrusion-based robotic printing, this work presents a vision-enabled supervisory loop that detects extrusion state online and triggers controller-side responses to reduce the risk of failing prints. Extending earlier in-process monitoring approaches, this study focuses on integrating near-nozzle RGB-D perception with controller-side intervention in active printing. Robot state variables are streamed from the KUKA KRC4 controller of a KUKA KR210 R3100 via KUKAVarProxy (KVP) and temporally aligned with extrusion-state estimates derived from an eye-in-hand Intel RealSense D405 RGB-D camera. A proof-of-concept experiment demonstrates that detected extrusion-loss events can be followed by adaptive reduction of the program speed override, slowing or pausing motion to enable recovery before major print degradation accumulates. The main contributions are:

- **KVP supervisory interface:** Time-stamped streaming of robot state and read/write access to selected robot variables.
- **Vision-based extrusion monitoring:** Confidence-based extrusion-state detection.
- **Fail-safe intervention policy:** Detect–slow/stop

program-override strategy to mitigate air-printing.

- **Loop-level characterization:** Measurement of detection delay and end-to-end latency (perception → KVP write → override effect) plus qualitative failure-mitigation outcomes.
- **User interface:** Runtime supervisory UI for tuning τ and logging.

As shown in Fig. 1, the experimental platform comprises a pump-fed material supply and a nozzle end-effector mounted on an industrial robot. An eye-in-hand vision sensor observes the nozzle and immediate deposition zone to enable in-process monitoring of extrusion continuity during execution.

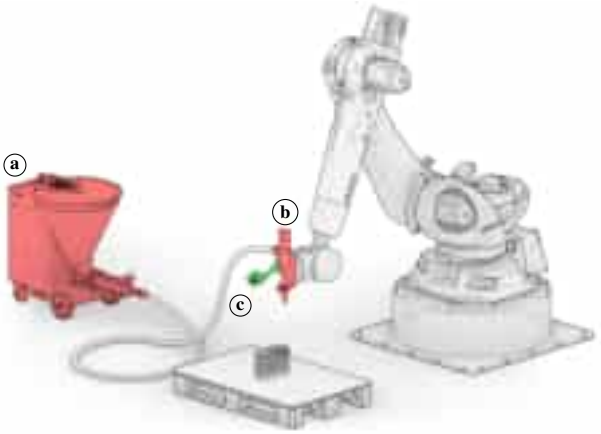


Figure 1. Robotic AM setup: (a) feeder/pump, (b) end-effector/ extruder, (c) eye-in-hand vision sensor.

2 Related work

Large-scale AM has evolved from geometry- and material-centric studies toward integrated pipelines that combine sensing, interpretation, and intervention. This section reviews (1) quality assessment and in-process monitoring for extrusion-based AM and (2) controller interfaces that enable supervisory feedback on industrial robots.

2.1 Quality assessment and in-process monitoring

Quality assessment in extrusion AM can be performed during fabrication (*in-process*) or after completion (*post-process*), and it can be embedded in the production flow (*in-line*) or carried out with separation in time/handling (*at-line/off-line*). In this work, we focus on *in-process, in-line* supervision to enable immediate intervention while the robot program is executing.

Table 1 summarizes the monitoring terms used in this paper. We use *in-process* to denote measurement during fabrication, and *in-line* for sensing integrated into the production flow without interrupting execution.

Table 1. Terminology used in this paper for monitoring and inspection placement.

Term	Description (as used in this work)
In-process	Measurement during fabrication; enables immediate detection/intervention.
Post-process	Inspection after fabrication; verification and documentation.
In-line	Embedded in the production flow; no interruption of execution.
At-/Off-line	Separated from production by handling and/or time delay; delayed feedback.

Extrusion-based AM comprises interdependent subprocesses whose timing and interaction strongly influence print quality, motivating standardized evaluation and monitoring strategies [1]. For construction-scale extrusion, monitoring is frequently framed as a multi-scale problem where no single indicator is sufficient; instead, assessment should cover material delivery, geometric consistency, and early-age behavior [7]. Early demonstrations of measurement-informed correction further indicate that real-time sensing can support corrective actions during deposition [8].

Vision-based monitoring complements geometric inspection by providing spatially resolved observations around the nozzle and deposited bead. Robot-vision methods have been used to support environment-aware printing and adaptive decisions under non-ideal boundary conditions [9]. For earthen extrusion, monitoring is particularly relevant because material variability (e.g., soil composition, moisture) can affect extrudability, build-up, and interlayer bonding [2, 5]. Frameworks for 3DPE emphasize both sustainability motivations and increased process variability introduced by earth-based feedstocks [3, 10].

Beyond monitoring placement, inspection can also be distinguished by *granularity* (e.g., nozzle-centric point/feature cues versus stage-level or post-print verification); in this work, point/feature cues are used to trigger stage-level supervisory intervention (slow-down or stop) during deposition [7].

Commercial desktop fused deposition modeling (FDM) systems further illustrate camera-based failure detection during printing, where detected anomalies can trigger a pause of the build [11, 12, 13].

2.2 Controller interfaces for supervisory feedback in robotic fabrication

Industrial robot-based fabrication typically relies on supervisory feedback rather than low-level closed-loop control: external sensing triggers discrete interventions such as slowing, pausing, or adjusting process parameters. In

construction printing, this trend is reflected in the shift from measurement-only systems toward monitoring-and-action loops where sensing is explicitly coupled to control signals [8, 7].

A practical barrier is the limited and vendor-specific access to controller state and intervention channels. Available interfaces differ in achievable update rates, integration effort, and required controller-side configuration. For KUKA systems, prior work contrasts KVP and the Robot Sensor Interface (RSI), highlighting timing behavior and trade-offs for feedback applications [14]. Lightweight and open solutions are therefore attractive for research prototypes, as they reduce reliance on licensed options and minimize invasive controller modifications.

Building on these insights, the present work implements a vision-enabled *supervisory* loop using KVP: controller state (e.g., axes, tool pose, process-related globals) is streamed and time-aligned with a vision-based extrusion-state estimate, while interventions are applied via program override. This positions KVP as a pragmatic bridge between in-process perception and controller action for extrusion-failure mitigation in robotic 3D printing [14, 7].

3 Methods

System responsibilities are distributed across the robot controller level, KVP-based variable streaming, RGB-D sensing, and a workstation that performs synchronization, inference, dataset recording, and visualization.

3.1 System workflow and architecture

During execution, the workstation acquires and timestamps RGB-D frames of the nozzle region while concurrently polling KRC4 controller variables via KVP, including joint axes, TCP data, and process-related globals such as `$OV_PRO`). Both streams are synchronized, stored in a session-structured dataset, and visualized in a custom interface (Figure 3).

The synchronized observations are subsequently used for extrusion-state classification within a nozzle-centric region of interest (ROI), yielding a binary printing/non-printing signal. Based on this signal, a fail-safe supervisory intervention policy maps sustained extrusion loss to a bounded controller-level action via `$OV_PRO` through KVP, thereby slowing or stopping motion to limit "air-printing" and enable timely recovery.

As shown in Figure 2, the system architecture is organized into a robot controller level and a supervisory level. At the robot controller level, the program executes the print motion while robot state variables are made accessible through KVP. At the supervisory level, an RGB-D camera feed is processed by the detection pipeline to determine the current printing state from the nozzle region. This decision is then mapped to controller action by writing `$OV_PRO` via KVP: printing continues under the nominal override, whereas sustained non-printing triggers a reduced override to slow down or effectively pause robot motion.

3.2 Robotic printing platform and sensing

The experimental setup uses a KUKA KR210 R3100 with an Erratic Extruder MS40 supplied by an Erratic Feeder DM80 through a 5 m hose. To observe deposition at the nozzle, an Intel RealSense D405 short-range RGB-D camera is mounted in an eye-in-hand configuration on the robot flange and oriented toward the nozzle and immediate deposition region (Figure 4). This viewpoint captures bead formation and extrusion continuity during motion.

An external workstation serves as the acquisition and logging node. The workstation is a Windows PC that runs acquisition, synchronization, inference, visualization, and dataset recording. RGB and depth frames are streamed from the D405 with per-frame timestamps, while robot state and selected controller and process variables are ac-

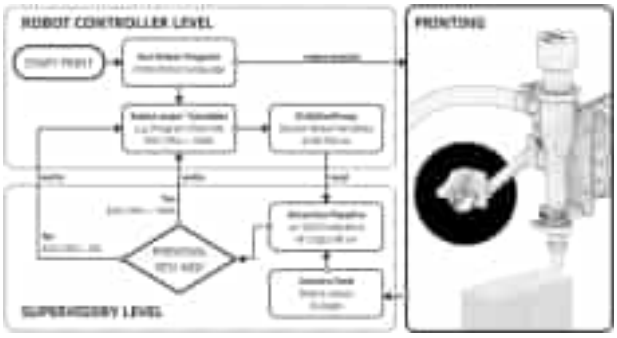


Figure 2. System workflow for vision-enabled feedback-loop: robot controller level, robot motion, KVP-based supervisory level intervention.



Figure 3. User interface with live extrusion-state inference, ROI overlay, robot-variable plots, logging, and adjustable stop/slow intervention parameters.

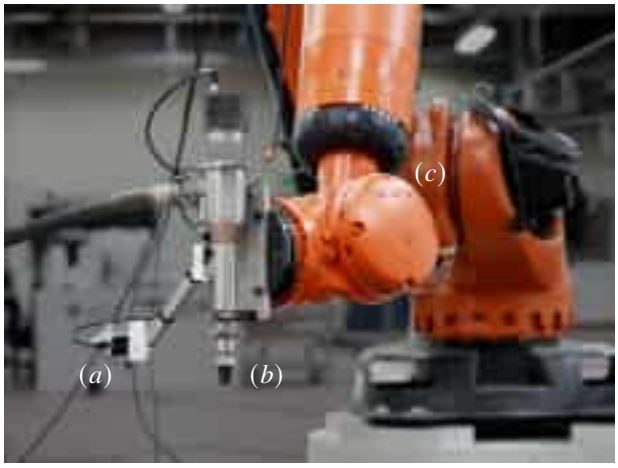


Figure 4. Eye-in-hand nozzle-centric sensing setup: RGB-D sensor (a) mounted to observe nozzle region (b) on the industrial robot (c) for in-process extrusion monitoring.

Table 2. Logged variables, corresponding KUKA controller variable semantics and Description.

Log key	KUKA variable	Description
t_wall	–	Host wall-clock timestamp (Unix s); session reference time.
t_mono	–	Host monotonic time (s); stable for latency measurements.
t_robot	–	KVP interface timestamp associated with each sample.
a1..a6	`$AXIS_ACT.A1..A6`	Actual joint angles (deg).
e1	`$AXIS_ACT.E1`	External axis position (unbounded / "infinite" axis). In this setup, E1 encodes the extruder axis position; its time derivative relates to extrusion rate.
ov_pro	`$OV_PRO`	Program override / global speed scaling (%).
e_rpm	`E_RPM`	Derived extruder speed computed from E1 progression (scaled to RPM for visualization/ monitoring).
e_mod	`EXTR_MOD`	Extrusion modification override (dimensionless scaling factor).

cessed from the KRC4 via KVP (TCP/IP). Both streams are recorded concurrently in a session-structured dataset, enabling timestamp-based alignment of RGB-D observations with robot kinematics and process state. The overall workflow, including sensing, controller streaming, logging, and the integration of extrusion-state detection with program override intervention, is summarized in Figure 2.

3.3 Controller interface via KVP

Controller telemetry and supervisory actuation are implemented via KVP, which exposes external read/write access to selected KRL global variables and KRC4 system variables over TCP/IP during program execution. Unlike low-level motion interfaces, KVP enables lightweight supervisory feedback through discrete controller-level interventions.

During printing, the workstation polls the controller to obtain time-stamped robot state for contextualization and temporal alignment with RGB-D frames. In the current setup, the logged variables include the joint configuration (A1–A6) and the tool state required to interpret nozzle motion; the interface is extensible to additional process signals (e.g., extrusion rpm, extruder command values) depending on the configured KRL globals. All KVP reads are recorded together with local receive timestamps to capture the as-executed motion state observed by the external supervisor.

Supervisory actions are issued through the same channel using the program speed override `$OV_PRO` as the intervention variable. Upon a sustained extrusion-loss decision, the workstation writes a bounded override value to slow down or effectively pause execution. To ensure stable behaviour and to avoid excessive controller traffic, override writes are rate-limited and only updated on state transitions (normal → intervention, intervention → normal). If automatic recovery is enabled, the nominal override value is cached prior to intervention and restored once extrusion is detected again. Overall, KVP provides a practical link between vision-based monitoring and controller-level motion intervention, while keeping the actuation explicit, bounded, and compatible with safety-oriented supervisory control.

3.4 Data synchronization and logging pipeline

All sensing, controller streaming, inference, and dataset logging are executed on a single workstation to ensure a shared time base. The D405 provides synchronized RGB-D frames of the nozzle region, while robot and process variables are polled from the KUKA controller via KVP (KVP) in parallel. The two streams are acquired asynchronously at different rates: KVP samples are queried at an effective update rate of approximately 9.1 Hz (about 110 ms per update), and the camera stream is recorded at 30 fps. To enable consistent multimodal fusion, every sample is timestamped upon arrival at the workstation using (i) a wall-clock timestamp for dataset indexing and (ii) a monotonic timestamp for latency evaluation.

The logger writes a session-structured dataset with three synchronized components: (1) RGB and depth image sequences, (2) a per-frame timestamp table linking frame indices to host and camera time, and (3) a controller-state log containing the KVP samples, like extrusion value `E1` (Table 2). Note that `E1` is an external-axis position signal that increases along the toolpath; the variable `E_RPM` represents a scaled speed estimate derived from the rate of change of `E1`. In post-processing, each camera frame is associated with the temporally nearest KVP sample in host

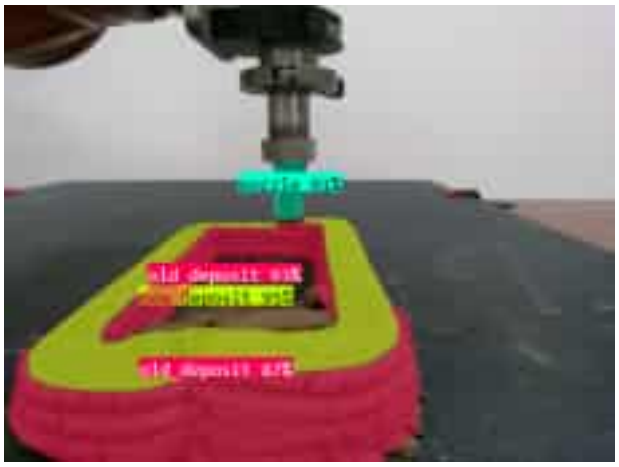


Figure 5. Example YOLOv11 instance segmentation outputs used for extrusion-state inference: nozzle, newly deposited material near the tip (*new_deposit*), and previously deposited layers (*old_deposit*).

time. Associations exceeding a configurable tolerance are optionally rejected to prevent spurious matches under transient delays. This logging design provides time-aligned nozzle observations and robot states for extrusion-state estimation and supervision.

3.5 Vision-based extrusion-state estimation

On top of the logging layer (Section 3.4), a learning-based perception module is trained to detect nozzle-adjacent deposition cues. Recordings were collected under two operating conditions—normal printing and induced extrusion interruption—to capture representative examples of both states. Annotations were managed in Roboflow, and a YOLOv11 instance segmentation model was trained on RGB images to predict three classes: *nozzle*, *new_deposit* (fresh material near the tip), and *old_deposit* (previously deposited layers). Figure 5 shows typical outputs used for subsequent decision logic. In total, 400 images were manually labeled and augmented using geometric and photometric transformations, resulting in 1536 images, which were then split into 92% training, 5% validation, and 2% test data. This setup was used for a fast proof-of-concept validation and to improve robustness under varying conditions, although some augmentation-related correlation risk cannot be fully excluded.

3.6 Supervisory intervention policy

Extrusion supervision is implemented as a rule-based policy that converts vision-derived nozzle-region cues into a controller-level intervention. A YOLOv11 instance-segmentation model predicts instance masks and confidence scores for *nozzle*, *new_deposit*, and *old_deposit* on RGB frames. At runtime, RGB inference is applied to a depth-filtered crop so that evaluation is focused on the near-nozzle working region rather than the wider surroundings. The decision is then restricted to a nozzle-centered ROI beneath the tool, where fresh deposition is expected.

The intervention logic operates on a binary extrusion-state estimate derived from threshold-based ROI evidence for fresh deposition relative to the nozzle tip. To suppress transient misdetections (e.g., brief occlusions or segmentation noise), the decision is filtered using a persistence window τ. In the current experiments, $\tau = 10\,\mathrm{s}$ bounds "air-printing" while reducing nuisance stops; τ can be adjusted at runtime in the developed user interface to match part geometry and acceptable defect tolerance (Fig. 2), trading faster interruption against a higher probability of false stops.

Printing-state decision. From each frame, compact indicators are computed inside the ROI, including (i) nozzle confidence, (ii) the normalized area of *new_deposit* within the ROI, and (iii) a spatial consistency cue that checks whether *new_deposit* occurs in the expected vicinity below the nozzle tip. A *non-printing* state is declared when the nozzle is confidently detected but *new_deposit* evidence remains below threshold or spatially inconsistent with the nozzle tip region.

Temporal buffering. To avoid spurious interventions from transient occlusions and segmentation noise, the decision is temporally filtered using a sustained-evidence criterion. Let $s_t \in \{\text{print}, \text{no-print}\}$ denote the per-frame decision. An intervention is triggered only if s_t = no-print persists for at least τ seconds (here $\tau = 10\,\mathrm{s}$, corresponding to $N = \lceil \tau f \rceil$ frames at frame rate f). This buffering prevents short false negatives from interrupting the process.

Controller intervention via override. When sustained non-printing is detected, the supervisor writes a bounded program speed override to the KUKA controller via KVP using the system variable `$OV_PRO`. Override updates are clamped to valid limits and rate-limited to avoid repeated writes in consecutive cycles. If recovery is enabled, the pre-intervention override value is stored and restored once printing evidence is re-established. Figure 6 illustrates the time-lapse behavior, including ROI evaluation, the printing/non-printing overlay, and the resulting override-based slow-down/stop response under extrusion loss.

4 Experiments and Results

4.1 Baseline vs. induced-fault comparison

To demonstrate the proposed supervisory safety layer, we report a proof-of-concept comparison between (i) a nominal reference run and (ii) an induced extrusion-loss



Figure 6. Time-lapse of the vision-based supervisor during printing, showing the RGB-D decision overlay (printing/not-printing) and tracked nozzle ROI; timestamps (t=00:00–06:40) indicate the evolution over the full run till detected failure and program override intervention.

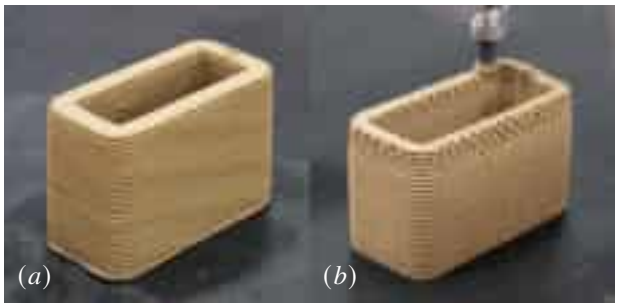


Figure 7. Extrusion-loss comparison: (a) uninterrupted extrusion; (b) induced pump-off event leading to under-extrusion until the supervisor intervenes and stops motion.



Figure 8. Front-view comparison emphasizing deposition quality differences for the same toolpath: (a) baseline deposition; (b) induced pump-off event (poe) leading to under-extrusion with intervention triggered after detected extrusion loss.

run in which the material pump was switched off approximately halfway through execution. The supervisor monitors the nozzle region and, upon sustained non-printing, writes a bounded program-override command via `$OV_PRO` through KVP to reduce motion, thereby limiting continued travel without deposition.

Figure 7 provides geometric context for the failure case. The baseline run (Figure 7a) maintains stable material delivery and produces consistent layer stacking. In the induced-failure run (Figure 7b), turning off the feeder results in progressively under-extruded layers; the final nozzle position indicates where the supervisory loop triggered an override-based stop to prevent further air-printing and to enable recovery actions (e.g., refill or unclogging) before degradation propagates.

While the perspective view highlights the overall geometry, Figure 8 emphasizes the resulting deposition quality. In the baseline case (Figure 8a), bead thickness and layer continuity remain visually consistent. In the failure case (Figure 8b), under-extrusion accumulates over successive layers prior to intervention. Together, the two views illustrate that the proposed supervision can prevent extended motion under extrusion loss, limiting the spatial extent of degraded deposition and supporting timely recovery.

4.2 Runtime performance and response behavior

We evaluate the proposed supervisor in terms of (i) runtime throughput and stability and (ii) override-based intervention behavior under an induced extrusion fault. During live operation, the effective end-to-end processing rate (capture → inference → UI/logging) was approximately 12 fps, below the configured camera capture rate, reflecting workstation compute constraints. Nevertheless, the system operated continuously during the reference and pump-off runs and produced complete, time-stamped logs without dropped-session failures.

In the induced pump-off run, the supervisor triggered a program-override intervention after $\tau = 10\,\mathrm{s}$ of sustained non-printing, reducing motion to limit further travel without deposition.

For the underlying perception module, the YOLOv11 instance segmentation achieved 99.0% precision, 83.9% recall, and 84.3% mAP@50 on the held-out evaluation set. The confusion matrix indicates low class confusion overall, with the dominant errors occurring between *new_deposit* and *old_deposit*, which exhibit similar appearance under the given lighting and surface conditions.

4.3 Validation of the supervisory feedback loop

To complement the proof-of-concept comparison, two small repeated-trial test conditions were evaluated, as summarized in Table 3: a non-stop print test and a stop-test with visible printing degradation. The non-stop test was used to verify that the supervisor does not trigger false interruptions during normal operation, while the stop-test was used to assess whether extrusion loss is detected consistently before visible quality degradation becomes pronounced. Three samples were evaluated per condition as a compact first repeatability check. The results indicate stable behaviour in the non-stop trials and consistent early detection in the stop-tests, supporting the intended role of the loop as an early warning and intervention layer.

Table 3. Repeated-trial timing results for false-interruption screening and early extrusion-loss detection.

Measure	S1	S2	S3
False interruption (non-stop test)	No	No	No
Quality-drop time [s]	26	25	27
YOLO detection time [s]	20	20	20
Lead time [s]	6	5	7

5 Conclusion

5.1 Key contribution and takeaway

This paper presented a vision-enabled supervisory feedback loop for extrusion-based robotic 3D printing that couples in-process perception with controller-level intervention using KVP. An eye-in-hand RGB-D sensor provides nozzle-centric observations, which are converted into an extrusion-state signal and mapped to a bounded intervention through the KUKA program override variable `$OV_PRO`. The loop enables a controller-compatible 'detect, slow/stop' strategy that limits air-printing while preserving lightweight integration with industrial robot controllers.

5.2 Constraints and limitations

This study is a proof of concept with limited experimental scope and fault coverage. All experiments were performed on a single robotic platform using one sensor configuration, and extrusion loss was introduced under controlled conditions; accordingly, generalization to other toolpaths, lighting conditions, and material behaviors remains to be established. Runtime performance was limited by the available computing hardware, resulting in an effective inference rate of approximately 12,fps on a Dell Precision 5680 equipped with an Intel i9-13900H, an RTX 4000 GPU, and 32,GB RAM; this reduces the temporal resolution of the perception stream and may lower sensitivity to brief failure events. While higher inference rates would be desirable for broader deployment, the achieved performance was sufficient for the present proof-of-concept demonstration and offered practical value compared with exclusive reliance on human supervision alone.

The supervisory policy is intentionally conservative, requiring extrusion loss to persist for a configurable time window τ before intervention; this improves robustness against transient occlusions and segmentation noise at the cost of a bounded detection delay. Moreover, the current intervention channel is restricted to program override via `$OV_PRO`, providing reliable slowdown/stop behavior but not yet exploiting additional process actuation variables (e.g., `EXTR_MOD` for increasing/ decreasing extrusion rate) for corrective actions such as extrusion-rate compensation.

5.3 Outlook and scaling path

Future work will expand data collection and evaluation by logging full RGB-D streams, controller variables, and intervention events across a broader set of toolpaths and failure modes. This will support quantitative characterization of loop timing, robustness across operating conditions, and sensitivity to the persistence window τ, as well as refinement of the vision model and decision logic.

Beyond pause/stop interventions, the same controller-access setup can be extended toward *corrective* measures. In particular, future iterations will investigate modulating extrusion commands online (e.g., via `EXTR_MOD`) to compensate short periods of reduced flow, and streaming the executed robot motion state to enable localized path-level recovery, such as re-tracing segments or re-depositing under-extruded layers when deviations are detected. These extensions must be designed in strict compliance with applicable safety regulations and risk assessments: once robot motion is altered autonomously, additional safeguards are required, including bounded interventions, limits on axis travel and rotational speeds, and independent safety mechanisms to prevent harm to personnel or equipment. This positions the presented KVP-enabled supervi-

sor as a scalable foundation for increasingly process-aware robotic extrusion AM.

6 Acknowledgement and Funding

The authors gratefully acknowledge the support of the industry partner ClayTec for providing the base materials used in the additive manufacturing experiments. The authors also acknowledge the financial support provided by the IDEA League Fund, which enabled international exchange and facilitated the collaboration between ETH Zurich and RWTH Aachen University. This work was supported by the German Federal Ministry for Economic Affairs and Energy (Bundesministerium für Wirtschaft und Energie, BMWE) through the Central Innovation Programme for SMEs (Zentrales Innovationsprogramm Mittelstand, ZIM), as part of the research project *ClayX* at RWTH Aachen University.

References

[1] R. A. Buswell, W. R. Leal de Silva, S. Z. Jones, and J. Dirrenberger. A process classification framework for defining and describing digital fabrication with concrete. *Cement and Concrete Research*, 134, 2020. doi:10.1016/j.cemconres.2020.106068.

[2] F. Faleschini, R. Angiuli, M. A. Zanini, S. Zanni, L. Hofer, and C. Pellegrino. Mechanical properties of sustainable mixes for 3d printing of earth-based constructions. *Construction and Building Materials*, 395, 2023. doi:10.1016/j.conbuildmat.2023.132496.

[3] Alexander Curth, Natalie Pearl, Angelica Castro-Salazar, Caitlin Mueller, and Lawrence Sass. 3d printing earth: Local, circular material processing, fabrication methods, and life cycle assessment. *Construction and Building Materials*, 2024. doi:10.1016/j.conbuildmat.2024.135714.

[4] Kunaljit Chadha, Alexandre Dubor, Edouard Cabay, Yara Tayoun, Lapo Naldoni, and Massimo Moretti. *Additive Manufacturing for the Circular Built Environment: Towards Circular Construction with Earth-Based Materials*. Springer International Publishing, 2024. doi:10.1007/978-3-031-39675-5_7.

[5] Alexandre Dubor, Edouard Cabay, Yara Tayoun, Oriol Carrasco, Elisabetta Carnevale, Eduardo Chamorro, and Vincent Huygues. 3d-printed earth architecture: Design approach for a performative habitat. In Phil Ayres, Mette Ramsgaard Thomsen, Bob Sheil, and Marilena Skavara, editors, *Fabricate 2024*. UCL Press, 2024. doi:10.2307/jj.11374766.27.

[6] Hendrik Benz, The Vinh Nguyen Trong, and Katharina Klemt-Albert. Real-time vision-based defect detection for large-scale on-site earthen additive manufacturing: Annotated dataset and dual-model framework. *Automation in Construction*, 181, 2026. doi:10.1016/j.autcon.2025.106666.

[7] Rob Wolfs, Erik Klooster, Matthijs Hebels, Elena I. Vitucci, Yves Petit, and Nicolas Roussel. On-line and in-line quality assessment across all scale levels of 3d concrete printing. *Cement and Concrete Research*, 185, 2024. doi:10.1016/j.cemconres.2024.107646.

[8] Rob J. M. Wolfs, Freek P. Bos, and Theo A. M. Salet. A real-time height measurement and feedback system for 3d concrete printing. In D. A. Hordijk and M. Lukovic, editors, *High Tech Concrete: Where Technology and Engineering Meet*. Springer International Publishing, 2018. doi:10.1007/978-3-319-59471-2_282.

[9] Roberto Naboni, Luca Breseghello, and Sandro Sanin. Environment-aware 3d concrete printing through robot-vision. In *Proceedings of eCAADe 40 – Co-creating the Future*, 2022.

[10] A. Perrot, E. Cailleux, N. Roussel, N. Binh, A. Bérard, J.-F. Caron, A. Cavalier, M. Huet, W.-C. Jau, C. Lanos, A. Nezamian, G. Orru, A. Razakamanantsoa, L. Vandewalle, and C. Ferraris. Snapshot on 3d printing with alternative binders for sustainable construction. *Cement and Concrete Research*, 185, 2024. doi:10.1016/j.cemconres.2024.107651.

[11] Douglas A. J. Brion and Sebastian W. Pattinson. Generalisable 3d printing error detection and correction via multi-head neural networks. *Nature Communications*, 13, 2022. doi:10.1038/s41467-022-31985-y.

[12] Hyungjung Kim, Hyunsu Lee, Ji-Soo Kim, and Sung-Hoon Ahn. Image-based failure detection for material extrusion process using a convolutional neural network. *The International Journal of Advanced Manufacturing Technology*, 111, 2020. doi:10.1007/s00170-020-06201-0.

[13] Hyungjung Kim, Hyunsu Lee, and Sung-Hoon Ahn. Systematic deep transfer learning method based on a small image dataset for spaghetti-shape defect monitoring of fused deposition modeling. *Journal of Manufacturing Systems*, 65, 2022. doi:10.1016/j.jmsy.2022.10.009.

[14] M. H. Arbo, I. Eriksen, F. Sanfilippo, and J. T. Gravdahl. Comparison of kvp and rsi for controlling kuka robots over ros. *IFAC-PapersOnLine*, 53(2), 2020. ISSN 2405-8963. doi:10.1016/j.ifacol.2020.12.268.

A Generative AI-Based Construction Safety Assistant Using Retrieval-Augmented Generation

Xiaoyu Hou[1], Aakash Walavalkar[2] and Bo Xiao[3]

[1]PhD Student, Dept. of Civil, Environmental, and Geospatial Engineering, Michigan Technological Univ., Houghton, MI. 49931, U.S.A. Email: xiaoyuho@mtu.edu
[2]Master Student, Dept. of Health Informatics, Michigan Technological Univ., Houghton, MI. 49931, U.S.A. Email: awalaval@mtu.edu
[3]Assistant Professor, Dept. of Civil, Environmental, and Geospatial Engineering, Michigan Technological Univ., Houghton, MI. 49931, U.S.A. Email: boxiao@mtu.edu (Corresponding Author)

Abstract –

Safety remains a central challenge in the construction industry, where OSHA standards define most regulatory requirements. Although comprehensive, these standards are dispersed across lengthy documents, making it difficult to locate provisions relevant to specific tasks or safety concerns. This complexity is especially evident in education and training settings, where instructors must interpret regulatory text and learners often lack familiarity with its structure and terminology. Recent advances in generative Artificial Intelligence (AI) offer new opportunities for improving access to technical information, but general-purpose models lack the grounding needed for authoritative regulations. This study presents a Retrieval-Augmented Generation (RAG) platform that provides clause-linked answers based on OSHA construction standards. The platform integrates document preprocessing, semantic retrieval, and citation-aware response generation through a web interface that enables direct inspection of referenced clauses. Automated evaluation using thirty representative safety questions showed strong performance, with average scores of 0.95 for context relevance, 0.99 for factual accuracy, and 0.90 for response relevance. The study contributes a verifiable and regulation-aligned approach for delivering accessible safety guidance, offering an adaptable framework for construction safety education and training.

Keywords-

Construction safety; OSHA; RAG; ISARC

1 Introduction

In 2023, the United States construction industry recorded 1,075 fatal work injuries, which represented nearly one fifth of all occupational deaths nationwide [1]. The consistently high number of fatalities highlights the persistent challenges associated with managing safety in an inherently hazardous sector [2]. Regulatory requirements issued under the Occupational Safety and Health Administration (OSHA) [3] construction standards therefore serve a central role in guiding hazard prevention across activities such as fall protection, scaffolding, electrical installation, excavation, equipment operation, and hazard communication. The standards offer comprehensive coverage of the risks commonly present on construction sites, yet the regulations are distributed across numerous subparts and contain extensive cross-referenced provisions [4]. The scale and organization of this regulatory corpus often make it difficult for workers, engineers, and students to identify the specific clauses relevant to a particular task or condition. Locating applicable requirements within lengthy documents requires substantial manual effort, which can hinder timely decision making and reduce the effectiveness of safety management practices.

Currently, users depend on keyword searches and personal familiarity with the structure of OSHA's construction regulations to locate relevant provisions, a process that often yields partial retrieval and inconsistent interpretation among individuals with limited regulatory experience [5]. The difficulty of identifying applicable clauses is evident across workforce training, onboarding programs, continuing professional development, and university instruction, where learners must connect specific safety questions to authoritative requirements without a clear mechanism for navigating the structure of the standards. Existing digital resources, including OSHA's eTool libraries and CPWR training materials [6, 7], improve access to regulatory information but remain primarily document centered and require substantial manual navigation. The absence of tools that can align user queries with the semantic content of regulatory text limits the role of OSHA standards as a foundation for

accurate, transferable, and instructionally meaningful safety knowledge. These limitations create a pressing need for a system capable of guiding users from practical safety questions to the corresponding OSHA clauses while providing transparent explanations that support both learning and decision making.

Advances in generative AI have created new opportunities for addressing the challenges associated with linking safety questions to relevant regulatory content. Recent developments in large language models have demonstrated strong capabilities in complex question answering, context interpretation, and the generation of instructional explanations, suggesting potential value for safety education and training [8]. Retrieval-Augmented Generation (RAG) has become particularly promising for applications that require the integration of authoritative sources with model-generated responses [9], since retrieved text can anchor explanations in verifiable regulatory clauses. Several studies have applied retrieval-based generative approaches to tasks such as building code interpretation, engineering documentation support, and domain-specific instructional assistance [10]. However, research focused on OSHA construction safety remains limited. Existing efforts rarely incorporate clause-level grounding in OSHA's construction regulations or provide mechanisms that align user queries with specific provisions in a manner suitable for education and training. The limited attention to OSHA-centered generative systems indicates the need for a solution capable of connecting user questions to regulatory clauses while delivering explanations that support accurate, transparent, and instructionally meaningful safety learning.

This study investigates how a retrieval-augmented generative AI framework can support accurate and transparent access to OSHA construction safety regulations through clause-grounded retrieval and explanation. To address this problem, a retrieval-augmented framework was developed to ground generative AI responses in authoritative OSHA construction regulations. The framework integrates document retrieval and controlled response generation to provide clause-level traceability and transparent evidence for each answer. A web-based platform was implemented to operationalize the framework and to present both the generated explanation and the corresponding regulatory text in an accessible interface. The platform was evaluated using automated metrics, including relevance, factual alignment, and citation correctness, to examine the extent to which system responses remain consistent with regulatory content. The results demonstrate that the proposed approach can effectively connect user questions with verifiable OSHA clauses while maintaining transparency in the response generation process.

2 Literature Review

This section reviews the state-of-the-art literature related to generative AI in construction education, which includes technologies in construction education, generative AI in intelligent tutoring, and generative AI in domain-specific education.

2.1 Construction Safety Management

Construction safety management has been widely examined due to the persistently high incidence of injuries and fatalities in the sector. Recent studies emphasize that structured safety management systems and systematic risk assessment play a critical role in reducing incident likelihood. Sadeghi et al. demonstrated that ensemble risk-assessment models can improve the identification of high-risk situations on construction projects [11], while Zhang and Mohandes developed a holistic occupational health and safety framework that links risk control measures to project-level decision making [12]. More recent work has shown that targeted management interventions and safety management measures can have measurable effects on workers' safety behaviors and incident outcomes [13, 14].

OSHA construction standards constitute the primary regulatory foundation governing hazard control in the United States and are frequently cited as the formal basis for site-level safety practices. However, prior research has indicated that regulatory requirements are difficult to explain and apply in practice. Semantic analyses of construction safety knowledge highlight that relevant information is highly fragmented and embedded in complex technical language, which complicates efforts to translate formal requirements into actionable guidance [15, 16]. Studies on construction safety knowledge systems similarly note that safety-related concepts are dispersed across documents and topics, and that workers often rely on local practices and informal guidance rather than directly consulting formal regulations [15]. Although digital resources, such as OSHA e-tools and online training portals, have broadened access to regulatory information, existing platforms still require substantial manual navigation and provide limited semantic support for locating clause-level requirements within lengthy regulatory documents [13].

Challenges related to regulatory comprehension also affect safety education and workforce training. Survey-based studies report that trainees and frontline workers frequently struggle to interpret technical terminology and to connect safety requirements with realistic site scenarios, which constrains the transfer of formal knowledge to field practice [17]. Eye-tracking and experimental training studies further show that limited safety knowledge undermines hazard-identification performance, even when workers have access to visual

examples and instructional materials [18]. The challenges highlight a continuing need for approaches that improve the accessibility and contextual alignment of safety regulations within both education and workforce training.

2.2 Generative AI Applications in Construction

Recent research has begun to explore how generative artificial intelligence can be incorporated into construction management, reflecting the broader digital transformation underway in the sector. Ghimire, Kim, and Acharya (2023) offered an early synthesis of the opportunities and challenges associated with generative AI [19], noting the potential to automate documentation, summarize technical information, and support managerial decision-making. Their analysis also emphasized that the rapid emergence of large language models has created new expectations for access to timely and interpretable information, while simultaneously raising concerns regarding reliability, data governance, and the need for domain-sensitive safeguards in safety-critical applications.

Industry-focused investigations provide complementary evidence regarding the growing relevance of generative AI. Nyqvist et al. (2024) examined how generative AI tools can be aligned with common construction management tasks and reported substantial perceived value in communication support [20], meeting administration, situational awareness, data analysis, and risk monitoring. Their evaluation, conducted through focus group discussions and practitioner surveys, suggests that generative AI can strengthen several routine managerial functions, although its effective use requires structured workflows and mechanisms that prevent excessive dependence on uncontrolled model behavior.

Applications at the project and design levels further illustrate the expanding role of generative AI across construction activities. Recent studies have shown that generative models can accelerate conceptual design, produce alternative layouts, and enhance sustainability assessments when integrated with BIM and digital twin environments, particularly by enabling rapid scenario exploration and reducing material waste [20, 22]. In project management settings, the use of large language models such as ChatGPT has demonstrated potential for automating routine planning descriptions, refining resource allocation, and supporting team coordination [23]. Emerging discussions also highlight generative AI as a candidate technology for advancing safety and sustainability outcomes, particularly by improving planning accuracy and facilitating data-driven decisions [24].

Although these developments indicate accelerating interest, current generative AI applications in construction largely operate as general-purpose assistants. Most systems do not incorporate domain-specific constraints, authoritative regulatory sources, or mechanisms that ensure traceable and verifiable outputs. Prior studies repeatedly point to the need for generative AI systems that can operate within clearly defined domain boundaries and provide transparent reasoning aligned with industry standards. The absence of such capabilities highlights a continuing gap, particularly for safety-critical contexts where the interpretation of formal regulatory requirements must be both accurate and verifiable.

3 System Design

3.1 Overview

The Construction Safety Assistant is implemented as a retrieval-augmented platform that supports OSHA-grounded question answering for construction safety education and training. The system consists of three primary components: OSHA document preprocessing, semantic retrieval, and controlled answer generation, all integrated into a web-based interface. The workflow begins with the preparation of OSHA construction safety documents, which are parsed, segmented, and transformed into an indexed knowledge base to enable efficient semantic search. When a user submits a query through the interface, the system encodes the input and matches it against the indexed corpus. Retrieved OSHA excerpts are then incorporated into a structured prompt for the generative model to ensure grounded and citation-constrained responses. The platform outputs an OSHA-referenced answer along with citation links pointing to the corresponding regulatory documents. Figure 1 presents the system architecture and the end-to-end flow from query submission to retrieval, generation, and citation display. Figure 1 presents the system architecture and the end-to-end workflow from query submission to retrieval, generation, and citation display.

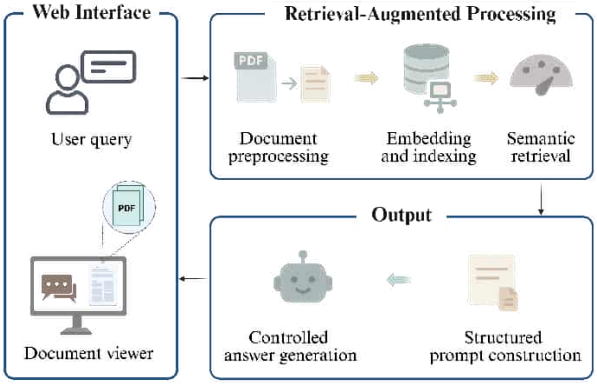


Figure 1. Overview of the construction safety assistant

3.2 OSHA Document Processing

The Construction Safety Assistant is built on OSHA construction safety regulations obtained in PDF format. Each document is converted into machine-readable text through an automated extraction procedure that preserves the regulatory reading order while removing layout artifacts inherent to PDF formatting. Non-informative elements such as headers, footers, page numbers, and residual formatting symbols are removed to ensure that only substantive regulatory content is retained.

After extraction, the regulatory text is segmented into fixed-length fragments to support efficient semantic retrieval. Each fragment is paired with metadata identifying the source OSHA document and its corresponding page index within the original file. These metadata elements enable transparent citation during answer presentation and allow users to inspect the regulatory basis of each response.

The processed fragments are then transformed into vector representations using a pretrained embedding model and stored in an indexed corpus that supports top-k similarity search. This indexing structure allows user queries to be mapped to relevant OSHA content with high precision, forming the foundation for verifiable and document-grounded question answering.

3.3 Retrieval-Augmented Answer Generation

The answer generation process integrates semantic retrieval with controlled natural-language synthesis to produce grounded explanations drawn from OSHA construction safety regulations. When a user submits a question, the system converts the input into a dense semantic representation that supports comparison with the indexed regulatory corpus. Let q denote the user query. The embedding function $f(\cdot)$ maps the query into a vector space as shown in Equation (1):

$$e_q = f(q) \tag{1}$$

This representation allows the system to evaluate conceptual similarity between the user query and the embedded OSHA fragments.

Each OSHA text fragment is represented by a vector e_i, and the semantic similarity between the query and each fragment is computed using cosine similarity. The similarity score is given in Equation (2):

$$Sim(e_q, e_i) = \frac{e_q \cdot e_i}{\| e_q \| \| e_i \|} \tag{2}$$

The retrieved fragments are ranked according to this similarity measure. The system selects the top-k fragments that exhibit the strongest alignment with the query, as expressed in Equation (3):

$$R_k(q) = TopK_{e_i \in E}\ Sim(e_q, e_i) \tag{3}$$

where E denotes the full embedded corpus. Each retrieved fragment includes metadata identifying the source OSHA document and page location, which is preserved throughout the subsequent generation process.

After the relevant fragments are identified, the system constructs a structured prompt that combines the original user query with the retrieved regulatory text. This prompt formulation establishes the context and boundaries within which the generative model operates, limiting the model's responses to material drawn from the retrieved fragments. When the retrieval stage fails to return fragments that meet the predefined relevance threshold, the system activates a fallback mechanism that guides the model to request clarification or to provide a neutral response rather than infer unsupported information. Figure 2 illustrates the structured prompt construction process.

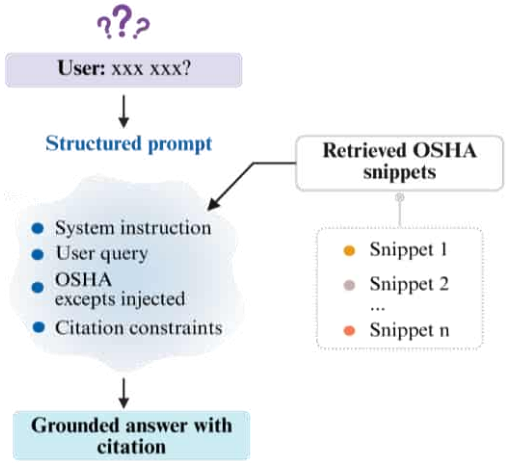


Figure 2. Structured prompt construction process

Using the retrieved content and structured prompt, the generative model synthesizes an answer that restates and interprets the OSHA provisions. Citation metadata is embedded directly into the response and linked to the original OSHA documents, allowing users to inspect the regulatory basis of each statement through the interface. This grounding mechanism supports transparent and verifiable use of OSHA standards in safety education and training.

3.4 Web Interface

Following the retrieval and generation pipeline, the Construction Safety Assistant presents the grounded responses through a web-based interface designed to support transparent review of OSHA regulatory material. The interface provides two coordinated components: a conversational panel for user interaction and a document viewer for inspecting cited OSHA content. In the conversational panel, users submit safety-related questions in natural language. The system's responses appear in the same panel and include embedded citation markers that identify the OSHA document and page from which the supporting content was retrieved. These markers are integrated directly into the response text to

maintain readability while indicating the regulatory source of each referenced statement. Selecting a citation opens the corresponding OSHA PDF page in a side panel, allowing users to review the original regulatory wording without leaving the ongoing interaction.

4 Implementation and Evaluation

4.1 Implementation

The platform is composed of a Python backend and a browser-based frontend. OSHA regulations are stored in PDF format and processed using a text extraction pipeline that removes headers, footers, page numbers, and formatting artifacts. The cleaned text is segmented into fragments of approximately 300 characters, each paired with metadata identifying the source PDF filename and page index.

The processed fragments and metadata are uploaded to Azure Blob Storage, where a Shared Access Signature (SAS) URL is generated for each fragment to provide direct access to its corresponding location within the original PDF. Text embeddings are created using the text-embedding-ada-003 model, and both the embeddings and their metadata are indexed through Azure Cognitive Search to enable vector-based similarity retrieval.

At runtime, the backend receives a user query, computes its embedding, and issues a vector search request to Cognitive Search. The retrieved OSHA fragments and their SAS URLs are assembled into a structured prompt that is sent to a GPT-4o model for grounded answer generation. Citation links are inserted into the output using the stored metadata so that each referenced clause can be opened directly.



Figure 3. Web interface of the platform

The backend exposes REST endpoints for query submission and citation retrieval. The application runs in a cloud environment that supports stateless request–response communication with the frontend, which is implemented using standard web technologies. As illustrated in Figure 3, the web interface allows users to submit queries, view generated responses, and interact with citation links to access the corresponding OSHA PDF pages in an embedded viewer, completing the end-to-end interaction between retrieval, generation, and document-level inspection. When users select a citation in the generated response, the front end loads the associated SAS URL and displays the referenced OSHA PDF page.

4.2 Automated Evaluation Framework

The performance of the platform was assessed using an automated evaluation framework configured with UpTrain [25], which enables systematic comparison between system-generated answers and authoritative OSHA regulatory text. UpTrain provides normalized metric scores between 0 and 1 for each evaluation criterion, supporting large-scale, clause-level assessment without manual scoring. A benchmark set of thirty representative safety questions was constructed to cover major regulatory domains such as fall protection, excavation and trenching, scaffolding, electrical hazards, confined spaces, and personal protective equipment. Each question was paired with a ground-truth citation from the corresponding OSHA construction standard, forming the reference set used in the evaluation. Table 1 summarizes the benchmark questions and their associated regulatory topics.

Table 1 Questions used for automated evaluation

No.	Question	Type
1	What is the minimum width for a scaffold walkway?	What
2	What does OSHA define as a “competent person” in excavation work?	
3	What is the permissible exposure limit (PEL) for noise in construction?	
4	What is the maximum allowable slope for a Type C soil trench?	
5	What class of hard hat offers protection from electrical hazards?	
6	What are the signage requirements near a trenching operation?	
7	What are the minimum illumination levels required in construction tunnels?	
8	What are the minimum spacing requirements for cross braces on	

	scaffolds?	
9	What is the definition of “confined space” in construction?	
10	When is fall protection required for workers on low-slope roofs?	
11	When should a safety harness be inspected?	
12	When should protective systems be installed in excavations?	When
13	How often must trenches be inspected by a competent person?	
14	How often must cranes be inspected?	
15	How should workers be protected from cave-ins during trenching?	
16	How does OSHA define “confined space” in construction?	
17	How do scaffolds and aerial lifts differ in terms of fall protection?	How
18	How should hard hats be worn in construction zones?	
19	How often must safety harnesses be inspected?	
20	Why is eye protection mandatory during welding operations?	
21	Why must hard hats be worn in construction zones?	
22	Why are GFCIs required on construction sites for power tools?	Why
23	Why is ventilation important in welding operations inside a confined space?	
24	Is it acceptable to stack bricks more than 7 feet high without restraint?	
25	Is training required before operating a scissor lift?	Is / Are
26	Is daily site inspection required for all excavation work regardless of depth?	
27	Are reflective vests mandatory for night work near roadways?	
28	Can a scaffold be used if it is partially disassembled?	
29	Can a ladder be used horizontally as a platform?	Can
30	Can a trench shield be used in unstable soil conditions?	

Three quantitative metrics were used to characterize system performance. Context relevance evaluates whether the retrieved OSHA passages correspond to the substantive focus of the user query. Factual accuracy assesses the correctness of statements in the generated answer relative to the regulatory source. Response relevance measures how well the answer addresses the intent of the question in a clear and semantically appropriate manner. In addition to these metrics, a Reference Match check was included to verify that the citation links embedded in the output point to the expected OSHA clauses, ensuring traceability and source transparency.

4.3 Evaluation Results

The automated evaluation produced consistent evidence of strong grounding performance across the selected metrics. The system achieved an average context relevance score of 0.95, indicating that retrieved OSHA passages typically aligned with the substantive focus of the questions. Factual accuracy reached 0.99, reflecting near-perfect agreement between generated statements and the corresponding regulatory clauses. The average response relevance score of 0.90 suggests that the generated answers generally addressed the user intent with appropriate specificity. All citation links passed the Reference Match check without errors. Table 2 summarizes the metric scores for all thirty benchmark questions.

Table 2 Automated evaluation results

No.	Context relevance	Factual accuracy	Response relevance
1	1	1	1
2	1	1	1
3	0.67	1	1
4	1	1	1
5	1	1	1
6	0.67	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	0.67	1	0.67
11	1	1	1
12	0.67	1	1
13	1	1	1
14	1	1	1
15	0.67	1	1
16	1	1	0.67
17	1	1	1
18	1	1	0.67
19	1	1	0.67
20	1	1	0.67
21	1	1	0.67
22	1	1	0.67
23	1	0.67	0.67
24	1	1	1
25	1	1	1
26	1	1	1
27	1	1	0.67
28	1	1	1
29	1	1	1
30	1	1	1
Ave. score	0.95	0.99	0.90

5 Discussion

The evaluation results highlight several characteristics of the system's behavior across different question types and regulatory topics. High factual accuracy reflects the constraint imposed by retrieval-augmented generation, which restricts responses to OSHA source text and minimizes reliance on model priors. Variations in context relevance are largely associated with questions whose answers span multiple regulatory clauses, particularly in areas such as confined spaces and ventilation where relevant information is distributed across several subparagraphs. Lower response relevance in some how and why questions stem from the broader interpretive nature of these items, which may not be fully represented within a single retrieved fragment.

The observed patterns underscore the influence of OSHA's document structure on retrieval performance. Standards that rely on explicit thresholds or enumerated requirements, such as fall protection and scaffolding, align well with fragment-level retrieval and produce consistent grounding results. Sections organized around descriptive or condition-dependent language present additional challenges, as semantically related content may reside in multiple locations. This separation affects the system's ability to assemble a complete context window for procedural or explanatory questions.

Several limitations arise from these characteristics. The retrieval process operates at the level of individual text fragments and does not aggregate related clauses unless they co-occur within the same segment. Citation links reflect the segmentation scheme rather than the full logical scope of a regulation, which may omit adjacent supporting text. Additionally, queries that rely on implicit hazard rationales or multi-step procedures may not fully align with the structure of the underlying standards, reducing the semantic match between the question and the retrieved passage. While the automated evaluation provides a consistent and scalable assessment of system performance, it does not fully capture aspects such as interpretability for practitioners or effectiveness in real-world decision-making contexts.

For future research, improvements to the retrieval process could focus on integrating multi-clause retrieval to better capture regulatory content that is distributed across several sections. Additional refinement of semantic clustering may also help group related provisions into coherent units, improving coverage for questions that require broader contextual grounding. Enhancing query processing with domain-informed expansion represents another opportunity to reduce mismatches between user phrasing and OSHA terminology. Future work will also incorporate expert-based and user-centered evaluation to complement automated assessment and to further examine the usability and practical effectiveness of the system in construction safety contexts.

6 Conclusion

This study developed a retrieval-augmented question-answering system to support construction safety education and training through clause-linked access to OSHA standards. The system integrates document processing, semantic retrieval, and citation-aware answer generation within a web-based interface, enabling users to examine both the generated response and its regulatory source. The findings provide a direct response to the research question by demonstrating that a retrieval-augmented framework can enable accurate, transparent, and clause-level access to OSHA construction safety regulations.

Automated evaluation across thirty representative safety questions showed consistently high factual accuracy and strong performance in relevance and citation correctness, with lower performance primarily observed in questions whose regulatory basis spans multiple sections of the standards. These results indicate that clause-grounded retrieval combined with citation-aware generation can support more reliable access to OSHA regulations by aligning user queries with verifiable regulatory content and preserving transparency in the response process.

The proposed approach provides a practical mechanism for interacting with prescriptive and authoritative documents such as OSHA standards, particularly in contexts where regulatory information is distributed across multiple sections and requires structured access. Future work may focus on improving retrieval coverage and contextual aggregation for complex regulatory queries. In addition, comparative evaluation with existing tools, such as keyword-based search systems and general-purpose language models, will be conducted to further assess the relative performance and practical advantages of the proposed approach. The framework can also be extended to additional regulatory corpora and specialized application contexts.

References

[1] U.S. Bureau of Labor Statistics. Fatal falls in the construction industry in 2023. On-line: https://www.bls.gov/opub/ted/2025/fatal-falls-in-the-construction-industry-in-2023.

[2] J. E. Dodoo and H. Al-Samarraie. A systematic review of factors leading to occupational injuries and fatalities. Journal of Public Health, 31(1):99–113, 2023.

[3] U.S. Department of Labor. Occupational Safety and

Health Administration. *Occupational Safety and Health Administration.* On-line: https://www.osha.gov/, Accessed: 30/11/2025.

[4] G. Chambers. Concise Guide to Workplace Safety and Health: What You Need to Know, When You Need It. *CRC Press*, 2011.

[5] H. Kim, H. S. Lee, M. Park, B. Chung, and S. Hwang. Information retrieval framework for hazard identification in construction. *Journal of Computing in Civil Engineering*, 29(3):04014052, 2015.

[6] U.S. Department of Labor. Occupational Safety and Health Administration. *Occupational Safety and Health Administration.* On-line: https://www.osha.gov/, Accessed: 30/11/2025.

[7] CPWR – The Center for Construction Research and Training. *Training Resources.* On-line: https://www.cpwr.com/, Accessed: 30/11/2025.

[8] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, ..., and G. Kasneci. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274, 2023.

[9] S. Zhao, Y. Yang, Z. Wang, Z. He, L. K. Qiu, and L. Qiu. Retrieval augmented generation (RAG) and beyond: A comprehensive survey on how to make your LLMs use external data more wisely. *arXiv preprint arXiv*:2409.14924, 2024.

[10] M. Aqib. Development of a Question Answering System Over Building Codes using Retrieval Augmented Generation, 2025.

[11] H. Sadeghi, S. R. Mohandes, M. R. Hosseini, S. Banihashemi, A. Mahdiyar, and A. Abdullah. Developing an ensemble predictive safety risk assessment model: Case of Malaysian construction projects. *International Journal of Environmental Research and Public Health*, 17(22):8395, 2020.

[12] X. Zhang and S. R. Mohandes. Occupational health and safety in green building construction projects: A holistic Z-numbers-based risk management framework. *Journal of Cleaner Production*, 275:122788, 2020.

[13] S. Tabatabaee, S. R. Mohandes, R. R. Ahmed, A. Mahdiyar, M. Arashpour, T. Zayed, and S. Ismail. Investigating the barriers to applying Internet-of-Things-based technologies to construction site safety management. *International Journal of Environmental Research and Public Health*, 19(2):868, 2022.

[14] X. Li, L. Zhang, and Y. Wang. Causal inference of construction safety management measures towards workers' safety behaviours: A multidimensional perspective. *Safety Science*, 172:106432, 2024.

[15] Y.-H. Huang and T.-R. Yang. Exploring on-site safety knowledge transfer in the construction industry. *Sustainability*, 11(22):6426, 2019.

[16] J. Li, Y. Zhou, and B. Guo. Enhancing knowledge of construction safety: A semantic network analysis approach. *Buildings*, 15(17):3036, 2025.

[17] Y. Mei, J. Huang, J. Liu, and L. Jia. A study of factors influencing construction workers' intention to share safety knowledge. *Buildings*, 14(2):440, 2024.

[18] Q. Huang, S. Zhang, and D. Fang. Effects of construction workers' safety knowledge on hazard-identification performance via eye-movement modeling examples training. *Safety Science*, 180:106653, 2024.

[19] P. Ghimire, K. Kim, and M. Acharya. Opportunities and challenges of generative AI in construction industry: Focusing on adoption of text-based models. *Buildings*, 14(1):220, 2024.

[20] R. Nyqvist, A. Peltokorpi, and O. Seppänen. Integration of generative artificial intelligence across construction management. *In Proceedings of the IOP Conference Series: Earth and Environmental Science*, volume 1389(1), pages 012011, 2024. IOP Publishing, August.

[21] K. K. Naji, M. Gunduz, A. Mohamed, and A. Alomari. Generative AI for sustainable project management in the built environment: Trends, challenges, and future directions. *Sustainability*, 17(20):9063, 2025.

[22] V. J. Gan. BIM-based building geometric modeling and automatic generative design for sustainable offsite construction. *Journal of Construction Engineering and Management*, 148(10):04022111, 2022.

[23] K. Cinkusz, J. A. Chudziak, and E. Niewiadomska-Szynkiewicz. Cognitive agents powered by large language models for agile software project management. *Electronics*, 14(1):87, 2024.

[24] N. Rane. (2023). Role of ChatGPT and similar generative artificial intelligence (AI) in construction industry. Available at SSRN 4598258.

[25] UpTrain: Open-source LLM evaluation framework. On-line: https://github.com/uptrain-ai/uptrain, Accessed: 05/13/2025.

OperaSimVR: An Immersive XR Interface for Cyber-Physical Systems in Construction

Kota Akinari[1], Yuichiro Kasahara[1,2], Tomoya Kouno[1], Akinosuke Tsutsumi[1],Genki Yamauchi[3], Daisuke Endo[3], Taro Abe[3], Takeshi Hashimoto[3],Keiji Nagatani[4] and Ryo Kurazume[5]

[1]Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
[2]Graduate School of Science and Technology, University of Tsukuba, Japan
[3]Construction Technology Research Department Advanced Technology Research Team, Public Works Research Institute, Japan
[4]Institute of Systems and Information Engineering, University of Tsukuba, Japan
[5]Faculty of Information Science and Electrical Engineering, Kyushu University, Japan

akinari, kasahara, kouno, tsutsumi@irvs.ait.kyushu-u.ac.jp,
yamauchi-g573bs, endou-d177cl, abe-t757co,t-hashimoto@pwri.go.jp,
nagatani@cs.tsukuba.ac.jp, kurazume@ait.kyushu-u.ac.jp

Abstract -

In recent years, the civil engineering and construction industry has faced a severe shortage of skilled workers due to a declining workforce and the aging of infrastructure. To address this challenge, we have been developing ROS2-TMS for Construction, a cyber-physical system (CPS) platform for earthwork sites that aims to automate earthwork operations using construction machinery. OperaSimVR serves as the cyber component of this CPS platform. It reconstructs earthwork sites in cyberspace, enabling real-time monitoring and remote operation. OperaSimVR has four modes: Standard Mode, Playback Mode, Teleoperation Mode, and Preview Mode. The Preview Mode visualizes the motions of autonomously operating construction machinery several seconds into the future, allowing operators to confirm and assess upcoming actions in advance. This paper presents an overview of OperaSimVR and reports the experimental results of the Preview Mode.

Keywords -

Previewed reality; Automated construction; Cyber-Physical System

1 Introduction

In the construction industry, the number of workers has been steadily declining due to the effects of a decreasing and aging population, as well as a higher incidence of occupational accidents compared with other industries. At the same time, the deterioration of infrastructure built during the period of rapid economic growth has become a serious issue, leading to an increasing demand for civil engineering and construction work. As a result, the shortage of skilled personnel has become increasingly severe.

To address these challenges, we have been developing a cyber-physical system (CPS) platform for earthwork sites, namely ROS2-TMS for Construction, as a system aimed at improving construction efficiency and safety [1]. ROS2-TMS for Construction enables the automatic operation of construction machinery using environmental data acquired from sensors installed in the work environment, thereby realizing the automation of earthwork operations. Furthermore, we have been developing OperaSimVR as the cyber component of the CPS, ROS2-TMS for Construction. OperaSimVR constructs a representation of the real-world earthwork site in a virtual (cyber) space using environmental information obtained from sensors. By immersing themselves in this cyberspace, users (site managers) can observe and monitor the conditions of the earthwork site. In this paper, we present an overview of OperaSimVR and report the experimental results of the Preview Mode.

2 Related work

In the civil engineering and construction industry, initiatives utilizing information technology have also been increasingly promoted. Virtual reality (VR) and augmented reality (AR) technologies are being used in conjunction with BIM (Building Information Modeling) during the design phase and for construction management. BIM is an information-rich building model that integrates a three-dimensional representation of a structure with data required for design, construction, and maintenance, such as material properties, costs, and schedules.

In studies that apply VR during the design phase, BIM is integrated to enable designers to immerse themselves in BIM-generated 3D models and visually verify design details [2]. In research on construction management using AR technology, AR-based management systems have been developed that link with BIM-based construction plans and overlay the designed geometry corresponding to user-specified times or construction stages onto the real

environment. This allows intuitive, on-site comparison between the planned construction process and the current site conditions using devices such as tablets [3].

In addition, studies using AR technologies to compensate for delays during teleoperation have also been actively conducted. Research aimed at mitigating the effects of teleoperation latency by estimating and overlaying the actual robot position includes studies in remote driving, where predictive wheel trajectories are displayed on the operator's head-mounted display (HMD) to improve steering stability degraded by communication delays and insufficient visual information [4]. Similarly, in robot-assisted remote surgery, communication delays between the surgical robot and the remote operator pose significant challenges; therefore, predictive displays based on AR have been proposed for teleoperated surgical robots [5]. However, since these studies primarily focus on compensating for teleoperation delays, they estimate and display the current pose of the robot and do not present future poses.

In environments where robots and humans coexist, such as those involving service robots, it is necessary to prevent collisions between humans and robots. To address this issue, systems have been studied that display a robot's motion several seconds into the future on an AR display, enabling humans to perceive the robot's future trajectory and thereby avoid collisions [6].

In this study, we developed a system that visualizes the future poses of autonomously operating construction machinery to predict collisions in advance at earthwork sites. The novelty of this system lies in its use of AR technology to present future actions to the operator during autonomous operation. Additionally, since it is implemented as part of the construction machinery control system, ROS2-TMS for Construction, both real and virtual construction machines can be operated using the same control commands, making it easy to implement the proposed Previewed Reality.

3 OPERA

OPERA (Open Platform for Earthwork with Robotics and Autonomy) is an open platform being developed by the Public Works Research Institute to standardize control signals for construction machinery and promote the automation of construction equipment [7]. OPERA aims to disseminate automated construction technologies by standardizing construction machinery control signals and adopting ROS/ROS2 as middleware [8]. OPERA includes electronically controlled construction machinery, such as hydraulic excavators and crawler dump trucks, as well as OperaSim, a simulator developed using Unity.

4 ROS2-TMS for Construction

ROS2-TMS for Construction is an automated construction cyber-physical system (CPS) platform developed based on OPERA, an autonomous construction technology framework. ROS2-TMS for Construction executes construction machinery operations based on a given sequence of tasks, using information acquired from sensors installed at the site.

Autonomous construction is performed following the workflow illustrated in Figure 1.

1. Environmental information is collected using sensors installed in the environment and on construction machinery (Measure).

2. Based on the collected environmental information, an earthwork site is reconstructed in real time within cyberspace. The site conditions are then analyzed, and construction machinery operations are planned (Analyze and Plan).

3. Control commands required to execute the planned operations are generated and transmitted to the construction machinery via OPERA (Control).

The architecture of ROS2-TMS for Construction is shown in Figure 2. Each function is modularized, and the system is formed through the coordination and interaction of these modules. To date, we have conducted automation experiments for earthwork operations, including excavation, loading, transportation, and dumping of soil using a hydraulic excavator and a crawler dump truck [9], as well as field deployment experiments for soil improvement construction [10] using ROS2-TMS for Construction.

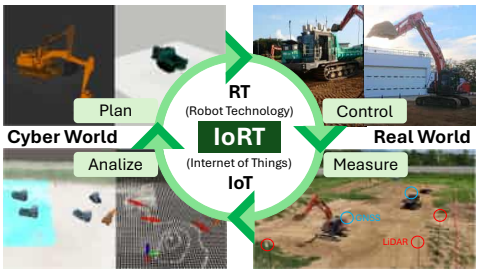


Figure 1. Schematic diagram of ROS2-TMS for Construction

5 OperaSimVR

5.1 Overview of OperaSimVR

OperaSimVR is the cyber component of ROS2-TMS for Construction. It was developed by extending OperaSim,

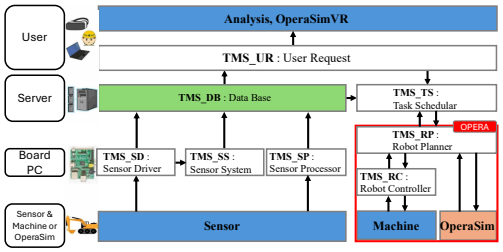


Figure 2. Architecture of ROS2-TMS for Construction

the simulator of OPERA, by adding several new functions in addition to its original simulation capability (Standard Mode). Specifically, visualization (Playback Mode), operation (Teleoperation Mode), and preview (Preview Mode) functionalities were incorporated. By deforming the ground surface of the simulator using terrain data from earthwork sites, OperaSimVR can reproduce earthwork operations such as excavation. In addition, buildings and autonomously operating construction machinery can be modeled, enabling the construction of a cyber representation of an actual earthwork site environment.

Users can immerse themselves in cyberspace by wearing a VR headset, allowing them to walk within the virtual environment, observe the site from a bird's-eye view, or experience views as if seated in the operator's cabin of construction machinery within cyberspace. When a VR headset is not used, the system can be operated via keyboard input through the Unity game view on a PC. Furthermore, real-time 360-degree images of the physical site can be acquired from omnidirectional cameras [11] installed on-site and switched with cyberspace view, enabling users to monitor the actual site conditions.

When using OperaSimVR, a control PC connected to the construction machinery, on which ROS2-TMS for Construction is running, launches the ROS-TCP Endpoint and establishes a connection with the PC used to run OperaSimVR, enabling operation and control through this communication link.

5.2 Standard Mode (OperaSim)

In the Standard Mode of OperaSimVR, the motions of construction machines can be simulated using the OperaSim simulator. OperaSim is the official simulator of OPERA, an autonomous construction technology platform developed by the Public Works Research Institute using Unity and the PhysX and AGX physics engines. In OperaSim, motion commands sent from an operation PC are processed in the same manner as those for actual OPERA-compliant construction machines, including the computation of gravity and friction. When excavation operations are performed, terrain deformation occurs, and soil behaviour can also be simulated.

5.3 Playback Mode

In the Playback Mode, the system acquires the current position information and joint angles transmitted from each construction machine and constructs and visualizes the state of the actual earthwork site within cyberspace. The construction machinery models in cyberspace operate in real time in synchronization with the motions of the corresponding machines in the physical environment. With respect to terrain representation, simulations of excavation operations can be performed in the same manner as in the Standard Mode. The Playback Mode enables remote monitoring of earthwork site conditions without the need to physically enter the site. In addition, it allows construction managers to observe the behaviour of construction machinery during autonomous construction even under conditions with limited visibility, such as nighttime or fog.

5.3.1 Machine location visualization

Position and orientation information are acquired using GNSS and IMU sensors mounted on construction machinery. The obtained position data are represented either in the world geodetic coordinate system (world coordinates) or in a map coordinate system, where the origin is translated to a reference point within the construction site. To visualize these data in cyberspace (Unity), coordinate transformation is required to align the coordinate systems and reference points.

Therefore, OperaSimVR performs internal coordinate transformations to align the reference point in cyberspace with the corresponding reference point of the actual construction site within the virtual construction environment. This process updates the position and orientation of the construction machinery models in cyberspace. Since coordinate systems, reference point locations, and offsets vary depending on the site and the construction machinery, these parameters—including the coordinate system, reference point coordinates, and offsets—can be configured for each site and machine.

As a result, the position and orientation of autonomously operating construction machinery in the physical environment can be visualized in real time within cyberspace.

5.3.2 Machine posture visualization

Joint angle information acquired from sensors such as IMUs mounted on the joints of construction machinery is received by the system. For an excavator, the swing, boom, arm, and bucket are actuated in cyberspace based on the joint angle data obtained from the physical machine. For

a crawler dump truck, the vessel and swing motions are driven according to the joint angle information from the real machine. As a result, the posture of autonomously operating construction machinery in the physical environment can be visualized in real time in cyberspace.

Figure 3 illustrates the visualization, using the Playback Mode of OperaSimVR, of the motions of construction machines autonomously operating with ROS2-TMS for Construction as well as the actual site conditions.



Bird's-eye views of cyberspace Bird's-eye views of real space

Figure 3. Real-time visualization of autonomously operating construction machinery in cyberspace using Playback Mode

5.4 Teleoperation Mode

In the Teleoperation Mode, individual construction machines can be remotely operated from OperaSimVR. According to the "Safety Rules for Automated Construction, Ver. 1.0" established by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan, construction sites must be divided into a "manned area," where operations are conducted by manned construction machinery, and an "unmanned area," where operations are performed by automated or remotely operated construction machinery [12] (Figure 4).

When unexpected events occur during autonomous construction and manual intervention is required, the construction machine operating autonomously within the unmanned area must either be stopped and operated by an on-site operator or controlled through remote operation. In such situations, OperaSimVR enables operators to respond by remotely controlling the construction machinery.

The Teleoperation Mode can be used when OperaSimVR is operating in either the Standard Mode or the Playback Mode. During teleoperation, the operator controls the machine while viewing imagery that simulates being seated in the operator's cabin of the construction machinery model in cyberspace. In the Standard Mode, the physical machine in the real environment is not actuated; instead, the operator controls the construction machinery model in cyberspace to simulate its behaviour. In contrast, in the Playback Mode, the operator can directly control the construction machinery in the physical environment.

The operations that can be performed in the Teleoperation Mode are listed below.

- Ground velocity commands (locomotion commands)
  Ground velocity commands are transmitted in response to inputs from a VR controller or a keyboard. In the Standard Mode, motion commands are directly applied to the construction machinery model in cyberspace. In the Playback Mode, the ground velocity command topic is published, and the construction machine in the real world receives the topic and executes the commanded motion. This enables operations such as forward and backward movement and turning.

- Joint angular velocity commands
  Joint angular velocity commands are transmitted in response to inputs from a VR controller or a keyboard. In the Standard Mode, the commands are directly applied to the construction machinery model in cyberspace. In the Playback Mode, the joint angular velocity command topic is published, and the construction machine in real world receives the topic and performs the corresponding motion. In the case of a hydraulic excavator, this allows the operator to control the swing, boom, arm, and bucket.

When the VR controller or keyboard is operated, motion command topics are published, and the physical construction machinery that receives these topics operates in real time. At the same time, the construction machinery model in cyberspace moves in synchronization with the physical machine, enabling the operator to control the real machine while visually monitoring its motion within cyberspace.

Figure 5 illustrates the teleoperation of a crawler dump truck using OperaSimVR. In each subfigure, the upper left shows the cyberspace view observed by the operator during teleoperation, the upper right shows the teleoperation interface in use, the lower left presents a view from the operator's cabin in the physical environment captured for comparison, and the lower left shows the physical site scene recorded for comparison.

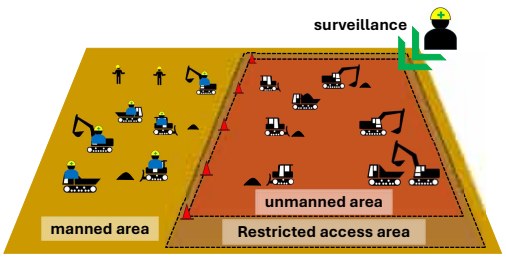


Figure 4. Manned and unmanned areas during autonomous construction [12]

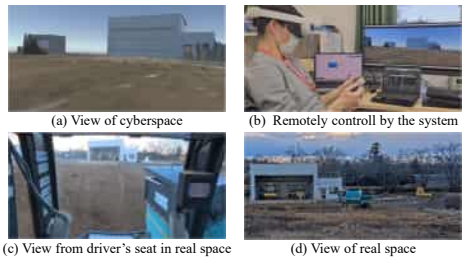


Figure 5. Teleoperation using OperaSimVR

5.5 Preview Mode

In the Preview Mode for autonomous operation, OperaSimVR receives motion plans from autonomously operating construction machinery and visualizes their future motions in cyberspace. According to the "Safety Rules for Automated Construction, Ver. 1.0" established by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan, during autonomous construction, "the operating status of automated construction and its surrounding conditions must be confirmed through visual observation, monitors, GNSS-based position information, and other means, and warnings or emergency stop systems must be activated when necessary." However, merely monitoring autonomously operating construction machinery in real time may be insufficient, as collisions or other hazardous events can still occur even if an emergency stop is triggered after a danger is recognized (Figure 6). To address this issue, the Preview Mode of OperaSimVR enables operators to examine future motions of construction machinery in advance, thereby preventing potential collisions before they occur. OPERA-compatible construction machinery performs path planning using Nav2 [13] and motion planning using MoveIt [14] on an onboard PC upon receiving motion commands, and executes motions based on these plans. Before the corresponding motions are executed on the actual construction machinery, the preview function of OperaSimVR visualizes the paths and motion plans generated by Nav2 and MoveIt in cyber space.

5.5.1 Machine future location visualization

The future positions of construction machinery are visualized as follows. First, the planned paths generated by Nav2 for each construction machine are acquired. Since the obtained paths are represented in a map coordinate system based on a reference point within the construction site, the coordinates are transformed into the Unity coordinate system in the same manner as in the Playback Mode.

Next, the current velocity of each construction machine is computed from its current position information. Using this velocity, the distance traveled over a specified future time interval is calculated. The construction machinery model is then displayed at a point along the planned path that is offset from the current position by this computed distance. If the predicted motion reaches the end of the planned path, the model stops at the terminal point. Through this process, the future positions of autonomously operating construction machinery several seconds ahead can be visualized. At present, future rotational motion is not considered; instead, the model orientation is determined based on the current orientation of the construction machinery. In addition, by visualizing the received planned paths in cyberspace, operators can inspect motions further ahead along the trajectory.

5.5.2 Machine future posture visualization

The future motions of the construction machinery are visualized as follows. First, motion plans generated by MoveIt for each construction machine are acquired. These motion plans are represented as a sequence of planned poses at fixed time intervals. From the acquired motion plans, the pose of the construction machinery at an arbitrary future time is extracted, and a model corresponding to the extracted pose is displayed in cyberspace. If the duration of the acquired motion plan is shorter than the specified future time to be visualized, the posture of the construction machinery model is set to the final pose of the motion plan. Through this process, the future posture of autonomously operating construction machinery several seconds ahead can be visualized.

In cyberspace, both a construction machine model that displays the current position and posture in real time and a construction machine model that represents its future motion are visualized. This enables the simultaneous visualization of both the current posture and the future motion within cyberspace.

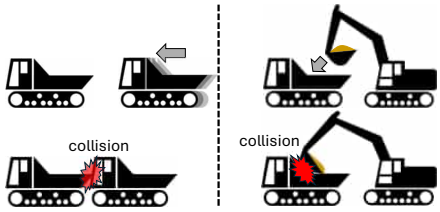


Figure 6. Example of a potential collision between construction vehicles

6 Experiments in Preview Mode

6.1 Experimental Setup

An experiment was conducted using actual construction machinery to evaluate the preview function for au-



Figure 7. IC120



Figure 8. ZX200

tonomous operation described in Chapter 5.5. Using ROS2-TMS for Construction, the construction machinery was operated autonomously to verify that future motions could be previewed. The experiment was carried out at the DX Field of the Public Works Research Institute using OPERA-compatible construction machines: a crawler dump truck (IC120, KATO Works) (Figure 7) and a hydraulic excavator (ZX200, Hitachi Construction Machinery) (Figure 8).

In this experiment, the construction machinery autonomously performed the following sequence of operations:

- Excavation of a soil pile using the hydraulic excavator
- Loading soil onto the crawler dump truck
- After repeated loading operations, the crawler dump truck traveled to the dumping location
- The crawler dump truck returned to the loading location

During the experiment, video streams were recorded from cameras installed in the environment and from a camera mounted in the operator's cabin of the construction machinery. Using these images, we verified the system behaviour when future motions of construction machines in cyberspace were overlaid onto real images using AR.

In OperaSimVR, one translucent model each of a hydraulic excavator (ZX200) and a crawler dump truck (IC120) was used to visualize the future motions of the construction machines. All models other than these two were made transparent, and the recorded real-world camera images were projected as the background, enabling an AR-style overlay of the construction machine models onto the real images.

A PC running ROS2-TMS for Construction was connected to the physical machines, and the ROS-TCP Endpoint was launched to establish communication with the PC running OperaSimVR.

In the experiment, the future motions three seconds ahead were overlaid and visualized.

6.2 Results

Figure 9 and Figure 10 show images in which the future motions of autonomously operating construction machines in the real environment are overlaid using AR. Figure 9 presents a bird's-eye view, while Figure 10 shows the view from the operator's cabin. The translucent models represent the machines visualizing the future motions of the backhoe and the crawler dump truck, respectively. In Figure 9, the excavator's digging and loading operations was visualized by displaying its posture predicted 3 seconds into the future, allowing the future motions to be confirmed. For the crawler dump truck, no future motion is displayed in images 1–7 because no path plan is generated while the excavator is performing the digging and loading operations. After the loading operation is completed, a path plan is received, and the future motion of the crawler dump truck is visualized.

Figure 10 shows the view from the excavator's operator cabin during the digging and loading operations, where the posture predicted 3 seconds into the future is displayed, enabling confirmation of the future motions.

Figure 11 and Figure 12 show images visualizing both the real-time motion and the motion 3 seconds into the future of autonomously operating construction machines in the real environment, as represented in cyberspace using VR.

Figure 13 shows a graph of the joint angles of the backhoe during the experiment, including the angles predicted by the proposed system, the predicted angles delayed by 3 seconds, and the actual joint angles of the backhoe. Since the visualization begins from the posture 3 seconds after receiving the motion plan, motions that are completed within 3 seconds are not shown in the graph. By comparing the predicted angles delayed by 3 seconds with the actual joint angles of the backhoe, it can be observed that, for all joints, the predicted angles generally agree with the actual joint angles. However, for the boom angle, an overshoot occurs in reality, but it is not visualized because it is not included in the motion plan. In future work, we aim to incorporate the overshoot into the visualization so that it can also be represented.

7 Conclusion

This paper proposes OperaSimVR and introduced the Preview Mode. In the experiments, future motions of autonomously operating construction machines were overlaid and visualized in cyberspace, enabling confirmation of their forthcoming behaviour.

In future work, we aim to utilize the passthrough function of VR headsets to overlay predicted future motions onto real-world images. This will enable supervisors to wear AR/VR goggles outside the unmanned area and visually confirm the future motions of autonomously operating construction machines by superimposing them onto the actual machines. However, the current system does not reflect the rotation of the crawler dump truck when indi-



Figure 9. AR view of autonomously operating construction machinery three seconds in the future

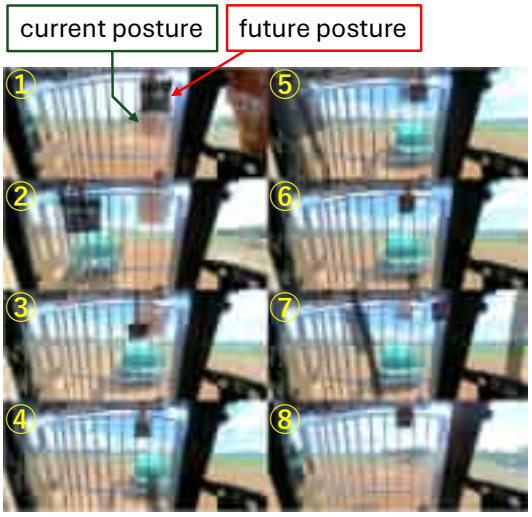


Figure 10. AR view from the cockpit of autonomously operating construction machinery three seconds in the future

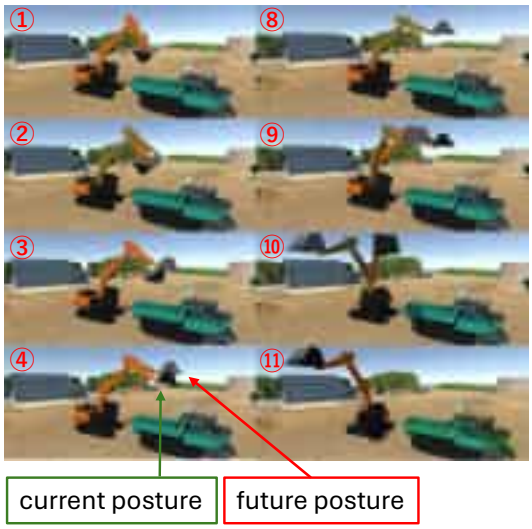


Figure 11. VR view of autonomously operating construction machinery three seconds in the future

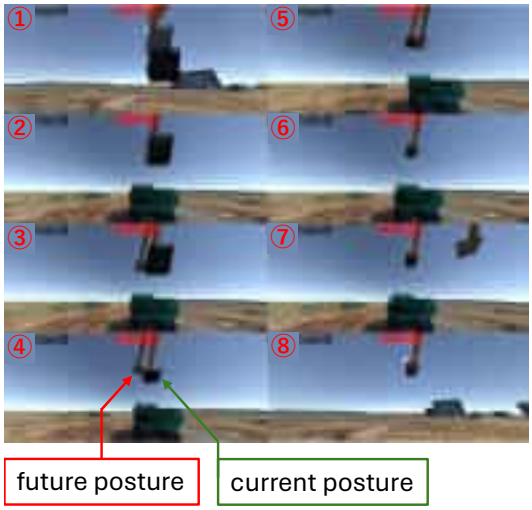


Figure 12. VR view from the cockpit of autonomously operating construction machinery three seconds in the future

cating its future position. Therefore, we plan to implement a method that orients the machinery model along the tangent direction of its path. In addition, we plan to develop a human–machine collaborative framework in which this system enables operators to preview the behaviour of construction machinery in advance. Within this framework, operators can intuitively and seamlessly perceive the predicted future behaviour of the machinery. If any issues are identified, the operator can halt the process; otherwise, the machine will only proceed with the task if no issues are detected. We also plan to add a function that triggers an emergency stop of the construction machine when the model representing its future motion collides with objects such as other construction machine models or structures, thereby preventing collisions. Furthermore, we plan to extend the visualization by adding information such as the final completed shape of earthwork operations.

Acknowledgments

This work was supported by Council for Science, Technology and Innovation(CSTI), Cross-ministerial Strategic

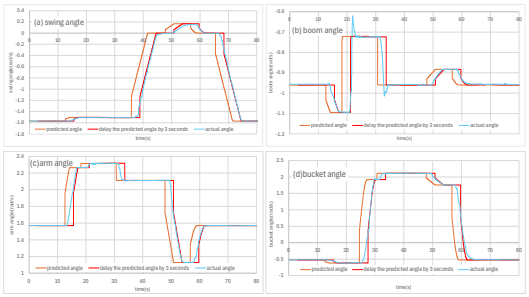


Figure 13. A graph comparing the actual and predicted joint angles of the backhoe during the experiment

Innovation Promotion Program(SIP), the 3rd period of SIP "Smart Infrastructure Management System" Grant Number JPJ012187 (Funding agency: Public Works Research Institute).

References

[1] IRVS. Ros2-tms for construction. On-line: `https://github.com/irvs/ros2_tms_for_construction`, Accessed: 12/17/2025.

[2] V. Rostamiasl and A. Jrade. A cloud-based integration of building information modeling and virtual reality through game engine to facilitate the design of age-in-place homes at the conceptual stage. *Journal of Information Technology in Construction*, 29: 377–399, 2024. doi:10.36680/j.itcon.2024.018.

[3] S. Arvikar, P. P. W. Aung, G. Cha, and S. Park. Augmented reality-based construction site management using optimized bim and project schedule integration. *Automation in Construction*, 175, 2025. doi:10.1016/j.autcon.2025.106204.

[4] Y. Tazaki, R. Marumoto, J. Ohgashira, H. Nagano, Y. Yokokohji, and S. Kameoka. Experimental evaluation of the effect of predictive display on steering stability in remote driving. *International Journal of Automotive Engineering*, 15(2), 2024. doi:10.20485/jsaeijae.15.2_90.

[5] R. Florian, Y. Zhang, Y. Zhi, R. O'Ryan, and M. Yip. Augmented reality predictive displays to help mitigate the effects of delayed telesurgery. In *Proceedings of 2019 International Conference on Robotics and Automation (ICRA)*, Montreal, Canada, 2019. doi:10.1109/ICRA.2019.8794051.

[6] A. Egashira, Y. Horikawa, T. Hayashi, A. Kawamura, and R. Kurazume. Near-future perception system: Previewed reality. *Advanced Robotics*, 35, 2021. doi:10.1080/01691864.2020.1829041.

[7] Public Works Research Institute(PWRI). Pwri-opera. On-line: `https://github.com/pwri-opera`, Accessed: 12/17/2025.

[8] G. Yamauchi, D. Endo, H. Suzuki, and T. Hashimoto. Proposal of an open platform for autonomous construction machinery development. In *Proceedings of 40nd International Symposium on Automation and Robotics in Construction (ISARC 2023)*, Chennai, India, 2023. doi:10.22260/ISARC2023/0027.

[9] Y. Kasahara, T. Istuka, K. Shibata, T. Kouno, R. Maeda, K. Matsumoto, S. Kimura, Y. Fukase, T. Yokoshima, G. Yamauchi, D. Endo, T. Hashimoto, and R. Kurazume. Task management system for construction machinery using the open platform opera. In *Proceedings of 2024 IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, Pasadena, USA, 2024. doi:10.1109/RO-MAN60168.2024.10731421.

[10] T. Kouno, K. Yuichiro, K. Akinari, A. Tsutsumi, T. Hachijo, S. Kimura, Y. Fukase, Y. Miyashita, T. Yokoshima, T. Abe, D. Endo, T. Hashimoto, K. Nagatani, G. Yamauchi, and R. Kurazume. Field implementation of an automated hydraulic excavator using ros2-tms for construction and opera. In *Proceedings of Proceedings of 2026 IEEE/SICE International Symposium on System Integration (SII)*, Cancun, Mexico, 2026.

[11] R. Maeda, T. Kouno, K. Matsumoto, Y. Kasahara, T. Itsuka, K. Nakashima, Y. Tamaishi, and R. Kurazume. Sensor pods and ros2-tms for construction for cyber-physical system at earthwork sites. In *Proceedings of IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, New York, USA, 2024. doi:10.1109/SSRR62954.2024.10770031.

[12] Transport Ministry of Land, Infrastructure and Tourism. Safety rules for automated construction, ver. 1.0". On-line: `https://www.mlit.go.jp/tec/constplan/content/001730920.pdf`, Accessed: 12/17/2025.

[13] Open Navigation. Nav2. On-line: `https://docs.nav2.org/`, Accessed: 12/25/2025.

[14] PickNik Robotics. Moveit. On-line: `https://moveit.picknik.ai/main/index.html`, Accessed: 12/25/2025.

Reconstructing Ground-Penetrating Radargrams into 3D Point Cloud Format for Building Inspection: A Web AR Approach Evaluated by Technology Acceptance Model

Meng Sun[1*], Longyong Wu[1], Ying Katherine Deng[1], Xiao Li[2], and Fan Xue[1]

[1]Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong, China
[2]Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China
sunmhku01@connect.hku.hk, wulongyong@connect.hku.hk, ydeng@hku.hk, shell.x.li@hku.hk, xuef@hku.hk

Abstract –

Non-destructive Testing (NDT) is important for building structure inspection. Ground Penetrating Radar (GPR) is an effective non-destructive testing technique for building inspections, yet the complexity of the radargram data has limited its widespread adoption in engineering practice. This paper presents an approach for reconstructing raw GPR radargrams into 3D point clouds for intuitive visualization and interactive analysis. To evaluate the usability and industry acceptance, this paper conducts a questionnaire survey based on the Technology Acceptance Model (TAM) among 44 industry professionals in Hong Kong. The results indicated that the 3D reconstruction approach received perceived usefulness of 3.26 (Belief that technology improves performance) and perceived ease of use of 3.55 (Belief that technology is easy to use), with the "see-through" function being the most highly regarded. Additional attitude and behavior intention research further confirmed the approach's advantages in terms of intuitiveness and cross-platform compatibility. This paper verified the feasibility and acceptance of reconstructing GPR radargrams into 3D point clouds, which can benefit digital transformation in building inspection.

Keywords –

Ground Penetrating Radar (GPR); Technology Acceptance Model (TAM); 3D reconstruction; Point cloud; Building inspection

1 Introduction

People spend "an average of 87% of their time in enclosed buildings," according to a U.S. national survey [1]. The aging and deterioration of buildings will inevitably lead to a degradation of structural performance, posing a potential threat to residents, public safety, and even cultural heritage buildings. Building inspection is crucial for identifying potential risks and related facility management issues [2, 3].

The verification of internal structures in the past relied on destructive tests, which could leave scars on buildings [4]. Such destructive tests are infeasible for certain types of buildings, such as protected heritage buildings [5]. As a Non-destructive Testing (NDT) method, Ground Penetrating Radar (GPR) has been introduced for subsurface building inspection using electromagnetic signals [6, 7]. GPR can detect subsurface features in as-built buildings and cultural heritage practices [8, 9, 10]. However, a persistent knowledge gap exists between inspectors and the interdisciplinary heritage conservation team, as complex raw radargrams (B-scan/C-scan) are difficult for inspectors to interpret and use in decision-making [4, 11]. Advanced 3D reconstruction is the key to filling the gap in processing the complexity of GPR results, and to enable complex shapes analysis using stereopsis vision [12, 13].

This paper presents a Web AR approach to reconstructing raw GPR radargrams into 3D point clouds (GPR-AR) to overcome subjectivity in conventional interpretations by intuitive, conclusive visual evidence. However, whether this approach can be accepted by industry professionals requires further evaluation and analysis.

The Technology Acceptance Model (TAM) is an information system theory in various fields to interpret and evaluate how users will accept and utilize new technology [2]. TAM clarifies the underlying mechanism of technology adoption by revealing the causal relationships between each external variable and actual system use. The external variables (EVs) encompass external environmental factors that affect potential users; EVs include respondents' backgrounds, technology characteristics, and usage context [14, 15]. The central conceptual component of TAM includes two key constructs, i.e., perceived usefulness (PU) and perceived ease of use (PEOU) [16, 17]. PU refers to the extent to which the respondents believe a particular technology

can enhance their work performance, while PEOU represents the degree to which the respondents believe using this technology is effortless [14].

This paper further presents a TAM-based analysis to assess the acceptance of the proposed GPR-AR. The GPR-AR employed advanced array GPR devices to acquire raw electromagnetic signals and reconstructed them into a 3D point cloud. A TAM-based questionnaire survey targeting professionals was conducted to assess the perceived potential for usability and ease of use. The contribution of this paper is twofold: (i) an intuitive and precise approach to assist the stakeholders in inspecting the building materials; (ii) a high degree of acceptance of reconstructing GPR radargrams into 3D point clouds in the Architecture, Engineering, and Construction (AEC) industry.

2 The 3D Reconstructed GPR in the Web AR (GPR-AR) Approach

2.1 Overview

Fig. 1 shows an overview of the 3D reconstructed GPR presented in the web AR (GPR-AR) approach. It comprises four layers: acquisition, processing, visualization, and interaction. The approach begins with the acquisition layer to collect on-site data. Then, the raw GPR radargram is converted to a 3D point cloud. Next, a 3D reconstruction of GPR scanning results is performed in the processing layer, followed by the visualization layer. The interaction layer provides valuable functions to construction stakeholders and facilitates building inspection.

2.2 Layer-wise technical implementation

The acquisition layer bridges real-world conditions into data. Regarding hardware control, the GPR device's parameters should result in sharp and clear electromagnetic signals. Then, a standard GPR grid scan is carried out to acquire a GPR project data file. Note that GPR data acquisition precedes any processing.

The processing layer generates C-scans and then transforms raw GPR radargrams into 3D point clouds. The Inverse Distance Weighted gridding method is used to convert B-scan data to C-scan data. Besides gridding, cell size, search radius, and moving average are necessary parameters. The spacing between C-scans is set as 0.01m. After the above processing, several C-scan slices can be generated and then passed into the point cloud reconstruction. The outcome point clouds serve for 3D visualization and enable 3D object detection. Each C-scan slice will be converted into a 2D point cloud slice, where the pixel coordinates (x, y) and slice depth (z) are mapped to spatial positions. Rather than RGB colors, the reflection intensity should be labeled at the discretized points. In visualization equalization of GPR-AR, logarithmic and normalization operations lead to define a new index: $RI = (\ln R - \ln R_{\min})/(\ln R_{\max} - \ln R_{\min})$, where R is the reflection intensity, $R_{\min}$ and $R_{\max}$ are the maximum and minimum reflection intensity in all slices. The labeled point cloud slice is finally stacked based on the spatial depth and merged into a structured point cloud.

The third layer visualizes the incoming point cloud in two modalities: isometric 3D and augmented reality (AR). The isometric 3D is convenient for offline visualization and analysis. AR mode aims to map an isometric 3D model for online rendering and visualization.

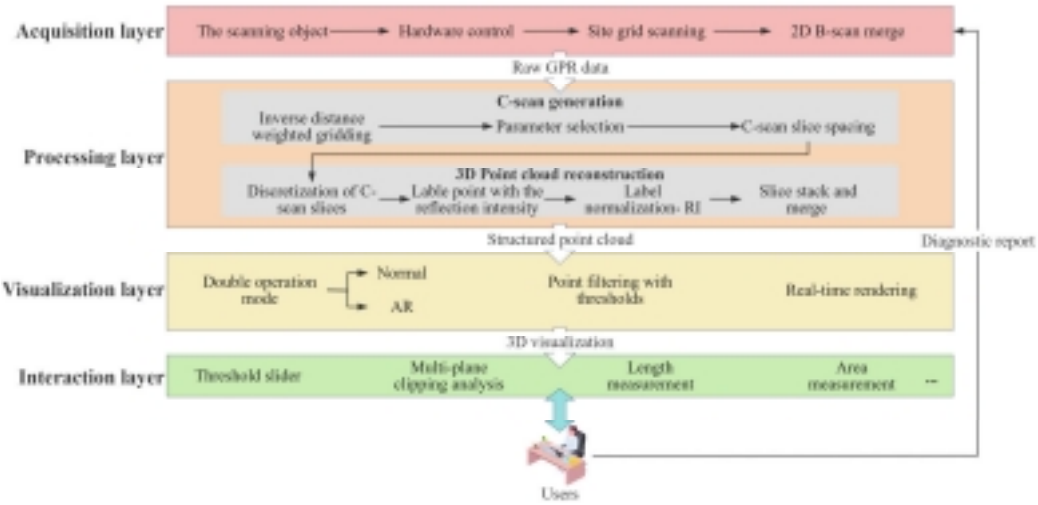


Fig. 1 The technical framework of GPR-AR

User functions in the interaction layer facilitate user operations, as shown in Fig. 2. A sliding filter controls point-level visibility based on the *RI* range. Adequately adjusting the filter hides low-value noises and highlights building elements. The 3D interface adopts adjustable XYZ-axis cutting planes for focused zones. Length and area measurements enable users to efficiently measure spatial data using mouse clicks.

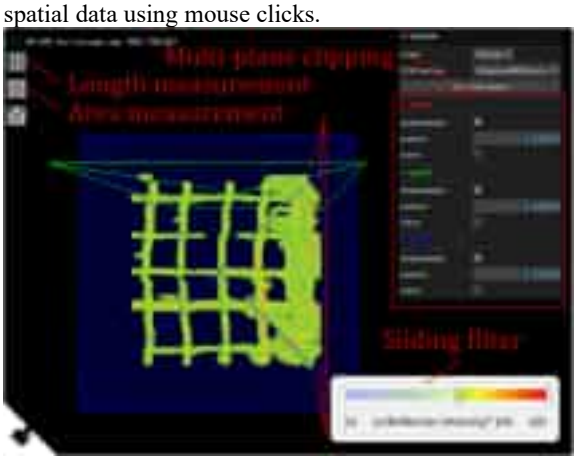


Fig. 2 Interface of the GPR-AR platform

2.3 A demonstration case of reconstruction

Fig. 3 shows a demonstration case of GPR-AR. In this case, a GPR scan was conducted on an aging stair of the MuiWo watchtower in Hong Kong, and the results were processed using GPR-AR. The raw B-scans are isolated and difficult for non-specialists to interpret. In contrast, GPR-AR provides a simple, easy-to-use 3D visualization of results. Using the length measurement function, the spacing between vertical and horizontal rebars was easily measured at 0.17m and 0.16m, respectively. The result was confirmed by the exposed rebar layout in Fig. 3. Two horizontal rebars are detected at a depth of 16 cm, spaced approximately 23 cm apart in the 3D visualization. The user-friendly GPR-AR interface and functions enabled stakeholders to measure and interpret the internal details.

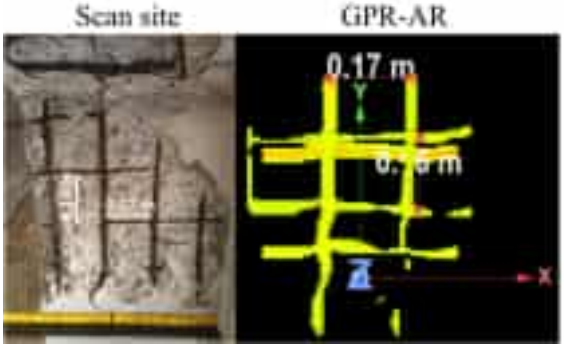


Fig. 3 A demonstration case of GPR-AR

3 Evaluation Using a Questionnaire Survey Based on TAM

3.1 Survey design

A questionnaire survey was designed based on the TAM, as shown in Fig. 4. The respondents were professionals in the AEC industry to reflect the acceptance levels of experts [16, 18, 17]. EVs use respondents' experience and their technology usage context. PU in TAM is measured by evaluating the core functions of the platform, while the overall usability rating measures PEOU. The outcomes of PU and PEOU surveys reflect the respondents' attitudes and behavioral intentions towards use (ABI). This paper employs TAM and designs a structured questionnaire to evaluate the acceptance of the GPR-AR platform among professionals.

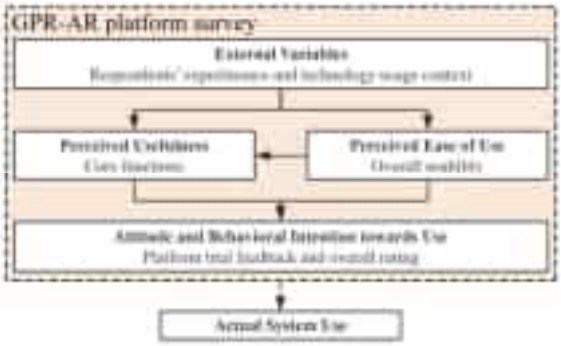


Fig. 4 GPR-AR platform survey based on the TAM [16, 14]

The questionnaire comprises three sections, as listed in Table 1: background information about the respondents, the GPR-AR platform evaluation and open-ended feedback on the GPR-AR platform. Each question corresponds to one specific TAM construct. The questionnaire was electronically distributed to industry professionals after two professional workshops on GPR.

Table 1 The structure of the GPR-AR evaluation questionnaire

Section	Questions	Constructs
Background	Q1 How long have you been working in the AEC industry?	EV
	Q2 How much do you know about GPR?	EV
Evaluation	Q3 How useful do you find GPR technology for building inspection?	PU
	Q4 How intuitive is the "see-through" visualization feature of the GPR-AR platform?	PEOU
	Q5 How useful do you find the length measurement tool in the GPR-AR platform to be for the building inspection?	PU
	Q6 How useful do you find the area measurement tool in the GPR-AR	PU

	platform to be for the building inspection?	
	Q7 How would you rate the ease of use of the GPR platform?	PEOU
	Q8 To what extent can the GPR-AR platform improve the building inspection?	PU
Feedback	Q9 Which aspect of the GPR-AR platform needs the most improvement?	ABI
	Q10 What other suggestions do you have for developing and applying the GPR-AR platform?	ABI

(Note: EV refers to external variables; PU refers to perceived usefulness; PEOU refers to perceived ease of use; ABI refers to attitudes and behavioral intentions towards use)

A five-point Likert scale was adopted in the evaluation [3]. Four questions correspond to PU (Q3, 5, 6, 8) to collect respondents' responses, including Not at all helpful = 1, Slightly helpful = 2, Moderately helpful = 3, Very helpful = 4, and Indispensable = 5. For the question corresponding to PEOU (Q4, 7), a five-point Likert scale was also used, where Very difficult = 1, Somewhat difficult = 2, Neutral = 3, Somewhat easy = 4, and Very easy = 5. Notably, although the "see-through" visualization is the prerequisite for platform use, it primarily serves to reduce the burden of data interpretation. Therefore, Q4 was classified as PEOU, which in turn has an indirect impact on PU.

3.2 Data collection

To ensure the accuracy and relevance of the investigation, the target population was limited to active professionals in the fields of construction, safety inspection, and heritage protection in Hong Kong. These professionals, selected from the Hong Kong Institute of Engineers (HKIE, 90.9%) and the Society of Registered Safety Officers (SRSO, 9.1%), were affiliated with contractors, developers, government agencies, and universities. All participants hold a high degree of professional standing and technical expertise. The survey was conducted during industry seminars and workshops from June to July 2025, ensuring that the feedback received was from practitioners familiar with on-site inspection challenges and emerging digital tools.

The collected data was processed into three categories. The data collected in the background section were categorical. The data in Section Evaluation are ordinal (e.g., Indispensable = 5 in the PU-related questions). The data in Section Feedback is open-ended qualitative data. All the ordinal data were analyzed using descriptive statistics to evaluate the GPR-AR platform. Means, variances, and standard deviations were calculated for each question to summarize central tendency and response dispersion.

4 Findings

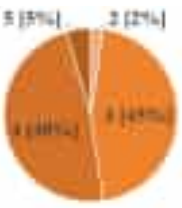
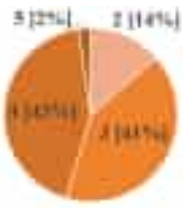
4.1 Questionnaire survey results

A thorough survey was carried out to evaluate the functions and applications of the GPR-AR platform. Among the 44 respondents, more than 75% have at least 1 year of professional experience in building construction or building safety, with 21 having over 10 years.

For those professional respondents, 45% had no prior knowledge of GPR, while others had only basic awareness (41%) or were familiar with general principles (14%), indicating that GPR is well-known but not widely applied within the construction and safety industry.

The evaluation results of three core functions are listed in Table 2. "See-through visualization" was the most recognized function, with more than half of the respondents finding it very helpful (score ⩾ 4). Length and area measurement are also considered essential

Table 2 Evaluation of core functions on the GPR-AR platform

Functions	"See-through" visualization	Length measurement	Area measurement
Construct	PEOU	PU	PU
Mean	3.55	3.30	3.34
Variance	0.38	0.39	0.54
Standard deviation	0.62	0.62	0.74
Distribution			

1: Not at all helpful; 2: Slightly helpful; 3: Moderately helpful; 4: Very helpful; 5: Indispensable

functions in engineering practice. Moreover, these industry stakeholders believed that the GPR-AR platform should be improved in several key functions of concern in the industry, including interface design (mentioned by 52%), functional design (50%), and processing speed (26%). Some professionals also suggested that conversion from a 3D scan to a 2D drawing/BIM is needed [19].

4.2 TAM model assessment

Four-item PU and two-item PEOU constructs were created to calculate composite scores (Table 2). The reliability of the PU was tested using Cronbach's alpha value (α), while the internal consistency of PEOU was evaluated through the correlation coefficient between the only two items [20]. For the ordered Likert-scale data, a nonparametric Spearman rank correlation analysis was employed to examine the influence of respondent EVs on platform evaluation [3].

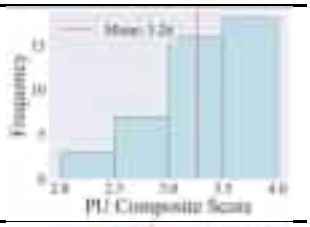
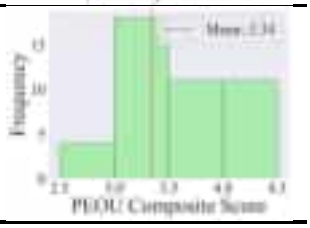
As shown in Table 3, PU demonstrated excellent reliability ($\alpha = 0.803$) based on its four-item scale. The PU composite was 3.26 (mean score of four items), with a distribution indicating a tendency towards higher scores. PEOU only included two items that significantly reduced α (0.434); therefore, an alternative 2-item correlation analysis was applied [21]. The 2-item construct demonstrated acceptable reliability for short scales, as indicated by the Spearman coefficient ($\rho > 0.30$, $p < 0.05$). Consequently, the 2-item composite score was calculated to provide preliminary evidence of internal consistency.

loading and composite reliability (CR). In Table 3, the items demonstrated significance and had factor loadings greater than 0.4 (Q4's was close to 0.4), which can be considered a sufficient variance-explained rate for their construct. CR of PU was greater than 0.7, indicating good aggregation validity; CR of PEOU was within (0.4, 0.7), indicating a moderate but acceptable aggregation effect.

The distributions of PU and PEOU by EVs (professional experience and GPR knowledge level) were elaborated in Fig. 5. PU of GPR-AR by professionals was significantly influenced by their professional background. PU had a "U-shaped relationship" with their professional experience, while PU was positively correlated with the GPR knowledge level. This indicates that GPR-AR is of great value to new practitioners and experienced professionals alike. On the other hand, PEOU was not significantly affected by their GPR knowledge level. This is because ease of use is an objective attribute of a technology, while usefulness is an evaluation that depends on the cognition and needs.

A Spearman correlation matrix was established to analyze the correlations among EV, PEOU, and PU. The correlation between EV and other constructs indicated the extent to which technology has become popularized. The work experience and GPR knowledge in EV were mutually independent. It was found that the correlation between the two items and other assessment variables was insignificant (Table 4). Organizations and stakeholders continue to maintain a conservative attitude, which aligns with the TAM. There was a significant correlation between PU and PEOU ($r = 0.654$), consistent

Table 3 Questionnaire analysis of evaluation items

Construct	Item	Factor Loading	α	CR	Distribution
Perceived usefulness (PU)	Q3	0.357	0.803	0.821	
	Q5	0.894			
	Q6	0.850			
	Q8	0.784			
Perceived ease of use (PEOU)	Q4	0.487	0.434	0.437	
	Q7	0.569			

Confirmatory factor analysis (CFA) assesses whether the measurement model accurately reflects the construct or variable being measured. Content validity was confirmed through classification using TAM. Convergent validity was evaluated by the extracted factor

with a prior TAM study [22]. This positive correlation reflects the interrelationship between PU and PEOU. Especially in the context of emerging technologies, PEOU has a significant impact on PU.

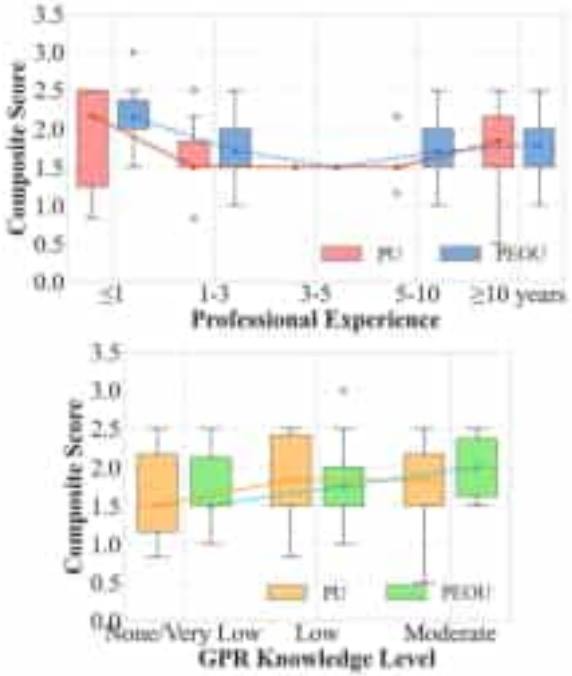


Fig. 5 Distribution of PU and PEOU by EVs

Table 4 Discriminant validity: Spearman correlation and AVE root value

	EV	PU	PEOU
EV	0.927		
PU	-0.077 (p=0.618)*	0.744	
PEOU	-0.165 (p=0.284)*	0.654 (p=0.001)*	0.53

*Values in parentheses represent the significance level, and the diagonal values are $\sqrt{AVE}$

4.3 Attitude and behavior intention from students' feedback

Given the limited practical experience with GPR among professionals in the above survey, it is valuable to gather additional comparative feedback from users with technical awareness. An exploratory study was conducted targeting 70 postgraduates with primary knowledge of GPR. Comparative evaluations and development suggestions for the GPR-AR platform and the mainstream commercial software *Geolitix* were collected. The thematic analysis method was employed to code and categorize the textual data, with a focus on respondents' attitudes and behavior intentions towards the platform. This provides an additional perspective for the TAM study, jointly forming a triangular validation framework for evaluating GPR-AR.

The live poll indicated that students recognized that GPR-AR can improve: (i) 3D presentation of subsurface over 2D sections (93%), (ii) uses for non-destructive testing (81%), and accurate rebar locations/details (76%). We received 15 pieces of feedback comparing GPR-AR with the mainstream GPR processing software *Geolitix*. Generally, it was agreed that GPR-AR has advantages in terms of intuitive 3D visualization, a lightweight user experience, an open-source and free nature, and cross-platform compatibility. However, the feedback also revealed deficiencies in functional depth and data analysis capabilities. It supplemented the ABI of the TAM-based survey and provided a clear direction for the functional optimization of the web AR approach.

5 Discussion and Limitations

5.1 Survey analysis

The questionnaire design in this paper was based on the TAM, aiming to structurally assess the potential application of the GPR-AR platform for building inspection. The survey results indicated that respondents had a high PU, especially in the "see-through" visualization aspect. The PEOU also reached a positive level, indicating that the platform has a strong foundation of user acceptance. Few correlations were identified between EV and the evaluation variables (Table 4), which demonstrates the ease of use of GPR-AR.

Although the Q3 description directly focuses on usability, limited familiarity with GPR and the wording of Q3 hinder respondents from clearly distinguishing between PU and PEOU for an emerging technology, resulting in a relatively low factor loading. Besides, correlated PU and PEOU ($r = 0.654$, $p < 0.001$) in the TAM results in PEOU items inadvertently capturing some of the PU construct [16, 22]. The boundary between "usability" and "ease of use" among these professionals, lacking sufficient experience, may be blurred [23]. Additionally, reliability analysis for PEOU involving a small number of items may limit the accuracy. Future research should adopt a more comprehensive PU and PEOU scale to clarify reliability and validity.

Based on the ABI feedback from students and the results of industry questionnaires, the platform is expected to be systematically enhanced. Expected improvements include optimizing user interface interactions, improving the intuitiveness of operations, and developing advanced analysis functions, such as machine-learning-based rebar recognition and intelligent defect diagnosis [24]. More case studies will also be used to verify engineering usability and ease of use.

In summary, the web AR approach (GPR-AR) has addressed the limitations of interpreting 1D and 2D GPR radargrams in building inspection. The investigation

results regarding the GPR-AR platform indicate that its visualization function effectively leverages augmented reality to make complex GPR data more comprehensible and accessible. This TAM analysis demonstrates that reconstructing raw GPR radargrams into 3D point clouds effectively reduces professional barriers to GPR, making it more accessible to diverse stakeholders.

5.2 Limitations and prospects of GPR-AR

While the TAM-based survey indicates that GPR-AR could contribute to AEC practice, it is limited to directly measuring improvements in inspection. Future research will focus on objective performance metrics using GPR-AR, such as accuracy and time efficiency, to verify perceptions.

Beyond the perceptual scope of this study, several technical challenges should be addressed for the wide adoption of GPR-AR. Firstly, the next phase of research should focus on the precise co-registration of volumetric point clouds between the surface and the subsurface. Secondly, the operation-led trajectory deviations will affect the accuracy of the results. Reducing the influence of operational errors during the analysis process is a research goal for future development. Moreover, the 3D model is currently an intuitive visualization without any annotations or segmentation. Future implementations aim to integrate semantic models and geometric algorithms to achieve a detailed digital twin model [25, 26, 27].

Despite these challenges, GPR-AR holds significant promise for a wide range of applications in the AEC industry. During building inspections, intuitive 3D visualization enables non-professional stakeholders to locate internal rebars and detect defects without destructive tests. In the field of heritage protection, the approach is helpful for conducting subsurface heritage documentation. In addition, the approach's intuitive, easy-to-use nature makes it more convenient for collaborative infrastructure monitoring. These application scenarios demonstrate the great potential of GPR-AR.

6 Conclusion

This paper integrates the reconstruction of Ground Penetrating Radar (GPR) radargrams into 3D point clouds within a web AR platform for building inspection. A questionnaire survey based on the Technology Acceptance Model (TAM) revealed that this integration was well accepted. The scores for perceived usefulness (PU) and perceived ease of use (PEOU) performed well, indicating potential application in engineering practice.

Future work can focus on the following aspects: achieving precise registration of surface and subsurface point clouds, optimizing operation-leading errors, and integrating machine learning algorithms to enable automatic object recognition. This research provides empirical evidence for the digital applications of GPR in building inspections, which align with Hong Kong's smart city development goals and the global digital transformation of the construction industry.

References

[1] N. E. Klepeis et al., "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants," *Journal of Exposure Science & Environmental Epidemiology*, vol. 11, no. 3, pp. 231-252, 2001. https://doi.org/10.1038/sj.jea.7500165

[2] Tsukasa Mizutani, Jingzi Chen, and Shuto Yotsumoto, "The 3D localization of subsurface pipes from Ground Penetrating Radar images using edge detection and point cloud segmentation," *Measurement*, vol. 236, p. 115102, 2024. [Online]. https://doi.org/10.1016/j.measurement.2024.115102

[3] Songbo Hu, Yihai Fang, Eric Windholz, and Kourosh Khoshelham, "Factors influencing the adoption of remote visual inspection in building regulatory activities," *Journal of Management in Engineering*, vol. 41, p. 05025002, 2025. [Online]. https://doi.org/10.1061/JMENEA.MEENG-6559

[4] Wenting Zhan, Wei Pan, Winnie Law, and Haipeng Shen, "Stakeholders' knowledge, attitudes, and intentions of adopting Modular Integrated Construction for sustainable development in Hong Kong," *Journal of Management in Engineering*, vol. 40, p. 04023070, 2024. [Online]. https://doi.org/10.1061/JMENEA.MEENG-5642

[5] D. M. McCann and M. C. Forde, "Review of NDT methods in the assessment of concrete and masonry structures," *NDT & E International*, vol. 34, pp. 71–84, 2001. [Online]. https://doi.org/10.1016/S0963-8695(00)00032-3

[6] Wallace Wai-Lok Lai, Xavier Dérobert, and Peter Annan, "A review of Ground Penetrating Radar application in civil engineering: A 30-year journey from Locating and Testing to Imaging and Diagnosis," *NDT & E International*, vol. 96, pp. 58–78, 2018. [Online]. https://doi.org/10.1016/j.ndteint.2017.04.002

[7] Kevin Lee and Chen Feng, "GPR-Former: Context-aware moisture detection," *International Symposium on Automation and Robotics in Construction (ISARC) Proceedings*, vol. 2025 Proceedings of the 42nd ISARC, Montreal, Canada, pp. 1–8, 2025. [Online]. https://doi.org/10.22260/ISARC2025/0002

[8] Eduardo Diz-Mellado et al., "Non-destructive testing and Finite Element Method integrated procedure for heritage diagnosis: The Seville

Cathedral case study," *Journal of Building Engineering*, vol. 37, p. 102134, 2021. [Online]. https://doi.org/10.1016/j.jobe.2020.102134

[9] Anne Lalagüe, Matthew A. Lebens, Inge Hoff, and Eivind Grøv, "Detection of rockfall on a tunnel concrete lining with Ground-Penetrating Radar (GPR)," *Rock Mechanics and Rock Engineering*, vol. 49, pp. 2811–2823, 2016. [Online]. https://doi.org/10.1007/s00603-016-0943-y

[10] Mercedes Solla et al., "A Building Information Modeling approach to integrate geomatic data for the documentation and preservation of cultural heritage," *Remote Sensing*, vol. 12, p. 4028, 2020. [Online]. https://doi.org/10.3390/rs12244028

[11] F. Hassani, "Documentation of cultural heritage; techniques, potentials, and constraints," , vol. XL-5-W7, 2015, pp. 207–214. [Online]. https://doi.org/10.5194/isprsarchives-XL-5-W7-207-2015

[12] Alexander Bornik and Wolfgang Neubauer, "3D Visualization techniques for analysis and archaeological interpretation of GPR data," *Remote Sensing*, vol. 14, p. 1709, 2022. [Online]. https://doi.org/10.3390/rs14071709

[13] Fan Xue, Weisheng Lu, Christopher J. Webster, and Ke Chen, "A derivative-free optimization-based approach for detecting architectural symmetries from 3D point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 32–40, 2019. [Online]. https://doi.org/10.1016/j.isprsjprs.2018.12.005

[14] Stanisław Fel, Jarosław Kozak, and Piotr Horodyski, "Responsibility and AI: Exploring technology acceptance models," *Journal of Innovation & Knowledge*, vol. 10, p. 100852, 2025. [Online]. https://doi.org/10.1016/j.jik.2025.100852

[15] Jiwoong Shin, Sungwoo Moon, Bong-ho Cho, Sungjoo Hwang, and Byungjoo Choi, "Extended technology acceptance model to explain the mechanism of modular construction adoption," *Journal of Cleaner Production*, vol. 342, p. 130963, 2022. [Online]. https://doi.org/10.1016/j.jclepro.2022.130963

[16] Fred D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, vol. 13, pp. 319–340, 1989. [Online]. https://doi.org/10.2307/249008

[17] Viswanath Venkatesh, "Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the Technology Acceptance Model," *Information Systems Research*, vol. 11, pp. 342–365, 2000. [Online]. https://doi.org/10.1287/isre.11.4.342.11872

[18] Nikola Marangunić and Andrina Granić, "Technology acceptance model: a literature review from 1986 to 2013," *Universal Access in the Information Society*, vol. 14, pp. 81–95, 2015. [Online]. https://doi.org/10.1007/s10209-014-0348-1

[19] Alexis Toledo Bórquez, Felipe Muñoz La Rivera, Rodrigo F. Herrera, and Javier Mora Serrano, "Augmented reality and ultrasonic sensor-based tool for rebar inspection on construction sites," *Automation in Construction*, vol. 176, p. 106288, 2025. [Online]. https://doi.org/10.1016/j.autcon.2025.106288

[20] Lee J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, pp. 297–334, 1951. [Online]. https://doi.org/10.1007/BF02310555

[21] Jose M. Cortina, "What is coefficient alpha? An examination of theory and applications," *Journal of Applied Psychology*, vol. 78, pp. 98–104, 1993. [Online]. https://doi.org/10.1037/0021-9010.78.1.98

[22] William R. King and Jun He, "A meta-analysis of the technology acceptance model," *Information & Management*, vol. 43, pp. 740–755, 2006. [Online]. https://doi.org/10.1016/j.im.2006.05.003

[23] Mark Turner, Barbara Kitchenham, Pearl Brereton, Stuart Charters, and David Budgen, "Does the technology acceptance model predict actual use? A systematic literature review," *Information and Software Technology*, vol. 52, pp. 463–479, 2010. [Online]. https://doi.org/10.1016/j.infsof.2009.11.005

[24] Meng Sun, Dong Liang, Jiajia Wang, Bate Bate, and Fan Xue, "Clarity in DEM cementation predictions: Integrating automated machine learning and interpretability analysis of shear wave velocity," *Advanced Engineering Informatics,* vol. 69, p. 104060, 2026. [Online]. https://doi.org/10.1016/j.aei.2025.104060

[25] Fan Xue, Wenjin Zhang, Guangji Xu, Qqianyun Zhou, and Yijie Wu, "Surface or skeleton? Automatic hierarchical clustering of 3D point clouds of bronze frog drums for heritage digital twins," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-M-1-2023, pp. 293–299, 2023. [Online]. https://doi.org/10.5194/isprs-annals-X-M-1-2023-293-2023

[26] Meng Sun, Rui Zhao, and Fan Xue, "The Opportunities and Challenges of Multimodal GenAI in the Construction Industry: A Brief Review," *The 29th International Symposium on Advancement of Construction Management and Real Estate (CRIOCM2024)*, 2024

[27] Dong Liang et al, "Transfer learning from building information model-based synthetic data for three-dimensional module detection in point clouds of modular-integrated construction hoisting," *Engineering Applications of Artificial Intelligence,* vol. 164, p. 113243, 2026. [Online]. https://doi.org/10.1016/j.engappai.2025.113243

The information backbone of robotized construction: Principles, procedures, and software architecture

A. Borrmann[1], T. Bruckmann[2], K. Dörfler[3], T. Hartmann[4], K. Smarsly[5], M. Brinkhoff[1], A. Das[3], S. Esser[1], M. Hassaan[1], C. Jeziorek[2], N. Kim[1], M.R. Kolani[1], A. Maurya[4], S. Nousias[1], P. Petropoulakis[1], A. Tandon[5], T. Wrabel[1]

[1]Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2]Chair of Mechatronics, University of Duisburg-Essen, Germany
[3]Digital Fabrication, Technical University of Munich, Germany
[4]Civil Systems Engineering, Technical University of Berlin, Germany
[5]Digital and Autonomous Construction, Hamburg University of Technology, Germany

andre.borrmann@tum.de

Abstract -

Advances in robotic systems, sensing technologies, and artificial intelligence offer substantial potential to automate construction processes. However, current construction robotics applications are limited to specific tasks in isolated systems, lacking the integration required for coordinated, large-scale robotized construction that is applicable in a wide spectrum of trades. A key challenge is the absence of digital frameworks integrating product information, construction processes, robotic capabilities, and execution feedback across planning, simulation, and on-site operation. This paper presents the concept, principles, and software architecture of the information backbone of robotized construction (IBRC), a unifying framework for planning, simulating, and executing construction projects using coordinated, heterogeneous robotic systems. The IBRC comprises two tightly coupled environments: an offline planning environment and an online execution environment. The planning environment supports successive refinement of product and process information, starting from a building information model (BIM), via a fabrication information model (FIM), to a robot execution model (REM). The refinement is realized through a graph-based product–process representation and rule-based construction recipes, enabling the derivation of atomic construction actions at a granularity suitable for robotic execution. The execution environment builds on the results and supports both simulated and real-world operation. It relies on middleware, such as ROS 2 for real-time robot communication or Apache Kafka for scalable integration with higher-level services. Central orchestration is achieved through behavior-tree-based multi-robot coordination.

Keywords -

Robotized Construction; Information Backbone; Construction Robotics; BIM-based Process Planning; Multi-Robot Coordination; Fabrication Information Model

1 Introduction

Recent advancements in mechatronics, sensor systems, and artificial intelligence have substantially enhanced robotic capabilities, revolutionizing industrial manufacturing and increasing automation across many industries. It is, therefore, not surprising that robotic systems have made significant advancements in construction environments over the past years [1]. However, the robotic automation in construction thus far has been limited to isolated systems and highly specialized processes. Hence, a paradigm shift towards realizing a comprehensive system-of-systems approach is required that facilitates heterogeneous fleets of autonomous and semi-autonomous robotic systems to collaboratively work alongside humans in a well-coordinated manner on-site [2]. Furthermore, to achieve reliable robotized construction, adequate and precise information about the buildings under construction, the construction methods, processes and progress, the position, action, and interaction of robots, as well as the status of the site as a whole, must be available in a coherent digital representation. On the basis of the information, the central coordination of all on-site robotic activities must be established. Today, appropriate information backbones for simulating and executing robotized construction processes do not yet exist. Therefore, this paper proposes the concept of an information backbone of robotized construction (IBRC) that provides these functionalities [3].

The IBRC aims to provide a framework that enables a virtual laboratory environment where new robotized construction methods can be tested and developed. The framework provides fundamental services, such as digital representations of the facility under construction and the involved processes, and, at the same time, provides flexible interfaces to allow the integration of any type of robotic system, sensors, and construction methods. This paper presents the general concepts underlying the IRBC. To this end, rather than describing isolated applications, robotized construction is framed as a two-phase workflow, encompassing (1) the configuration and planning of a largely robotized construction project and (2) its simulation and on-site execution through the orchestration of a fleet of heterogeneous robots across diverse construction tasks.

Regarding (1), the paper defines a digital workflow in

which product and process information are represented as interdependent graph data structures. These graph representations are co-evolved and mutually constrained, enabling the consistent formulation of constructability, sequencing, and system capabilities. Most importantly, the workflow results in a set of atomic, executable actions that can be assigned to robotic systems. Regarding (2), the simulation environments, communication protocols, and interactions with a central coordination system are discussed.

2 State of the art

A crucial aspect of construction robot task and motion planning is the effective use of BIM and construction schedules to generate executable plans. In early work on integrating BIM with robotics, Meschini et al. [4] highlighted the need to link BIM with the Robot Operating System (ROS) [5] to support construction automation. Their study demonstrated scenarios such as precast timber module installation and robotic assistance for elderly care, showing how BIM-based models can be used to evaluate alternative robotization strategies. However, it remained at a high level and did not provide a concrete BIM–ROS integration method. Addressing this limitation, Ibrahim et al. [6] proposed mapping the as-built state of an indoor construction site by generating waypoints in BIM, transforming them into robot navigation coordinates, and performing LiDAR-based mapping during execution.

Similarly, Mantha et al. [7] proposed task allocation and route planning for multiple service robots using simplified 2D building representations, reflecting the broader limitation of insufficient detail in construction schedules. In parallel, CAD-based and geometry-driven frameworks such as compas-fab [8] provide more direct links between design models and robotic task planning, simulation, and execution.

According to the process hierarchy introduced by Everett and Slocum [9], manufacturing-oriented robot process planning operates at the level of elemental motions, whereas construction schedules are typically specified only at the level of coarse-grained activities. This mismatch creates a fundamental gap between the information available in construction plans and the level of detail required for robotic execution. Even with BIM, it is impractical for planners to manually define robot-level tasks at this granularity, and recent approaches [10, 11, 6] have not yet closed this gap.

One of the few approaches addressing this issue is Kim et al. [12], who extend BIM with task-planning information to generate motion-level robot behaviors. Their ROS-based prototype for indoor wall painting includes conversion from IFC to a ROS-compliant world model and integrated localization, navigation, and motion planning.

Related work by Zhu et al. [13, 14, 15] focuses on robotic assembly of prefabricated elements. The smart construction object (SCO) concept assigns state and requirements to components to coordinate multiple robots, and subsequent work enables automatic conversion from IFC to simulation description format (SDF) to create smart building components for simulation.

Despite these advances, most approaches remain restricted to comparatively simple assembly scenarios and lack generality. A comprehensive framework capable of planning and executing diverse robotized construction processes in a coordinated manner is therefore still required.

3 The information backbone: Overall architecture and components

The proposed framework for planning and executing robotized construction projects is organized around an information backbone that integrates product and process representations across planning and execution stages. The backbone provides a consistent data model and set of interfaces through which design information, construction logic, and robotic task specifications are incrementally developed, validated, and executed. While the detailed implementation is described in [16], this section outlines the underlying principles, the overall software architecture, and the interaction between the core components of the IBRC.

At the architectural level, the proposed system consists of two tightly coupled environments:

1. The **planning environment** (“offline” components) that allows for the iterative development of the construction project, starting from the overall building design down to precise instructions for individual robotic systems.
2. The **execution environment** (“online” components) that allows for the steering and coordination of a heterogeneous fleet of robotic systems, implementing the assigned tasks to complete the construction project, while being able to react to unforeseen events. The execution environment is designed to ensure a seamless transition between simulation-based validation and execution in physical environments.

Both environments operate on a shared information backbone that encodes the evolving relationship between the construction product and process, and may employ model-based methods and process simulations.

3.1 Planning environment

The planning module of the IBRC enables the successive and parallel refinement of both the product (= building components) and the process representation. The module

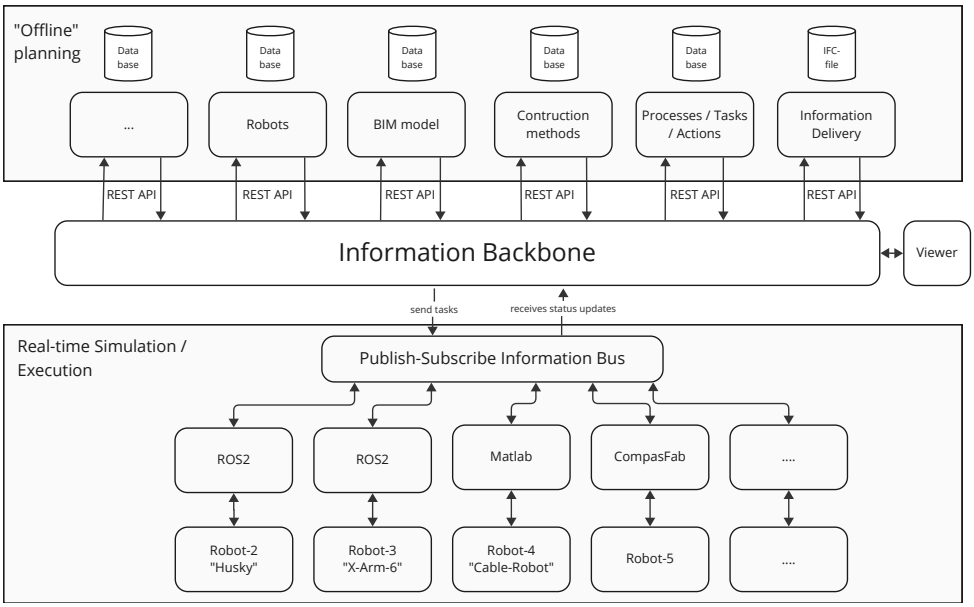


Figure 1. The architecture of the information backbone consists of two major parts: The planning environment and the execution environment

consists of several modules that provide or process data in a service-oriented manner [16].

3.1.1 BIM to FIM to REM: Graph-based refinement of products and processes

The product-process elaboration process is outlined in Fig. 2 and described in detail in [17]. The starting point is a relatively coarse BIM model of the building to be constructed, comparable to an as-designed LOD300 model. The model is represented as a graph that combines two trees, with one tree representing the hierarchical breakdown of the product side and the other representing the hierarchical breakdown of the process side. The process tree is automatically derived from the product graph. Product nodes are linked to the set of process nodes that realize the construction.

By assigning construction methods to individual building components, such as *in-situ concreting, bricklaying,* or *additive manufacturing*, both the product and process graphs are refined. After refinement, the model is referred to as a Fabrication Information Model (FIM) and has a much finer resolution than typical BIM models [18]. FIM contains and represents all atomic – in the sense of indivisible – elements that have to be manipulated for the production of a building element. For example, assigning the construction method *bricklaying* to a wall generates the FIM that contains all individual bricks and the mortar; while assigning the construction method cast-in-situ creates a FIM that represents the concrete, reinforcement, and temporary formwork elements. At the same time, the process graph is further refined and now contains individual activities such as *erect formwork*, *insert reinforcement*, *fill concrete*, and *remove formwork*.

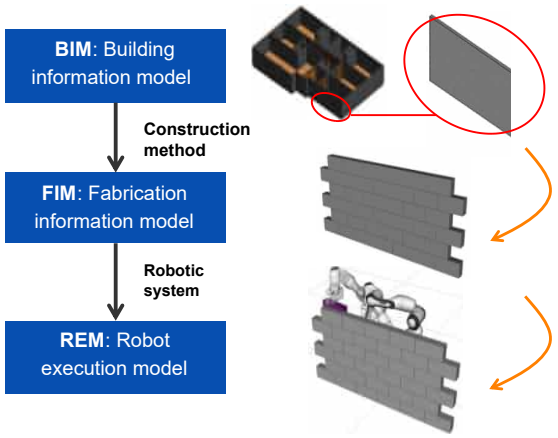


Figure 2. From BIM to FIM to REM

The simultaneous refinement of product and process graph is performed by applying *recipes* that encapsulate the knowledge about specific construction methods [19]. Recipes not only represent the rules for creating sub-nodes for individual process nodes and product nodes, but also define precedence relationships and the required resources (including robots, cranes, construction machines, or human workers). For example, to start "*Fill Concrete*", the activity "*Erect Formwork*" must be completed. For placing formwork and inserting reinforcement, a (robotized) crane is required. The information might later be used e.g. in an event simulation to experiment with the effects that various robot allocation strategies in terms of the amount

and type of robots will have on the overall construction schedule, as well as on construction costs.

The product-process graph is refined down to a level where leaf nodes represent atomic elements (e.g., bucket) and atomic actions (e.g., FillBucket), as shown in Fig. 3. Atomic actions are later transformed into manipulation instructions sent to the individual robotic systems. The instructions are described at a level of abstraction that is executable by robotic systems while omitting low-level motion details, allowing a certain degree of interpretation and thus granting robotic systems the autonomy to plan execution. For example, the *transport formwork* action, sent to a robotized crane, specifies only start and end positions, not the exact crane hook trajectory or jib movement.

This kind of fine-grained information – such as trajectory data or actuator commands needed for a certain manipulation task – is represented by a robot execution model (REM). REMs are created in the next step, and typically reside outside the information backbone, in components closer to robot control. While FIM is robot-type-agnostic, REMs are created for specific robot types and contain all set motion data for the respective robot type to complete a task. REM computation employs robot models (of potentially varying granularities) that might be used to e.g. kinematically compute a simple trajectory or even optimize a motion within a dynamical simulation, considering e.g. collisions, motion time or energy consumption. Accordingly, choosing another robot type for the same manipulation task leads to different motions and thus a different REM. Still, the methods and interfaces for REM computation might be generalized: a BIM-FIM-REM toolchain has been demonstrated to generate a REM model for automated masonry with cable-driven parallel robots [20]. Finding a suitable robot involves robot task allocation (section 3.1.3), but nevertheless, there is feedback from the robotic systems on previous steps, as different robots have different characteristics. For example, robots with higher payloads can place larger brick formats.

The Backbone is accessible mainly through APIs, but also a User Interface, consisting of a modular model viewer and a robot planning viewer. The model viewer offers a visual inspection of the building, status tracking, and enables task planning and execution. The robot planning viewer is used to display simulation results and, during execution, the robot's next planned actions in real-time.

3.1.2 Discrete event simulation

The precedence relationships of the process graph define the general sequence of atomic tasks, but sequence alone does not capture the probabilistic time distribution a specific robot needs for each task. Such knowledge about the overall stochastic distribution is, however, important to plan for the resource-intensive robotic equipment and the equally important supply chain that is required to provide the robotic equipment with the required material.

To determine the stochastic duration of overall robotic work processes on-site, we propose to employ constraint-based discrete event simulation [21, 19]. In these simulations, the duration of a directed graph network of tasks is simulated. Constraints determine when a single task in the network can be started: all preceding tasks in the network need to be finished, all required materials need to be available, and all robotic and non-robotic equipment needs to be available. When all prerequisites are met, the simulation starts a task that executes it for a duration drawn from a specific probability distribution representing the stochastic duration of the atomic task, ideally defined based on empirical data collected from previous construction work. During the duration of a task, all resources are bound and not available for starting other tasks.

A stochastic Monte-Carlo repetition of various individual discrete-event simulations then provides important information to strategically plan the site. Variations in the amount and provision of robotic and non-robotic equipment can provide stochastic durations and costs of specific resource allocations. For each provided resource allocation scenario, simulation results provide information about when specific material resources are required, which, in turn, allows for planning material supply chain aspects for the entire construction site.

3.1.3 Robotic skills and task allocation

To enable the selection of suitable entities, such as robotic systems, construction machinery, or humans, for dedicated construction tasks, a formal method for describing the skills, constraints, and requirements is developed. The *skills* section includes an enumeration of tasks such as *lifting, placing, drilling, demolition*, etc. The *constraints* section includes the maximum load as well as the operational range of the system, among others. The *requirements* section includes the necessary space, supply (electricity, material), and the weight of the system to be considered as loads for the underlying structure.

The formal description of robotic skills, constraints, and requirements forms the basis for multi-robot task allocation. Construction tasks derived from the FIM model are assigned to individual robots capable of performing the task. For example, transport before the assembly of a prefabricated component requires a mobile robot capable of grabbing, lifting, and transporting elements of certain sizes and weights. If multiple robots are qualified, they are ranked by performance factors such as speed, required space, etc. The actual task allocation procedure is based on a free-market allocation approach as proposed by [22].

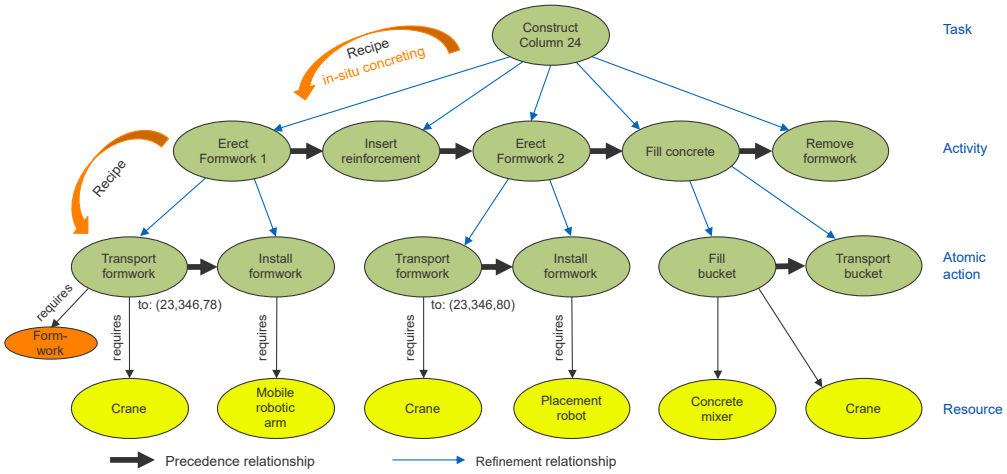


Figure 3. Process graph representing the hierarchical breakdown of the construction processes. The leaf nodes represent instructions for the individual robotic systems.

3.2 Execution environment

The purpose of the second module of the information backbone lies in the execution of the construction processes prepared in the planning environment, primarily involving steering and coordinating the diverse robotic systems. As real-world experimentation is cumbersome and expensive, the proposed approach supports performing robotic simulations with a high level of realism. At the same time, communication protocols are defined to facilitate an easy transition from controlling simulated robots to controlling real robots.

3.2.1 Middleware

Communication between the central coordination module and the robotic systems is handled by middleware, with an implementation using the Robot Operating System 2 (ROS 2) as the primary framework. ROS 2 is a modular, open-source framework for developing, deploying, and operating robotic systems. It provides standardized communication mechanisms, tooling, and abstractions that support distributed, heterogeneous, and real-time-capable robotic applications.

ROS 2 builds on Data Distribution Service (DDS) implementations such as eProsima Fast DDS [1], or Eclipse Cyclone DDS[2], which use the DDS-RTPS wire protocol[3] for message transport over standard network protocols, including UDP and TCP. ROS 2 communication follows a message-oriented paradigm. Data is exchanged through strongly typed messages defined in an interface description language. Messages are published and subscribed via named topics, enabling loosely coupled, asynchronous communication between robot components and processes.

Alternatively, Apache Kafka[4] is available as an additional abstraction layer. While ROS 2 provides a very complete toolset with extensive libraries, computational methods and robot-centric communication, Kafka provides a scalable and fault-tolerant event-streaming that supports integration of external systems such as digital twins, BIM services, or analytics platforms. Bridging ROS topics to Kafka streams allows robot-generated events to be consumed by higher-level coordination and monitoring services, while process-level commands can be reliably disseminated to the robotic fleet. The separation supports heterogeneous system landscapes and enhances scalability without affecting time-critical robot control. Furthermore, Apache Kafka provides an interface to industry-standard bus systems such as EtherCAT. In this context, an interface to MATLAB Simulink is created to enable the simulation of robotic systems outside of the ROS 2 environment.

Exemplarily, this is used for the simulation and control of a cable-driven parallel robot for construction.Due to its size, payload and sensitivity towards stable control frequencies, industrial hard real-time control is required. Accordingly, control is performed on a Simulink target in a Beckhof TwinCAT PLC, while hardware communication is performed via an EtherCAT bus.

3.2.2 Shared model hosting and hierarchical information flow

Beyond communication protocols, the system architecture must define how shared product and process representations are hosted and maintained. In the proposed framework, the BIM and the FIM serve as authoritative,

[1] ttps://github.com/eProsima/Fast-DDS
[2] https://github.com/eclipse-cyclonedds/cyclonedds
[3] https://www.omg.org/spec/DDSI-RTPS/
[4] https://kafka.apache.org/

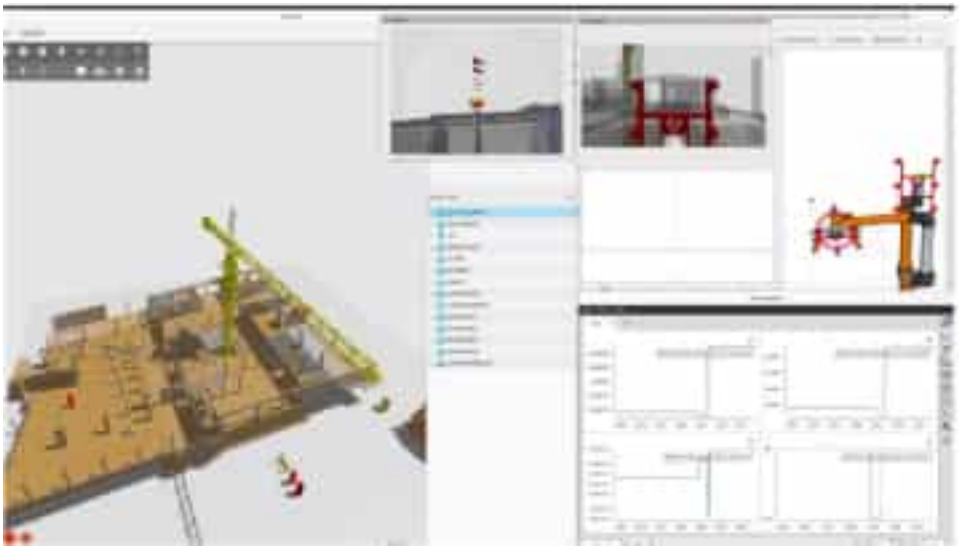


Figure 4. Gazebo is one of the employed simulation environments. Here, the interaction between a robotized crane and a mobile manipulator on a wheeled base is simulated.

global descriptions of the construction state and therefore cannot be treated as local or transient data structures.

Both models are hosted in a cloud environment and are accessible to all distributed agents participating in the construction process, including planning services, simulation components, and robotic execution systems. Shared hosting enables synchronized access to up-to-date product and process information, supporting continuous updates that reflect construction progress, execution outcomes, and detected deviations.

3.2.3 Behavior tree for the central orchestration of the robotic fleet

For executing robotic processes in a coordinated manner, process graphs resulting from the planning environment are transformed into behavior trees (Fig. 5) [17]. Behavior trees are hierarchical, graph-based task control structures for modeling and executing robot behaviors in a modular and reactive manner [23, 24]. The authors decompose complex tasks into reusable nodes, developed in the form of robotic capability subtrees [24], that encapsulate actions, conditions, and control flow, enabling clear separation of decision-making and execution. Due to the scalability, readability, and support for real-time reactivity, behavior trees are widely adopted for robotic task execution and control in both research and industrial applications.

In contrast to controlling individual robotic systems, behavior trees are employed as a centralized coordination mechanism to orchestrate the tasks of an entire heterogeneous robotic fleet. In the IBRC architecture, the leaf nodes of the behavior tree issue messages to the respective robotic systems, enabling coherent and synchronized multi-robot behavior. The messages mirror the level of (process) detail provided by the product-process graph and allocate tasks based on robotic capabilities, leaving a certain freedom of interpretation and degree of independence to the individual robotic systems.

Individual robotic systems may additionally implement local behavior trees for fine-grained control over robotic behavior. In the hierarchical organization of the IBRC, behavior trees do not directly manipulate shared models; instead consume task assignments and constraints derived and report execution results back to the information backbone, enabling closed-loop updates of the global construction state. More details are provided in [17].

3.2.4 Robotic simulation environments

Robotic simulation environments are a central component of the proposed architecture, as construction robots must operate in complex, unstructured, and continuously evolving environments. Reliable interaction with the built environment, human workers, and other robots requires extensive validation of perception, control, and coordination strategies prior to on-site deployment. Simulation therefore provides a controlled setting in which these capabilities can be developed, tested, and refined before being transferred to physical systems.

Within the proposed framework, simulation environments act as first-class execution entities that mirror the physical world within the same software architecture. Simulated robots receive control commands and publish state and sensor data via middleware interfaces, allowing them to function as substitutes for physical robots during planning, validation, and testing (Fig. 4), and enabling seam-

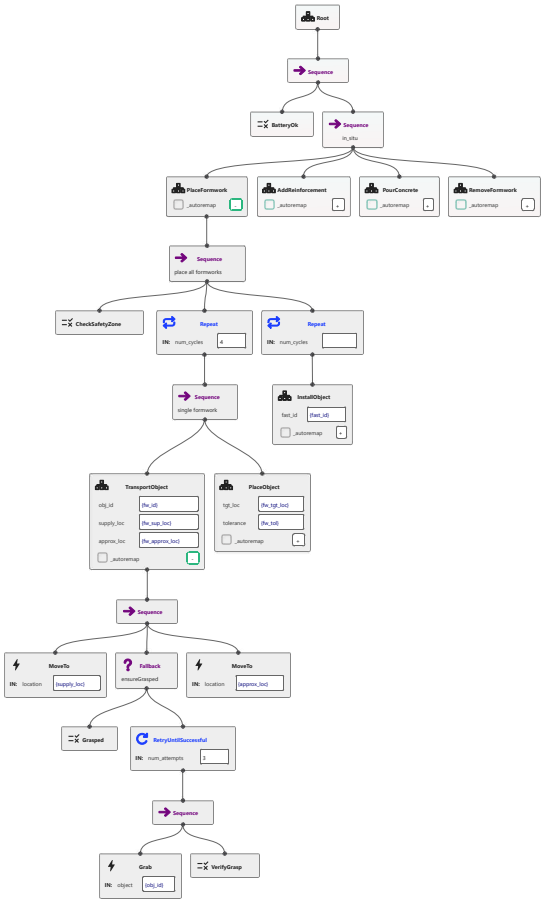


Figure 5. Behavior tree created from the process tree, serving as an executable representation for steering and coordinating the fleet of robotic systems on-site.

less switching between simulated and real execution while preserving identical orchestration and control logic.

Simulation also supports data-driven development workflows. Real-world data acquired through teleoperation, such as camera images, joint states, or encoder readings, can be used to train, calibrate, and refine algorithms that are then evaluated in simulation, which is particularly relevant for construction-specific challenges such as traversing uneven terrain. Core subsystems including controllers, navigation stacks, and SLAM algorithms are therefore developed and validated in simulation prior to deployment. Although simulations cannot fully replicate real-world conditions, noise-aware training and robust algorithm design enable effective transfer to physical robots.

Our framework employs Gazebo, Isaac Sim, and PyBullet as simulation backends, all providing high-fidelity rigid-body physics and tight ROS integration (Fig. 4). Simulation environments are initialized using product and process information from the planning environment. BIM and FIM models are translated into simulator-specific geometry and assets, e.g., SDF entities in Gazebo [25] or NVIDIA-internal formats in Isaac Sim. Robots are instantiated from URDF models provided by the information backbone, allowing the planning layer to remain independent of simulator-specific data structures while supporting multiple simulation platforms. Alternatively, MATLAB Simulink provides powerful tools for modeling construction robots, enabling high-granularity representations of kinematics, dynamics, and physical subsystems such as electronics, pneumatics, and hydraulics.

3.2.5 Human–robot collaboration

Although the proposed framework targets a high degree of robotization, construction remains dependent on human labor and decision-making. Human activities must therefore be explicitly represented in planning and execution, treating the site as a shared workspace where humans and robots perform tasks independently, sequentially, or in close physical interaction.

Human representation is a prerequisite for planning and simulation of collaboration. In simulation, humans can be modeled at different abstraction levels, e.g., kinematic chains, motion-capture trajectories, or avatar-based motion primitives. These models enable analysis of human–robot tasks derived from the process graph and behavior tree, allowing spatial conflicts, interference, and safety-critical situations to be detected before deployment.

Construction activities are classified into robot-only, human-only, and collaborative tasks. Robot-only and human-only tasks can proceed under sequencing constraints, whereas collaborative tasks require coordinated actions. This classification supports task allocation, synchronization, and monitoring across heterogeneous agents.

Coordination is implemented via synchronization constructs embedded in the centralized behavior tree. Robots can wait for human actions, react to changes in human state or position, signal completion, and assist during joint operations. Thus, behavior trees encode both robotic actions and interaction logic for handovers, sequential execution, and parallel work on shared workspaces or workpieces.

Given complementary strengths – robots in repetitive, heavy, and precision-demanding operations; humans in dexterity and situational decision-making – the framework incorporates human-in-the-loop control. Human input can initiate execution, approve transitions, or override robot behavior for supervision and safety.

All human actions and human–robot interactions are reported to the information backbone and integrated into the digital twin, capturing progress, contributions, and interaction states. Simulation further enables validation of collaborative scenarios with respect to safety, efficiency, and usability prior to real-world deployment.

4 Summary

The paper has presented overall concepts, procedures, and the software architecture of the Information Backbone for Robotized Construction. Further details are provided in accompanying papers [16, 17, 25]. The concept has been implemented in a proof-of-concept prototype combining the described modules through a micro-service platform [16]. In the next step, a comprehensive simulation experiment of a close-to-real construction project will be performed to further refine and elaborate the components.

Acknowledgments

The authors gratefully acknowledge the support provided by the German Research Foundation (DFG) for funding the research unit under grant FOR 5672 (grant number 517965147) as well SM 281/30-1 and SM 281/32-1. Furthermore, the authors acknowledge the support of the TUM Innovation Network "CoConstruct", the DFG Transregio 277 "Additive Manufacturing for Construction" (grant number 414265976), and the EU Pathfinder project "AM2PM".

References

[1] J. Stührenberg and K. Smarsly. LIO-BIM – coupling LiDAR inertial odometry with building information modeling for robot localization and mapping. *Advanced Engineering Informatics*, 66:103477, 2025.

[2] K. Smarsly and K. Dragos. Advancing civil infrastructure assessment through robotic fleets. *Internet of Things and Cyber-Physical Systems*, 4:138–140, 2024.

[3] A. Borrmann, T. Bruckmann, K. Dörfler, T. Hartmann, and K. Smarsly. Towards realizing the information backbone of robotized construction–computational methods and cyber-physical architectures for collaborative robotic fleets. In *Proc. of the CIB W78 Conf. 2021*, 2021.

[4] S. Meschini, K. Iturralde, T. Linner, and T. Bock. Novel applications offered by integration of robotic tools in BIM-based design workflow for automation in construction processes. In *Proc. of the CIB*IAARC W119 CIC 2016 Workshop*, 2016.

[5] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng. Ros: An open-source robot operating system. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA) Workshop on Open Source Software*, 2009.

[6] A. Ibrahim, A. Sabet, and M. Golparvar-Fard. BIM-driven mission planning and navigation for automatic indoor construction progress detection using robotic ground platform. In *Proc. of the 2019 European Conf. on Computing in Construction*, pages 182–189, 2019.

[7] B.R. Mantha, M.K. Jung, B.G. de Soto, C.C. Carol, and V.R. Kamat. Generalized task allocation and route planning for robots with multiple depots in indoor building environments. *Automation in Construction*, 119, 2020.

[8] R. Rust, G. Casas, S. Parascho, D. Jenny, K. Dörfler, M. Helmreich, A. Gandia, Z. Ma, I. Ariza, M. Pacher, B. Lytle, Y. Huang, C. Kasirer, E. Bruun, and P.Y.V. Leung. COMPAS FAB: Robotic fabrication package for the compas framework. https://github.com/compas-dev/compas_fab/, 2018. Gramazio Kohler Research, ETH Zürich.

[9] J.G. Everett and A.H. Slocum. Automation and robotics opportunities: Construction versus manufacturing. *Journal of Construction Engineering and Management*, 120, 1994.

[10] K.M. Lundeen, V.R. Kamat, C.C. Menassa, and W. McGee. Autonomous motion planning and task execution in geometrically adaptive robotized construction work. *Automation in Construction*, 100, 2019.

[11] L. Ding, W. Jiang, C. Zhou, and S. Liu. BIM-based task-level planning for robotic brick assembly through image-based 3d modeling. *Advanced Engineering Informatics*, 43, 2020.

[12] S. Kim, M. Peavy, P. Huang, and K. Kim. Development of BIM-integrated construction robot task planning and simulation system. *Automation in Construction*, 127, 2021.

[13] A. Zhu, P. Pauwels, and B. De Vries. Smart component-oriented method of construction robot coordination for prefabricated housing. *Automation in Construction*, 129: 103778, 2021.

[14] A. Zhu, P. Pauwels, and B. De Vries. Component-based robot prefabricated construction simulation using ifc-based building information models. *Automation in construction*, 152:104899, 2023.

[15] A. Zhu, T. Dai, G. Xu, P. Pauwels, B. De Vries, and M. Fang. Deep reinforcement learning for real-time assembly planning in robot-based prefabricated construction. *IEEE Transactions on Automation Science and Engineering*, 20(3):1515–1526, 2023.

[16] T. Wrabel, S. Esser, and A. Borrmann. A framework architecture for the information backbone for robotic construction. In *Proc. of the 43rd Int. Symp. on Automation and Robotics in Construction (ISARC 2026)*, 2026.

[17] M. Brinkhoff, S. Esser, and A. Borrmann. An integrated product–process representation for deriving robotic task structures using graph transformations. In *Proc. of the 43rd Int. Symp. on Automation and Robotics in Construction (ISARC 2026)*, 2026.

[18] M. Slepicka, S. Vilgertshofer, and A. Borrmann. Fabrication information modeling: Interfacing building information modeling with digital fabrication. *Construction Robotics*, 6(2):87–99, 2022.

[19] I-C. Wu, A. Borrmann, U. Beissert, M. König, and E. Rank. Bridge construction schedule generation with pattern-based construction models and constraint-based simulation. *Advanced Engineering Informatics*, 24(4):379–388, 2010.

[20] T. Bruckmann and R. Boumann. Simulation and optimization of automated masonry construction using cable robots. *Advanced Engineering Informatics*, 50:101388, 2021.

[21] U. Beißert, M. König, and H.-J. Bargstädt. Constraint-based simulation of outfitting processes in building engineering. In *Proc. of the 24th CIB-W78 Conf.*, 2007.

[22] R. Zlot and A. Stentz. Market-based multirobot coordination for complex tasks. *The Int. Journal of Robotics Research*, 25(1):73–101, 2006.

[23] M. Colledanchise and P. Ögren. *Behavior Trees in Robotics and AI: An Introduction*. CRC Press, Clermont, Florida, 2018.

[24] A. Tandon, J. Stührenberg, and K. Smarsly. Automated building inspections coupling building information modeling and behavior trees. In *Proc. of the 2025 European Conf. on Computing in Construction*, 2025.

[25] M.R. Kolani, A. Mirhaghgoo, S. Nousias, and A. Borrmann. Model-based adaptive controller for fast and restricted swing control in tower cranes. In *Proc. of the 43rd Int. Symp. on Automation and Robotics in Construction (ISARC 2026)*, 2026.

Beyond Geometric Conversion: Enriching Industrial CAD Models for BIM-Based Design

Davide Avogaro[1], Maximilian Maria Wurm[2], Johan Wickström[2] and Carlo Zanchetta[3]

[1]Department of Management and Engineering, University of Padova, 36100 Vicenza, Italy
[2]Paul Scherrer Institut PSI, 5232 Villigen PSI, Switzerland
[3]Department of Civil, Environmental and Architectural Engineering, University of Padova, 35131 Padova, Italy

davide.avogaro.2@phd.unipd.it, maximilian.wurm@psi.ch, johan.wickstroem@psi.ch, carlo.zanchetta@unipd.it

Abstract –

Interoperability between industrial and construction digital models remains a persistent challenge, particularly in projects that require the integration of manufacturing machinery within building environments. Manufacturing workflows typically rely on CAD models based on the STEP standard, whereas the Architecture, Engineering, and Construction (AEC) sector adopts BIM methodologies founded on the IFC standard, which integrates geometry with semantic, relational, and contextual information. Although both standards are mature and widely adopted, interoperability remains limited, with the main bottleneck arising from the transformation of industrial CAD models into semantically meaningful BIM representations.

This paper addresses this gap by proposing a lightweight, open-source framework for STEP-to-IFC integration that goes beyond simple geometric conversion. Instead, the framework adapts industrial product logic to satisfy the information requirements of the AEC domain. The proposed procedure reduces assembly complexity by removing unnecessary components, simplifying overly detailed geometries, limiting the number of objects to those strictly required, and enriching them with appropriate semantic properties.

The framework is implemented as a Blender add-on using the Bonsai extension and features an intuitive user interface, making it accessible to practitioners without advanced programming skills. Preliminary results demonstrate its effectiveness in enriching industrial CAD models with BIM-compatible semantics, highlighting its potential for both research and professional applications in complex industrial–building integration scenarios.

Keywords –

STEP to IFC; CAD to BIM; industry and civil interoperability; BIM; Blender; Bonsai

1 Introduction

Interoperability between digital models remains one of the most critical challenges in both the Architecture, Engineering, and Construction (AEC) and industrial domains. The manufacturing industry typically relies on Computer-Aided Design (CAD) systems to design machinery, products, and furniture that will be produced by industries. In contrast, the civil sector generally adopts Building Information Modeling (BIM) methodologies, which integrate geometric representations with structured product and semantic information. In the design of industrial or production plants, as well as in certain specialized projects, machinery must be integrated within buildings that are required to accommodate, protect, and support all systems necessary for their operation. Consequently, it is essential for civil engineers to import models into their authoring software that provide comprehensive information necessary for designing buildings, systems, supports, and related infrastructure. By integrating these domains, the design process can be accelerated, achieve higher-quality outcomes, and reduce the likelihood of errors. Such projects are commonly classified as complex, as they are not only complicated, multifaceted, and composed of numerous interconnected components [1], but also characterized by an additional layer of complexity arising from CAD–BIM integration.

Within their respective domains, manufacturing and construction rely on well-established standards for representing three-dimensional and information-rich models: the Standard for the Exchange of Product Data (STEP) format [2] in manufacturing, and the Industry Foundation Classes (IFC) format [3] in construction. Although both standards are based on explicit and formally defined data structures, effective interoperability between the two domains remains limited due to substantial differences in conceptual modeling paradigms, levels of detail, and intended use

cases. Notably, the primary challenge does not lie in data transfer from the civil to the industrial domain (i.e., IFC to STEP), as several tools are available to convert IFC models into simplified STEP geometries, typically renouncing product and semantic data. Instead, the more significant issue concerns the reverse process—from STEP to IFC—because IFC models are required to represent not only geometry but also rich product, relational, and contextual data.

This study proposes a framework that should not be regarded merely as a geometric conversion tool, but rather as a system capable of adapting industrial product logic to the information requirements of the AEC domain based on open-source tools and formats. The proposed approach demonstrates strong potential for application in both academic research and professional practice.

2 Background

As demonstrated in previous work [4], the existing literature on CAD–BIM interoperability between the industrial and AEC domains is limited, with most relevant contributions dating to the 2006–2008 period and lacking alignment with recent standard developments such as STEP AP242 and IFC4.3_ADD2. This situation highlights a clear gap in the current body of knowledge. Moreover, the reviewed studies consistently emphasize the absence of practical, deployable solutions and underscore the urgent need for interoperable, open-source tools capable of mitigating the fragmentation within the AEC sector [5–8].

At present, the most common approaches adopted to compensate for the lack of robust conversion tools include the following practices:

- Importing geometries as meshes, resulting in excessively complex representations that require the manual addition of product properties and relationships. Such meshes are often derived from low-level CAD models [9] that provide only viewable representations, characterized by faceted geometries or "frozen" solids, such as the Stereolithography (STL) format, the Initial Graphics Exchange Specification (IGES), or CATIA Graphics Representation (CGR). Each manual step is prone to error, and the use of highly detailed geometries leads to unnecessarily models.
- Manual redrawing of simplified components across different BIM authoring software, with properties and relationships added manually. This process is time-consuming and highly error-prone.
- Using converters that map all elements to *IfcBuildingElementProxy* or *IfcElementAssembly* entities, which causes semantic degradation, does not address excessive geometric complexity, and generates instances for all parts of a product—including components irrelevant for civil purposes (e.g., screws, bolts, labels). This results in IFC files with an excessive number of objects and redundant information.
- Relying on proprietary ecosystems managing both CAD and BIM environments, which does not resolve data exchange issues with external stakeholders and may conflict with regulatory requirements that mandate open formats, as is often the case in public administration projects. Additionally, large models can cause performance issues, making it difficult to upload and obtain a complete project view.

In summary, current procedures focus primarily on conversion tools without considering the fundamental differences between industrial and civil design logics. Industrial models are typically highly detailed, representing every part of a product to the level required for industrial production. Such models contain a large number of objects (e.g., screws, bolts) and detailed geometries that exceed the information needs of the AEC domain and can significantly compromise BIM performance. In contrast, the AEC sector requires models with a lower level of geometric detail, omitting minor components while preserving overall dimensional information. Moreover, industrial models often lack product properties, requirements, and inter-object relationships, which are typically stored in separate documents or managed via other systems. Conversely, in the IFC format, such data are represented as object properties, enabling professionals to design more efficiently and implement validation procedures to improve quality and save time. Thus, a simple conversion tool is insufficient. What is required is a procedure capable of adapting the logic of industrial products to the information requirements of the AEC domain. An initial attempt to address this challenge was proposed by [4], introducing a framework that leverages open-source technologies to adapt industrial product logic for AEC use. The present work extends this framework, presenting a new version designed to overcome limitations identified in the previous implementation.

3 Materials

A market analysis was conducted to identify open-source tools capable of importing industrial geometries and exporting IFC models suitable for implementing the proposed procedure. This analysis identified two primary candidates: FreeCAD [10] and Blender [11] equipped with the Bonsai extension [12]. Blender and Bonsai were selected due to the availability of documentation and because these tools are specifically developed and widely adopted for civil and BIM-oriented workflows.

Blender does not natively support the direct import of STEP files. To address this limitation, an intermediate exchange format was required that would preserve geometric integrity and structural information. Among the formats supported by Blender, glTF 2.0 (Graphics Library Transmission Format) was selected, as it is an efficient, open, and neutral standard well aligned with Blender's internal architecture. Developed and maintained by the Khronos Group, glTF 2.0 is designed for the efficient transmission and representation of three-dimensional content. It consists of a structured set of files, including a JSON file (.gltf) describing the scene, node hierarchy, materials, cameras, and references to geometry and texture data; one or more binary files (.bin) storing high-density buffered data such as meshes and animations; and image files (.jpg, .png) defining textures associated with materials.

The glTF format was chosen not only for its performance and ability to preserve product hierarchies, but also for its support of custom metadata through the *extras* field. While STEP files typically convey only geometric information and do not include semantic or property data, it is possible to develop a procedure to encode any additional properties associated with industrial components within the *extras* field, thereby enabling their transfer into the BIM environment. The open-source software Mayo [13] was used to convert STEP files into glTF format, which can then be imported into Blender without loss of geometric fidelity.

Blender supports the development of custom add-ons developed in Python, including the creation of tailored user interfaces. This capability enables the implementation of structured and intuitive workflows that are accessible even to users without advanced programming skills. The proposed framework is entirely based on open-source tools, ensuring independence from proprietary software ecosystems and broad accessibility. The overall framework and its operational steps are described in detail in the following section.

4 Framework

4.1 Case study: Paul Scherrer Institut (PSI)

The Paul Scherrer Institut (PSI) is a research center that conducts internationally recognized, cutting-edge research in the natural and engineering sciences. PSI develops, builds, and operates highly complex particle accelerator–based large-scale research facilities, which are made available to both the scientific community and industry. Among its most prominent facilities are the Swiss X-ray Free Electron Laser (SwissFEL) and the Swiss Light Source (SLS) [14].

Due to its extensive building portfolio and the continuous development and upgrading of facilities to incorporate new technologies and enhance its scientific offerings, PSI routinely faces challenges related to the integration of BIM models used for civil design with CAD models employed for the industrial design of its research infrastructure. For clarity, the following sections present a girder from the SLS facility as an illustrative example (Case Study 1, CS1). The framework has also been tested on additional models from the same facility, including a linear accelerator (CS2), a transfer line (CS3), and an undulator (CS4).

4.2 Add-on

The framework is presented step by step in the following sections. The user interface of the Blender add-on developed for this framework is shown in Figure 1 and is available on GitHub [15]. The arrangement of commands in the interface follows the sequential workflow of the procedure, guiding the user through the proper use of the framework. The framework is not limited to the STEP format but is generally applicable to all file formats that can be imported into Blender and that provide a hierarchical decomposition of components.



Figure 1. User interface of the proposed Blender add-on that reflects the methodological workflow

4.2.1 Make meshes unique

This operation is required to prevent unexpected behavior and to ensure the correct execution of the procedure. The command standardizes object names by removing unsupported or problematic characters that could cause ambiguities or errors during script execution. In addition, modeling software often employs multi-user meshes when multiple identical objects are present to reduce memory usage, meaning that a single mesh dataset is stored and referenced by multiple distinct objects. While this strategy is effective for optimizing storage, it can lead to errors when mesh data are accessed programmatically, potentially causing script failures. To avoid such issues, all meshes are made unique so that each object has its own mesh, ensuring that it references independent mesh data. This step is required solely to

guarantee the robustness and reliability of the code and does not affect the geometric outcome of the model.

4.2.2 Select fundamental nodes for IFC conversion

This step aims to adapt the complexity of the hierarchical structure of the imported STEP model to meet the requirements of the AEC domain. Well-structured industrial models often feature deeply nested hierarchies to manage and organize the numerous components that make up a product. This hierarchy is expressed through parent nodes containing multiple child nodes. As shown in Figure 2, when an industrial model is imported into the Scene Container, the hierarchy is displayed using two main types of icons: a triangle icon representing mesh objects and an axes icon representing empty objects. Mesh objects consist of vertices, edges, and polygonal faces, whereas empty objects are void transformation nodes used solely for grouping and manipulating child objects. Typically, empty objects act as parent nodes that contain other empty objects and mesh objects. When a parent node is collapsed in the tree view, the number of its child elements is displayed next to its name, as illustrated in Figure 2.



Figure 2. Hierarchical structure of the imported model in its initial assembly tree

For civil purposes, distinguishing every individual part of an industrial product is generally unnecessary. Instead, the focus is on representing the overall product at an appropriate level of decomposition. Figure 3 illustrates a simplified hierarchy in which only the elements relevant for BIM representation are retained, representing the outcome of the proposed procedure.



Figure 3. Final assembly tree

Accordingly, the user's first task is to select the fundamental nodes that will correspond to independent IFC objects. Within the assembly tree, a fundamental node represents an object that will be converted into an IFC element. If a fundamental node contains other fundamental nodes as children, it corresponds to an *IfcElementAssembly* aggregating multiple IFC elements. Conversely, if a fundamental node is a leaf in the assembly tree, all of its child mesh components are merged into a single object, as this represents the maximum level of decomposition required for civil applications. So, if a leaf fundamental node is an empty object, all mesh objects contained within its hierarchy are merged into a single mesh object.

To perform this operation, the user selects the desired object, chooses the appropriate Level of Detail (Low, Medium, or High) from the corresponding menu, and then presses the *Select nodes for IFC conversion* button. The concept of Level of Detail is discussed in Section 4.2.5. Once the button is pressed, a custom property named *Node for IFC conversion* is created and associated with the selected object. This property states that the node is a *fundamental node* and is essential for the correct functioning of the subsequent processing scripts.

4.2.3 Delete small parts

This step is optional but highly recommended, as it allows the removal of elements that are unnecessary for civil purposes and typically characterized by a high number of vertices, faces, and triangles, such as screws, bolts, and similar components. The command deletes all meshes that do not have the *Node for IFC conversion* property assigned and whose size is below a user-defined threshold. Meshes marked with the *Node for IFC conversion* property are preserved, as some small components—such as plugs, valves, or comparable elements—may still be relevant and required in the IFC model. By defining an appropriate threshold and activating the *Delete parts smaller than* command, the user can efficiently eliminate insignificant elements while retaining those essential for IFC conversion. Figure 4 shows the difference between the initial state and the model after deleting objects smaller than 0.05 × 0.05 × 0.05 m. As reported in Table 2, in the considered case studies, approximately the 61.2% of the total number of objects are removed through this operation.

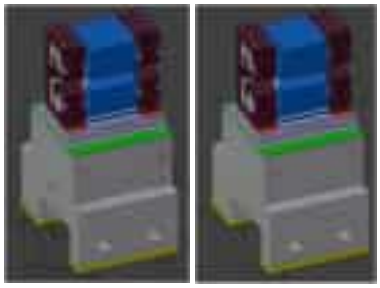


Figure 4. Comparison of the initial state (left) and the girder after removing elements below 0.05 × 0.05 × 0.05 m (right)

4.2.4 Database and ODS printing

For each fundamental node, an OpenDocument Spreadsheet (ODS) file is generated. The ODS format is an open-source spreadsheet standard. The user must specify a directory on their local system where these ODS files will be stored. Each ODS file is named after the corresponding fundamental node, excluding any automatically generated progressive suffixes. In modeling software such as Blender, progressive suffixes (e.g., ASS. AS2B QUAD_DIPOLE.1, ASS. AS2B QUAD_DIPOLE.002) are appended to distinguish objects with identical base names. Since these objects are geometrically and hierarchically identical, generating multiple ODS files with the same content is unnecessary; therefore, a single ODS file is created using only the base name. Each ODS file contains a list of the elements in the hierarchy of the fundamental node after which the file is named. If the fundamental node contains other fundamental nodes as children, as shown in Figure 5, the user is required to define the data for the corresponding IFC objects, including the *IfcClass*, *PredefinedType*, *ObjectType*, and any required property sets (Psets) and properties. If the fundamental node is a leaf node, as illustrated in Figure 6, the user must instead specify data related to the mesh components that will constitute that node. Specifically, for each mesh, the user indicates whether it should be deleted—for example, if it represents an internal or otherwise irrelevant component—or retained. For retained meshes, the user may also specify whether the geometry should be simplified. When geometries are excessively detailed for civil applications, simplification replaces the original mesh with its bounding box, thereby significantly reducing the overall model size.

As illustrated in Figures 5 and 6, the cells requiring user input within the ODS files are highlighted to guide the data-entry process: grey cells must not be filled, dark green cells are mandatory, and light green cells are optional.

Based on the directory specified by the user, the script automatically creates two subfolders—*Completed* and *To be completed*—if they are not already present. For each fundamental node, the script checks whether an ODS file with the corresponding name exists in the *Completed* folder. If no such file is found, a new ODS file is generated in the *To be completed* folder.

After completing the data entry in the ODS files, the user must manually move the files from the *To be completed* folder to the *Completed* folder. This workflow supports the reuse of ODS files across different conversion projects, as multiple models may share identical components—for example, different STEP files of girder models may reference the same *ASS. AS2B QUAD_DIPOLE* magnet, making it unnecessary to compile the corresponding ODS file more than once. By maintaining a single, centralized repository of completed ODS files, redundant data entry is avoided, resulting in improved efficiency and reduced preparation time. In addition, storing manually entered data in a centralized database prevents the need for repeated compilation, which is particularly beneficial in work-in-progress scenarios where the design of industrial products both affects and is affected by the design of the building. This approach enables the procedure to be repeated and updated versions to be shared with minimal or no additional manual effort.

4.2.5 Simplify assembly

Once all parameters have been defined, the geometry can be simplified according to the information specified in the ODS files. For each leaf fundamental node, the system determines the appropriate degree of simplification based on the Level of Detail (LOD) selected by the user. When the LOD is set to *Low*, the script reads the corresponding ODS file, deletes all child mesh components marked for removal by the user, and replaces the geometry of all remaining child components with their respective bounding boxes. When the LOD is set to *Medium*, the script deletes all child components marked for removal and simplifies only those child meshes explicitly flagged for simplification in the ODS file by replacing them with their bounding boxes. When the LOD is set to High, no child mesh components are deleted or simplified. In all cases, the remaining child



Figure 5. ODS file of the fundamental node ASM-30050.12



Figure 6. ODS file of the leaf fundamental node ASS. AS2B SUPER BEND (BS2)

mesh components associated with a fundamental node are merged into their parent node. This process significantly reduces the number of objects in the scene, limiting it to the number of fundamental nodes while preserving the level of geometric abstraction required for the selected LOD. Results of the simplification process for the different LOD levels are shown in Figure 7.

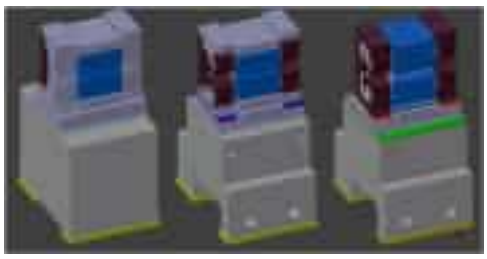


Figure 7. Simplified model outcomes for low (left), medium (center), and high (right) LOD

4.2.6 Compress geometries

To further reduce geometric complexity and file size, the geometries can be optimized using the *Compress geometries* command. This operation merges vertices that are within a defined proximity threshold, removes duplicate vertices, converts triangular faces into quadrilateral faces where possible, deletes vertices and edges not associated with any faces, and enforces geometric planarity. These optimizations reduce the overall complexity of the geometry during IFC export, resulting in significantly lighter IFC files.

4.2.7 IFC conversion

The final step converts the processed geometries into IFC elements using Bonsai capabilities. For each fundamental node, the script reads the IFC attributes and properties defined by the user in the corresponding ODS file and generates the appropriate IFC entities, as shown in Figure 3. Decomposition relationships between parent and child elements are established through *IfcRelAggregates* relationships, thereby explicitly representing the hierarchical structure of the product.

The conversion process is compatible with all IFC versions supported by Bonsai. However, the user must ensure that the selected IFC classes, attributes, and property sets (Psets) are valid for the chosen IFC schema version, as inconsistencies or unsupported definitions may cause the script to fail. For this reason, the use of IFC4 or later versions is recommended. The results of the simplification process at the different LOD levels are presented in Figure 8. Although the geometric level of detail varies, the product data associated with the objects remains unchanged across all representations. As shown, some unexpected behaviour can be observed in the colouring of elements.

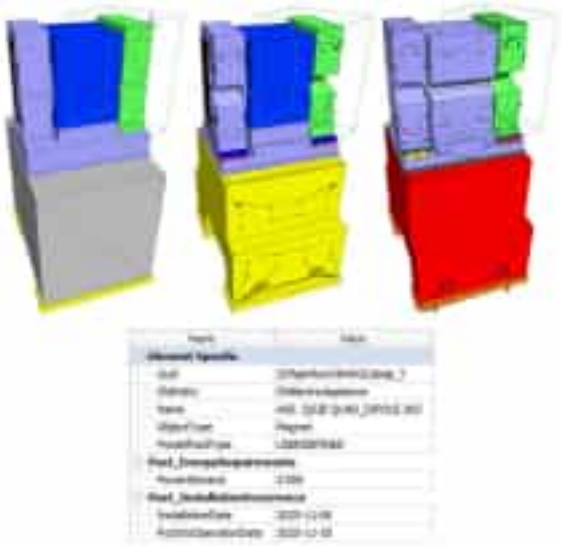


Figure 8. Resulting IFC models viewed in BIMvision [16] with different levels of detail: low (left), medium (center), and high (right). The same magnet selected has the same properties (bottom) in all models

5 Discussion

At the beginning of the present work, the primary objective was to overcome some of the limitations identified in the previous framework [4]. One of the main limitations of that version was the excessive number of rows required to be filled in the spreadsheet. Using the same case study files, the current version of the framework reduced the number of rows to complete by 86% for CS1, 74% for CS2, 81% for CS3, and 82% for CS4, resulting in an average reduction of 81% across all case studies.

However, the most relevant results emerge from the comparison with the original STEP file, which enables a meaningful evaluation of the current status of the tool against other solutions. The first metric of interest is the file size relative to the original STEP model. As shown at Table 1, IFC models generated with a high Level of Detail (LOD) occupy more storage space than the original STEP files. It is important to note that, even at this high LOD, IFC models were processed by removing small elements, as described in Section 4.2.3. This indicates that the same tessellated geometries encoded in IFC require more storage space than when represented in the STEP format. This observation further enhances the significance of the results obtained with lower LODs, which demonstrate encouraging compression ratios and confirm the effectiveness of the proposed approach in reducing model size while maintaining an acceptable level of detail.

Table 1 File size comparison for different case studies, including mean compression percentage

	CS1 [MB]	CS2 [MB]	CS3 [MB]	CS4 [MB]	Mean of %
STEP	12.0	39.8	7.9	314.4	100.0%
IFC high LOD	45.0	140.7	10.9	531.6	258.7 %
IFC medium LOD	4.0	16.5	3.8	147.1	42.4 %
IFC low LOD	0.1	0.8	0.1	0.3	0.9 %

This issue is not solely related to file size. Several systems can handle large IFC files without significant performance loss, making geometric simplification unnecessary. The key factor is instead selecting only the required objects. Some tools convert all STEP entities into *IfcBuildingElementProxy* or *IfcElementAssembly* objects. While this enables geometry import into BIM or civil environments, it is ineffective for product data management. Large numbers of objects without associated product information add complexity while providing only generic classifications. Without linked data, the benefits of the IFC standard are largely lost, reducing the model to a collection of heavy 3D geometries. Table 2 shows the number of objects removed or merged during conversion. The resulting compression rate is encouraging and confirms that the proposed framework effectively adapts an industrial model for civil engineering applications.

Table 2 Object count across process phases for different case studies, including mean compression percentage

	CS1	CS2	CS3	CS4	Mean of %
Initial stage	1633	3729	501	7632	100.0%
Delete small objects [<0.05m]	839	1241	213	2148	38.8%
Simplify geometries	10	325	6	52	2.8%
Convert IFC meshes	6	256	4	43	2.2%

Another key efficiency metric is the workload for both the user and the computational process, as shown in Figure 9. The main factor affecting the total duration is model complexity, measured by the number of objects. On average, processing rates are about 0.58 elements per second for low LOD, 0.63 for medium LOD, and 0.72 for high LOD. As the number of objects increases, both user effort and processing time grow, regardless of the selected LOD. Performance differences between LOD levels mainly reflect the computational cost of handling geometries of varying complexity, and this gap becomes more evident as the number of objects increases.

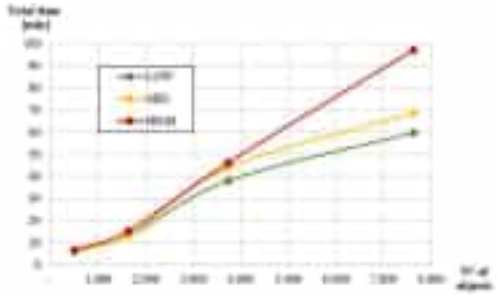


Figure 9. Procedure duration (minutes) relative to the number of objects contained in the STEP model.

5.1 Limitations

A key limitation of the framework is that, despite being designed for general professionals, it remains cumbersome and requires targeted training for effective use. Users must also possess knowledge of the IFC standard to correctly assign properties. Errors can occur during ODS file compilation (e.g., typographical mistakes), highlighting the need for more error-resistant mechanisms such as dropdown menus or predefined inputs. Additionally, manual file transfers introduce further risk of user error.

Another limitation lies in the reliance on STEP files with a specific decomposition logic, assuming a well-structured hierarchy of products and parts. In practice, such files may contain inconsistencies, merged components, or unexpected configurations. Therefore, broader testing on real-world cases is needed to assess robustness and identify improvements for professional use.

5.2 Future works

Future work will address the identified limitations through more extensive testing, providing clearer insights into application time and performance. Enhancements are planned to reduce manual effort, including features to automatically populate the *Node for IFC conversion* property using existing ODS database entries, enabling recognition of previously defined components and streamlining the workflow.

The framework will also be extended to support additional IFC classes, such as *IfcDistributionElement* and *IfcPort*, and to incorporate system management properties. Furthermore, it will be expanded to handle material data that require dedicated IFC classes rather than simple attributes or property sets.

The current need for expertise in IFC and Blender limits scalability and adoption. To mitigate this, a more

user-friendly solution should be developed to assist in selecting IFC classes and properties while reducing direct interaction with Blender. A potential approach is a web-based platform that runs Blender processes in the background, simplifying the workflow and improving accessibility.

6 Conclusion

This work examines the challenges of integrating industrial and civil domains through open standards such as STEP and IFC. Despite relying on mature data models, their differing paradigms and information requirements hinder interoperability, particularly when embedding industrial machinery within building models. The primary bottleneck does not lie in simplifying BIM data for industrial use, but rather in enriching industrial CAD models to satisfy the semantic, relational, and contextual requirements of the BIM environment. To address this gap, the study introduces a lightweight, open-source framework that goes beyond mere geometric translation. The framework demonstrates the feasibility of aligning industrial product logic with AEC information structures. The entire procedure is based on open-source technologies and is implemented as a Blender add-on using the Bonsai extension. Its intuitive user interface makes the workflow accessible even to users without specialized programming skills.

The procedure presented in this paper focuses on reducing the assembly complexity of industrial models by removing unnecessary components, simplifying excessively detailed geometries, reducing the number of objects to those strictly required, and attaching the appropriate semantic properties to them. In addition, mechanisms for storing and reusing manually inserted data are incorporated to minimize repetitive tasks. These strategies are intended to make the framework suitable for professional use by reducing manual effort and saving time during the conversion process. Preliminary results indicate the effectiveness of the proposed approach in adapting industrial product logic to AEC requirements.

7 Acknowledgements

The authors would like to thank Paul Scherrer Institut (PSI) for their technical support, for providing the materials used in this case study, and for granting permission to publish selected images of mechanical components from the Swiss Light Source (SLS) project.

References

1. Xia, B.; Chan, A.P. Measuring Complexity for Building Projects: A Delphi Study. *Eng. Constr. Archit. Manag.* **2012**, *19*, 7–24.
2. *ISO 10303-21*; **2016**.
3. *ISO 16739-1*; **2024**.
4. Avogaro, D.; Zanchetta, C. Bridging Information from Manufacturing to the AEC Domain: The Development of a Conversion Framework from STEP to IFC. *Systems* **2025**, *13*, 421.
5. Farinha, F.; Gonçalves, R.J.; Garção, A.S. Integration of Cooperative Production and Distributed Design in AEC. *Adv. Eng. Softw.* **2007**, *38*, 772–779, doi:https://doi.org/10.1016/j.advengsoft.2006.08.038.
6. Gielingh, W. An Assessment of the Current State of Product Data Technologies. *Comput.-Aided Des.* **2008**, *40*, 750–759, doi:https://doi.org/10.1016/j.cad.2008.06.003.
7. Yang, D.; Eastman, C.M. A Rule-Based Subset Generation Method for Product Data Models. *Comput.-Aided Civ. Infrastruct. Eng.* **2007**, *22*, 133–148, doi:https://doi.org/10.1111/j.1467-8667.2006.00476.x.
8. Yang, Q.Z.; Zhang, Y. Semantic Interoperability in Building Design: Methods and Tools. *Comput.-Aided Des.* **2006**, *38*, 1099–1112, doi:https://doi.org/10.1016/j.cad.2006.06.003.
9. Nzetchou, S.; Durupt, A.; Remy, S.; Eynard, B. Semantic Enrichment Approach for Low-Level CAD Models Managed in PLM Context: Literature Review and Research Prospect. *Comput. Ind.* **2022**, *135*, 103575, doi:https://doi.org/10.1016/j.compind.2021.103575.
10. © The FreeCAD Team FreeCAD Available online: https://www.freecad.org/.
11. Blender Foundation Blender Available online: https://www.blender.org/ (accessed on 18 September 2025).
12. IfcOpenShell Bonsai Available online: https://bonsaibim.org/index.html (accessed on 18 September 2025).
13. Delorme, H. Mayo Available online: https://github.com/fougue/mayo (accessed on 9 April 2025).
14. Paul Scherrer Institute PSI PSI - Large Research Facilities Available online: https://www.psi.ch/en/research/large-research-facilities (accessed on 14 January 2026).
15. Avogaro, D. STEP-to-IFC Available online: https://github.com/Daavogaro/STEP-to-IFC (accessed on 14 January 2026).
16. BIMvision Available online: https://bimvision.eu/download/.

A Two-Stage Blind Source Separation Method for Structural Health Monitoring Signals Integrating Deep Kernel Regression and Wavelet Multiresolution Analysis

Boqiang Xu[1,2], Xiuzhen Tian[3], Junhui Han[3], Chao Liu[2], Qian-Cheng Wang[1,*]

[1]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, China
[2]Department of Bridge Engineering, College of Civil Engineering, Tongji University, China
[3]Handan Municipal Transportation Bureau, Hebei Province, China
bq.xu@cityu.edu.hk, ggcbgs2025@163.com, hdsjtj@163.com, lctj@tongji.edu.cn, qc.wang@cityu.edu.hk

Abstract

In structural health monitoring (SHM) systems, sensor signals are often affected by multiple external load effects that are strongly coupled across different frequency bands, posing significant challenges for accurate signal interpretation. To address this issue, this paper proposes a two-stage blind source separation method aimed at effectively decoupling sensor signals across multiple frequency ranges. In the first stage, a deep kernel regression (DKR) model integrating a long short-term memory (LSTM) network with Gaussian process regression (GPR) is developed. By adaptively learning the complex mapping between time-varying environmental factors and kernel hyperparameters, the model enables accurate prediction and removal of low-frequency temperature-induced strain components. In the second stage, wavelet multiresolution analysis (WMRA) is applied to the residual signal. Through optimized selection of the wavelet basis and key parameters, the mid-frequency quasi-static vehicle-induced response and high-frequency signal components are effectively separated. The proposed method is validated using 227 days of monitoring data collected from a real long-span bridge. The results demonstrate that the proposed approach can progressively extract signal components with clear physical interpretations from the original mixed signals. The separated vehicle-induced responses exhibit smooth curves consistent with mechanical principles, while time-frequency analysis confirms the effective suppression of high-frequency interference. Overall, the proposed method provides an effective tool for fine-grained signal decomposition in structural health monitoring, significantly improving the accuracy and reliability of load effect separation, and shows strong potential for engineering applications.

Keywords –

Structural Health Monitoring, Sensor Signal Processing, Blind Source Separation, Deep Kernel Regression, Wavelet Multiresolution Analysis, Gaussian Process Regression

1 Introduction

For a health monitoring system of a bridge or building structure, sensor measurements contain crucial information about the structural load-bearing condition and constitute one of the most critical components of the monitoring system [1]. However, sensor signals result from the combined action of multiple external loads. For effective condition assessment and damage early warning, it is necessary to precisely decompose sensor data to isolate effects caused by different load types. To better characterize structural response components, the field of structural health monitoring has developed a series of solutions to eliminate or reduce the influence of interfering responses and separate the desired structural response for health state evaluation. Typically, researchers categorize the information in the total structural response of a bridge into three parts: i) high-frequency components dominated by vibration effects; ii) long-term, low-frequency components induced by factors like temperature; and iii) vehicle-induced quasi-static effects, which lie between the two in frequency. To perform refined signal analysis, blind source separation of measured signals is required.

In theory, modern signal processing methods can separate effects caused by temperature variation, vehicle loads, and noise based on their distinct frequency characteristics. For example, previous studies have leveraged temperature-induced strain for bridge condition assessment [2] and developed automatic uncoupling methods to isolate vehicle- and temperature-induced strains [3]. Subsequent signal reconstruction

algorithms can then recover the original distributed effects. However, signals in different frequency bands often exhibit substantial frequency disparities, making direct decomposition using a single unified framework challenging. For instance, while high-frequency signals can be decomposed using time-frequency methods, these approaches struggle with low-frequency components like diurnal (24-hour period) temperature-induced effects [4]. To address this issue, the optimal strategy is to design tailored blind source separation algorithms according to the characteristics of each signal component. Therefore, this paper proposes a two-stage sensor signal blind source separation algorithm that integrates deep kernel regression and wavelet transform.

The algorithm first constructs a deep kernel regression-based time-series prediction model to extract the temperature-induced component from the signal [5]. This model employs a sequential feature extraction module and a non-sequential feature fusion module, utilizing multi-source, multi-format historical and real-time data to predict the kernel parameters of a Gaussian Process Regression model. Using these predicted parameters, the model forecasts the low-frequency thermal response. Subsequently, the residual signal is processed using wavelet multiresolution analysis to separate the mid-frequency vehicle-induced response from the remaining high-frequency components. Through this proposed dual-stage signal decomposition approach, efficient sensor signal processing is achieved.

This study presents three principal contributions: (1) a novel two-stage framework that decouples temperature-induced (low-frequency) and vehicle-induced (high-frequency) strain components using Deep Kernel Regression (DKR) and Wavelet Multiresolution Analysis (WMRA), respectively; (2) the DKR model synergistically integrates deep learning adaptability with Gaussian process uncertainty quantification and physical interpretability—overcoming limitations of ARIMA in nonlinear thermal dynamics and LSTM in confidence estimation; and (3) WMRA ensures mathematically rigorous, adaptive time-frequency decomposition robust to mode-mixing artifacts of empirical mode decomposition variants, enabling precise isolation of transient vehicle responses. Collectively, this physics-informed, uncertainty-aware methodology advances structural health monitoring through interpretable, reliable multi-component strain separation.

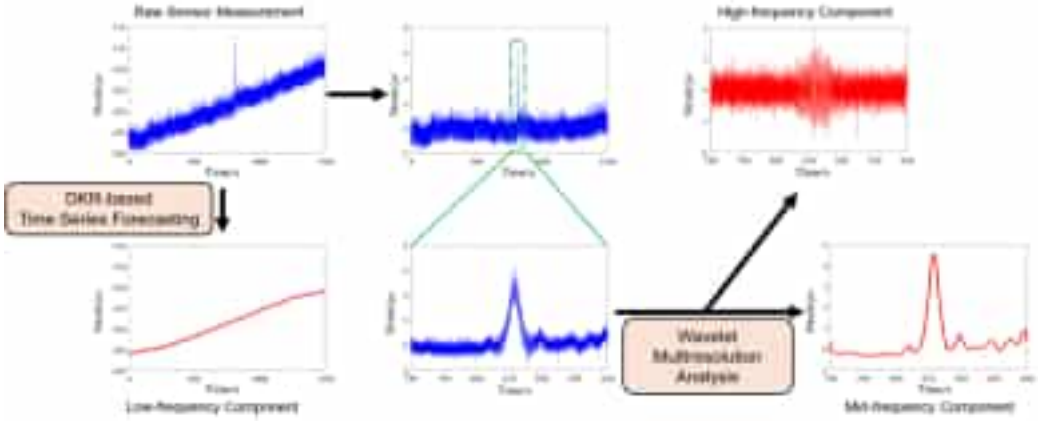


Figure 1. Framework of the two-stage signal decomposition method

2 Data Acquisition and Preprocessing

2.1 Raw Data Collection

The raw data is derived from the health monitoring system of a long-span suspension bridge in Hebei Province, China. The monitoring system installed four strain sensors on the bottom surface at the mid-span and the three-quarter span of the bridge's main span respectively. In addition, a temperature sensor was mounted at the pylon to detect the environmental conditions. Therefore, a total of nine sensors were used for data collection in this task, including eight strain sensors and one temperature sensor, as shown in Figure 2. All sensors had a sampling frequency of 200 Hz, and the final dataset included all available data collected over 227 days from February 22 to October 6, 2022.



Figure 2. Layout of strain sensors and temperature sensors

2.2 Offline Extraction of Temperature-Induced Strain

Since the low-frequency signal processing in the proposed two-stage method relies on a data-driven approach, data preprocessing is required to construct the dataset for deep kernel regression. The raw strain data contain outliers caused by heavy vehicle loads. To robustly extract temperature-induced strain components, the robust locally weighted scatterplot smoothing (RLOESS) algorithm is adopted [6].

RLOESS minimizes the influence of outliers by iteratively adjusting data point weights. Its core steps include neighborhood selection, weight calculation, weighted quadratic polynomial regression, and smoothed value estimation. Using this method, temperature-induced strain data over the entire 227-day monitoring period were successfully extracted.".

3 Stage I: Deep Kernel Regression Model for Low-Frequency Signal Extraction

To address the challenges of predicting temperature-induced strain in complex structures, a deep kernel regression (DKR) model is proposed. This model integrates a long short-term memory (LSTM) [7] network with Gaussian process regression (GPR) [8], enabling deep learning–based estimation of GPR kernel hyperparameters that vary with environmental conditions. This approach overcomes the limitation of conventional GPR models that rely on fixed kernel parameters, while combining the feature learning capability of deep learning with the interpretability and uncertainty quantification advantages of Bayesian methods. A detailed discussion of this algorithm can be found by Xu and Liu's work [9].

3.1 Gaussian Process Regression-based Time Series Forecasting

Compared to other methods, GPR has three critical advantages: (i) GPR provides a full Bayesian framework that quantifies prediction uncertainty via posterior variance—a non-negotiable requirement in structural health monitoring where safety-critical decisions depend on confidence intervals; (ii) The kernel function explicitly encodes prior physical knowledge (e.g., periodic temperature effects through the Periodic kernel), enhancing model interpretability and alignment with engineering principles; (iii) Unlike black-box models, GPR offers transparent hyperparameter interpretation (e.g., length-scale parameters directly reflect signal smoothness and periodicity), facilitating domain expert validation.

GPR is a non-parametric Bayesian regression method. For an input time series $X = \{x_1, x_2, ..., x_N\}$ and its corresponding temperature-induced component output $Y = \{y_1, y_2, ..., y_N\}$, GPR assumes the target function $f(x)$ follows a Gaussian process prior: $f(x) \sim GP(m(x), k(x, x'))$, where $m(x)$ is the mean function and $k(x, x')$ is the kernel (covariance) function. Given training data (X, y), the posterior distribution of the predicted values f^* at test points X^* remains Gaussian, with mean μ^* and variance Σ^* given by Equations (1) and (2), respectively.

$$\mu^* = K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \tag{1}$$

$$\Sigma^* = K(X^*, X^*) - K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X^*) + \sigma_n^2 I \tag{2}$$

To capture the composite characteristics of short-term monotonicity and long-term periodicity in temperature-induced strain data, a combined kernel function is chosen as the product of a Radial Basis Function (RBF) kernel (which is suitable for modeling gradual temperature-induced drift) and a Periodic kernel (which encodes repeating temperature-induced patterns). This combined kernel contains three hyperparameters $\Theta = (\Theta_1, \Theta_2, \Theta_3)$, controlling the function's rate of change, period length, and weight, respectively, as shown in Equation (3):

$$\begin{aligned} K(\Theta_1, \Theta_2, \Theta_3) &= \Psi\left[k_{RBF}(x_i, x_j), k_{Periodic}(x_i, x_j)\right] \\ &= \left[exp\left(-\frac{\|x_i - x_j\|^2}{2\Theta_1^2}\right)\right]^{\Theta_3} \left[\left(-\frac{2\,sin^2\left(\frac{\pi}{p}\|x_i - x_j\|\right)}{\Theta_2^2}\right)\right]^{1-\Theta_3} \end{aligned} \tag{3}$$

As hyperparameters significantly impact prediction results in GPR-based time series tasks, they must be dynamically adjusted based on the data.

3.2 Forecasting Framework

Based on the constructed dataset and environmental variables, a hybrid deep learning–Bayesian regression framework is developed:

1. **Sequential feature extraction:** An LSTM network is used to extract temporal features from historical temperature-induced strain data.
2. **Non-sequential feature fusion:** Fully connected layers integrate LSTM features with key environmental variables, including date, time, and temperature.

3. **Deep kernel regression:** The fused features are used to predict GPR kernel hyperparameters via a deep learning module, followed by real-time temperature-induced strain prediction using the parameterized GPR model.
4. **Model training:** The entire framework is trained by maximizing the marginal likelihood of the training dataset.

The algorithmic framework of this model based on deep kernel regression is shown in Figure 3.

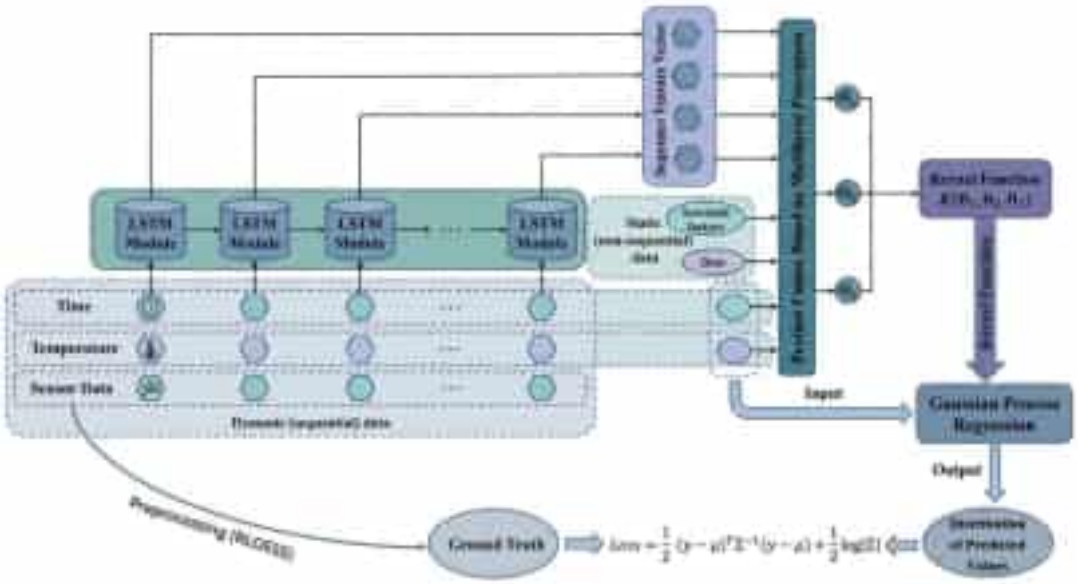


Figure 3. Framework of the deep kernel regression method

3.3 Forecasting Framework

The training objective is to find the optimal hyperparameters Θ_{opt} that maximize the probability of the true strain values under the predictive distribution. The loss function is defined as the negative log marginal likelihood, as shown in Equation (4):

$$\mathcal{L} = \frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu) + \frac{1}{2}\log\ |\Sigma| \qquad (4)$$

This loss function consists of two terms: the first is a weighted prediction error, and the second relates to prediction uncertainty (determinant of the covariance matrix). Optimising this loss function not only drives predictions closer to true values but also constrains output uncertainty, enhancing model robustness.

4 Stage II: Wavelet Multiresolution Analysis for High-Frequency Signal Separation

After successfully extracting the low-frequency temperature-induced components in Stage I, the residual signal still contains mixed mid- to high-frequency components, primarily consisting of vehicle-induced quasi-static effects and vibration-related noise. To address this, Stage II employs Wavelet Multiresolution Analysis (WMRA) for further signal decomposition [10]. This stage focuses on separating the vehicle-induced quasi-static response (mid-frequency) from high-frequency noise and vibrations. Unlike conventional filtering techniques, WMRA provides localized time-frequency characterization, making it particularly suitable for processing non-stationary signals under varying operational conditions. This section details the theoretical foundations, parameter selection strategies, and implementation procedures of the WMRA-based separation framework.

4.1 Wavelet Transform and Multiresolution Analysis

Following the extraction of low-frequency thermal signals in the first stage, and considering the significant frequency differences between vibration-induced high-frequency effects and vehicle-induced quasi-static responses, this section employs Wavelet Multiresolution Analysis (WMRA) for signal separation. Wavelet transform possesses time-frequency localization capabilities, enabling effective separation of signal components with different frequencies.

Wavelet transform overcomes the lack of temporal locality in Fourier transform. For a function $f(t)$, its continuous wavelet transform is defined as

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = |a|^{-1/2} \int_R f(t) \psi \left(\frac{t-b}{a} \right) \mathrm{d}t \quad (5)$$

where $\psi_{a,b}(t)$ is the wavelet basis derived from the mother wavelet $\psi(t)$ through scaling factor a and translation factor b. For discrete signals, the discrete wavelet transform is used:

$$\psi_{j,k}(t) = a_0^{-j/2} \psi \left(\frac{t - k a_0^j b_0}{a_0^j} \right) = a_0^{-j/2} \psi \left(a_0^{-j} t - k b_0 \right) \quad (6)$$

Multiresolution Analysis (MRA) is central to discrete wavelet transform, approximating the signal space $L^2(R)$ through a series of nested subspaces $\{V_j\}$ that satisfy conditions of monotonicity, approximation, scaling, and translation invariance [11, 12]. The signal can be decomposed layer by layer into low-frequency approximation (A_j) and high-frequency detail (D_j) components, as illustrated in Figure 4.



Figure 4. Signal decomposition method based on wavelet multi-resolution analysis

4.2 Selection of Key Parameters

Many wavelet functions exist with differing properties, leading to varied separation results for structural monitoring data. Therefore, targeted analysis of wavelet function characteristics is crucial for selecting an appropriate function. Generally, the following aspects guide the selection of wavelet functions and parameters:

1. **Wavelet Function Selection:** The choice depends on properties such as support length, orthogonality, vanishing moments, and symmetry. Commonly used wavelets include Daubechies (dbN) and Symlets (symN). Selection follows the principle of self-similarity, considers discriminant functions and support length, and balances properties based on signal characteristics.
2. **Decomposition Level Determination:** The number of decomposition levels n is determined by the signal sampling frequency f_n and the upper-frequency limit f_m of the target component to be separated. Too few levels lead to insufficient separation; too many may lose information and increase computational load.
3. **Thresholding:** To eliminate noise and irrelevant high-frequency components, thresholding is applied to the decomposed high-frequency coefficients. Common threshold functions include:

a. Hard Thresholding: Coefficients with absolute values greater than the threshold λ are retained; others are set to zero. This preserves local signal features but may cause reconstruction oscillations.

b. Soft Thresholding: Coefficients with absolute values greater than λ are shrunk towards zero (subtracting λ). This yields smoother reconstructed signals but may blur edge information. Soft thresholding is more widely used in practice.

4.3 Signal Separation and Denoising Procedure

Assuming the observed noisy signal model is $s(t) = f(t) + e(t)$, where $f(t)$ is the effective signal and $e(t)$ is noise (typically assumed to be Gaussian white noise), the objective of wavelet-based signal separation/denoising is to suppress noise while preserving or separating target frequency components (e.g., live load effects). The basic procedure is as follows:

1. **Decomposition:** Select an appropriate wavelet function and decomposition level, and perform discrete wavelet decomposition on signal $s(t)$ to obtain low-frequency approximation coefficients and high-frequency detail coefficients at each level.
2. **Thresholding:** Apply the selected threshold function to the high-frequency detail coefficients to suppress noise and irrelevant high-frequency components.
3. **Reconstruction:** Perform inverse wavelet transform using the processed coefficients (including low-frequency coefficients and thresholded high-frequency coefficients) to reconstruct the separated or denoised signal.

For the bridge monitoring data processing task in this paper, the key wavelet analysis parameters are selected as follows: The Daubechies wavelet (db4) is chosen due to its compact support and orthogonality, which provide good time-frequency localization suitable for analyzing non-stationary strain signals. Its appropriate number of vanishing moments (4) effectively extracts high-frequency transient components while smoothing low-frequency trends, and it offers high computational efficiency.

The decomposition level is determined to be 7. Based on a 200 Hz sampling frequency and signal characteristics, a 7-level decomposition finely divides the frequency bands. Empirical verification shows that 7 levels achieve an optimal balance between separation effectiveness and computational complexity. Soft thresholding is employed because its coefficient shrinkage enables smoother processing, avoiding potential signal discontinuities (Gibbs phenomenon)

associated with hard thresholding, making it more suitable for subsequent analysis of slowly varying components.

In summary, for this stage, the algorithm uses the db4 wavelet for 7-level decomposition, applies soft thresholding to detail coefficients D1-D7, and reconstructs the signal to obtain high-frequency vibration signals and mid-frequency vehicle-induced signals. This method effectively separates load effects of different frequencies.

5 Algorithm Training and Testing

5.1 Deep Kernel Regression Model Training and Testing

The processed 227 days of temperature-induced strain data for each strain gauge were segmented into 1093 data sequences of 300-minute length. Data from 8 sensors resulted in a total of 8744 sequences. The dataset was split into training, validation, and test sets with a 70%/15%/15% ratio. Model hyperparameters were optimised based on the trained model's performance on the validation set. After obtaining the optimal model, its performance was evaluated by comparing it with classical algorithms (AR, ARIMA) [13] and deep learning algorithms (RNN [14], GRU [15], LSTM [7]) on the same test set. The prediction results of various methods on the test set are shown in Figure 5.

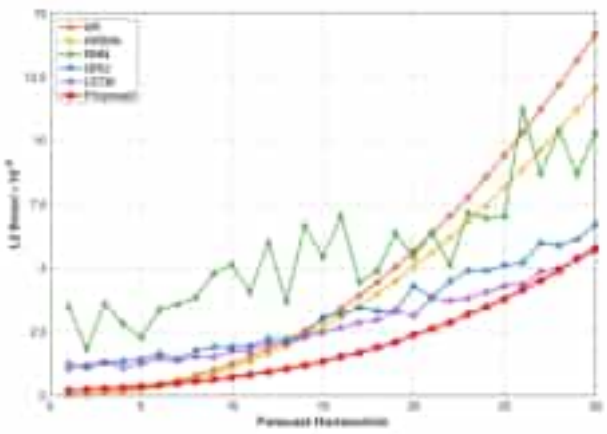


Figure 5. Error over time

Overall, traditional algorithms (AR/ARIMA) exhibited smaller errors and smoother prediction trends in the initial prediction period (approximately the first 5 minutes), aligning with the minor fluctuation characteristics of temperature-induced strain. However, as the prediction horizon increased, errors grew significantly, deviating from true values, indicating their poor adaptability for long-term prediction. Deep learning algorithms (RNN/GRU/LSTM) generally followed the actual trend, but predicted values showed substantial volatility and instability, significantly differing from the actual slowly varying state of temperature-induced strain.

Testing demonstrates that the proposed model maintains the interpretability and output stability of Bayesian methods while achieving superior medium-to-long-term prediction accuracy compared to the benchmark algorithms. Furthermore, the model framework exhibits good extensibility, allowing for the incorporation of additional sensor information to further enhance predictive performance.

5.2 Wavelet Multiresolution Analysis Algorithm Testing

To validate the proposed second-stage signal separation method, a segment of measured strain data was selected for testing. First, the Deep Kernel Regression (DKR) algorithm proposed in Section 3 was applied to extract the low-frequency temperature-induced strain. The original data were then subtracted by this thermal component to obtain strain data with temperature effects removed.

The signal after temperature effect removal still contained noticeable noise and high-frequency vibration induced by vehicle-bridge coupling. To further extract the vehicle-induced quasi-static effect, a 6,000-sample data segment (corresponding to 30s) containing a vehicle crossing event was selected, and wavelet multiresolution analysis was applied for high-frequency separation. After signal reconstruction, the vehicle-induced quasi-static effect was obtained, as shown in Figure 6. The smoothness of the reconstructed signal was significantly improved. The signal shapes from all four sensors were similar, each presenting a unimodal smooth curve consistent with the theoretical bridge influence line. This verifies that the separation method effectively extracts the quasi-static response induced by vehicle loads while suppressing high-frequency noise and vibration.

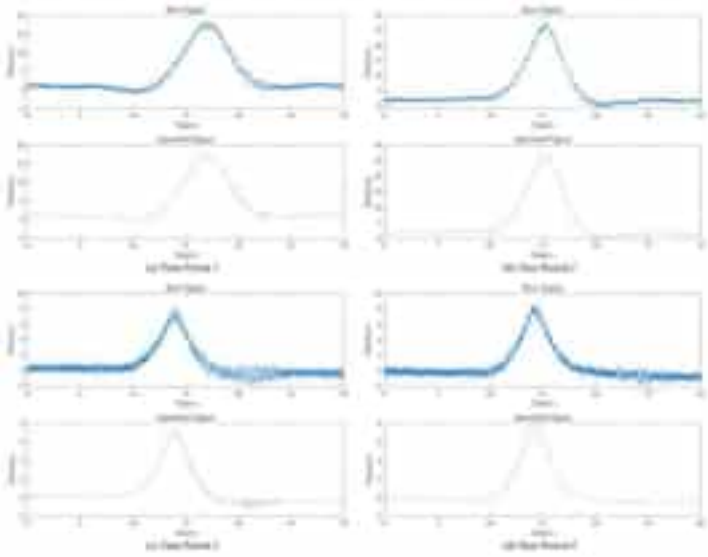


Figure 6. Signal Decomposition Method Based on Wavelet Multi-Resolution Analysis

To further analyse the separation effectiveness, wavelet time-frequency diagrams of the signal before and after high-frequency separation were compared (Figure 7). It can be observed that before separation (Raw), the time-frequency diagram contained numerous scattered energy points in the high-frequency region, primarily originating from sensor noise and vehicle-bridge coupling high-frequency vibrations. After separation (Denoised), the energy distribution of high-frequency components almost completely disappeared, with the time-frequency diagram concentrated in the low-frequency region. Time-frequency analysis indicates that the proposed method effectively filters high-frequency components from the signal and successfully extracts the low-frequency-dominated physical response of vehicle-induced quasi-static effects, validating the effectiveness of the signal separation algorithm. Combining the demarcated algorithm with some existing vehicle perception algorithms enables the refined research on vehicle-bridge correlation [16-19].

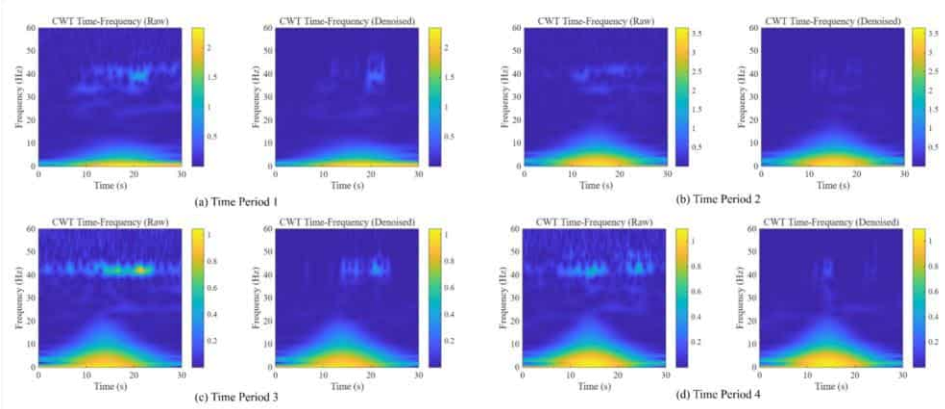


Figure 7. Wavelet time-frequency diagrams of strain before and after high-frequency signal separation

6 Conclusion

This paper addresses the challenge of mixed multi-source signals in bridge health monitoring by proposing an innovative two-stage blind source separation method. Employing a staged processing strategy, it integrates deep kernel regression with wavelet multiresolution analysis, effectively tackling the difficulty traditional methods face in simultaneously handling signal components with large frequency disparities.

In the first stage, we constructed an adaptive model based on deep kernel regression. By fusing the sequential feature extraction capability of LSTM networks with the probabilistic framework advantages of GPR, precise separation of temperature-induced effects was achieved. This model dynamically adjusts kernel parameters based on environmental factors, demonstrating performance superior to traditional methods in medium-to-long-term

prediction. In the second stage, we optimized the wavelet multiresolution analysis method, determining parameter configurations suitable for bridge monitoring signals. This method effectively separated vehicle-induced quasi-static effects from high-frequency vibration and noise, reconstructing smooth response curves that align with theoretical expectations.

Experimental validation based on 227 days of measured data demonstrates that the proposed two-stage method can clearly separate three types of physical effect components from complex raw signals: low-frequency thermal trends, mid-frequency vehicle-induced responses, and high-frequency noise. This method provides an approach to decouple components induced by highly random human activities from infrastructure responses [20] and an effective solution for the fine-grained processing of bridge health monitoring data.

Acknowledgement

This work was supported by CityU PAP Start-Up Grant 9382011 and Science and Technology Plan Project of Hebei Provincial Department of Transport "Intelligent Vehicle-Bridge Correlation Monitoring Technology for the Long-Span Corrugated Steel Web Continuous Girder Bridge of National Highway G107 Zhanghe River Grand Bridge".

References

[1] K. Y. Wong, Design of a structural health monitoring system for long-span bridges, *Structure and Infrastructure Engineering*, vol. 3, no. 2, pp. 169–185, Jun. 2007.

[2] Q. Xia, Y. Cheng, J. Zhang, et al., In-service condition assessment of a long-span suspension bridge using temperature-induced strain data, *Journal of Bridge Engineering*, vol. 22, no. 3, Art. no. 04016124, 2017.

[3] S. Li, H. Xu, X. Zhang, et al., Automatic uncoupling of massive dynamic strains induced by vehicle-and temperature-loads for monitoring of operating bridges, *Mechanical Systems and Signal Processing*, vol. 166, pp. 108332, 2022.

[4] L. Zhou, Y. Xia, J. M. Brownjohn, et al., Temperature analysis of a long-span suspension bridge based on field monitoring and numerical simulation, *Journal of Bridge Engineering*, vol. 21, no. 1, Art. no. 04015027, 2016.

[5] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, Deep kernel learning, in *Artificial Intelligence and Statistics*, pp. 370–378. PMLR, 2016.

[6] Cleveland, W.S.. Robust locally weighted regression and smoothing scatterplots. Journal of the *American statistical association*, 74(368), pp.829-836, 1979.

[7] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] W. Fu, Y. Chen, Y. Luo, H.-P. Wan, Z. Ma and Y. Shen, Deformation prediction model of large-span prestressed structure for health monitoring based on robust Gaussian process regression, *Engineering Structures*, vol. 318, pp. 118597, 2024.

[9] B. Xu and C. Liu, A deep kernel regression-based forecasting framework for temperature-induced strain in large-span bridges, *Engineering Structures*, vol. 323, pp. 119259, 2025.

[10] A. Ouahabi, Ed., Signal and image multiresolution analysis. Hoboken, NJ: John Wiley & Sons, 2012.

[11] S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7), pp.674-693, 2002.

[12] I. Daubechies. Ten lectures on wavelets. *Society for industrial and applied mathematics*, 1992.

[13] G.E. Box, G.M. Jenkins, G.C. Reinsel and G.M. Ljung. Time series analysis: forecasting and control. *John Wiley & Sons*, 2015.

[14] J.L. Elman. Finding structure in time. *Cognitive science*, 14(2), pp.179-211, 1990.

[15] J. Chung, C. Gulcehre, K. Cho , et al., Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint* arXiv:1412.3555, 2014.

[16] B. Xu and C. Liu, Keypoint detection-based and multi-deep learning model integrated method for identifying vehicle axle load spatial-temporal distribution, *Advanced Engineering Informatics*, vol. 62, pp. 102688, 2024.

[17] Y. Wu, T. Wang, R. Gu, et al., Nighttime vehicle detection algorithm based on image translation technology, *Journal of Intelligent & Fuzzy Systems*, vol. 46, no. 2, pp. 5377–5389, 2024.

[18] B. Xu, X. Liu, G. Feng, et al., A monocular-based framework for accurate identification of spatial-temporal distribution of vehicle wheel loads under occlusion scenarios. *Engineering Applications of Artificial Intelligence*, 133, p.107972, 2024.

[19] B. Xu, X. Jiao, Q Wang, et al., A Review of Computer Vision-Based Algorithms for Vehicle Load Spatial-Temporal Distribution Identification. *Measurement Science and Technology*, 2026.

[20] Q.C. Wang and W, Li. Activity-based models for smart and sustainable urban environment. *Routledge Handbook of Smart Built Environment*, pp. 220-242. Routledge, 2025.

Construction Site Layout Monitoring Using Instance Segmentation

Vimal Bharathi and Jochen Teizer

Department of Civil and Mechanical Engineering, Technical University of Denmark
vvebh@dtu.dk, teizerj@dtu.dk

Abstract

Construction sites are complex and dynamic work environments with frequent changes in layout due to material handling and a diverse workforce presence, necessitating adaptable safety measures. Falls from height remain the primary cause of fatalities in construction, particularly impacting small firms with limited resources. Conventional manual construction site layout inspections are subjective, relative infrequent, and difficult to scale, often resulting in noncompliance with fall protection standards. This paper introduces an automated, geo-aware monitoring system that leverages high-resolution orthomosaic images from consumer drones and a novel Adaptive Multi-Scale DETR (AMS-DETR) model for panoptic segmentation. AMS-DETR enhances traditional DETR performance through dynamic query generation and multi-scale attention yielding lower inference time and segmentation Quality. This approach enables proactive, AI-driven site monitoring, improves regulatory adherence, and offers scalable safety management suitable for resource-constrained projects, advancing intelligent construction safety practices.

Keywords –
DETR, instance segmentation, orthomosaic images, site layout monitoring, unmanned aerial vehicle.

1 Introduction

Construction sites are among the most variable work environments, where complex operations lead to frequent shifts in personnel, materials, and equipment [1]. This constant change poses major challenges in keeping safety measures current and aligned with standards. Unlike stable factory settings, construction sites evolve daily through stages like excavation, framing, and finishing, requiring quick setup of temporary protections such as guardrails to address fall risks at heights over 1.8 m [2]. In 2023, falls caused 421 deaths (39.2% of all U.S. construction fatalities, totaling 1,075) [3], alongside more than 20,000 nonfatal fall-related cases annually [4]; and over 70% of fatal falls happened at companies with fewer than 11 employees, highlighting gaps in monitoring for smaller operations. On a global scale, the International Labor Organization projects over 60,000 construction deaths yearly, with falls responsible for 30–50%, varying by region due to differences in regulations and site conditions [5]. These trends reveal a key problem: even with clear rules (e.g., OSHA's 1.07 m guardrail requirements), unexpected removal of safety gear—due to task demands, worker fatigue, or weather issues quickly makes plans outdated, and over 70% of violations involve absent or incorrectly placed barriers.

As a result, regular compliance checks are essential, but current manual methods using checklists and measurements by safety staff are reactive, subject to human error, and limited by site size and access difficulties. Research shows that up to 65% of fall incidents arise from gaps missed during routine (e.g., weekly) checks [6], worsened by emerging threats like proximity issues: for instance, cranes within 2 m of walkways (violating ANSI Z359.2 buffers) or unguarded gaps between paths and pits, which can significantly increase the risk of secondary impacts [7].

Tools like digital twins, remote sensing, and AI can transform the above issues with real-time tracking that could spot issues early, link to evolving 4D-BIM models for predictions, and fit into standard workflows. However, past efforts are often isolated for example, 2D image detectors miss depth details, while basic Orthomosaic classifiers overlook the fine distinctions needed for tracking specific "things" (e.g., individual guardrail sections) versus backgrounds ("stuff" like debris floors) limiting geospatial tracking and risk calculations [8].

This paper addresses these limitations by proposing a location-aware monitoring pipeline. It integrates high-resolution orthomosaic maps generated from overhead imagery captured by DJI Mini 4 Pro drone and processed via photogrammetry with an advanced Adaptive Multi-Scale DETR (AMS-DETR) model for instance segmentation of safety features (e.g., guardrails, Safety poles, etc.) and site elements (e.g., cranes, excavations, steel plates, etc.). AMS-DETR overcomes key DETR

limitations, including rigid queries, scale mismatches, and boundary inaccuracies, through adaptive query generation and multi-level attention mechanism, delivering higher segmentation quality while enabling real-time inference.

It enables the generation of predictive compliance scores and interactive hazard visualizations within a user-friendly SafeBIM platform [36]. The proposed framework provides a scalable, proactive approach to AI-supported safety oversight on highly dynamic construction sites, with strong potential to enhance fall prevention, regulatory compliance, and overall site management practices.

2 Background

Construction sites present unique challenges for safety monitoring due to their dynamic nature, with frequent changes in layout, equipment, and personnel that heighten risks such as falls from height. Traditional manual inspections are inadequate for real-time compliance, prompting the adoption of AI-driven technologies like Unmanned Aerial Vehicles (UAV), deep learning for image segmentation, and Building Information Modeling (BIM) integration. This section reviews existing DEtection TRansformer (DETR) models for segmentation a critical tool for analyzing site images and related work on AI-based site layout monitoring, highlighting gaps addressed by our proposed Adaptive Multi-Scale DETR (AMS-DETR) framework.

2.1 Existing DETR Models for Panoptic Segmentation

The DETR family has revolutionized object detection by formulating it as a set prediction task using transformer architectures, bypassing traditional components like non-maximum suppression [9]. Rooted in the original transformer model [10], DETR variants have been adapted for panoptic segmentation, which combines instance segmentation of "things" with semantic segmentation of "stuff" [11]. This is essential for processing drone-generated orthomosaic images in construction, where variability in scale, occlusion, and object density demands robust, real-time performance.

Despite advancements, DETR models exhibit persistent limitations in segmentation tasks, including slow convergence, inefficient attention, poor multi-scale handling, and query mismatches. The base transformer [10] lacks spatial hierarchies, leading to high computational costs ($O(n^2)$ complexity) and memory issues for high-resolution images (>1024×1024 pixels), impeding edge-device deployment. DETR [9] introduces fixed queries and bipartite matching for end-to-end detection but requires 300–500 epochs for convergence and struggles with small or occluded objects; panoptic extensions [12] add mask heads, yielding Panoptic Quality ≈43.4 on COCO due to heuristic merging errors.

Deformable DETR [13] sparsifies attention for faster training and better small-object detection but fails in boundary accuracy for crowded scenes and incurs latency from convolutional decoders. DAB-DETR [14] uses dynamic anchors for density adaptation but limits query flexibility, causing over/under-prediction in sparse/dense layouts. DN-DETR [15] employs denoising for quicker convergence but generalizes poorly to sparse scenes and multi-modal data. Lite-DETR [16] prioritizes lightweight design but sacrifices accuracy on textures and small objects. RT-DETR [17] achieves real-time speeds with intra-scale attention but trades PQ for efficiency.

Common flaws include inefficiency (>100 ms inference), mask inaccuracies, fixed queries and weak multi-scale adaptation. Our AMS-DETR mitigates these through adaptive query generation (dynamic counts for crowded and less crowded images) and hierarchical multi-scale attention, achieving higher segmentation quality and less inference time for construction orthomosaic images.

2.2 Site Layout Monitoring via Image Segmentation

Recent studies have advanced site layout monitoring by integrating UAV, AI segmentation, and BIM to enable real-time hazard detection and compliance in dynamic environments. Drone-based photogrammetry generates orthomosaic images for scalable analysis, surpassing manual methods prone to errors in large sites.

Tárrago Garay et al. [18] fused drone point clouds with 4D-BIM for progress/deviation detection (70% reduced verification), supporting unguarded edge identification but without AI segmentation. Han et al. [19] used SAM-based panoramic segmentation for indoor layout, highlighting adaptive segmentation value but limited to indoors. Akinsemoyin et al. [20] applied Faster R-CNN on UAV imagery for hardhat/activity detection, enabling machinery-proximity rerouting but instance-only. Alexandre et al. [21] combined AI-drones with LiDAR for anomaly/progress tracking, focusing on compliance without panoptic depth. Recent YOLOv8-seg improvements have advanced construction site monitoring Bhai et al. [37], offering good real-time performance on standard images. However, they face challenges with extreme scale variations and global context in large orthomosaics. Our method addresses these via adaptive queries and hierarchical multi-scale attention, enabling superior generalization for drone-based orthomosaic analysis.

In recent studies Hong and Teizer [22] introduced RAM module in DTCS, integrating SafeBIM rule-checking with RTK-GNSS trajectories for real-time

struck-by hazard detection, proximity analysis, and dynamic walkway optimization, though sensor-based rather than vision-segmentation. Pfitzner et al. [23] developed KG-enhanced CV (YOLOv8x + GNN) for concrete pouring progress, linking BIM volumes to site images via spatial-temporal graphs for bottleneck identification, but activity-specific without broad safety. Pfitzner et al. [24] proposed grid-based spatial complexity analysis with ViTPoseActivity and BIM projection for rebar duration prediction, enabling localized productivity insights but focused on labor efficiency over fall hazards.

These approaches show progress in BIM-CV fusion and real-time tracking but rely on isolated detectors, lack unified instance segmentation for multi-scale features in orthomosaics, or require extra sensors, limiting geospatial risk mapping (e.g., precise guardrail/pathway buffers).Our AMS-DETR framework advances by delivering real-time instance segmentation of drone orthomosaics (guardrails, cranes, edges, pathways) with geo-spatial information for real time 4D-BIM comparison for the onsite safety Managers.

3 Methodology

To address persistent limitations in existing DETR-based instance segmentation models particularly fixed query cardinality and inefficient multi-scale feature fusion we propose an Adaptive Multi-Scale extension to the SOLQ network (AMS-DETR) for real-time analysis of high-resolution construction orthomosaic images. Building upon SOLQ's end-to-end query-based instance segmentation formulation [25], the proposed approach integrates adaptive query modulation and scale-aware hierarchical attention to improve robustness across highly variable construction scenes. Unlike earlier DETR variants that rely on a fixed number of object queries (e.g., 100 queries in [9, 17]) or convolution-heavy mask heads [26], AMS-DETR preserves SOLQ's fully transformer-based decoding while enabling dynamic allocation of representational capacity based on scene complexity. This design is particularly suited for drone-based site monitoring, where images range from sparse excavation zones to dense equipment clusters within the same orthomosaic images.

The novelty of AMS-DETR lies in its targeted architectural extensions to the SOLQ framework rather than introducing a new segmentation paradigm. Similarly, approaches such as [27] focus on global mask transformers but introduce redundant memory paths that scale poorly in occluded, high-resolution scenes. In contrast, the proposed method augments SOLQ with complexity-driven query adaptation and hierarchical multi-scale attention, allowing the decoder to adjust both the number and scale sensitivity of active queries without introducing heuristic post-processing or convolutional bottlenecks. These additions improve instance separation and boundary accuracy for small and overlapping objects, yielding measurable gains over baseline SOLQ and Mask DINO [26], while maintaining sub-50 ms inference on 1024×1024 inputs. AMS-SOLQ further reduces memory consumption by approximately 40%, making it suitable for edge deployment in construction site monitoring scenarios.

3.1 Adaptive Query Mechanism

DETR-based instance segmentation models employ a fixed number of object queries (e.g., 100 learnable embeddings in DETR [9] and DAB-DETR [14]), which leads to inefficiencies when processing scenes with large variations in object density. In construction orthomosaic imagery, this manifests as missed detections in densely

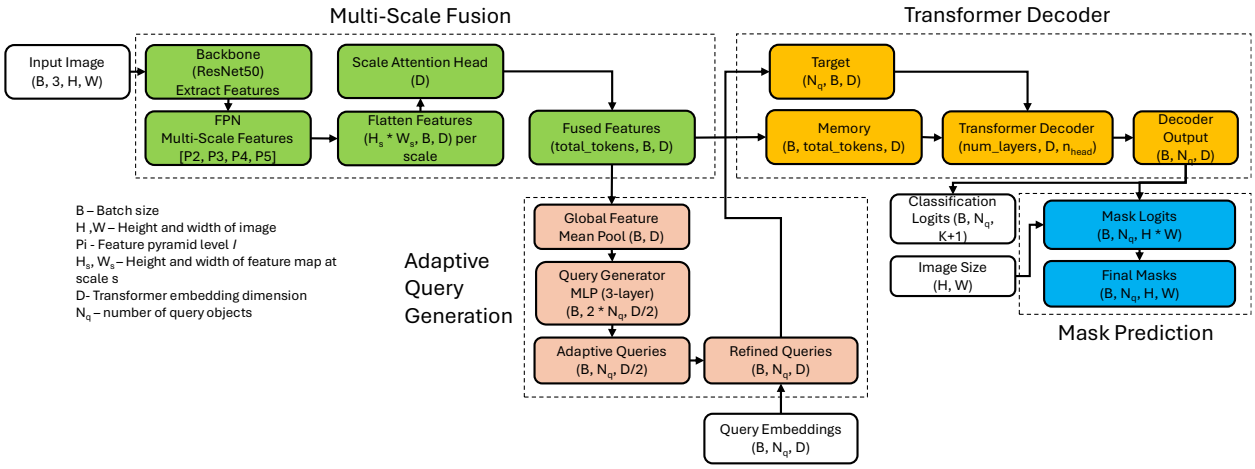


Figure 1: AMSDETR network architecture

populated regions (e.g., clustered equipment) and redundant predictions in sparse areas (e.g., open work zones), as also observed in Co-DETR [28] and RT-DETR [17].

To address this limitation within the SOLQ framework, AMS-DETR introduces an adaptive query mechanism that dynamically modulates the effective number of queries based on scene complexity. Encoder feature maps are globally aggregated and passed through a lightweight multi-layer perceptron (MLP) to estimate image complexity using normalized feature entropy:

$$H(I) = -\sum_i pi \, log \, pi \quad (1)$$

Where p_i represents normalized activation responses. The number of active queries is then determined as:

$$N *= clip(N0 + \alpha H(I), Nmin, Nmax) \quad (2)$$

With N_0=150 base queries and N_{max}=300. Queries beyond N^* are progressively suppressed during decoding using sigmoid-gated residual refinement to ensure stable optimization.

Unlike prior approaches that rely on fixed query sets [9], [14], the proposed mechanism allows the SOLQ decoder to allocate representational capacity in proportion to scene complexity, improving instance separation in crowded regions while preventing over-prediction in sparse layouts. This adaptive strategy enhances matching efficiency and reduces query redundancy without introducing heuristic filtering or post-processing. Empirically, the mechanism improves recall for small and overlapping objects compared to DAB-DETR [14] and Co-DETR [28], while maintaining real-time inference performance comparable to RT-DETR [17].

3.2 Multi-Scale Hierarchical Attention

DETR-based models exhibit limited robustness to scale variation due to rigid global attention mechanisms in DETR [9] or insufficient cross-scale fusion in Deformable DETR [13], leading to degraded performance for small objects and occluded instances in high-density construction scenes. This limitation exists mainly in large orthomosaic images, where objects span multiple orders of magnitude in spatial scale.

To address this, AMS-DETR augments the SOLQ feature encoding stage that combines deformable attention [13], [17] with global context-guided scale aggregation inspired by GM-DETR [29]. Feature maps extracted from a multi-scale backbone (e.g., ResNet-50) at strides $s \in \{8,16,32\}$ are processed in parallel and fused through scale-aware attention.

Each scale is first projected using a lightweight 1×1convolution to obtain scale-specific embeddings. A global context vector q, obtained via global feature aggregation, is then used to compute adaptive attention weights across scales. Specifically, for each scale s, a scale attention weight $\alpha_s$is computed as:

$$\alpha_s = Softmax\left(\frac{\mathrm{QK_s}}{\sqrt{d}}\right) \quad (3)$$

where $\mathrm{k}_s$denotes the pooled key embedding of scale s, and $d$is the embedding dimension. The final scale-aware representation is obtained via a weighted aggregation of scale-specific features:

$$F = \sum_s \alpha_s \mathrm{V}_s \quad (4)$$

with $\mathrm{V}_s$denoting the projected feature map at scale s. This formulation enables the encoder to dynamically emphasize high-resolution features for small objects (e.g., safety posts) and lower-resolution features for larger structures (e.g., work zones), thereby improving instance localization across a wide range of object scales.

Overall, the proposed hierarchical attention improves small-object recall and occlusion robustness compared to standard DETR [9] and Deformable DETR [13], while maintaining efficient inference through sparsified attention similar to RT-DETR [17]. Compared to convolution-heavy multi-scale fusion approaches such as Mask DINO [26], the proposed method achieves improved boundary precision without introducing high-resolution bottlenecks, making it suitable for real-time orthomosaic analysis.

4 Training Dataset

The data set used for this paper was collected weekly over a period of 1 year using DJI Mini 4 Pro. The data was collected in a time-series using manual path planning. The images collected per flight are processed into Orthomosaic images using Agisoft Metashape Pro [30] which can be seen in Figure 3. This resulted in approximately 45 Orthoimages with a resolution of approx. 14.000 x 14.000 pixels which was then split into 1024x1024 pixels and processed for training. The objects in the images were labelled using LabelMe software.

The labeled classes in Figure 2 were selected based on their direct relevance to fall-from-height hazards and site layout compliance monitoring, as guided by OSHA [2] and ANSI [7] standards. Key safety elements include guardrails, poles, and excavations (to detect unguarded edges), while site elements such as cranes, steel plates, and work areas support geospatial risk analysis (e.g., proximity to pathways) and integration with 4D-BIM models. These classes represent the most frequent and critical objects observed across the one-year dataset.

The resulting dataset contains approximately 1500 images which are filled with objects used for this work after filtering out the images that do not contain any objects. The dataset is split into 80% training and 20% validation data.



Figure 2: Labelled objects used for this work to monitor the construction site layout activity.



Figure 3: Cropped orthoimages from drone flights.

5 Results

Performance is evaluated using standard COCO-style Average Precision (AP) metrics. Bounding box AP and mask AP are reported at IoU thresholds from 0.5 to 0.95 (AP@0.5:0.95), following common practice in instance segmentation benchmarks. The performance of our model during training can be seen in the Table 1.

The overall Box AP and Mask AP are primarily affected by the relatively low performance on smaller object classes such as safety poles, cranes, and certain equipment categories. These objects occupy fewer pixels in high-resolution orthomosaic images and often appear with limited visual detail, making them more difficult for the model to detect and segment accurately.

Per-class results in Table 2 confirm that larger and well-defined objects (e.g., office containers, rollers, storage containers) achieve high AP scores, while smaller or thin-structured objects (e.g., safety poles) exhibit significantly lower performance. This imbalance explains the gap between overall metrics and the strong qualitative segmentation results observed for larger and safety-critical structures in Figure 5.

This shows that the instance segmentation over a large orthoimages of approx. 14000x14000 pixels has proved to be a fruitful way for monitoring site layout of construction site. The images also show the number of instances of each object at the bottom right. The training metrics can be seen in Figure 4 which shows the Box loss

Table 1: Performance metrics.

Model	Epochs	Box AP (0.50:0.95)	Mask AP (0.50:0.95)
AMS-DETR	140	0.54	0.42
SOLQ	170	0.42	0.29

Table 2: Per-Class AP

Class	Box AP (0.5:0.95)	Mask AP (0.5:0.95)
Articulated truck	0.54	0.47
Crane	0.37	0.29
Fence	0.61	0.41
Gate	0.45	0.23
Hydraulic excavator	0.57	0.39
Office container	0.89	0.79
Protective barrier	0.54	0.49
Roller	0.90	0.90
Safety pole	0.34	0.27
Stairs	0.68	0.58
Steel plates	0.38	0.36
Storage container	0.77	0.71
Work area	0.76	0.74

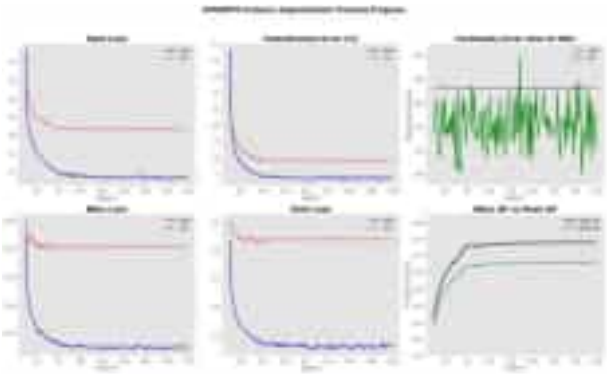


Figure 4: Training metrics.

and GIoU loss along with the BBox AP and Mask AP. The classes that most significantly affect the overall metrics are small-scale and thin-structured objects, particularly safety poles and similar site elements. Due to their limited spatial footprint in large orthomosaic images (~14,000 × 14,000 pixels), these objects are harder to localize and segment precisely, leading to lower AP values and a disproportionate impact on the aggregated performance metrics.

6 Limitations and Conclusion

The results demonstrate the efficacy of the proposed AMS-DETR model in addressing the challenges of real-time instance segmentation for construction site orthomosaic images, particularly in dynamic environments where fall hazards evolve rapidly. As shown in Table 1, the model's performance metrics highlight significant improvements in segmentation quality and inference time compared to baseline DETR variant. This mechanism dynamically adjusts query counts based on scene entropy (as quantified in the methodology), enabling the model to efficiently process sparse scenes (e.g., isolated excavations) without redundant predictions and dense clusters (e.g., equipment-crowded zones) without under-detection.

Figure 4 training metrics further illustrate the model's convergence advantages: the box loss and generalized IoU (GIoU) loss stabilize within 70–100 epochs, faster than traditional DETR [9] or Deformable DETR [13], Mask AP improvements are evident in Figure 5, where boundary precision is maintained even in crowded predictions, with a confidence threshold of 80% effectively filtering noise while preserving accurate instance separation. However, the overall metrics are impacted by challenges in detecting and segmenting small-scale objects, such as safety poles and certain equipment instances. These objects often occupy very few pixels in orthomosaic images and may suffer from low contrast, occlusion, or scale variation, making them difficult for transformer-based models to capture effectively. This limitation is consistent with known challenges in instance segmentation, where small objects contribute disproportionately to performance degradation despite accurate detection of larger structures which align with observations in related works [23, 24], where knowledge graph enhancements or grid-based analyses mitigate similar issues in activity-specific segmentation, suggesting potential synergies for future iterations. This observation highlights a known limitation of transformer-based detectors in high-resolution imagery, where small-object representation and feature resolution remain critical bottlenecks despite improvements in multi-scale attention.



Figure 5: Full ortho instance segmentation for construction site layout status on 2025-05-17.

Practically, the full orthomosaic instance segmentation in Figure 5 exemplifies the system's geospatial utility by geo-aligning segmented features with 4D-BIM models, deviations (e.g., unguarded edges <10 cm from pathways) can be detected with cm-level accuracy, facilitating Voronoi-based risk heatmaps and proximity analyses (e.g., cranes within 2 m of walkways) which is part of the future works. This proactive capability surpasses reactive manual inspections [6], offering scalable compliance monitoring for small firms. Integration with the SafeBIM platform [36] enables interactive visualizations, such as predictive compliance scores derived from temporal changes across the 1-year dataset, empowering site managers to prioritize interventions and reduce secondary risks.

Despite these strengths there are certain limitations. The dataset, while comprehensive (1500 labelled images from 45 orthomosaics), is site-specific and collected via manual drone paths, potentially introducing biases in viewpoint and weather conditions. Model inference, though real-time (<50 ms/frame), may degrade on ultra-high-resolution inputs without further optimization, and the reliance on LabelMe for annotations could amplify human error in ambiguous regions like "work areas." Compared to sensor-fusion approaches [22, 31] our vision-only pipeline is cost-effective but lacks depth cues, limiting 3D hazard assessment. Future work could incorporate multimodal data (e.g., LiDAR) or advanced refinements like diffusion-based boundary sharpening [32] to enhance amorphous class handling. Additionally, expanding the dataset with diverse global sites could improve generalization, aligning with ILO's call for region-specific safety adaptations [5].

Furthermore, aligning construction site layout monitoring using instance segmentation and using it in a Digital Twin for Construction Safety (DTCS) framework has potential for practitioners to detect and track relevant progress while comparing as-performed vs. as-planned status effortlessly [33-36]. Providing previously not available information in safety meetings (e.g., toolbox and safety planning or review meetings) adds value to select and apply the best control measures.

Acknowledgment

The research presented in this paper was funded by the European Union Horizon 2020 research and innovation program under grant agreement no. 101162318 (AM2PM) and Erhvervsfyrtårn Femern (FF6S). The Business Lighthouse Femern is co-financed by the European Union and the Danish Business Development Agency.

References

[1] J. Teizer, "Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 225–238, Apr. 2015. https://doi.org/10.1016/j.aei.2015.03.006

[2] OSHA, "Fall protection standards," 2023. https://www.osha.gov/fall-protection/standards

[3] "National census of fatal occupational injuries in 2023", U.S. Bureau of Labor Statistics (BLS), 2024. https://www.bls.gov/news.release/cfoi.nr0.htm

[4] "Nonfatal occupational injuries and illnesses," U.S. Bureau of Labor Statistics (BLS), 2023. https://www.bls.gov/iif/

[5] "Work-related fatalities in construction," International Labour Organization (ILO), 2023. https://www.ilo.org

[6] H. J. Lipscomb, A. L. Schoenfisch, W. Cameron, K. L. Kucera, D. Adams, and B. A. Silverstein, "How well are we controlling falls from height in construction? Experiences of union carpenters in Washington State, 1989–2008," *American J Industrial Med*, vol. 57, no. 1, pp. 69–77, Jan. 2014, https://doi.org/10.1002/ajim.22234

[7] "Z359.2 Fall protection code," American National Standards Institute (ANSI), 2023. https://webstore.ansi.org

[8] M. Golparvar-Fard, F. Peña-Mora, C. A. Arboleda, and S. Lee, "Visualization of Construction Progress Monitoring with 4D Simulation Model Overlaid on Time-Lapsed Photographs," *J. Comput. Civ. Eng.*, vol. 23, no. 6, pp. 391–404, Nov. 2009. https://doi.org/10.1061/(ASCE)0887-3801(2009)23:6(391)

[9] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," May 28, 2020, *arXiv*. https://arxiv.org/abs/2005.12872

[10] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 02, 2023, *arXiv*. https://arxiv.org/abs/1706.03762

[11] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic Segmentation," Apr. 10, 2019, *arXiv*: arXiv. https://arxiv.org/abs/1801.00868

[12] Z. Li *et al.*, "Panoptic SegFormer: Delving Deeper into Panoptic Segmentation with Transformers," Mar. 18, 2022, *arXiv*. https://arxiv.org/abs/2109.03814

[13] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection," Mar. 18, 2021, *arXiv*: arXiv. https://arxiv.org/abs/2010.04159

[14] S. Liu *et al.*, "DAB-DETR: Dynamic Anchor Boxes are Better Queries for DETR," Mar. 30, 2022, *arXiv*: arXiv. https://arxiv.org/abs/2201.12329

[15] F. Li, H. Zhang, S. Liu, J. Guo, L. M. Ni, and L.

Zhang, “DN-DETR: Accelerate DETR Training by Introducing Query DeNoising,” Dec. 08, 2022, *arXiv*. https://arxiv.org/abs/2203.01305

[16] F. Li *et al.*, “Lite DETR : An Interleaved Multi-Scale Encoder for Efficient DETR,” Mar. 13, 2023, *arXiv*. https://arxiv.org/abs/2303.07335

[17] Y. Zhao *et al.*, “DETRs Beat YOLOs on Real-time Object Detection,” Apr. 03, 2024, *arXiv*. https://arxiv.org/abs/2304.08069

[18] N. Tárrago Garay, J. C. Jimenez Fernandez, R. San Mateos Carreton, M. A. Montes Grova, O. Kruth, and P. Elguezabal, “Integration of Drone-Based 3D Scanning and BIM for Automated Construction Progress Control,” *Buildings*, vol. 15, no. 19, p. 3487, Sep. 2025. https://doi.org/10.3390/buildings15193487

[19] J. Han, S. Yoon, M. Kang, and T. Kim, “Approach to Enhancing Panoramic Segmentation in Indoor Construction Sites Based on a Perspective Image Segmentation Foundation Model,” *Applied Sciences*, vol. 15, no. 9, p. 4875, Apr. 2025. https://doi.org/10.3390/app15094875

[20] A. Akinsemoyin, I. Awolusi, D. Chakraborty, A. J. Al-Bayati, and A. Akanmu, “Unmanned Aerial Systems and Deep Learning for Safety and Health Activity Monitoring on Construction Sites,” *Sensors*, vol. 23, no. 15, p. 6690, Jul. 2023. https://doi.org/10.3390/s23156690

[21] A. Oliveira, A. Dias, T. Santos, P. Rodrigues, A. Martins, and J. Almeida, “LiDAR-Based Unmanned Aerial Vehicle Offshore Wind Blade Inspection and Modeling,” *Drones*, vol. 8, no. 11, p. 617, 2024. https://doi.org/10.3390/drones8110617

[22] K. Hong and J. Teizer, “Digital construction site layout planning and real-time trajectory analysis for proactive safety monitoring and control of struck-by hazards,” *Automation in Construction*, vol. 177, p. 106353, Sep. 2025. https://doi.org/10.1016/j.autcon.2025.106353

[23] F. Pfitzner, S. Hu, A. Braun, A. Borrmann, and Y. Fang, “Monitoring concrete pouring progress using knowledge graph-enhanced computer vision,” *Automation in Construction*, 174, p. 106117, Jun. 2025. https://doi.org/10.1016/j.autcon.2025.106117

[24] F. Pfitzner, A. Braun, and A. Borrmann, “Object Detection-Based Knowledge Graph Creation: Enabling Insight into Construction Processes,” in *Computing in Civil Engineering 2023*, Corvallis, Oregon: American Society of Civil Engineers, Jan. 2024, pp. 186–193. https://doi.org/10.1061/9780784485224.023

[25] B. Dong, F. Zeng, T. Wang, X. Zhang, and Y. Wei, “SOLQ: Segmenting Objects by Learning Queries,” Sep. 30, 2021, *arXiv*. https://arxiv.org/abs/2106.02351

[26] F. Li *et al.*, “Mask DINO: Towards A Unified Transformer-based Framework for Object Detection and Segmentation,” Dec. 12, 2022, *arXiv*. https://arxiv.org/abs/2206.02777

[27] A. Kirillov *et al.*, “Segment Anything,” Apr. 05, 2023, *arXiv*. https://arxiv.org/abs/2304.02643

[28] Z. Zong, G. Song, and Y. Liu, “DETRs with Collaborative Hybrid Assignments Training,” Aug. 10, 2023, *arXiv*. https://arxiv.org/abs/2211.12860

[29] X. Liu, X. Yang, L. Shao, X. Wang, Q. Gao, and H. Shi, “GM-DETR: Research on a Defect Detection Method Based on Improved DETR,” *Sensors*, vol. 24, no. 11, p. 3610, Jun. 2024. https://doi.org/10.3390/s24113610

[30] Agisoft LLC, *Agisoft Metashape professional*. (2023). https://www.agisoft.com/pdf/metashape-pro_2_0_en.pdf

[31] W. Choi, S. Na, and S. Heo, “Integrating Drone Imagery and AI for Improved Construction Site Management through Building Information Modeling,” *Buildings*, vol. 14, no. 4, p. 1106, Apr. 2024. https://doi.org/10.3390/buildings14041106

[32] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” Dec. 16, 2020, *arXiv*. https://arxiv.org/abs/2006.11239

[33] S. Siebert, J. Teizer, “Mobile 3D Mapping for Surveying Earthwork Projects using an Unmanned Aerial Vehicle (UAV) System,” Automation in Construction, 2014, vol. 41, pp. 1–14, http://dx.doi.org/10.1016/j.autcon.2014.01.004

[34] V. Bharathi, J. Teizer, “Construction site layout planning and progress monitoring using computer vision on drone images,” European Conference on Computing in Construction, CIB W78 Conference on IT in Construction, Porto, Portugal, 2025, https://doi.org/10.35490/EC3.2025.313

[35] K.W. Johansen, R. Pimentel de Figueiredo, O. Golovina, J. Teizer, “Autonomous Safety Barrier Inspection in Construction: An Approach Using Unmanned Aerial Vehicles and SafeBIM,” 38th International Symposium on Automation and Robotics in Construction, Dubai, UAE, 2021, https://doi.org/10.22260/ISARC2021/0085

[36] J. Teizer, K.W. Johansen, C.L. Schultz, K. Speiser, K. Hong, O. Golovina, “A Digital Twin Model for Advancing Construction Safety,” J. Fottner et al. (Eds.): CLEaR 2023, Lecture Notes in Civil Engineering (LNCE 390), 2024, ISBN: 978-3-031-44020-5, pp. 201–212, https://doi.org/10.1007/978-3-031-44021-2_22

[37] R. Bai, M. Wang, Z. Zhang, J. Lu, and F. Shen, "Automated Construction Site Monitoring Based on Improved YOLOv8-seg Instance Segmentation Algorithm," IEEE Access, vol. 11, pp. 139082-97, 2023, https://ieeexplore.ieee.org/document/10348569

Trajectory-Based Activity Recognition for Earthmoving Operations Using Finite-State Machines

Kepeng Hong and Jochen Teizer

Department of Civil and Mechanical Engineering, Technical University of Denmark, Denmark
keho@dtu.dk, teizerj@dtu.dk

Abstract

Identifying construction equipment activity states is essential for understanding and optimizing earthmoving operations. Trajectory data can be used to infer semantic equipment activities, while it often leads to inconsistent results when simple rule-based methods are applied. This paper presents a trajectory-based activity recognition approach using finite-state machines (FSMs) for earthmoving operations. The method combines kinematic features from Real-Time Kinematic Global Navigation Satellite Systems (RTK-GNSS) trajectories with spatial constraints and predecessor rules to infer activity states of excavators and dump trucks. A real-world earthmoving case study demonstrates that the proposed FSM-based approach captures cyclic operational patterns. The identified activity states provide a reliable basis for downstream operational analysis and simulation, supporting data-driven construction productivity and emission management.

Keywords

Trajectory tracking; Earthmoving operations; Finite-state machine; Activity recognition; RTK-GNSS; Construction automation

1 Introduction

Earthmoving operations are fundamental to infrastructure and building construction, characterized by repetitive and cyclic interactions among heavy construction equipment, such as excavators and dump trucks. The performance of these operations is strongly influenced by how equipment activities are coordinated over time and space. As a result, accurate identification of equipment activity states is required for analyzing productivity, utilization, and operational inefficiencies in earthmoving processes, and for supporting downstream applications such as simulation-based planning and performance assessment. Effective site management requires continuous monitoring of productivity metrics, such as cycle times and utilization rates, to identify bottlenecks [1,2].

Recent advances in Real-Time Kinematic Global Navigation Satellite Systems (RTK-GNSS) have enabled the continuous collection of high-precision trajectory data from construction equipment [3,4]. Such data provides detailed information to infer equipment movement patterns and interactions on site. However, raw trajectory data do not directly represent semantic activity states of equipment, such as *Load*, *Haul*, *Dump*, *Idle*, or *Wait* for dump trucks. Simple rule-based approaches that rely solely on speed or position thresholds often suffer from noisy state transitions and spatiotemporal inconsistencies, especially in congested and dynamic construction environments [5].

To address these challenges, various activity recognition methods have been explored, including rule-based logic, machine learning models, and vision-based approaches [6,7]. While data-driven methods can achieve high classification accuracy, they typically require large labeled datasets and lack interpretability, which limits their practical deployment on construction sites. In contrast, finite-state machines (FSMs) provide an interpretable and structured way to model cyclic equipment operations by explicitly encoding allowable states and transitions. An FSM is a computational model that represents a system as a set of discrete states and allowable transitions between them. By incorporating predecessor constraints, FSMs improve temporal consistency and robustness in activity recognition for construction equipment [5]. In the context of earthmoving operations, FSMs are well suited to describe repetitive operational cycles, where equipment typically transitions through sequential states such as loading, hauling, dumping, and returning.

This paper presents a trajectory-based activity recognition method for earthmoving operations using FSMs. The proposed approach integrates kinematic features derived from RTK-GNSS trajectories with spatial context and predecessor rules to infer temporally consistent activity states of excavators and dump trucks. The identified activity states directly support operation-level analysis and simulation-based applications, enabling data-driven evaluation of earthmoving performance under real-world site conditions.

2 Related Work

Existing research on construction equipment monitoring has primarily focused on collecting and analyzing sensory data to characterize on-site movement and utilization. While trajectory data offers rich insights into the spatiotemporal patterns of construction equipment, effective operational analysis requires the capability to translate raw motion into meaningful activity states [8]. Consequently, this section reviews the literature across two aspects: trajectory-based monitoring of earthmoving operations and activity recognition methods for inferring semantic states.

2.1 Trajectory-based Earthmoving Operation Monitoring

Tracking construction equipment using Global Positioning System (GPS) has been widely investigated to support productivity analysis and operational monitoring. GPS-based systems are extensively applied to capture the spatiotemporal movement of earthmoving equipment, enabling the inference of critical operational patterns, such as cycle times, travel distances, and equipment utilization [9,10]. With the advent of RTK-GNSS, the positioning accuracy of these systems has improved significantly, making trajectory data increasingly suitable for detailed analysis of equipment behavior on construction sites [4].

Recent studies have further explored data-driven and machine learning–based approaches to extract higher-level semantic information from equipment trajectories and sensor data. By leveraging supervised learning, probabilistic models, or deep learning techniques, these methods aim to recognize equipment activities or operational states [10,11], which have demonstrated promising classification accuracy under controlled conditions and highlight the potential of learning-based models for capturing complex motion patterns.

However, inferring detailed equipment activities solely from trajectory data remains challenging. Several studies have attempted to identify activities using speed thresholds or proximity-based rules. While these rule-based methods are computationally efficient and easy to implement, they lack robustness in complex scenarios.

In congested and dynamic construction environments, equipment frequently operates at low speeds, performs short repositioning movements, or temporarily pauses within active zones due to traffic or delays. Consequently, purely kinematic-based approaches often struggle to distinguish between functionally distinct states, leading to noisy state transitions and misclassification. These limitations highlight the critical need for activity recognition methods that account for the cyclic and sequential logic of earthmoving operations, rather than relying exclusively on instantaneous spatial or kinematic threshold rules.

2.2 Activity Recognition Approaches for Construction Equipment

Automated activity recognition serves as the analytical core for evaluating construction productivity and emissions. By analyzing spatiotemporal data, researchers can infer operational patterns at a high level, such as cycle times, haul distances, and overall equipment utilization rates. The integration of high-precision RTK-GNSS has further refined this capability, allowing for the detection of subtle changes based on positioning data.

Armin et al. [11] categorized existing approaches of equipment activity recognition based on their input data sources, including vision-based, audio-based, and motion-based methods. Vision-based approaches rely on camera data and have demonstrated high accuracy, but are sensitive to lighting conditions, occlusions, and site complexity. Audio-based methods utilize acoustic signals and can be cost-effective, but their performance may vary depending on environmental noise. Motion-based approaches, which use trajectory and kinematic data from sensors (e.g., GPS, IMU, or UWB), are more robust to environmental conditions and are well suited for large-scale outdoor construction sites.

However, they require effective modeling of temporal dynamics to infer meaningful activity states from raw motion data. Translating raw kinematic data into semantic activity states remains a complex challenge. Existing methodologies can be broadly categorized into heuristic rule-based approaches [12], classical machine learning methods (e.g., Support Vector Machines [13, 14], Random Forests [15]), and deep learning models (e.g., LSTM [16], CNN [17]). Heuristic rule-based methods infer activities using predefined heuristics, such as speed thresholds or geofence-based proximity rules. These methods are favored for their low computational cost and ease of deployment. But they often lack flexibility in handling complex and irregular operational patterns. In contrast, classical machine learning methods can learn complex patterns from multi-sensor data (e.g., fusing location data with accelerometers), aiming to improve classification accuracy for complex maneuvers [6, 11]. While they improve classification performance over purely heuristic approaches, they typically rely on carefully designed features and overlook temporal dependencies. Deep learning models further enhance classification capability by learning features directly from raw or minimally processed data. However, these methods require substantial labeled data and often lack interpretability, which limits their practical deployment in construction environments.

All these methods have inherent limitations in dynamic site environments. Machine learning models

often suffer from overfitting and lack explainability. On the other hand, purely rule-based approaches are often too rigid. In congested sites, equipment frequently exhibits irregular movement patterns, such as creeping at low speeds due to traffic or performing short repositioning maneuvers, which do not align with static kinematic thresholds. This underscores the necessity for a method that explicitly accounts for the cyclic, sequential nature of earthmoving operations.

In this context, FSM-based approach can be viewed as a structured modeling framework within motion-based methods, explicitly incorporating state-transition logic to improve temporal consistency while maintaining interpretability. Compared with rule-based methods that rely on isolated heuristics, FSM-based approaches explicitly model temporal state transitions for more consistent activity interpretation, while remaining more interpretable and data-efficient than machine learning-based methods.

3 Methods

This section describes the proposed trajectory-based activity recognition method for earthmoving operations. The method integrates RTK-GNSS trajectory data with FSM to infer temporally consistent equipment activity states. The overall workflow consists of trajectory preprocessing, FSM-based state modeling, rule-based candidate state detection, and activity timeline generation. Figure 1 illustrates the overall workflow of the proposed trajectory-based FSM approach for equipment activity recognition.

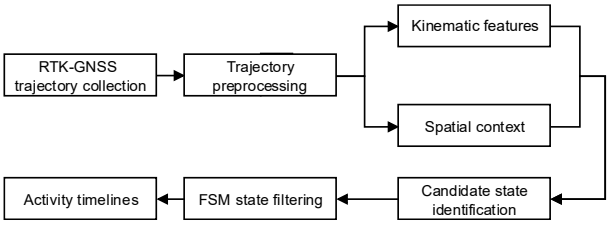


Figure 1. Overview of the trajectory-based FSM workflow for equipment activity recognition in earthmoving operations.

3.1 Trajectory Data Acquisition and Preprocessing

To capture the precise movement patterns of the fleet, equipment trajectories are collected using RTK-GNSS sensors mounted on excavators and dump trucks. These sensors record positioning data, including latitude, longitude, and heading, at a frequency of 1 Hz. Since raw coordinates lack operational context, they are transformed into a site-aligned local coordinate system that matches the digital site layout. This alignment enables the system to perform spatial reasoning, automatically determining whether a machine is located within specific operational zones such as excavation pits, loading points, dumping areas, or haul roads.

From the processed trajectories, two fundamental kinematic features are derived to describe equipment behavior: translational speed and rotational velocity. Translational speed is primarily used to distinguish between stationary states (e.g., *Load*, *Idle*) and mobile states (e.g., *Haul*, *Return*). Meanwhile, rotational velocity, calculated from the change in heading, is a critical indicator for identifying excavator cycles, as these machines frequently rotate their upper structure while the tracks remain static during digging and swinging operations.

However, raw field data inevitably contains noise due to signal multipath effects or temporary obstructions. To prevent the activity recognition model from reacting to these false fluctuations, a temporal smoothing filter is applied to the kinematic data streams. This preprocessing step reduces short-term jitter and satellite signal drift, ensuring that the inputs for the FSM are stable and reliable.

3.2 FSM Design for Earthmoving Equipment

To capture the cyclic workflows of earthmoving operations, separate FSMs are designed for excavators and dump trucks, as illustrated in Figure 2. The state identification integrates instantaneous kinematic features with spatial context. Candidate states are first proposed based on speed and heading, then validated against the equipment's location within defined operational zones (e.g., loading or dumping areas).

For dump trucks, the FSM defines states such as *Load, Haul, Dump, Return,* and *Wait*. Transitions are triggered by specific thresholds, for example, speeds above 2.0 m/s indicate driving modes, while low speeds in specific zones indicate service or queuing. For excavators, the *Excavating* state is distinctively identified by low translational speed ($v < 0.5\ m/s$) combined with high rotational velocity ($\omega > 5\ °/s$).

The FSM enforces predecessor logic to maintain temporal consistency. By explicitly coding allowable transition sequences, such as preventing a truck from entering *Load* without a preceding *Return* state, the system effectively filters out unrealistic state switches caused by signal noise or short interruptions, ensuring robust activity recognition.

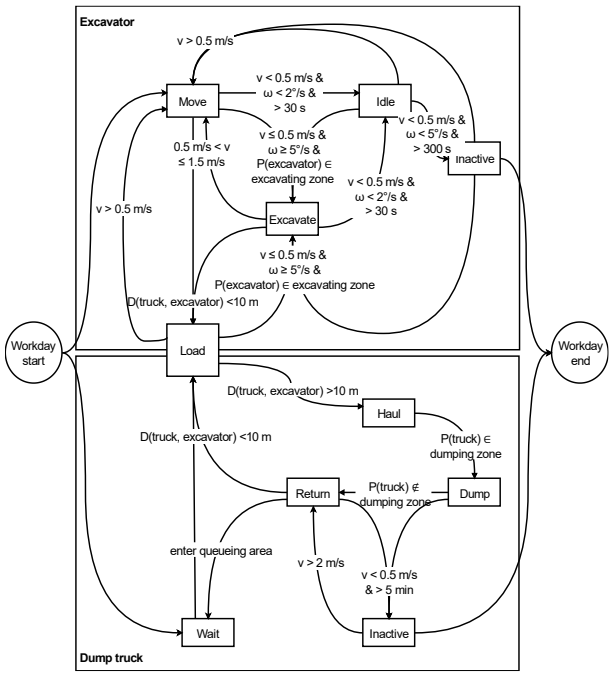


Figure 2. Finite-state machine for excavator and dump truck activity recognition. States (rectangles) and transitions (arrows) are determined by kinematic, spatial, and temporal conditions derived from RTK-GNSS trajectories.

3.3 Rule-based Candidate State Identification

At each discrete time step, candidate activity states are preliminarily identified by evaluating instantaneous kinematic features against predefined thresholds. The primary differentiator is translational speed (v), which serves to classify equipment behavior into stationary, low-speed maneuvering, or high-speed hauling modes. For instance, a speed threshold of 0.5 m/s is selected to distinguish productive stationary activities, such as *Load* and *Dump*, from active transport, while speeds exceeding 2.0 m/s generally indicate a driving state. However, kinematic data alone is often insufficient to determine a stationary state; for example, a truck stops due to traffic looks identical to one stopping for loading. Therefore, spatial constraints are applied as a necessary secondary filter to resolve these ambiguities.

For dump trucks, the identification logic relies heavily on geofenced operational zones to interpret low-speed behaviors. Specifically, a low-speed condition ($v < 0.5\ m/s$) is mapped to a specific activity based on the equipment's containment within defined areas: if the truck is located within an excavator's loading radius (10 m), the state is identified as *Load*; if it is within a designated dumping zone, it is labeled as *Dump*; and if positioned in a queueing area, it is classified as *Wait*. Conversely, high-speed driving periods are categorized into *Haul* or *Return* depending on the truck's position within the haul corridors and its direction of travel relative to the excavation zones. The detailed rule set for dump trucks is summarized in Table 1.

Table 1. State identification rules for dump trucks.

State	Kinematic condition	Spatial condition	Predecessor
Load	v ≤ 0.5 m/s	Distance to excavator ≤ 10 m	*Wait*, *Return*
Dump	v ≤ 0.5 m/s	Within dumping zone	*Haul*
Haul	v ≥ 2.0 m/s	Along haul corridor	*Load*
Return	v ≥ 2.0 m/s	Along return corridor	*Dump*
Wait	v < 0.5 m/s	Within queue area	Duration ≥ 30 s
Inactive	v < 0.2 m/s	Anywhere	Duration ≥ 300 s

Excavators require a more complex rule set due to their stationary working nature, where the chassis remains fixed while the upper structure performs the work. Consequently, rotational velocity becomes a critical indicator alongside translational speed. A combination of low translational speed ($(v < 0.5\ m/s)$ and significant upper-structure rotation ($\omega > 5\ °/s$) within the excavation zone is defined as the *Excavate* state. When it is in close proximity to a dump truck, it is reclassified as *Load*. Repositioning movements are captured when the translational speed falls between 0.5 and 1.5 m/s. Finally, prolonged periods of low motion, both in translation and rotation, are categorized as *Idling* or *Inactive* based on specific duration thresholds, as detailed in Table 2.

The threshold values used in the rule-based candidate state identification were determined based on a combination of domain knowledge, exploratory analysis of the collected trajectory data, and iterative calibration. Specifically, speed thresholds were selected to distinguish between stationary, low-speed maneuvering, and active hauling states observed in the dataset. The rotational velocity threshold reflects typical excavator swing behavior during excavation. The spatial threshold (e.g., 10 m radius) corresponds to practical interaction distances between excavators and dump trucks on site. Duration thresholds were introduced to differentiate short-term pauses from prolonged inactivity.

These parameters can be iteratively refined to ensure temporally consistent activity transitions while avoiding excessive sensitivity to noise in the trajectory data. In practice, moderate variations in these thresholds were found to have limited impact on the overall cyclic activity patterns. These rules generate a set of candidate states that represent plausible equipment activities at each timestep but may still contain short-term inconsistencies.

Table 2. State identification rules for excavators.

State	Kinematic condition	Spatial condition	Predecessor
Load	$v \leq 0.5$ m/s	Distance to truck ≤ 10 m	*Move, Excavate*
Excavate	$v \leq 0.5$ m/s $\omega > 5°$ /s	Within excavating zone	*Move, Load*
Move	$0.5 < v \leq 1.5$ m/s	Anywhere	*Excavate, Load*
Idle	$v \leq 0.5$ m/s $\omega \leq 2°$ /s	Anywhere	Duration > 30 s
Inactive	$v \leq 0.5$ m/s $\omega \leq 2°$ /s	Anywhere	Duration > 300 s

3.4 FSM-Based State Filtering and Activity Timeline Generation

The candidate states are filtered through the FSM transition rules to ensure temporal consistency. Only transitions that are allowed by the FSM are accepted, while invalid transitions are rejected or delayed until valid conditions are satisfied. This process prevents rapid switching between states caused by measurement noise, brief pauses, or short repositioning movements.

The final output of the method is a continuous activity timeline for each equipment unit, describing the sequence and duration of operational states over time. These timelines capture the cyclic structure of earthmoving operations and enable quantitative analysis of duty cycles, utilization, queueing behavior, and non-productive time.

The proposed FSM-based trajectory-driven method provides an interpretable solution for equipment activity recognition in earthmoving operations. By combining kinematic features, spatial context, and FSM-based temporal constraints, the method connects raw trajectory data and semantic operational states. The resulting activity timelines can be directly integrated into downstream applications, such as performance analysis and simulation of earthmoving operations.

4 Case Study and Results

The proposed trajectory-based activity recognition method was evaluated using data collected from a real-world earthmoving operation. The case study involves excavation and hauling activities carried out on an active construction site, where multiple excavators and dump trucks operated simultaneously within a confined workspace.

The study area includes predefined operational zones such as excavation areas, loading zones, dumping areas, haul corridors, return paths, and queueing areas. These zones were derived from the site layout and used to provide spatial context for activity recognition. Trajectory data from one representative workday were selected for analysis to capture typical operational patterns while avoiding interruptions caused by weather or major schedule changes. Figure 3 illustrates the spatial distribution of tracked equipment trajectories overlaid on the site layout and orthophoto.

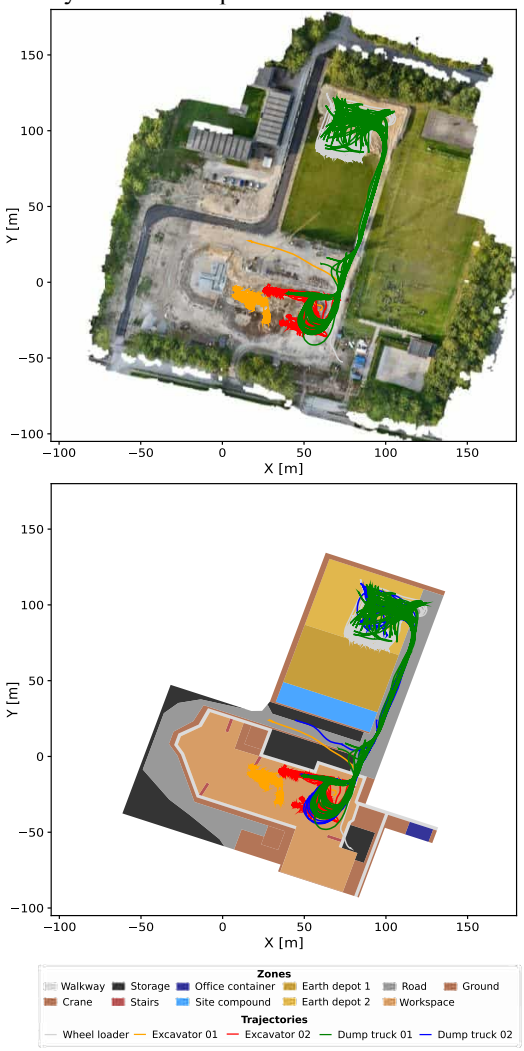


Figure 3. Trajectories of tracked equipment against the 2D site layout (upper image) and orthophoto (lower image) for one workday.

To distinguish the semantic activities visually, Figure 4 shows the typical dumper duty cycle observed from on-site drone images, illustrating the dump trucks' states of *Load*, *Haul*, *Dump*, *Return*, and *Wait* that the algorithm aims to identify.



Figure 4. Observed dump truck duty cycle: (a) Load, (b) Haul, (c) Dump, (d) Return, and (e) Wait.

4.1 Activity Recognition Results

The FSM-based activity recognition method was applied to the processed RTK-GNSS trajectories to infer equipment activity states over time. For each excavator and dump truck, the method generated a continuous activity timeline describing the sequence and duration of operational states. Figure 5 presents the identified activity timelines for the tracked equipment on the selected workday. The results show clear cyclic patterns consistent with typical earthmoving operations, dump trucks alternate between loading at the excavation area, hauling material to the dumping location, dumping, and returning to the excavation site. Compared with simple rule-based classification, the FSM-based approach produces temporally consistent state sequences without rapid fluctuations. Short interruptions, low-speed maneuvers, and brief pauses within operational zones are effectively filtered out by the FSM transition logic.

To quantitatively validate the FSM method, we established a rigorous second-by-second ground truth by manually annotating one hours of synchronized video footage as the upper diagram shown in Figure 5, where two dumpers worked in parallel.

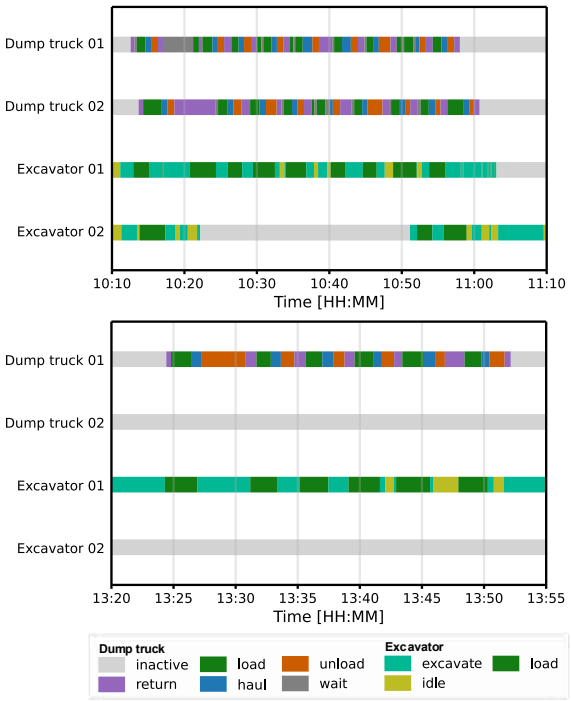


Figure 5. Timeline of identified working stages of the excavators and trucks on one selected working day.

To account for these natural phase shifts without unfairly penalizing the model, we evaluated the recognition accuracy using a standard point-by-point metric with a ± 10-second temporal tolerance window. As detailed in Table 3, the FSM approach demonstrates exceptional reliability for key duty cycle activities. It achieved a satisfying Recall and an outstanding F1-Score of 1.00 for Unload, combined with a high Precision of 0.90 for Haul. The high performance for Unload is mainly due to its clear spatial constraint, as it is defined within a fixed dumping zone. In contrast, other activities depend more on the relative positioning and interactions with excavators, and are more susceptible to interference from surrounding equipment, which introduces ambiguity and leads to comparatively lower performance.

Table 3. Quantitative evaluation of FSM-based activity recognition for dumpers against manual annotation.

Activity	Precision	Recall	F1-Score
Load	1.00	0.39	0.56
Haul	0.90	0.51	0.66
Unload	1.00	1.00	1.00
Return	0.61	0.95	0.74
Wait	0.56	0.59	0.58

4.2 Analysis of Operational Patterns

The extracted activity timelines enable quantitative analysis of equipment duty cycles and interaction patterns. For dump trucks, a substantial portion of non-productive time is associated with waiting and idling near the excavation area, reflecting queue formation when multiple trucks compete for limited excavator capacity. These queueing periods can be distinguished from active loading and hauling states due to the integration of spatial constraints and predecessor logic in the FSM.

For excavators, the activity timelines reveal dominant *Excavate* and *Load* states, with idling primarily occurring during periods of truck unavailability or short operational interruptions. The temporal alignment between excavator *Load* states and truck *Load* states further confirms the consistency of the identified interactions between equipment units.

For this specific analysis, the study focuses on the utilization rate of the dump trucks because the on-site excavators were intermittently tasked with loading external trucks, which introduces variability to the excavator's operation independent of the tracked trucks. Table 4 summarizes the observed work statistics for the selected workday. It presents the total duration and percentage of "Work" (active states: Load, Haul, Dump, Return) vs. "Idle" (Wait) for the monitored equipment.

Table 4. Summary of working cycles of two trucks on one selected workday.

Machine	Work (min)	Idle (min)	Utilization (%)
Truck 01	238.16	11.67	95.33
Truck 02	114.45	29.10	79.73

Overall, the case study confirms that the trajectory-based FSM approach can robustly infer meaningful activity states from RTK-GNSS data under realistic site conditions. The resulting activity timelines provide a representation of earthmoving operations, forming a reliable foundation for future applications such as performance analysis and bottleneck identification.

5 Discussion

This study demonstrates that the FSM-based approach provides an effective and interpretable approach for recognizing equipment activity states from RTK-GNSS trajectory data in earthmoving operations. As evidenced by our quantitative validation against a second-by-second manual ground truth, the algorithm reliably captures critical operational cycles, achieving perfect recall for unloading events and high precision (0.90) for hauling. Compared with simple rule-based methods that rely solely on instantaneous thresholds, the FSM-based approach introduces temporal memory through predecessor constraints. This logic significantly reduces unrealistic state oscillations caused by GPS signal noise, short interruptions, or low-speed maneuvering. Furthermore, unlike data-driven classification models (e.g., neural networks), the FSM logic does not require substantial labeled training datasets and remains transparent to domain experts, allowing for easier debugging and adaptation to different site layouts.

Beyond activity recognition, the generated timelines provide critical insights into fleet performance. The case study revealed that operational inefficiencies can be quantified. As observed in the results, the overlap of truck waiting times highlights a temporary resource imbalance between the excavator's loading capacity and the truck fleet's arrival rate. By visualizing these hidden non-productive times, the proposed framework allows site managers to move from reactive observations to proactive resource leveling, for instance, by adjusting the number of trucks to reduce idling.

Several limitations should be acknowledged. First, the method relies on predefined spatial zones and threshold parameters, which currently require site-specific tuning based on the site layout. Second, equipment performing side tasks or operating outside predefined workflows may lead to temporary misclassification. These limitations could be addressed by incorporating additional sensing modalities to provide complementary information [7, 18]. Nevertheless, the current approach achieves a practical balance between robustness, interpretability, and deployment effort for outdoor earthmoving environments.

6 Conclusions

This paper presented a robust trajectory-based activity recognition method for earthmoving operations using finite-state machines. By integrating RTK-GNSS trajectory data with kinematic features, spatial context, and predecessor constraints, the proposed method infers temporally consistent activity states of excavators and dump trucks. Validation in a real-world tunnel construction project demonstrated that the method effectively captures cyclic operational patterns and produces interpretable activity timelines under realistic site conditions.

The contributions of this study extend beyond algorithm development. By transforming raw position data into semantic operational metrics, the framework provides a reliable foundation for Digital Twin applications. It enables project managers to pinpoint bottlenecks in near real-time and make informed decisions regarding process optimization.

Future research will focus on automating the

generation of site maps and geofences to reduce manual setup time. Additionally, we aim to integrate the output of this productivity analysis into real-time simulation models to predict future project performance and enable proactive schedule adjustments.

Acknowledgment

The research presented in this paper was funded by the European Union Horizon 2020 research and innovation program under grant agreement no. 101162318 (AM2PM) and Erhvervsfyrtårn Femern (FF6S). The Business Lighthouse Femern is co-financed by the European Union and the Danish Business Development Agency.

References

[1] L. Song and N.N. Eldin. Adaptive real-time tracking and simulation of heavy construction operations for look-ahead scheduling. *Automation in Construction*, 27:32–39, 2012. https://doi.org/10.1016/j.autcon.2012.05.007

[2] A. Heydarian and M. Golparvar-Fard. A visual monitoring framework for integrated productivity and carbon footprint control of construction operations. In *Proceedings of the Computing in Civil Engineering Conference*, pages 504–511, Miami, FL, USA, 2011. https://doi.org/10.1061/41182(416)62

[3] K. Hong and J. Teizer. Digital construction site layout planning and real-time trajectory analysis for proactive safety monitoring and control of struck-by hazards. *Automation in Construction*, 177:106353, 2025. https://doi.org/10.1016/j.autcon.2025.106353

[4] N. Wielgocka, T. Hadas, A. Kaczmarek, and G. Marut. Feasibility of using low-cost dual-frequency GNSS receivers for land surveying. *Sensors*, 21(6):1956, 2021. https://doi.org/10.3390/s21061956

[5] J. Louis and P.S. Dunston. Methodology for real-time monitoring of construction operations using finite state machines and discrete-event operation models. *Journal of Construction Engineering and Management*, 143:04016106, 2017. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001243

[6] F. Shahnavaz and R. Akhavian. Automated estimation of construction equipment emission using inertial sensors and machine learning models. *Sustainability*, 14:2750, 2022. https://doi.org/10.3390/su14052750

[7] V. Bharathi, K. Hong, and J. Teizer. Time-lapse image processing for estimating construction equipment emissions. In *Proceedings of the 33rd Annual Conference of the International Group for Lean Construction*, pages 1102–1113, Osaka, Japan, 2025. https://doi.org/10.24928/2025/0168

[8] Y. Zheng. Trajectory data mining: An overview. *ACM Trans. on Intelligent Systems and Technology*, 6:29, 2015. https://doi.org/10.1145/2743025

[9] N. Pradhananga and J. Teizer. Cell-based construction site simulation model for earthmoving operations using real-time equipment location data. *Visualization in Engineering*, 3:12, 2015. https://doi.org/10.1186/s40327-015-0025-3

[10] R. Akhavian and A.H. Behzadan. Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. *Advanced Engineering Informatics*, 29:867–877, 2015. https://doi.org/10.1016/j.aei.2015.03.001

[11] A. Langroodi, F. Vahdatikhaki, and A. Doree. Activity recognition of construction equipment using fractional random forest. *Automation in Construction*, 122:103465, 2021. https://doi.org/10.1016/j.autcon.2020.103465

[12] F. Vahdatikhaki and A. Hammad. Framework for near real-time simulation of earthmoving projects using location tracking technologies. *Automation in Construction*, 42, pp. 50-67, 2014. https://doi.org/10.1016/j.autcon.2014.02.018

[13] M. Golparvar-Fard, A. Heydarian, J.C. Niebles. Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers. *Advanced Engineering Informatics*, 27, pp. 652-663, 2013. https://doi.org/10.1016/j.aei.2013.09.001

[14] B. Sherafat, et al. A hybrid kinematic-acoustic system for automated activity detection of construction equipment, Sensors, 19, 2019. https://doi.org/10.3390/s19194286

[15] D. Wass, H. Axelsson, Machine Learning for Activity Recognition of Dumpers, 2019. http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1354995&dswid=-2037

[16] K.M. Rashid, J. Louis. Times-series data augmentation and deep learning for construction equipment activity recognition, *Advanced Engineering Informatics*, 42, 2019. https://doi.org/10.1016/j.aei.2019.100944

[17] T. Slaton, C. Hernandez, R. Akhavian, Construction activity recognition with convolutional recurrent networks, *Automation in Construction*, 113, 2020. https://doi.org/10.1016/j.autcon.2020.103138

[18] V. Bharathi and J. Teizer. Construction site layout planning and progress monitoring using computer vision on drone images. In *Proceedings of the 2025 European Conference on Computing in Construction*, Porto, Portugal, 2025. https://doi.org/10.35490/EC3.2025.313

A Care-Centered Vision–Language Model Framework for Evaluating Urban Under-Bridge Spaces for Older Adults

Zian Huang[1], Siwei Zhang[2], and Xingyu Tao[3]*

[1] The Hong Kong Polytechnic University, Hong Kong, China
[2] The University of Hong Kong, Hong Kong, China
[3] The Hong Kong Polytechnic University, Hong Kong, China
zi-an.huang@connect.polyu.hk, siweiz@connect.hku.hk, xingyu.tao@polyu.edu.hk

Abstract

In many urban areas, under-bridge spaces have become common gathering spots for older adults, offering venues for leisure, social interaction, and physical activity. Various policies and community initiatives have been introduced to improve the quality and usability of these spaces. However, efficiently and automatically assessing the elder-friendliness of such environments remains an open challenge, particularly for supporting data-informed spatial design and policy decisions. To this end, this study proposes a Care-Centered Vision–Language (CC-VLM) framework with three contributions: (i) Develop a care-theory-centered evaluation metric model, using this as the basis for subsequent labeling and assessment; (ii) Design a multi-elder-agent workflow for automatically generating labels to solve the problem of data scarcity from older adults; and (iii) Build a XGboost-based classification model for automated under-bridge spaces scene assessment. The proposed framework is applied to 5 under-bridge spaces in Hong Kong. Experimental results demonstrate an accuracy of 65.5% and a recall rate of 71.4%. This framework offers a scalable and interpretable approach to support the inclusive and age-friendly transformation of overlooked urban spaces.

Keywords –

Care Theory; Older Adults; VLM; Under-Bridge;

1 Introduction

Globally, under-bridge spaces in high-density cities have emerged as vital gathering spots for older adults, providing venues for leisure, social interaction, and physical activity. Many regions face a pressing contradiction: a growing and aging population drives sustained demand for public recreational spaces, while existing provision remains insufficient and is projected to fall further behind. For instance, Hong Kong—representative of such high-density urban contexts—has seen its elderly population proportion projected to rise from 18.4% in 2029 to 36.3% in 2064, alongside a long-standing shortfall in per capita open space (failing to meet its own 2m ²standard since 2008) [1-3]. The current use of under-bridge spaces is subject to government regulations [4] that limit permissible uses. However, official policy exhibits a duality of control and activation: while protecting safety, traffic, environmental, and structural concerns, it also actively encourages greening, beautification, and adaptive reuse. Notable examples include the "Fly the Flyover Operation," organised by the Development Bureau's Energising Kowloon East Office [5] and statements in the Official Record of Proceedings from the Legislative Council [6]. This situation indicates the value of utilizing remaining urban spaces, such as under-bridge areas. But because they are so numerous, automated assessment tools become particularly important.

Existing research on evaluating urban spaces has followed two main paradigms: qualitative approaches and quantitative data-driven methods. Qualitative studies provide rich contextual insights into user experiences, urban space design theory, and elderly-specific needs, emphasising Inclusive, Pleasurable, Safe, Comfortable, and Meaningful Activities [7]. There are numerous qualitative assessment tools for public space quality, among which the UN-Habitat and the Good Public Space Index (GPSI) are considered the most comprehensive; most of these tools rely on on-site observation, questionnaire surveys, and site visits [8]. Those methods are subjective, labour-intensive, and limited in scale, and these drawbacks are amplified when studying older adults due to their physical, cognitive, and

communication challenges. Quantitative methods, largely enabled by street-view imagery and computer vision, can partially compensate for the shortcomings of qualitative research. Common quantitative methods employ semantic segmentation (e.g., PSPNet, SegNet, Swin Transformer) to extract features, such as using a transformer-based framework for urban physical disorder detection [9]. Quantitative methods also include using statistical analysis, machine learning, or deep learning to predict some socioeconomic profiles, like using the spatial distribution of notable urban features identified through the street view images to predict poverty status, crime, and health behaviour [10]. However, these quantitative studies typically require extremely large sample sizes—often involving datasets with over a million street-view images sourced from platforms like Google Street View or Baidu Maps [9, 10]. Such data demands pose significant barriers for niche, overlooked spaces like under-bridges, where imagery is sparse, site-specific, and difficult to collect at scale due to accessibility constraints, variable lighting, and limited coverage in mapping services. Furthermore, the qualitative method cannot elicit people's subjective perceptions of space. Although qualitative analysis can achieve this, its limitations have been discussed previously.

Vision-Language Models (VLMs) represent a paradigm shift in multimodal AI. By jointly encoding images and text through aligned embeddings (typically involving an image encoder, text encoder, fusion mechanism, and generative language component), VLMs enable advanced tasks such as image captioning, visual question answering (VQA), and zero/few-shot reasoning on novel concepts [11, 12]. VLMs can provide language-guided interpretability. VLMs can also serve as simulated "elderly agents" through targeted prompting, automating subjective evaluations from an older adult's perspective and mitigating the inherent drawbacks of qualitative methods noted earlier. Recent applications show the effectiveness of VLMs in urban scene recognition and street-view analytics. Systematic reviews emphasize VLMs' role in assessing street-level perceptions, including safety, walkability, greenery, and socioeconomic indicators from imagery sources like Google Street View [13]. The spatial reasoning of VLM can be further enhanced by tuning particularly for challenging tasks involving negation or counterfactuals [14]. In addition, VLM can be integrated with broader frameworks for urban renewal diagnostics, combining perceptual scoring with spatial autocorrelation to inform planning decisions [15]. However, off-the-shelf VLMs are not directly suited to evaluating elderly suitability in under-bridge spaces. Pre-training datasets under-represent neglected residual areas, leading to inconsistent performance on atypical layouts and fine details [13, 14]. General-purpose VLMs also lack grounding in elderly-specific care concepts and remain largely vision-centric, with limited handling of specific factors critical for older adults.

This paper, therefore, proposes a Care-Centered Vision–Language Model (CC-VLM) for comprehensive, interpretable, scalable, and elderly-specific assessment of under-bridge spaces. Three sub-objectives are:

1. Develop a care-theory-centered evaluation metric model through a literature review on care and urban design theory, combined with on-site observation, to provide theoretical grounding and a novel perspective on under-bridge spaces assessment.

2. Design a multi-elder-agent workflow to automatically generate diverse visual labels from an elderly perspective, aiming to reflect older adults' perception of the environment. This can overcome subjectivity and labour intensity, particularly acute when elderly participants face physical, cognitive, or communication barriers.

3. Build an XGboost-based classification model for automated scene assessment, integrating care-theory metrics and multi-elder-agent labels to deliver quantitative elderly-friendliness scores and explanations.

2 Methodology

CC-VLM consists of three stages:

(1) Care-Theory-Centered Evaluation Metric Development and Feature Extraction. A hierarchical indicator system is built from care ethics literature and field observation, yielding 27 measurable spatial indicators grouped into three layers (Physical Shelter, Psychological Belonging, Social Activity). Then, a Detectron2 model with a ResNet-101-FPN backbone pre-trained on COCO will be established to perform Panoptic segmentation. According to the Evaluation Metric, Custom class merging and post-processing yield pixel-level ratios, instance densities, and morphological metrics for the 27 indicators.

(2) Multi-Elder-Agent Scoring and Demographic Weighting. Based on the metric developed in step 1, four distinct elderly personas (differing in age, mobility, and sensory profile) are simulated via structured prompts to MiniCPM. Each agent scores the three care layers on a 1–100 scale. Final image scores are computed as a weighted average based on Hong Kong's official elderly age distribution, yielding a continuous elderly-friendliness score.

(3) Assessment Model Building and Interpretation. The extracted features and weighted scores are used to train an XGBoost binary classifier. The CC-VLM framework is shown in Figure 1.

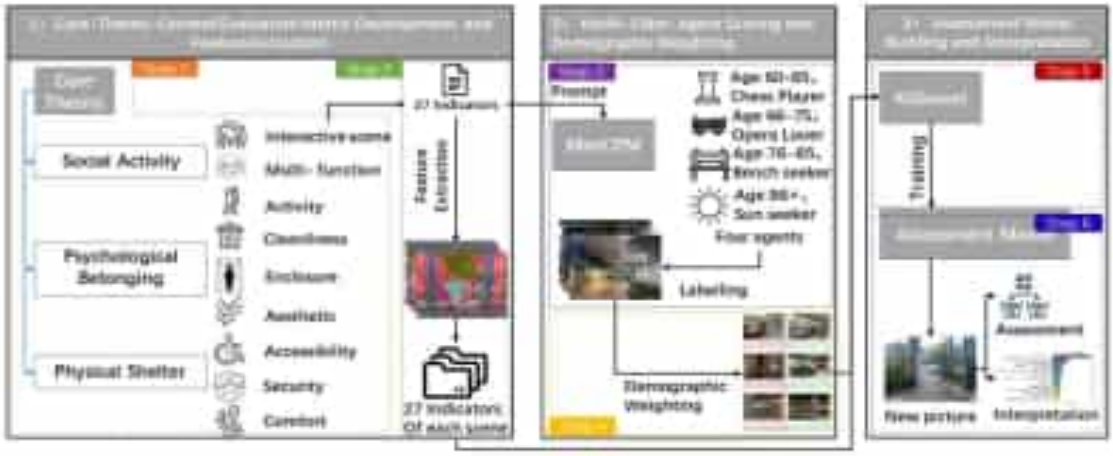


Figure 1. CC-VLM framework.

2.1 Care-Theory-Centered Evaluation Metric Development and Feature Extraction

2.1.1 Care-Theory-Centered Evaluation Metric Development

To address the limitations of existing public space evaluation tools, this study constructs a dedicated Care-Theory-Centered Evaluation Metric through theoretical and methodological synthesis. Drawing on the care ethics in urban contexts, the framework integrates three progressive phases: (1) political ethics, rooted in foundational manifestos emphasizing interdependence and universal care as a societal organizing principle [16]. (2) empirical studies, informed by care infrastructures in high-density cities, such as Boston's labor-focused activations and relational interventions [17], framework for a revised approach to intergenerational public space design, pocket park in Los Angeles [18]. and (3) design practice, exemplified by relational and caring city models in London that translate ethics into actionable placemaking [19]. This evolutionary trajectory is complemented by established urban design standards, including environmental psychology principles, WHO Age-Friendly Cities guidelines, and Hong Kong-specific urban design directives, which collectively emphasize multidimensional human needs in public realms [7, 8]. The structure of the metric is shown in Table 1.

2.1.2 Feature Extraction

This step contains two parts that transform raw street-view images into quantifiable indicators.

Part 1: Visual Analysis Leverages Detectron2's panoptic segmentation model with a ResNet-101 Feature Pyramid Network (FPN) backbone ShuffleNet), ResNet demonstrated superior overall performance [20]. Besides, it can also cover common item categories.

Table 1. Care-Theory-Centered Evaluation Metric

Primary	Secondary	Tertiary
Physical Shelter	Comfort	Seat, Green rate, Microclimate.
	Security	Physical disorder, Visual transparency, Safety protection facilities.
	Accessibility	Accessible space, Accessibility facilities, pedestrian paths.
Psychological Belonging	Aesthetic experience	Disorderliness, Color richness, Green rate,
	Enclosure sense	semi-enclosed spaces, Visibility of sky, Boundary closure.
	Cleanliness	Garbage coverage rate, Cleanliness, Ground pavement.
Social Activity	Activity Support	Activity space area, Social furniture, Pedestrian activity.
	Multiple function	Convenience facilities, Non-motorized Transport, Convenience of Access.
	Interactive scene	Recreational facilities, Stay friendliness, Openness of Public Spaces.

Part 2: Logical Extension Transform visual outputs into the indicators through systematic post-processing and mapping. Three kinds of representative indicators across distinct calculation categories are as follows:

Greenery Coverage Rate (turn pixel data into ratio indicators). This indicator reflects the proportion of vegetation in spaces and serves as a core metric for evaluating ecological comfort for older adults. Its calculation formula is shown as equation 1:

$$G = 100 \times \frac{M_{tree} \cup M_{grass}}{N} \quad (1)$$

Where $M_{tree} \cup M_{grass}$ represents the total pixel count of the merged "tree " and "grass " categories in the panoptic segmentation mask. N represents the total pixel count of the image.

Bench Density (turn count into density indicators). This indicator measures the supply of elderly-friendly resting facilities. It can be shown as equation 2:

$$P = 100 \times \frac{count(bench)}{A}, A = N \times 0.01^2 \quad (2)$$

Where P represents the bench density in the space per 100 m^2; $count(bench)$ represents the total instance count of the "bench" category in the panoptic segmentation result. A is physical area of space. 0.01 is the conversion coefficient from pixel to meter.

Semi-Enclosed Space Quantity (turn spatial features into morphological indicators). This indicator characterizes the scale of enclosed local areas in under-bridge spaces. It can be shown as equation 3:

$$E = num\{S | S \geq ENCLOSED_AREA_THRESH\} \quad (3)$$

Where E is the number of semi-enclosed spaces. S is the area of a single enclosed region in the under-bridge space. $ENCLOSED_AREA_THRESH$ is the area threshold for defining a semi-enclosed space and was set to 1.5 m^2. According to field surveys, enclosed areas smaller than 1.5 m^2 are mostly structural gaps with no actual rest value, so this threshold filters invalid regions. OpenCV is used as the core computer vision tool, it extracts the outer contours of discrete spatial regions and computes the pixel area of each region.

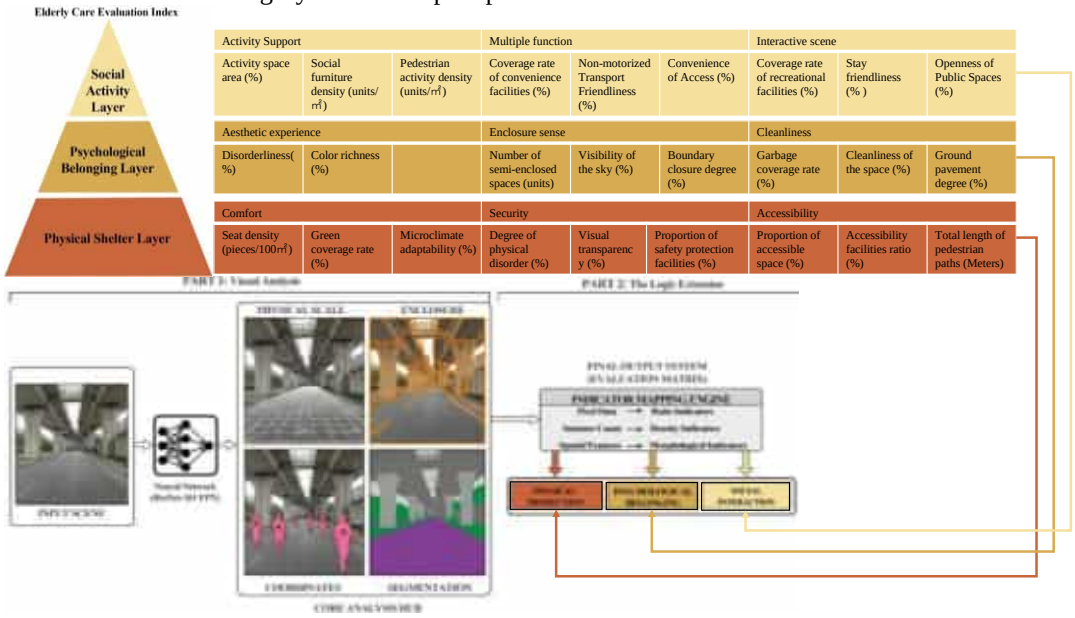


Figure 2. Feature extraction based on care-theory-centered evaluation metric

2.2 Multi-Elderly-Agent Scoring and Demographic Weighting

Existing research has realized task splitting and collaborative execution of BIM spatial queries through an LLM-based agent framework [21]. Drawing on this idea, this study designs multi-elderly agents to simulate the subjective evaluation perspectives of elderly groups.

This step addresses the subjectivity and small-scale limitations of traditional qualitative evaluations by simulating diverse elderly perspectives through AI agents and ensuring that scores reflect the needs of the target population via demographic weighting.

2.2.1 Design of Multi-Elderly Agents

To comprehensively cover the heterogeneous needs of older adults across different age groups and physical conditions, 4 elderly agents were designed through a prompt engineering approach.

Prompt Engineering: Each persona receives a structured prompt template, assigned a realistic background and key concerns derived from care theory, to ensure the agents' evaluations mirrored real-world elderly needs:

'You are a [age]-year-old elderly person with [mobility/sensory profile]. Evaluate the attached under-bridge space image from your perspective according to

three dimensions: (1) Physical Shelter (accessibility, safety, comfort), (2) Psychological Belonging (aesthetics, enclosure, cleanliness), (3) Social Activity (activity support, functional diversity, interaction affordance). Provide three scores from 1–100.'

Model Deployment and Optimization: The MiniCPM model was adopted for agent reasoning, optimized for batch processing with 8-bit NF4 quantization and image resizing to balance efficiency and accuracy. Each agent processed 302 images, generating three-dimensional scores and an average score per image.

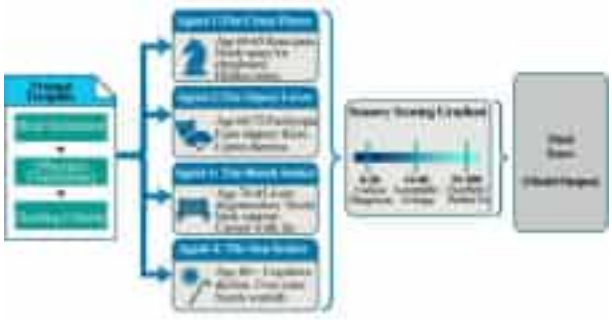


Figure 3. Multi-Elderly-Agent Scoring Framework.

2.2.2 Demographic Weighting

To reflect the actual age structure of Hong Kong's elderly population, weighted aggregation was performed on the four agents' total scores. The weighted total score for each under-bridge space was calculated as equation 4:

$$Weighted\ Score = \sum_{i=1}^{4} (S_i \times W_i) \quad (4)$$

In which S_i is the average total score of the i-th elderly agent, and W_i is the corresponding demographic weight. According to the statistics from the Census and Statistics Department of Hong Kong, the age structure of the elderly population in Hong Kong is shown in Table 2.

Table 2. The age structure of the elderly population in Hong Kong

Ages	Ratio (%)
60-65	26
66-75	43.7
76-85	19.8
Above 85	10.5

2.3 XGboost-Based Classification Model And Model Performance Analysis

This section integrates objective spatial features and demographic-weighted elderly satisfaction scores to construct a robust, interpretable, and scalable automated assessment model. The core goal is to learn the mapping relationship between under-bridge space attributes and elderly-friendly degrees, while providing data-driven decision support for age-friendly urban design or renewal.

As shown in equation 5, given the input feature matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ (where $n = 302$, the number of under-bridge image samples, $m = 27$, the quantifiable spatial indicators) and the demographic-weighted satisfaction score $S_\omega \in [1, 100]$ for each sample, the task is formulated as a binary classification problem to predict elderly satisfaction levels. The binary label y_i is defined using a calibrated threshold $\tau = 68$ (the average score determined by elderly agent). Similar urban application employs XGBoost on street-view features to predict public-space vitality scores and then reference the dataset mean to derive high or low classifications for age-inclusive urban design interventions [23].

$$y_i = \begin{cases} 0, S_{\omega,i} < \tau (Low\ Elderly - Care) \\ 1, S_{\omega,i} \geq \tau\ (Low\ Elderly - Care) \end{cases} \quad (5)$$

2.3.1 Model Selection and Building

Extreme Gradient Boosting (XGBoost) is selected as the base model due to its superior performance in handling high-dimensional tabular data, nonlinear feature interactions, and class imbalance. Its regularized objective function balances fitting capability and generalization. It can be expressed as equation 6.

$$L(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (6)$$

Where $\sum_{i=1}^{n} \boldsymbol{l(y_i, \hat{y}_i)}$ stands for the empirical loss of the predicting model. It measures the prediction error between the model's output ($\hat{y}_i$, for under-bridge space samples) and the true binary label (y_i, 0=low elderly-care, 1=high elderly-care), using logistic loss to fit the binary classification task. $\sum_{k=1}^{K} \boldsymbol{\Omega(f_k)}$ stands for regularization loss.

When constructing the model, we encountered three other issues: (1) Class imbalance in labeled samples, (2) Noise sensitivity of spatial features due to diverse under-bridge layouts, and (3) Need for interpretable predictions to guide urban planning decisions. So we added four interconnected modules: (1) Feature preprocessing with outlier-resilient normalization; (2) Dual-stage data augmentation for class imbalance mitigation; (3) Bayesian optimized hyperparameter tuning; (4) Interpretability enhancement.

1. Feature preprocessing with outlier-resilient normalization. To mitigate the impact of outliers in spatial feature measurements, we employ RobustScaler for normalization to ensure stability. It is expressed as equation 7.

$$X_{ij}^{scaled} = \frac{X_{ij} - median(X_j)}{IQR(X_j)} \quad (7)$$

Where $\mathrm{IQR}(X_j) = Q_3(X_j) - Q1(X_j)$ (25th and 75th percentiles of feature j).

2. Dual-stage data augmentation for class imbalance mitigation. To address class imbalance without distorting feature semantics, we propose a dual-stage augmentation strategy, including SMOTE Oversampling and Gaussian Noise Augmentation.

3. Given the dataset imbalance and the high-stakes urban renewal context in Hong Kong, the model is optimized using a weighted multi-metric score via Optuna. This process can be expressed as equation 8.

$$\text{Score} = 0.45\text{Acc} + 0.30\text{F1} + 0.125\text{Prec} + 0.125\text{Rec} \quad (8)$$

Where equation 9

$$\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (9)$$

is the correctness of the model. It assesses overall classification correctness, weighted most heavily. Because bridges, as basic city infrastructure, are very difficult to change or renew. A high accuracy ensures the model's baseline credibility and avoids resource misallocation. Equation 10

$$\text{F1} = 2 \times \frac{Prec \times Rec}{Prec + Rec} \quad (10)$$

Serves as the primary balance metric for our imbalanced dataset. F1 mitigates the accuracy bias toward the majority class, ensuring that the model neither over-misses high-potential regions nor over-labels low-quality regions. Equation 11

$$\text{Prec} = \frac{TP}{TP + FP} \quad (11)$$

Measures the reliability of "high elderly-friendliness" predictions. A high precision avoids unnecessary renovation costs, for misclassifying a low-friendly space would waste urban renewal budgets. Equation 12

$$\text{Rec} = \frac{TP}{TP + FN} \quad (12)$$

Evaluate the ability to identify all high-potential spaces. A high recall rate can prevent missed opportunities for optimization; failing to detect spaces with high potential will result in neglecting available space for the elderly in underutilized urban areas.

To further enhance model robustness and reduce generalization error, we construct a 9-model soft-voting ensemble based on the Bayesian-optimized parameter set.

4. To bridge the gap between model predictions and urban planning practice, we integrate complementary interpretability tool, Gain-Based Feature Importance, to quantify the contribution of each spatial indicator by cumulative loss reduction during tree splitting. This process can be expressed as equation 13.

$$Gain_j = \frac{\sum_{k=1}^{K} Gain_{jk}}{\sum_{l=1}^{m}\sum_{k=1}^{K} Gain_{lk}} \quad (13)$$

Where $Gain_{jk}$ represents the loss reduction from splitting feature j in tree k. The numerator $\sum_{k=1}^{K} Gain_{jk}$ accumulates the total loss reduction contributed by indicator j across all K decision trees in the XGBoost ensemble; the denominator $\sum_{l=1}^{m}\sum_{k=1}^{K} Gain_{lk}$ sums the total loss reductions of all m spatial indicators, ensuring $0 \ll Gain_j \ll 1$ and $\sum_{j=1}^{m} Gain_{jk} = 1$. Then we select the top 15 indicators with the largest normalized importance scores.

3 Proof of Concept

3.1 Experiment Environment

For the selection of the survey sites for the underpass space in Hong Kong, five research sites were selected, namely: Tsuen Wan-MTR Station Flyover, Sham Shui Po-Tung Chau Street Flyover, Tai Po-Tung Lo Wan Road Flyover, Causeway Bay-Canal Road East, To Kwa Wan- Flyover. Their locations are shown in Figure 4.

The Experiments were conducted on a high-performance computing server equipped with dual NVIDIA GeForce RTX 3090 GPUs (each with 24 GB GDDR6X memory, totaling 48 GB VRAM), an Intel Xeon Gold 6248R CPU @ 3.00 GHz (24 cores/48 threads), 256 GB DDR4 RAM, and 2 TB NVMe SSD storage. The operating model Ubuntu 22.04 LTS, with CUDA 12.1 and cuDNN 8.9 installed for GPU acceleration. All codes were implemented in Python 3.10.



Figure 4. Map of the locations for on-site research.

3.2 Results of CC-VLM

As shown in Figure 5, high scoring samples are more likely from spaces with high green coverage rates, complete resting facilities, and moderate social vitality. These attributes align with the core needs of older adults emphasized in care theory and in on-site observations. In contrast, low-scoring samples are mostly concentrated in spaces characterized by inadequate safety facilities, high garbage coverage rates, and limited semi-enclosed spaces.



Figure 5. Multi-Elderly-Agent Scoring example.

3.3 Results Evaluation

To assess the performance of the classification model, we employ a set of metrics, including accuracy, recall, F1 Score, area under the receiver operating characteristic curve. These indicators are selected for their ability to provide a balanced evaluation. Besides, these metrics are consistent with the research objective, as they not only quantify predictive efficacy but also emphasizing sensitivity to adverse conditions. For instance, high recall ensures that the model reliably flags spaces requiring intervention, supporting equitable transformations of overlooked urban areas for aging populations. Baselines for these indicators are established relative to random guessing and prior benchmarks in analogous street-view-based urban analytics. For binary classifier, random performance yields accuracy, F1 and AUC-ROC are all of 0.5. When comparing with similar study, Hu et al. reported 79.9% accuracy and 75.51% mean average precision in urban physical disorder detection [9]. In our experiments, recall and AUC-ROC values exceeding 0.7 and 0.6 respectively are considered adequate. As shown in Table 3, overall performance metrics confirm balanced yet modest results across standard measures.

Table 3. Binary Classification Model Performance Metrics

Indicator Name	Score
AUC	0.615
Recall	0.714
Precision	0.536
F1-Score	0.612
Accuracy	0.655

As shown in Figure 6, the model demonstrated moderate discriminative ability, with an Area Under the ROC Curve (AUC) of 0.615. This AUC value, while above random chance, reflects the limited variability in the 302-image dataset, indicating potential for improvement with expanded data or refined features.

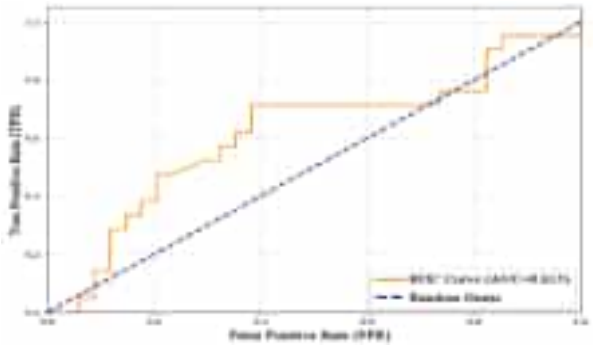


Figure 6. ROC Curve.

As shown in Figure 7, feature importance rankings offer interpretable insights directly tied to the care theory. Indicators from the Physical Shelter Layer predominated, led by green coverage rate and facility completeness, emphasizing greenery and basic infrastructure as primary drivers of elderly satisfaction. Psychological Belonging Layer features ranked prominently, supporting the influence of emotional comfort. Social Activity Layer metrics also contributed substantially.

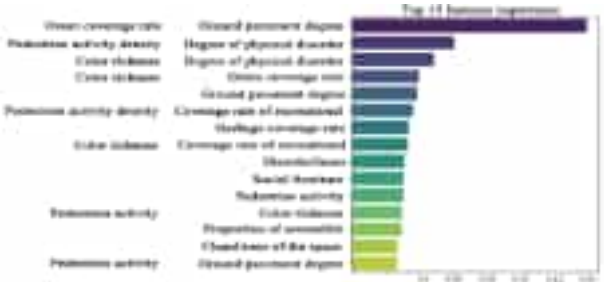


Figure 7. The importance of the top 15 features.

4 Conclusion

This study addresses the critical challenge of efficiently assessing urban under-bridge spaces. By proposing the CC-VLM framework, we achieve three core objectives: (1) a theory-grounded hierarchical metric model synthesizing care ethics into practice; (2) an automated workflow that simulates diverse elderly perspectives to generate subjective visual labels; and (3) a multi-model-enhanced VLM integrated with XGBoost for scene assessment and prediction. Experimental results validate the framework's effectiveness, confirming that greenery, facility completeness, and activity density are dominant drivers.

Limitations include modest predictive performance due to the small dataset size, overreliance on visual imagery alone, and reliance on simulated agents. In particular, the XGBoost classifier is trained using scores generated by the VLM agents as target labels, therefore it is not directly reflecting real-world elderly experiences.

Future work could expand to larger datasets, integrate more data from real older adults, and incorporate multi-modal environmental sensors for evaluations.

5 Acknowledgement

This research is supported by the Hong Kong PolyU (UGC) Start-up Fund for New Recruits (P0056204) and Zhejiang Provincial "Pioneer" and "Leading Goose" Science and Technology Plan Project (2026C04040)

Reference

[1] Census and Statistics Department The Government of the Hong Kong Special Administrative Region. "Table 110-01001 : Population by Sex and Age Group."https://www.censtatd.gov.hk/en/web_table.html?id=110-01001

[2] Development Bureau of Hong Kong. "Hong Kong 2030+."https://www.pland.gov.hk/file/planning_studies/comp_s/hk2030plus/TC/index.htm

[3] Planning Department of Hong Kong. "Hong Kong Planning Standards and Guidelines." https://www.pland.gov.hk/file/tech_doc/hkpsg/full/pdf/ch4_en.pdf

[4] R. ROHINIKUMAR, "Design Guidelines for urban Public Spaces for the Elderly," Tesis de maestría en Arquitectura. Uttarakhand: Indian Institute of …, 2017.

[5] Kowloon East Office. "Promoting the "Spirit of Creation" Fly the Flyover Operation." https://www.ekeo.gov.hk/en/the-spirit-of-creation/innovative-placemaking-urban-design/fly-the-flyover-operation/index.html

[6] LEGISLATIVE COUNCIL OF HONG KONG. "OFFICIAL RECORD OF PROCEEDINGS." https://www.legco.gov.hk/yr2025/chinese/counmtg/floor/cm20250521-confirm-ec.pdf#nameddest=wrq11

[7] V. Mehta, "Evaluating Public Space," *Journal of Urban Design*, vol. 19, no. 1, pp. 53–88, 2013, doi: 10.1080/13574809.2013.854698.

[8] S. Tawfik Mohamed, A. Mandour, and H. Baker, "A critical review of quality assessment tools for public spaces," *Engineering Research Journal*, vol. 177, pp. 255–274, 2023.

[9] C. Hu *et al.*, "UPDExplainer: An interpretable transformer-based framework for urban physical disorder detection using street view imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 204, pp. 209–222, 2023.

[10] Z. Fan, F. Zhang, B. P. Loo, and C. Ratti, "Urban visual intelligence: Uncovering hidden city profiles with street view images," *Proceedings of the National Academy of Sciences*, vol. 120, no. 27, p. e2220417120, 2023.

[11] A. Ghosh, A. Acharya, S. Saha, V. Jain, and A. Chadha, "Exploring the frontier of vision-language models: A survey of current methodologies and future directions," *arXiv preprint arXiv:2404.07214*, 2024.

[12] J. Zhang, J. Huang, S. Jin, and S. Lu, "Vision-language models for vision tasks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 46, no. 8, pp. 5625–5644, 2024.

[13] Z. Peng *et al.*, "Vision language model (VLM)-enabled street view analytics: a systematic literature review," *Engineering, Construction and Architectural Management*, pp. 1–19, 2025.

[14] J. Ro, N. Kim, and Y. Yoon, "How Well Do Vision-Language Models Understand Cities? A Comparative Study on Spatial Reasoning from Street-View Images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 6476–6485.

[15] Y. Yao, D. Ò. Giuliano, and F. Lu, "Urban Street-Scene Perception and Renewal Strategies Powered by Vision–Language Models," 2025.

[16] A. Chatzidakis, J. Hakim, J. Litter, C. Rottenberg, and C. Collective, *The care manifesto: The politics of interdependence*. Verso Books, 2020.

[17] A. Binet *et al.*, "The urban infrastructure of care: Planning for equitable social reproduction," *Journal of the American Planning Association*, vol. 89, no. 3, pp. 282–294, 2023.

[18] C. Nelischer, A. Loukaitou-Sideris, and G. Wendel, "Caring public space: Advancing justice through intergenerational public space design and planning," *Journal of Urban Affairs*, vol. 47, no. 8, pp. 2789–2808, 2025.

[19] A. Gabauer *et al.*, *Care and the city: encounters with urban studies*. Taylor & Francis, 2022.

[20] S. Zhang, X. Yu, and J. Ma, "Vision AI reveals waste-safety perception relationships in urban environments," *npj Urban Sustainability*, vol. 5, no. 1, p. 83, 2025.

[21] C.-F. Chan *et al.*, "Context-aware vision-language model agent enriched with domain-specific ontology for construction site safety monitoring," *Automation in Construction*, vol. 177, p. 106305, 2025.

[22] A. Li, P. K.-Y. Wong, X. Tao, J. Ma, and J. C. Cheng, "An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent," *Advanced Engineering Informatics*, vol. 66, p. 103375, 2025.

[23] Ruan Y, Zhang X, Wang S, Chen X, Chen Q. Street View-Enabled Explainable Machine Learning for Spatial Optimization of Non-Motorized Transportation-Oriented Urban Design. Land. 2025.

A Digital Twin Simulation Framework for Conveyor-Based Construction Waste Sorting Using RGB-Derived Mesh Assets

Shanuka Dodampegama[1], Avish Singh[1], Lei Hou[1], Ehsan Asadi[1], Kevin Zhang[1], and Sujeeva Setunge[1]

[1]Centre for Future Construction, School of Engineering, RMIT University, Melbourne, VIC 3001, Australia
s3975925@student.rmit.edu.au, s3898103@student.rmit.edu.au, lei.hou@rmit.edu.au, ehsan.asadi@rmit.edu.au, kevin.zhang@rmit.edu.au, sujeeva.setunge@rmit.edu.au

Abstract

Construction and demolition waste (CDW) sorting is increasingly being automated using robotics and AI to improve safety, efficiency, and material recovery. However, developing and validating perception and control algorithms directly on physical conveyor systems is costly, time consuming, and difficult to reproduce due to variability in object layouts and operating conditions. The lack of scalable and repeatable benchmarking environments limits rapid experimentation and fair comparison of alternative approaches. This paper presents a digital twin (DT) simulation framework for conveyor-based CDW sorting that enables systematic and data-driven evaluation of robotic strategies. The proposed system integrates three main components: an RGB annotation-driven mesh generation pipeline that converts real waste images into 3D object assets, a Unity based physics simulation that models the conveyor and robotic workspace, and a ROS 2 communication layer that enables real time interaction with external perception and control modules. The generated meshes preserve object footprint and spatial relationships while remaining computationally efficient, allowing realistic cluttered scenes to be reproduced automatically. The framework also provides structured ground truth and event-based signals to support quantitative evaluation. Experimental results show that mesh generation requires approximately 24 ms per image and 5.3 ms per object, while the simulation maintains stable real time performance at around 46–47 frames per second across varying conveyor speeds. These results indicate that the platform is suitable for large scale experimentation. The proposed DT offers a practical and extensible testbed for developing, benchmarking, and transferring robotic waste sorting algorithms from simulation to real world deployment.

Keywords –

Digital Twin; Robotics; Construction and Demolition Wastes; ROS 2

1 Introduction

Construction and demolition waste (CDW) represents a large proportion of global solid waste and poses significant environmental and economic challenges due to its volume and heterogeneity. Automated robotic systems have the potential to improve CDW sorting efficiency by reducing manual labour, increasing material recovery, and supporting circular economy goals. However, the irregular shapes of waste objects, high inter-class variability, and dynamic conveyor conditions make perception, manipulation, and control particularly challenging for robotic systems [1].

Digital twin (DT) technologies have recently emerged as a promising approach for accelerating development and testing of robotic systems in simulation before real world deployment. A DT refers to a virtual representation of a physical system that enables real time communication and interaction between the digital model and its physical counterpart, supporting design, optimisation, and testing tasks without the cost and risk associated with physical experimentation [2]. In robotics, DT platforms integrated with robotic middleware such as the Robot Operating System (ROS) enable evaluation of perception, planning, and control algorithms in controlled environments.

Despite these advances, most existing DT frameworks target structured industrial tasks or generic manipulation scenarios. In contrast, CDW sorting requires handling irregular objects, cluttered conveyor streams, and variable spatial layouts derived from real waste scenes. Existing research on robotic waste sorting largely focuses on object detection and classification,

with limited infrastructure for reproducible evaluation of complete perception-to-action pipelines.

This paper presents a data-driven DT framework for conveyor-based CDW sorting that integrates ROS 2 with Unity and generates simulation assets from annotated RGB waste images. The framework enables realistic reproducible scene generation, object state publishing, and event-based evaluation within a controllable environment. The contributions of this work are threefold: (i) an automated pipeline for generating 3D mesh assets from annotated RGB images of CDW while preserving object footprint, spatial layout, and key geometric attributes relevant for simulation, (ii) a synchronised Unity–ROS 2 DT environment capable of representing conveyor-based CDW sorting scenarios with dynamically generated waste streams derived from real images, and (iii) a configurable benchmarking framework including scenario definitions and performance metrics that enable systematic evaluation of sorting systems in simulation under varying object layouts and conveyor operating conditions.

2 Related Work

2.1 DTs and Simulation Environments for Robotics

DTs are increasingly used for developing and validating robotic systems, enabling safe, repeatable experimentation and rapid iteration of perception, planning, and control algorithms. In construction and built environment applications, recent reviews highlight rapid growth in DT research while identifying challenges in interoperability, real time data integration, and scalable cyber physical architectures that combine sensing, simulation, and decision making [3].

Within robotics, Unity is increasingly adopted as a practical simulation platform due to its high-quality visualisation, flexible scene design, and interactive modelling capabilities. It is commonly integrated with ROS or ROS 2 to connect virtual scenes with external robotic software. Prior work has demonstrated Unity-based DT pipelines and evaluated aspects such as communication latency, synchronisation, and real time interaction [2]. Comparative studies with robotics focused simulators such as Gazebo indicate that Unity supports rapid scene prototyping and visual realism, while Gazebo provides tighter native ROS integration and physics modelling depending on the application [4].

Researchers have also explored more extensible simulation architectures that combine Unity with external computation or messaging layers to support collaborative experiments and modular system design. For example, multi-user simulation environments with Unity-MATLAB bridge have been proposed to enable colloborative experiments and evaluate alternative communication mechanisms [5]. Similarly, DT-driven control platforms have been developed in which Unity-based visual environments interact with industrial robot controllers through bidirectional state and command synchronisation [6]. These studies demonstrate that the DT frameworks can significantly reduce system integration effort and accelerate the development of robotic applications.

However, most existing DT platforms are designed for structured industrial tasks or general manipulation scenarios. They rarely consider domain specific challenges such as conveyor-based handling of irregular waste objects, cluttered object streams, and dynamically generated scenes derived from real observations. As a result, current DT environments provide limited support for benchmarking robotic sorting algorithms under realistic CDW conditions.

2.2 Robotic and AI for CDW Sorting

Robotic sorting has been widely investigated as a means of improving recycling efficiency and reducing manual labour in hazardous and labour-intensive environments. Despite these advances, automated waste sorting remains challenging because CDW streams typically contain cluttered scenes, irregular object shapes, and significant material variability, requiring robust perception and decision making capabilities [1]. Recent research in CDW automation has primarily focused on vision-based detection and classification methods. Advances in deep learning have enabled improved recognition and segmentation of waste categories, supporting automated material recovery systems. They also highlight important practical considerations for real world deployment, including distributed processing between edge and cloud systems and coordination between multiple robotic units in large facilities [7].

Several works have extended these perception capabilities by integrating object detection with robotic manipulation. In such systems, detected objects are localised and then handled using robot kinematic control to perform pick and place or transfer operations [8]. At the same time, sensitivity analyses and field studies emphasise that real world viability depends not only on recognition accuracy, but also on consistent operation under varying throughput, conveyor speeds, and labour or operational constraints [9].

These observations indicate that evaluation of robotic sorting system must consider the full perception-to-action workflow rather than isolated algorithmic components. However, most existing studies validate their approaches using limited experimental setups or fixed test scenarios. Infrastructure that enables controlled, repeatable, and configurable benchmarking of sorting strategies remains relatively limited.

2.3 Data-driven Geometry Generation for Simulation Assets

Realistic simulation environments require representative object geometry to accurately reflect real world operating conditions. In many robotic applications, the quality of object models strongly influences the validity of experimental results, particularly when evaluating perception and manipulation strategies. Manually modelling large numbers of irregular objects is time consuming and often impractical, motivating the development of automated asset generation methods.

In the computer vision community, significant research has focused on reconstructing 3D structure from a single RGB image. These approaches include methods that predict depth maps, point clouds, or mesh representations directly from image features. Survey studies highlight the diversity of reconstruction techniques as well as the inherent difficulties caused by ambiguity, occlusion, and incomplete observations [10]. Although such learning-based methods can produce detailed geometry, they are computationally demanding and may not generalise well to the diverse and irregular shapes commonly found in CDW streams.

Consequently, many robotics and DT simulations instead rely on predefined CAD models or manually curated asset libraries. Although these assets are suitable for structured industrial components, they do not capture the variability and randomness observed in real waste streams. Existing Unity-ROS DT setups have been used to simulate manipulation tasks, but they typically assume that object models are known in advance rather than derived directly from real images [11].

As a result, current approaches either focus on high fidelity reconstruction for visual realism or depend on handcrafted object libraries. Neither direction adequately supports lightweight and scalable geometry generation that can rapidly create large numbers of representative assets for benchmarking cluttered waste sorting scenarios.

2.4 Research Gap and Positioning of This Work

The literature highlights three complementary research directions; DT frameworks for robotic simulation, AI-based perception and manipulation methods for waste sorting, and data-driven approaches for generating 3D geometry from images. However, these areas remain largely disconnected in the context of conveyor-based CDW sorting. Existing DT platforms provide general simulation capabilities but rarely incorporate data-driven scene generation from real waste observations. Meanwhile, robotic CDW sorting research primarily focuses on perception or manipulation algorithms without providing infrastructure for systematic benchmarking under controlled conditions. In parallel, RGB-based 3D reconstruction methods aim for detailed geometry recovery but are often computationally demanding and unsuitable for rapid asset generation.

Consequently, there remains a gap between realistic scene modelling and practical benchmarking environment for evaluating robotic sorting systems. To address this gap, this work proposes a data-driven DT framework that integrates RGB annotation-based mesh generation with a synchronised Unity-ROS 2 simulation environment tailored to conveyor-based CDW sorting scenario. The framework enables automatic generation of realistic cluttered scenes, controlled scenario definition, and event-based performance evaluation, providing a reproducible platform for benchmarking robotic sorting strategies.

3 System Architecture

This section presents the architecture of the proposed DT framework for conveyor-based CDW sorting. The system integrates an RGB driven mesh generation module, a Unity-based simulation environment, and a ROS 2 communication layer. Together, these components enable controllable scenario generation and real time interaction with external robotic software, allowing perception, planning, and control algorithms to be evaluated in a consistent and repeatable environment.

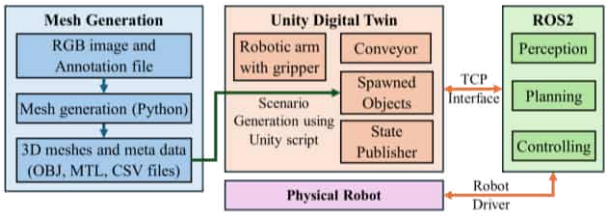


Figure 1. Overall system architecture and data flow between modules.

3.1 Overall Framework

Figure 1 illustrates the architecture of the proposed DT system, which comprises four components: an offline mesh generation module, a Unity based simulation environment, a ROS 2 coordination layer, and an optional physical robot interface. These components are loosely coupled so that scenario generation and simulation remain independent from external control algorithms. The mesh generation module converts annotated RGB images into lightweight 3D mesh proxies. Object boundaries extracted from polygon annotations are exported as OBJ files together with metadata describing class, dimensions, orientation, centroid, and estimated mass. The generated assets are stored locally and imported into Unity for conveyor sorting scene creation.

The Unity environment includes a conveyor, a simulated robot arm, and dynamically spawned waste

objects positioned according to the source image layout using metric scaling. Communication between Unity and ROS 2 is achieved via a TCP interface that supports bidirectional data exchange. Unity publishes object states and scene variables, while ROS 2 nodes provide control inputs such as belt speed and task priorities. This architecture allows the same interfaces to be used in both simulation and physical robotic system.

3.2 Mesh Generation

The mesh generation module converts annotated RGB images into lightweight 3D object assets used to instantiate realistic waste scenes in the Unity DT. Instead of reconstructing exact geometry, the goal is to generate physically plausible mesh proxies that preserve object footprint, class identity, spatial layout, and key geometric attributes required for benchmarking. The overall pipeline is illustrated in Figure 2. The dataset used for asset generation consists of 800 RGB images from a public dataset, containing approximately 1077 concrete, 1146 brick, 1655 wood, and 931 rubber objects [12]. Polygon-based instance annotations compatible with the YOLO format were generated in previous work and reused here to extract object boundaries.

For each image, segmentation annotations are read to obtain polygon boundaries of individual objects. Polygon vertices are converted from image coordinates to metric units using a scale derived from the conveyor belt width, preserving relative spacing in the simulation. Each footprint is then extruded vertically to form a closed triangular mesh, with the height sampled from class specific ranges to introduce realistic thickness variation. The resulting mesh is centred at its centroid to simplify spawning within Unity.

Algorithm 1. RGB annotation-based mesh and metadata generation

Input: RGB image, polygon annotation file, belt width, class height ranges, material densities

Output: Mesh files and metadata for all objects in the image

1. Compute pixel to metric scale from belt width and image width
2. Load polygon annotations and convert to pixel coordinates
3. Remove objects that are truncated near image boundaries
4. For each polygon instance do
   a. Convert polygon vertices to metric coordinates
   b. Estimate geometric properties including area, centroid, and orientation
   c. Sample extrusion height from the class specific height range
   d. Triangulate polygon and extrude to generate a closed mesh
   e. Compute volume and derived mass using material density
   f. Shift mesh so centroid becomes local origin
   g. Export mesh as OBJ and MTL files with unit and axis conversion
   h. Save metadata including class label, centroid position, dimensions, orientation, area, volume, height, and estimated weight
5. Save all object records to a scene level CSV file

Figure 2. Algorithm used for mesh generation.

Most waste objects exhibit approximately convex shapes, while rubber fragments frequently contain concave boundaries. To support both cases, polygon triangulation is performed using an ear-clipping algorithm capable of generating stable meshes from arbitrary simple polygons. Basic geometric validation steps are also applied during preprocessing, including polygon orientation correction and triangulation checks, ensuring robust mesh generation even when annotations contain irregular vertex distributions or minor noise.

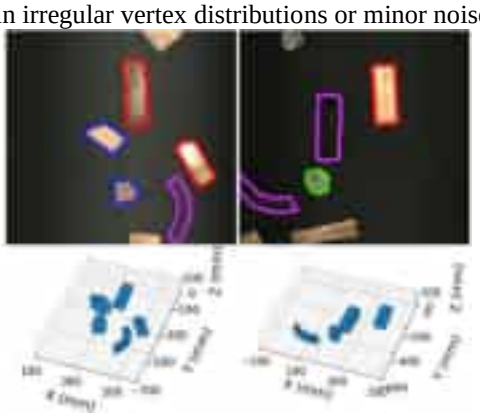


Figure 3. Example of RGB driven mesh generation. Top row shows the annotated RGB image with polygon boundaries and class labels. Bottom row shows the generated 3D meshes obtained by polygon extrusion and metric scaling

For compatibility with Unity, vertices are converted to metres and aligned with the simulation coordinate convention before export. The final assets are saved as OBJ and MTL files, together with a CSV file storing object metadata such as class, position, dimensions, orientation, and estimated mass. This automated process enables rapid and repeatable generation of simulation scenes without manual modelling. Figure 3 shows examples of the generated meshes alongside their annotated RGB images, demonstrating that the resulting assets preserve object layout while providing efficient volumetric representations for physics-based simulation.

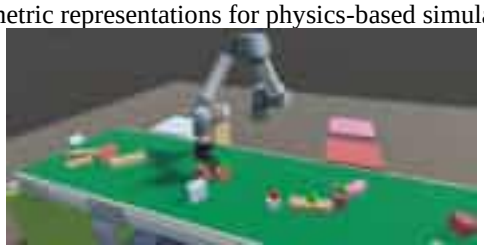


Figure 4. Unity based conveyor DT with dynamically spawned waste objects generated from RGB annotations. The environment includes the conveyor, simulated robot arm, and physics-based object interactions.

3.3 Conveyor DT in Unity

Unity serves as the DT of the physical conveyor sorting system, providing a controllable and repeatable environment for testing perception and decision making algorithms. The scene includes a conveyor belt, a simulated robotic arm with a gripper, and dynamically spawned waste objects generated by the mesh pipeline.

Unity is adopted for its flexible scene configuration, real time rendering, and built in physics capabilities.

At the start of each simulation episode, mesh assets and their metadata are loaded and used to spawn objects on the conveyor as shown in **Error! Reference source not found.**. Object attributes such as class, centroid, dimensions, and orientation are applied to preserving the relative layout of objects through metric scaling, enabling realistic cluttered scenes to be reproduced without manual placement.

Conveyor motion is simulated within Unity, while belt speed is controlled externally through ROS 2 commands and can include smooth variations to emulate realistic operating conditions. The environment continuously publishes object states and receives control inputs, allowing Unity to function as an interactive and measurable testbed for benchmarking rather than a passive visualisation tool.

3.4 ROS2 Interface and Topics

ROS 2 acts as the coordination layer linking the Unity DT with external perception, planning, and control modules. Data exchange is performed through ROS 2 topics via a TCP based interface, enabling bidirectional real time communication between the simulation and higher-level algorithms.



Figure 5. Main ROS 2 topics used for bidirectional communication between Unity and external modules. Arrows indicate data flow direction between Unity and ROS 2.

During operation, Unity publishes object states and global variables such as conveyor speed, while external nodes subscribe to this information for tracking, scheduling, and decision making. Control commands from ROS 2 are then sent back to Unity to influence the simulation, allowing the DT to operate as an interactive and modular experimental platform. Figure 5 summarises the main topics used in the current implementation, providing the necessary state and control information while maintaining a lightweight interface suitable for real time operation.

3.5 Event Detection via Colliders

To enable quantitative evaluation, the Unity DT uses physics-based colliders to automatically detect interaction events during simulation. Instead of relying on visual inspection, these triggers generate discrete signals for contacts, workspace transitions, and object removal, which are published to ROS 2 for real time monitoring and analysis. In addition, a collider at the tool centre point (TCP) is used to approximate grasp interaction, enabling evaluation of picking alignment without explicitly simulating detailed gripping physics.

Four types of colliders are used. Triggers attached to the gripper fingers detect contact with waste objects. A TCP collider provides a simplified representation of grasp interaction. Boundary colliders define the active conveyor workspace and record object entry and exit events. An additional trigger below the belt detects fallen objects and removes them from the scene to maintain stable simulation performance.

All events are transmitted through ROS 2 topics, allowing external nodes to compute statistics such as processed counts, missed interactions, and failures. This enables the DT to function as a measurable and repeatable benchmarking testbed.

4 Scenario Generation and Benchmarking Metrics

The proposed DT functions as both a simulation environment and a benchmarking platform for evaluating perception, scheduling, and manipulation strategies. It enables controlled experimentation through automatic scenario generation, structured ground truth, and quantitative performance metrics, supporting repeatable and fair comparison of algorithms under varying conveyor conditions.

4.1 Scenario Randomisation

To reflect real world variability and avoid overfitting to fixed layouts, the DT supports automatic randomisation of scene parameters at the start of each simulation episode. Randomisation is applied at both object and system levels. At the object level, scenes are generated by selecting mesh assets from different annotated RGB images and spawning them using their recorded spatial layout. Parameters such as extrusion height and material thickness are sampled within predefined ranges to introduce realistic variations in volume and mass while preserving object footprints.

At the system level, conveyor dynamics are varied by adjusting belt speed and adding smooth stochastic perturbations. Additional factors, including object ordering, spawn timing, and scene density, further diversify operating conditions. Because scenarios are generated programmatically, large numbers of repeatable test cases can be created efficiently, enabling statistical evaluation across multiple runs and reducing bias from

fixed layouts.

4.2 Available Ground Truth

A key advantage of the DT is the availability of precise ground truth information for every object and event in the scene. Unlike physical experiments, where sensing noise and occlusion may limit measurement accuracy, the simulation environment provides direct access to exact geometric and state variables.

For each object spawned, the system maintains attributes including class label, centroid position, orientation, footprint dimensions, area, volume, height, and estimated mass. During runtime, Unity continuously publishes object poses and scene states to ROS 2, enabling external algorithms to access accurate and synchronised measurements. In addition, collider-based triggers generate event signals such as object contact, workspace entry or exit, and object removal.

This comprehensive ground truth supports detailed analysis of algorithm behaviour, debugging, and objective evaluation of success and failure conditions without manual annotation.

4.3 Evaluation Metrics

The DT facilitates quantitative assessment of sorting and manipulation strategies through measurable metrics derived from ground truth and event signals. These metrics can be computed online via ROS 2 nodes or offline from logged data.

Typical indicators include the number of processed objects, task completion rate, and the proportion of correctly selected or prioritised targets. Timing related metrics such as processing time per object and throughput are obtained from timestamps associated with entry and exit events. Contact and removal signals enable detection of failed interactions, such as missed picks or unintended collisions. Example metrics are defined as

$$\text{Throughput} = n_{success} / T \quad (1)$$

$$\text{Success rate} = n_{success} / N \quad (2)$$

$$\text{Miss rate} = (N - n_{success}) / N \quad (3)$$

where $n_{success}$ denotes the number of successfully handled objects, T is the total time, and N is the total number of objects. As these measures are computed directly from ground truth rather than sensor estimates, they provide consistent and repeatable benchmarks across different algorithms and experimental settings.

Overall, the combination of automated scenario generation, structured ground truth, and quantitative metrics enables the framework to function as a controlled testbed for developing and comparing robotic sorting strategies prior to deployment on physical hardware.

5 Experimental Demonstration and Evaluation Workflow

This section presents system-level demonstrations to validate the computational efficiency and runtime stability of the proposed DT framework. The focus is not on evaluating specific picking strategies, but on confirming that the mesh generation pipeline and Unity-based simulation operate in real time and support large scale experimentation under varying conveyor conditions.

5.1 Asset Generation Performance

The computational cost of the mesh generation stage was evaluated to assess the feasibility of creating simulation assets directly from annotated RGB images. For each image, the pipeline extracts polygon instances, constructs extruded meshes, computes geometric descriptors, and exports mesh and metadata files without manual intervention.

Across multiple images, the average processing time was approximately 24 ms per image. At the object level, average mesh construction time was 5.3 ms per instance. These results indicate that multiple objects can be generated within a few milliseconds, enabling rapid batch creation of large numbers of simulation scenarios. The low computational overhead demonstrates that the proposed RGB driven geometry generation approach is efficient and scalable, making it practical for repeated asset generation during experimentation.

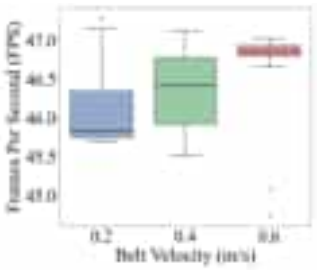


Figure 6. Variation of FPS for three different conveyor speeds measured for 1 minute.

5.2 Simulation Stability and Throughput

The runtime performance of the Unity-based DT was evaluated under cluttered scenes and varying conveyor speeds. Generated mesh assets were dynamically spawned using recorded spatial layout, resulting in multiple interacting objects on the conveyor.

Across tested speeds of 0.2 m/s, 0.4 m/s, and 0.6 m/s, the simulation maintained stable frame rates of approximately 46-47 frames per second (FPS), as shown in Figure 6. Only minor fluctuations were observed, with no noticeable degradation at higher belt velocities. This confirms that the lightweight mesh representation and physics configuration impose minimal computational

overhead and enable real time execution on a standard desktop workstation. Communication performance was also evaluated, showing stable round-trip latency over time with an average round-trip time of approximately 100 ms between Unity and ROS 2. This indicates that the system supports reliable real time interaction for control and state feedback.

To emulate realistic operating conditions, conveyor velocity can be modulated with smooth stochastic variations around a nominal speed. Figure 7 shows an example profile with a base speed of 0.2 m/s and small perturbations. This variability enables testing of perception and scheduling algorithms under non constant motion while maintaining stable simulation performance. Overall, these results demonstrate that DT supports efficient asset generation, stable real time simulation, and reliable communication, providing a practical foundation for large scale benchmarking of robotic sorting strategies.

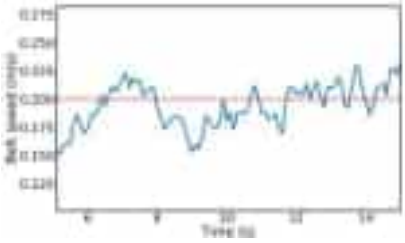


Figure 7. Example belt velocity profile generated from ROS2 side with a base speed of 0.2m/s.

5.3 Evaluation Workflow and Impact of Mesh Approximation

To illustrate how the proposed DT supports benchmarking of robotic sorting systems, a simplified picking experiment was conducted using a velocity feed-forward and feedback-based controller. The objective is not to evaluate the controller itself, but to demonstrate the interaction between perception, control, and the DT, and how performance metrics are computed.

In this framework, detailed physics-based gripping is not simulated. Instead, grasp interaction is approximated using colliders attached to the gripper fingers and tool centre point (TCP). A pick is considered successful when the gripper aligns on opposing sides of an object and the TCP overlaps the object region for a short dwell period ($\approx$0.6 s) before moving to the sorting bin. Based on these events, metrics such as throughput, success rate, and miss rate can be computed using n_success as defined in Equations (1)–(3). To further demonstrate the evaluation workflow, tracking performance was measured by recording the time required for the end-effector to align with a moving target such that the positional error falls below 4 mm. Across multiple trials, an average tracking time of approximately 2.3 s was observed. This result indicates that DT enables consistent evaluation of perception-to-action behaviour under dynamic conveyor conditions.

Under this evaluation, the influence of mesh approximation primarily relates to object footprint, centroid location, and effective grasp width. As meshes are generated directly from annotated polygon boundaries, these planar characteristics are preserved. While fine-scale surface details and complex concavities are not explicitly modelled, their impact is limited for the system-level interaction criteria used in benchmarking.

The mesh generation pipeline is also robust to minor annotation noise. Variations in polygon vertices mainly affect local contour smoothness without significantly altering object footprint or layout. Geometric validation steps, including orientation correction and triangulation checks, ensure stable mesh construction from imperfect annotations. Consequently, the DT is designed for system-level evaluation of perception-to-action behaviour, object localisation, and motion alignment rather than detailed contact dynamics. The use of lightweight proxy meshes enables efficient generation of diverse CDW objects while maintaining sufficient geometric fidelity for reproducible benchmarking.

6 Discussion and Future Integration

The proposed DT framework provides a practical approach for bridging real world waste observations with controllable simulation-based experimentation. By combining RGB driven mesh generation, a Unity-based physics environment, and ROS 2 integration, the system enables repeatable and instrumented evaluation of robotic sorting strategies. Automatic conversion of annotated images into 3D assets eliminates the need for manual modelling and supports rapid generation of diverse scenarios. Together with structured ground truth and event signals, this enables systematic benchmarking that is difficult to achieve in physical experiments.

A key strength of the framework lies in its modular architecture. Asset generation, simulation, and control components are loosely coupled, allowing independent modification of perception, planning, or control modules without altering the environment. This reduces development time and enables fair comparison of alternative approaches under consistent conditions and supports large scale experimentation across varying object layouts, densities, and conveyor dynamics.

Despite these advantages, several limitations remain. The generated meshes are simplified geometric proxies based on polygon extrusion and do not capture fine scale surface details or complex material properties. While this is sufficient for system-level evaluation of perception, localisation, and motion planning, it may limit applicability for detailed grasp contact modelling. In addition, simulation physics cannot fully represent real

world effects such as friction variability or deformability. Therefore, results obtained in the DT should be interpreted as indicative of system-level performance rather than exact predictions of real world behaviour.

Future work will focus on integrating the framework with physical robotic systems and perception pipelines. As both simulated and real setups share the same ROS 2 interfaces, control algorithms developed and tested in the DT can be transferred to the physical robot with minimal modification. Planned extensions include camera-based detection and tracking, implementing grasp planning and scheduling strategies, and conducting closed loop picking experiments on the conveyor. Additional improvements will explore more advanced geometry modelling, domain randomisation of physical parameters, and expansion of the asset library using larger and more diverse datasets.

Overall, the framework provides a scalable foundation for iterative development of AI-driven robotic sorting systems. By enabling safe, repeatable, and data rich experimentation in simulation prior to hardware deployment, the DT reduces integration risk and accelerates the pathway from algorithm design to real world application.

7 Conclusions

This paper presented a DT simulation framework for conveyor-based CDW sorting that integrates RGB driven mesh generation, a Unity-based physics environment, and a ROS 2 communication layer. The system automatically converts annotated RGB images into 3D mesh assets to generate realistic and repeatable sorting scenarios. By coupling data-driven asset generation with real time interaction and structured ground truth, the framework enables systematic benchmarking of perception, scheduling, and manipulation strategies without the cost and risk of physical experimentation. Experimental results confirmed that the mesh generation pipeline operates efficiently, requiring only milliseconds per object, while the Unity simulation maintains stable real time performance across varying conveyor speeds. The framework also supports the evaluation of perception-to-action behaviour and motion alignment under dynamic conditions. Although the meshes are simplified geometric proxies, they preserve key planar characteristics required for benchmarking and are sufficient for system-level evaluation. The modular design supports integration with ROS 2 based algorithms and future deployment on physical robotic platforms, providing a practical and extensible testbed that facilitates rapid development and evaluation of AI-enabled robotic waste sorting systems.

References

[1] S. Dodampegama, L. Hou, E. Asadi, G. Zhang, and S. Setunge, ‘Revolutionizing construction and demolition waste sorting: Insights from artificial intelligence and robotic applications’, *Resources, Conservation and Recycling*, vol. 202, p. 107375, Jan. 2024.

[2] M. Singh *et al.*, ‘Unity and ROS as a Digital and Communication Layer for Digital Twin Application: Case Study of Robotic Arm in a Smart Manufacturing Cell’, *Sensors (Basel)*, vol. 24, no. 17, p. 5680, Aug. 2024.

[3] V. V. Tuhaise, J. H. M. Tah, and F. H. Abanda, ‘Technologies for digital twin applications in construction’, *Automation in Construction*, vol. 152, p. 104931, Aug. 2023.

[4] M. Singh *et al.*, ‘Comparative Study of Digital Twin Developed in Unity and Gazebo’, *Electronics*, vol. 14, no. 2, p. 276, Jan. 2025.

[5] A. Haces-Garcia and W. Zhu, ‘Building an Accessible and Flexible Multi-User Robotic Simulation Framework with Unity-MATLAB Bridge’, *Computers*, vol. 13, no. 11, p. 282, Nov. 2024.

[6] W. Chang, W. Sun, P. Chen, and H. Xu, ‘Construction and Validation of a Digital Twin-Driven Virtual-Reality Fusion Control Platform for Industrial Robots’, *Sensors*, vol. 25, no. 13, p. 4153, Jan. 2025.

[7] Z. Wang, X. Yang, X. Zheng, D. Huang, and B. Jiang, ‘Computer Vision System for Multi-Robot Construction Waste Management: Integrating Cloud and Edge Computing’, *Buildings*, vol. 14, no. 12, p. 3999, Dec. 2024.

[8] H. T. N. Le and H. Q. T. Ngo, ‘Application of the vision-based deep learning technique for waste classification using the robotic manipulation system’, *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 391–400, Dec. 2025.

[9] X. Liu, Z. Farshadfar, and S. H. Khajavi, ‘Computer Vision-Enabled Construction Waste Sorting: A Sensitivity Analysis’, *Applied Sciences*, vol. 15, no. 19, p. 10550, Jan. 2025.

[10] M. S. U. Khan, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal, ‘Three-Dimensional Reconstruction from a Single RGB Image Using Deep Learning: A Review’, *J Imaging*, vol. 8, no. 9, p. 225, Aug. 2022.

[11] J. Yun *et al.*, ‘Digital twin model construction of robot and multi-object under stacking environment for grasping planning’, *Applied Soft Computing*, vol. 149, p. 111005, Dec. 2023.

[12] T. Ji, J. Li, H. Fang, R. Zhang, J. Yang, and L. Fan, ‘Rapid dataset generation methods for stacked

construction solid waste based on machine vision and deep learning', *PLOS ONE*, vol. 19, no. 1, p. e0296666, Jan. 2024.

From 2D Drawings to BIM Models: An LLM-Assisted Approach

Jean Viaunel Victor[1] and Pavan Kumar[2] and Shang-Hsien Hsieh[3]

[1]Master Student, Department of Civil Engineering, National Taiwan University, Taiwan
[2]PhD Candidate, Department of Civil Engineering, National Taiwan University, Taiwan
[3]Professor, Department of Civil Engineering, National Taiwan University, Taiwan
viocitovictor@caece.net

Abstract

Building Information Modeling (BIM) has become a central platform for digital building representation; however, the creation and manipulation of BIM models remain largely manual, and accessible primarily to domain experts. At the same time, architectural information is still commonly produced in 2D drawing formats, creating inefficiencies in translating design intent into semantically rich 3D BIM models. Recent advances in large language models (LLMs) offer new possibilities for natural language–based interaction with complex digital systems, yet their integration with BIM for automated and scalable building modeling is still limited. This research addresses this gap by proposing an automated framework that integrates LLMs with BIM systems to support the transformation of 2D architectural drawings into parametric model generation and manipulation. The framework's capability is demonstrated by using a single-floor architectural plan as input; the system enables the scalable generation of multi-storey buildings. The results indicate that LLM-assisted BIM workflows can significantly reduce manual modeling effort and improve efficiency, while remaining computationally tractable for generating multi-storey buildings. This study demonstrates the potential of LLM–BIM integration as a foundation for more intuitive, efficient, and scalable building modeling workflows.

Keywords –
image2BIM, img2BIM, LLM-BIM integration

1 Introduction

Building Information Modeling (BIM) has become a foundational technology for digital building representation and management across the architecture, engineering, and construction (AEC) industry. It is widely regarded as an intelligent, three-dimensional, model-centric process that enables the comprehensive integration of geometric and semantic information across the entire building lifecycle**[1].** By integrating geometric, semantic, and relational information within a unified digital environment, BIM supports improved design coordination, construction planning, and lifecycle management of built assets. As such, BIM has been widely adopted as a central platform for advancing digital transformation in the built environment.

Despite its widespread adoption, the creation and manipulation of BIM models remain largely dependent on manual operations performed within specialized authoring tools. Furthermore, heavily reliant on 2D CAD formats for design communication [2], these processes are time-consuming, require significant domain expertise, and familiarity with software-specific modeling paradigms. Consequently, BIM authoring remains time-consuming, error-prone, and inaccessible to non-expert stakeholders, particularly during the early design stages, where rapid exploration and iteration are crucial.

At the same time, architectural information is still predominantly produced and communicated through two-dimensional (2D) drawings, such as floor plans, sections, and elevations. While these representations effectively convey design intent, they lack the explicit semantic structure and parametric relationships required for downstream BIM-based analysis and automation**[2].** The persistent reliance on 2D documentation creates a disconnect between early-stage design practices and BIM-centric workflows.

Large Language Models (LLMs) provide an opportunity, and a recent review highlights that LLMs in the construction sector demonstrate their high potential, with median accuracies of 89% for building information detection and 91% for auxiliary modeling automation [3]. While earlier studies leveraged LLMs primarily for building information detection[3] and querying 3D spatial relationships through tree-based indexing [4], the contemporary landscape is increasingly defined by generative workflows. Notably, the workflow proposed by[2] orchestrates a multi-agent system to transform abstract textual instructions into imperative Python code for BIM authoring software APIs, enabling the creation

of natively editable models with internal layouts and semantic metadata. Parallel efforts have integrated conversational AI with Visual Programming Languages (VPLs), such as Grasshopper, to allow designers to iterate on forms and materials through natural-language prompts[5][6]. Similarly, [7] proposed the integration of LangGraph and LLM to generate BIM models, and the LSVR-SE workflow demonstrates text-to-BIM conversion for generating compliant Industry Foundation Classes (IFC) files directly from language commands[8].

However, as the field matures, there remains a valuable opportunity to refine the workflows at the technical grounding and orchestration of these generative processes. Existing workflows primarily excel at engaging natural language as input; they often rely on abstract linguistic prompts that may lack the inherent geometric precision and structured data found in existing technical documentation. Furthermore, recent studies indicate that agent-based BIM interactions can remain fragmented, often requiring repetitive re-implementation of core capabilities across different proprietary APIs[9]. This study seeks to build a workflow integrating LLM and BIM to generate BIM models, taking 2D DXF architectural drawings, which provide necessary geometric constraints for professional-grade modelling. Central to this approach is the adoption of the Model Context Protocol (MCP). By utilizing MCP to orchestrate a staged transformation pipeline, this study enables the systematic parsing of 2D primitives and their translation into semantically rich, multi-storey IFC models. The proposed workflow provides an interoperable foundation that moves beyond single-story constraints, offering a scalable bridge between traditional 2D documentation and automated 3D parametric modelling and manipulation.

This research study aims to develop a necessary technical proof-of-concept of a workflow that integrates LLMs with BIM systems to transform the 2D architectural drawings into parametric models. By focusing on the technical feasibility of the proposed integrated workflow, the research study establishes a scalable baseline for automating complex BIM authoring tasks. The proposed framework enables natural language–driven building generation and manipulation, with a specific emphasis on achieving scalability in multi-storey building modeling.

2 Methodology

This study proposes a unified framework integrating LLM with a parametric model from a 2D architectural floor plan. The LLM-driven pipeline for converting two-dimensional DXF architectural drawings into a semantically enriched three-dimensional IFC Building Information Model consists of the LLM orchestrating four core tools: Parse DXF, detect room, Generate IFC, and Create multi-storey building, as shown in Figure 1. Each of these core tools consists of sub-tools such as Read DXF file, followed by Detecting units, normalizing coordinates, and extracting elements, forming a tool Parse DXF as shown in Figure 2.

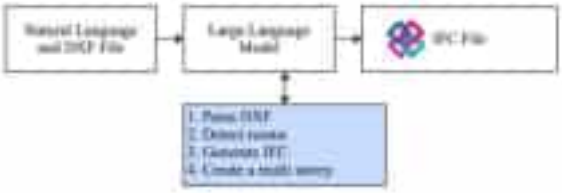


Figure 1. Overview of LLM-driven BIM to automate 2D floor plans into BIM Models

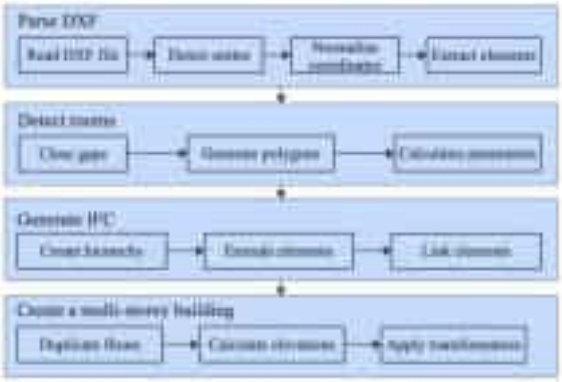


Figure 2. A Detailed Workflow for Generating Multi-Storey BIM Models from DXF Floor Plans

2.1 Parse DXF

This is the first tool that LLM executes in the stack of 4 tools in sequence. In the Parse DXF, the provided DXF file by the user is read with “ezdxf.readfile()” to access the header and model space. Units are detected from $INSUNITS and scaled via “get_dxf_unit_scale()” to millimeters. Layers are mapped using a configurable alias system through “detect_wall_layer_structure()”. Elements (walls, columns, windows, doors, slabs) are extracted with specialized functions, preserving semantic XData (e.g., Revit wall types).

2.2 Detect Rooms

Once the elements are extracted from the parsed DXF, the LLM verifies the outputs and calls wall endpoints, which are connected across small gaps (< 100 mm) at doors/windows, using “_find_and_close_openings()”. Walls are merged and polygonised with Shapely to form a room. Parameters such as wall start and end points, and locations of doors and windows are then returned to the LLM.

2.3 Generate IFC

Standard IFC files are generated through this tool. A standard IFC4 hierarchy (IfcProject → IfcSite → IfcBuilding → IfcBuildingStorey) is created via "create_combined_ifc()". Furthermore, walls are extruded to floor height (default 3000 mm) through IfcWallStandardCase objects. Doors and windows are modelled as IfcOpeningElement voids. All elements are linked to stories to maintain BIM validity.

2.4 Create Multi-storey Building

The generated IFC file is then used to create a multi-storey building as per the user's request. Parameters, such as the number of walls, windows, doors, and the final IFC model, are returned. The ground floor is treated as a reference floor and can be duplicated for the nth stories. The user can also instruct to specifically choose a certain floor to be the reference floor and duplicate it for the nth storey. Each copy is placed at elevation `$Z_i = i \times H\_f$`, with all elements shifted vertically. Specific stories can optionally be replaced with alternate DXF files (e.g., a lobby or penthouse).

2.5 Intent Interpretation Layer

In the intent interpretation layer, the LLM engages in interpreting the user inputs and orchestrates the tools in sequence through the Model Context Protocol (MCP) Standard. The natural-language requests are converted into structured parameters (floor indices, feature flags, heights). LLM callable tools execute functions such as "convert_dxf_to_ifc()" and "create_multi_storey_building()", using the Model Context Protocol (MCP) standard. Inputs are then validated, and structured results with statistics are returned by the LLM to the user.

3 Example demonstration

This section describes the demonstration of the framework proposed to integrate LLM and BIM to generate an IFC file from the 2D floor plan. A DXF file, as shown in Figures 3 and 4 is used as input along with the user request in Figure 5. The DXF file is a layout on an orthogonal alphanumeric grid and uses standard CAD layers for walls, doors, windows, furniture, sanitary fixtures, and stairs. This simple, structured representation is extremely important as it makes it easier for both humans and algorithms to read.

The framework accepts natural language requests alongside 2D DXF floor plans as input. An LLM interprets the user's intent, extracting structured parameters (e.g., number of storeys, floor height, floor swapping specifications) and mapping them to predefined tool functions through MCP configured for Claude code with Sonnet 4.5. From these tools, architectural elements (walls, columns, windows, doors) are extracted through layer-based classification. The final output is an IFC-compliant 3D BIM model, which is compatible with software such as Revit and ArchiCAD, as shown in Figure 5.

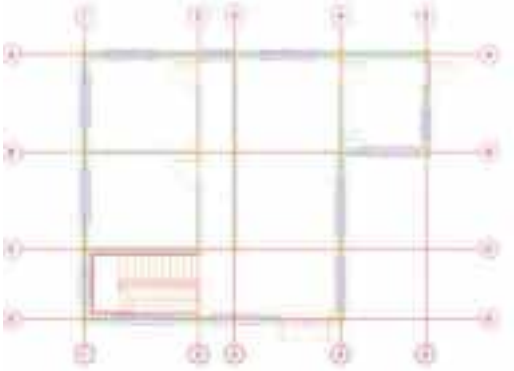


Figure 3. Floor plan 1

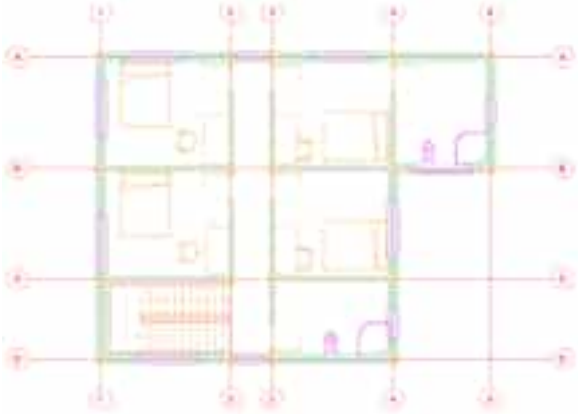


Figure 4. Floor plan 2



Figure 5. User request as input to the LLM to generate the 3D model.

4 Results & Discussions

This section lays out the result of the example demonstrated using the proposed framework to integrate BIM and LLM to generate IFC files from the 2D floor plans. The DXF file (figures 3 and 4), along with the user request, Create a 5-story building with @"../floors/Level 0.dxf" as ground floor and @"../floors/Level 1.dxf" as typical floors. is provided to the LLM to generate an IFC file. The generated IFC file can be viewed through tools such as Revit, ArchiCAD, etc. The IFC file generated in the example demonstration is shown in Figure 6.



Figure 6. IFC 3D Model Viewed In Revit



Figure 7. Three-Stage Generative Modeling Process

The transformation follows a staged pipeline in which geometric primitives are extracted and normalized, processed into element-level representations, and then translated into IFC entities with spatial relationships. This separation enables each element type to be handled using element-specific rules while maintaining IFC schema compliance.

Column entities are extracted during Parse DXF through layer-based classification and converted to IFC columns during IFC generation. The intermediate output (Figure 7(a)) visualizes the extracted vertical elements as isolated volumes at their detected plan locations, providing an immediate check that the layer-based extraction has correctly recovered structural placement before enclosure elements are assembled.

Wall geometry is then reconstructed from the extracted DXF wall elements and is then instantiated in IFC using IfcWallStandardCase objects. In the demonstrated input, the DXF plan is an orthogonal layout using standard CAD naming layers (AIA), which supports reliable extraction and simplifies downstream reconstruction. Table 1 shows how layers are matched to IFC classes. Walls are then extruded to the floor height parameter (default 3000 mm), and all generated elements are linked to their respective storeys to maintain BIM validity. The intermediate wall assembly shown in Figure 7(b) illustrates that the system recovers the external perimeter and internal partitions of the layout with minor inconsistencies with the source drawings.

Openings are handled as semantic voids to preserve IFC validity while delegating solid subtraction to BIM applications. Doors and windows are modelled as IfcOpeningElement voids, and the resulting openings are associated with host walls through standard IFC relationships during IFC generation. The Detect Rooms stage further supports topological consistency by closing wall endpoints across small gaps (< 100 mm) at door and window locations, improving enclosure continuity for downstream interpretation. The final stage shown in Figure 7(c) reflects this semantic modelling choice: openings appear consistently placed within the wall system, and BIM viewers render these voids during visualization.

Table 1. Mapping from AIA CAD naming layer to the Architectural Elements

DXF Layer Name	Architectural Element	IFC Class Generated
A-WALL	Wall	IfcWall
A-COLS	Column	IfcColumn
A-DOOR	Door	IfcDoor
A-WIND	Window	IfcWindow
A-SLAB	Floor slab	IfcSlab
A-STAIR	Stair	Not Generated
A-FURN	Furniture	Not Generated

The integration of natural language input with MCP-based orchestration provides a practical interface for specifying conversion parameters (e.g., storey count and height) without requiring direct interaction with BIM authoring tools. In this example, the produced BIM model preserves the orthogonal layout and storey structure of the source drawings, and the resulting IFC file is usable in standard BIM environments for inspection and further editing.

The proposed proof of concept workflow demonstrates architectural scalability through the automated generation of multi-storey models from 2D floor plan data and text, highlighting that efficiency and accessibility are currently qualitative, based on the reduction of discrete manual modelling steps.

At the same time, the results highlight limitations consistent with the current scope of the pipeline. Performance and accuracy are sensitive to the drawing being structured and layer-disciplined, as assumed by the

layer-based extraction approach used in Parse DXF.

In addition, the current implementation prioritizes orthogonal plan reconstruction (as used in the demonstration), and more complex drafting cases (e.g., diagonal walls, irregular junctions, and inconsistent drafting gaps) require additional geometric reasoning and automated quality checks beyond the current gap-closing strategy. Future work shall focus on expanding geometric coverage (e.g., non-orthogonal walls and junction resolution), increasing semantic enrichment of the exported model while preserving IFC4 compliance, and quantitatively comparing this LLM-assisted approach against traditional BIM authoring workflows for both expert and non-expert users.

5 Conclusion

This research study demonstrates a framework that integrates Large Language Models (LLMs) with Building Information Modeling (BIM) to bridge the gap between 2D architectural floor plans and semantically rich 3D parametric models. By automating the transformation of floor plans into accurate multi-storey representations, the proposed system can significantly reduce time-consuming manual work of building 3D models and reduce reliance on manual labour and specialized expertise traditionally required for BIM authoring. The proposed framework utilizes the capabilities of LLM and orchestrates multiple tools based on the inputs, such as DXF files alongside natural language inputs. The framework automates a staged transformation pipeline—extracting geometric primitives, normalizing element-level representations, and instantiating IFC-compliant entities.

The demonstration results confirm that the system can reliably transform standard 2D layouts into multi-storey IFC models that preserve structural placement, wall perimeters, and semantic relationships such as door and window openings. While the current implementation shows high fidelity for orthogonal, layer-disciplined drawings, it also highlights the necessity for advanced geometric reasoning to handle non-orthogonal layouts and drafting inconsistencies. Furthermore, this study establishes technical feasibility, and quantitative performance metrics shall be employed in future work of this study, with a focus on benchmarking the framework using element detection precision and geometric deviation to provide a statistically significant assessment of its robustness compared to existing non-LLM baselines. Ultimately, this study demonstrates that LLM-assisted BIM workflows can significantly lower the barrier to entry for complex building modeling, providing a scalable and intuitive foundation for future advancements in automated AEC digital transformation.

References

[1] S. Lee, M. Uhm, H. D. Jeong, and G. Lee, "Automated conversion and semantic gap filling of material layer annotations from CAD drawings to semantically rich BIM objects," *Advanced Engineering Informatics*, vol. 70, Mar. 2026, doi: 10.1016/j.aei.2025.104199.

[2] C. Du, S. Esser, S. Nousias, and A. Borrmann, "Text2BIM: Generating Building Models Using a Large Language Model-Based Multiagent Framework," *Journal of Computing in Civil Engineering*, vol. 40, no. 2, Mar. 2026, doi: 10.1061/jccee5.cpeng-6386.

[3] J. Liang, Y. Lin, Y. Liu, Y. Huang, and H. Wu, "Pre-training large language models based on Transformer architecture for building industry application: A review," *Tsinghua University*. Nov. 01, 2025 doi: 10.1007/s12273-025-1324-9.

[4] A. Li, P. K. Y. Wong, X. Tao, J. Ma, and J. C. P. Cheng, "An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent," *Advanced Engineering Informatics*, vol. 66, Jul. 2025, doi: 10.1016/j.aei.2025.103375.

[5] J. Ko, J. Ajibefun, and W. Yan, "Generative AI-powered parametric modeling and BIM for architectural design and visualization," in *Proceedings of the Design Society*, Cambridge University Press, pp. 1943–1952. Aug. 2025 doi: 10.1017/pds.2025.10208.

[6] B. El Hizmi, Y. Sterman, and G. Austern, "LLMto3D - Generation of parametric, 3D printable objects using large language models," *International Journal of Architectural Computing*, vol. 23, no. 3, pp. 701–719, Sep. 2025, doi: 10.1177/14780771251353792.

[7] M. Selim, K. Nassar, and O. Hosny, "Automating BIM (IFC) Data Analysis Using LangGraph," in *World Congress on Civil, Structural, and Environmental Engineering*, Avestia Publishing, 2025. doi: 10.11159/icsect25.182.

[8] M. Xu, G. Qi, N. He, and F. Ju, "Language-guided semantic editing in single-view 3D reconstruction," *Visual Computer*, vol. 42, no. 3, Feb. 2026, doi: 10.1007/s00371-025-04349-y.

[9] Hellin, N., Du, C., & Borrmann, A. Natural language information retrieval from BIM models: An LLM-based multi-agent system approach. *Proceedings of the 2025 European Conference on Computing in Construction (EC3).* 2025

Reliable Social Media Flood Image Localization for Urban Flood Situational Awareness

Zihua Zhu[1], Xianfei Yin[1,*] and Qihua Chen[1]

[1]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR
*Corresponding Author
zihuazhu2-c@my.cityu.edu.hk, xianfyin@cityu.edu.hk, qi-hua.chen@my.cityu.edu.hk

Abstract –

Precise geolocation of social media is vital for real-time urban flood monitoring. However, researchers are frequently hindered by the scarcity of geotags due to privacy protections and the limitations of coarse localization based on posting frequency, which fails to pinpoint street-level flood sites. We present a robust end-to-end framework that addresses these critical gaps by integrating state-of-the-art Visual Place Recognition (VPR) with a visual evidence verification mechanism. Our pipeline leverages vision foundation models to maintain operational resilience against extreme visual occlusion caused by disasters. To ensure reliability, an inlier-based visual evidence layer is introduced to verify the top-ranked prediction and provide an interpretable confidence cue. This approach establishes a dependable, city-scale monitoring system that bridges the gap between raw social media data and the precise demands of municipal emergency response.

Keywords –

Urban flood; Visual geo-localization; Social media; Situational awareness

1 Introduction

Urban flooding is becoming a persistent stress test for cities as climate change intensifies heavy precipitation events and pushes drainage and transport systems closer to failure [1]. The increasing frequency of intense rainfall heightens the likelihood of cascading disruptions across critical infrastructure and services [2]. Furthermore, current risk management frameworks often struggle when hazards exceed previously experienced magnitudes, leaving emergency operations exposed to unprecedented conditions [3]. From an infrastructure and civil engineering perspective, such events are not only disaster-management issues but also operational challenges for urban resilience. Consequently, rapid, city-wide flood situational awareness is essential to the development of sensing systems and smart, sustainable, resilient cities. While fixed monitoring approaches, such as physical gauges, provide precise measurements, they are spatially sparse and costly to deploy and maintain at scale. Similarly, although satellite remote sensing can expand spatial coverage, optical revisit time constraints often limit the ability to map urban inundation in real time during rapidly evolving events [4,5].

These limitations underscore the need for multi-source data fusion, which supplements traditional monitoring with complementary data streams. Social media serves as a high-velocity stream of Volunteered Geographic Information (VGI), embodying the "citizens as sensors" paradigm, in which individuals act as observers [6]. During disasters, crowdsourced content provides near real-time ground-truth verification, filling critical gaps in sparse official sensor networks [7]. For flood management, visual data complements numerical metrics by capturing the physical reality of hazards, enabling engineers to directly evaluate inundation severity and infrastructure functionality [8, 9]. Therefore, crowdsourced imagery represents a form of opportunistic urban sensing that facilitates infrastructure-aware decision workflows.

However, a critical barrier to integrating this visual data into engineering workflows is the lack of precise geolocation [10]. Due to strict platform policies and growing concerns regarding user privacy, geospatial metadata is often removed from public uploads [11]. Without reliable coordinates, even highly informative photographs cannot be anchored to specific infrastructure assets, isolating visual evidence from location-driven operations such as emergency dispatch and pump deployment [12]. Transforming raw visual information into spatially actionable intelligence is, therefore, a prerequisite for operationalizing social media data within professional infrastructure management systems [13].

Advances in computer vision offer a practical route to operational geolocation via Visual Place Recognition (VPR), which infers coordinates by retrieving the most similar views from a georeferenced street-view database and transferring their associated location priors [14]. The field was notably advanced by NetVLAD, which

introduced trainable aggregation layers for compact global descriptors [15], and more recently by CosPlace, which reframes geo-localization as a classification task to achieve city-wide efficiency without expensive mining [16]. However, these benchmarks predominantly focus on cyclic environmental changes, such as illumination or seasons [17]. In contrast, flooding introduces abrupt visual domain shifts, ranging from submerged roads to complex reflections [18]. Furthermore, raw retrieval accuracy alone cannot support engineering adoption; emergency response requires quantifiable reliability metrics to ensure that location predictions are trustworthy enough to guide operations [19].

This research addresses these gaps by presenting an end-to-end framework for flood image geo-localization from social media imagery, with a specific focus on reliable sensing for resilient urban infrastructure operations. Rather than proposing a new backbone architecture, this study asks a practically important question: which existing VPR paradigms remain reliable under severe flood-induced visual corruption, and how can their predictions be verified before use in human decision workflows? To answer this question, we first benchmark state-of-the-art VPR models against extreme flood scenarios to assess their reliability for street-level localization. To support human-in-the-loop management, a geometric verification layer is added to produce verifiable visual evidence alongside location predictions. This work demonstrates how AI-enabled sensing can be effectively harnessed to support resilient city operations.

2 Methodology

The proposed end-to-end framework for flood situational awareness is illustrated in Figure. 1. This pipeline is designed to transform unstructured social media imagery into actionable engineering intelligence through three interconnected stages. Firstly, the process begins with the data acquisition phase (left panel), where real-time crowdsourced images are captured and referenced against a citywide street-view database. Secondly, the core flood localization module (middle panel) utilizes VPR algorithms to extract query embeddings and perform vector similarity searches, identifying the most probable geographic coordinates from the reference dataset. Lastly, the workflow culminates in the situational awareness stage (right panel), where retrieved locations undergo visual verification to map the flood extent and support municipal decision-making. The specific implementation details of each module are described in the Section 3.1

2.1 Dataset Acquisition and Processing

The city-scale geo-localization framework uses two datasets: a georeferenced database of street-view imagery and a query set of external test images whose locations must be inferred. This study utilizes the San Francisco eXtra Large (SF-XL) dataset as the reference database and evaluates retrieval performance across three separate query datasets whose statistics are summarized in Table 1. Among these query sets, SF-Flood is a newly curated dataset in this study.

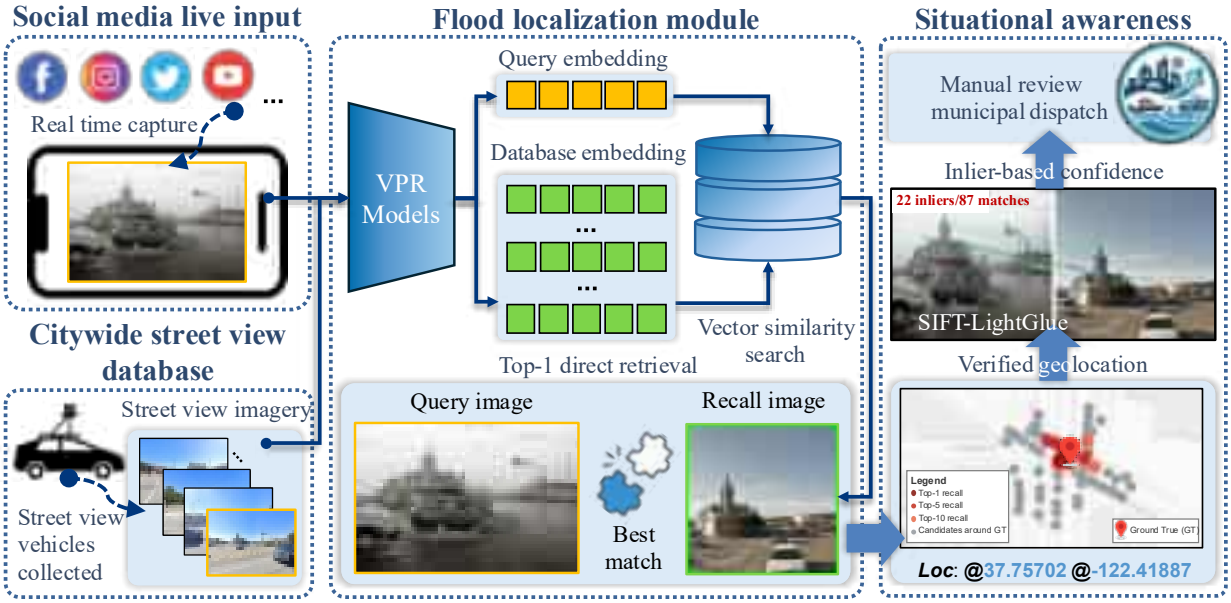


Figure 1. The schematic of the proposed end-to-end framework for urban flood situational awareness. A flood query image is matched against a geo-referenced street-view database using VPR-based descriptor retrieval. The top-1 recalled image provides an initial geolocation estimate, then verified by SIFT-LightGlue matching.

SF-XL, originally introduced by CosPlace [16], is a large-scale open-source street-view benchmark covering San Francisco. The full SF-XL release contains 41.2 million Google Street View panoramas collected across multiple years. In this study, we employ a 2.8 million-image subset as the query set for large-scale retrieval. This choice preserves city-wide spatial coverage and temporal variation, while keeping large-scale vector search computationally viable.

Table 1. Statistics for each subset used in the study.

	SF-XL (Original)	SF-XL test v1 (Test)	SF-XL test v2 (Test)	SF-Flood (Test)
Database	41.2M	2.8M	2.8M	2.8M
Queries		1000	598	400

These sets are designed to benchmark capabilities ranging from general environmental adaptability to extreme disaster resilience:

- **SF-XL Test v1 & v2 :** These serve as baselines for routine localization. v1 contains 1,000 Flickr images representing unconstrained photography [16], while v2 comprises 598 queries from the San Francisco Landmark Dataset [20] with accurate 6-DoF ground truth labels.
- **SF-Flood Dataset:** This specialized query set was curated in this study to evaluate localization under flood-induced domain shifts. Images were crawled from social media and strictly filtered to ensure they contain both visible flood evidence and identifiable geographic anchors (e.g., building façades or road layouts). To ensure the reliability of the 50 m evaluation criterion, each image was manually georeferenced through image-by-image matching in Google Street View. By explicitly incorporating extreme conditions such as water glare, reflections, and nighttime scenes, as illustrated in Figure 2, SF-Flood provides a rigorous stress test for VPR models under severe flood-related appearance changes.

We categorize the query images by their dominant localization challenge into four groups: Rich texture, Low texture, Viewpoint change, and Illumination change. Each image is assigned to the most prominent difficulty source for reporting, as details are shown in Table 2.

Table 2. The index system for crawling and extracting social media images related to flood.

Categories	Number	Definition
Rich texture	146	Repeated or visually similar urban structure; easy to confuse with look-alike locations.
Low texture	108	Low-density or peripheral scene with fewer strong localization anchors.
Viewpoint change	92	Major camera angle or framing change relative to street-view perspective.
Illumination change	54	Appearance dominated by lighting, night scenes, rain haze, or strong reflections.
Total	400	



Figure 2. Examples of images of SF-Flood dataset.

2.2 VPR Mechanism and Evaluated Models

VPR) is formulated as a large-scale image retrieval task. Given a query image I_q and a geo-referenced gallery image I_i, a VPR model $f(\cdot)$ maps each image to a d-dimensional normalized global descriptor:

$$z = f(I), \qquad z \in \mathbb{R}^{\mathbb{d}}, \qquad |z|_2 = 1 \tag{1}$$

Localization is then performed by nearest-neighbor search in the descriptor space using cosine similarity. The best-matched gallery image is obtained by

$$i^* = \arg\max_i \mathrm{z_q^\top z} \tag{2}$$

Where z_q and z_i denote the descriptors of the query image and the i-th gallery image, respectively. The location associated with the best-matched gallery image is taken as the predicted geolocation of the query.

This study evaluates a representative set of global-descriptor VPR methods spanning classic aggregation-based models, scalable classification-based training strategies, modern feature-mixing architectures, and foundation-model backbones. All methods are evaluated using the authors' released architectures and publicly available pretrained weights.

2.3 Reliability Verification and Visualization

While global-descriptor retrieval provides candidate locations, raw similarity alone can be unreliable in flood scenarios because severe appearance changes may lead to visually plausible but spatially incorrect matches.

Therefore, we introduce a post-retrieval verification stage for the top-ranked prediction.

Specifically, local correspondences between the query image and the retrieved image are established through a hybrid matching pipeline. First, Scale-Invariant Feature Transform (SIFT) is used to detect and describe local keypoints [21]. Second, the extracted features are matched using LightGlue [22], a neural feature matcher designed to improve correspondence quality and suppress outliers under challenging visual conditions. This choice is motivated by recent VPR literature, which shows that SIFT-LightGlue is a strong-performing option for post-retrieval verification and that inlier support is an effective indicator of retrieval confidence [23].

In this framework, inlier counts provide an interpretable reliability metric. Verified correspondences are visualized as direct connections, offering a verifiable link that supports manual audits of AI outputs. This stage acts as a confidence-aware module, bridging the gap between automated retrieval and accountable human-in-the-loop decision-making.

3 Experiment and Result

3.1 Experiment setup

To evaluate localization performance, we use Recall@K (R@K) with a success threshold of 50 m. A query is counted as successfully localized at rank K if at least one of the top-K retrieved gallery images lies within 50 m of the ground-truth location. The 50 m threshold is used to represent street-level localization accuracy. In practical terms, this means that the predicted location falls within the correct street or flood vicinity, which is sufficiently precise to help municipal responders perform on-site confirmation. Under this definition, R@1 is the proportion of cases in which the system places the query within 50 m of the true flood location on the first retrieval, while R@10 indicates whether the correct location is included within a short candidate list for follow-up review.

$$R@K = \frac{1}{|Q|}\sum_{q\in Q} 1!\left(\min_{1\le j\le K} d(q, I_j) \le 50\right) \tag{3}$$

Where Q is the query set, I_j is the j-th retrieved gallery image, $d(\cdot)$ is the geodetic distance in meter.

All experiments were conducted in a Linux environment (Ubuntu 22.04, x86-64) using Python 3.12.6, PyTorch 2.5.1, CUDA 12.4, and a single NVIDIA RTX 4090 GPU (24 GB). For global retrieval, images were resized to 224 × 224. For geometric verification, images were resized to 512 × 512 to preserve local structural details for keypoint matching. The number of verified inliers is used as an indicator of prediction reliability. Unless otherwise stated, default hyperparameter settings are used in the geometric verification stage.

Although the present study focuses on retrieval-based evaluation, the ranked candidate list may also support downstream spatial aggregation or map-based clustering analysis for practical deployment. We leave such application-oriented extensions to future work.

Table 3. Recall metrics on general-purpose datasets: SF-XL test v1 and v2. Detail Rcall@K on flood image dataset: SF-Flood. Best overall results on each dataset are in **bold**, second-best results <u>underlined</u>.

Methods	Year	Backbone	Descriptor Dimension	SF-XL test v1		SF-XL test v2		SF-Flood			
				R@1	R@10	R@1	R@10	R@1	R@5	R@10	R@20
NetVLAD [15]	2016	VGG16	4096	40.1	57.7	76.9	91.1	4.2	8.2	12.2	16.2
AP-GeM [24]	2017	ResNet101	2048	37.9	54.1	66.4	88.6	5.5	11.0	12.8	16.5
SFRS [25]	2020	VGG16	4096	51.2	66.6	83.1	93.5	7.8	15.0	17.2	21.0
Cosplace [16]	2022	ResNet50	2048	76.6	85.5	88.8	96.8	16.5	26.8	32.5	38.5
EigenPlaces [26]	2023	ResNet50	2048	84.0	90.7	90.8	96.7	20.8	31.0	36.2	42.4
Conv-AP [27]	2022	ResNet50	4096	47.5	63.8	74.4	89.0	20.5	30.8	36.5	42.2
MixVPR [28]	2023	ResNet50	4096	72.5	80.9	88.6	95.0	35.2	47.2	52.2	56.0
ProGEO [29]	2024	ResNet50	1024	84.7	90.3	93.0	96.7	53.8	64.0	70.5	73.8
SALAD [30]	2024	DINOv2	8448	88.7	94.4	94.6	98.2	56.0	71.8	75.5	79.5
CliqueMining[31]	2024	DINOv2	8448	85.5	92.6	94.5	98.3	55.5	68.2	74.0	76.8
DINO-Mix [32]	2024	DINOv2	4096	81.7	90.7	<u>95.2</u>	<u>98.7</u>	53.2	63.7	71.0	76.2
BoQ [33]	2024	DINOv2	12288	<u>91.9</u>	<u>96.6</u>	**95.2**	**98.7**	<u>70.5</u>	<u>79.5</u>	<u>81.8</u>	<u>84.5</u>
Megaloc [34]	2025	DINOv2	8448	**95.3**	**98.0**	94.8	98.5	**76.8**	**82.8**	**86.5**	**89.0**

3.2 Benchmarking on Flood Scenarios

Table 3 compares VPR performance on two general-purpose query sets (SF-XL test v1/v2) and the flood-specific query set (SF-Flood). On the standard sets, most recent methods achieve high recall, indicating that general urban localization is already relatively mature. For example, Megaloc reaches 95.3% R@1 on SF-XL test v1, while BoQ and DINO-Mix both reach 95.2% R@1 on SF-XL test v2. Even several CNN-based methods remain competitive under these routine conditions.

The picture changes substantially on SF-Flood. Under flood-induced appearance change, conventional global-descriptor methods degrade sharply. NetVLAD drops to 4.2% R@1, while CosPlace and EigenPlaces reach only 16.5% and 20.8%, respectively. MixVPR improves to 35.2%, but still remains far below the best-performing models. These results show that methods that perform well under normal urban conditions may fail when the scene is heavily altered by water coverage, reflections, occlusions, and adverse illumination.

In contrast, methods built on DINOv2 features remain much more robust. On SF-Flood, Megaloc achieves the best performance with 76.8% R@1, 82.8% R@5, 86.5% R@10, and 89.0% R@20, followed by BoQ with 70.5% R@1 and 84.5% R@20. This margin over the strongest CNN-based competitor (53.8% R@1 for ProGEO) suggests that foundation-model descriptors are better able to preserve stable structural cues under severe flood-related appearance change.

From an operational perspective, the result of 76.8% R@1 means that the system can place the query within 50 m of the true flood location in about three out of four cases at the first prediction. In addition, the strong R@10 and R@20 results indicate that the correct location is often contained within a short candidate list, which is useful for human review when the first prediction is uncertain

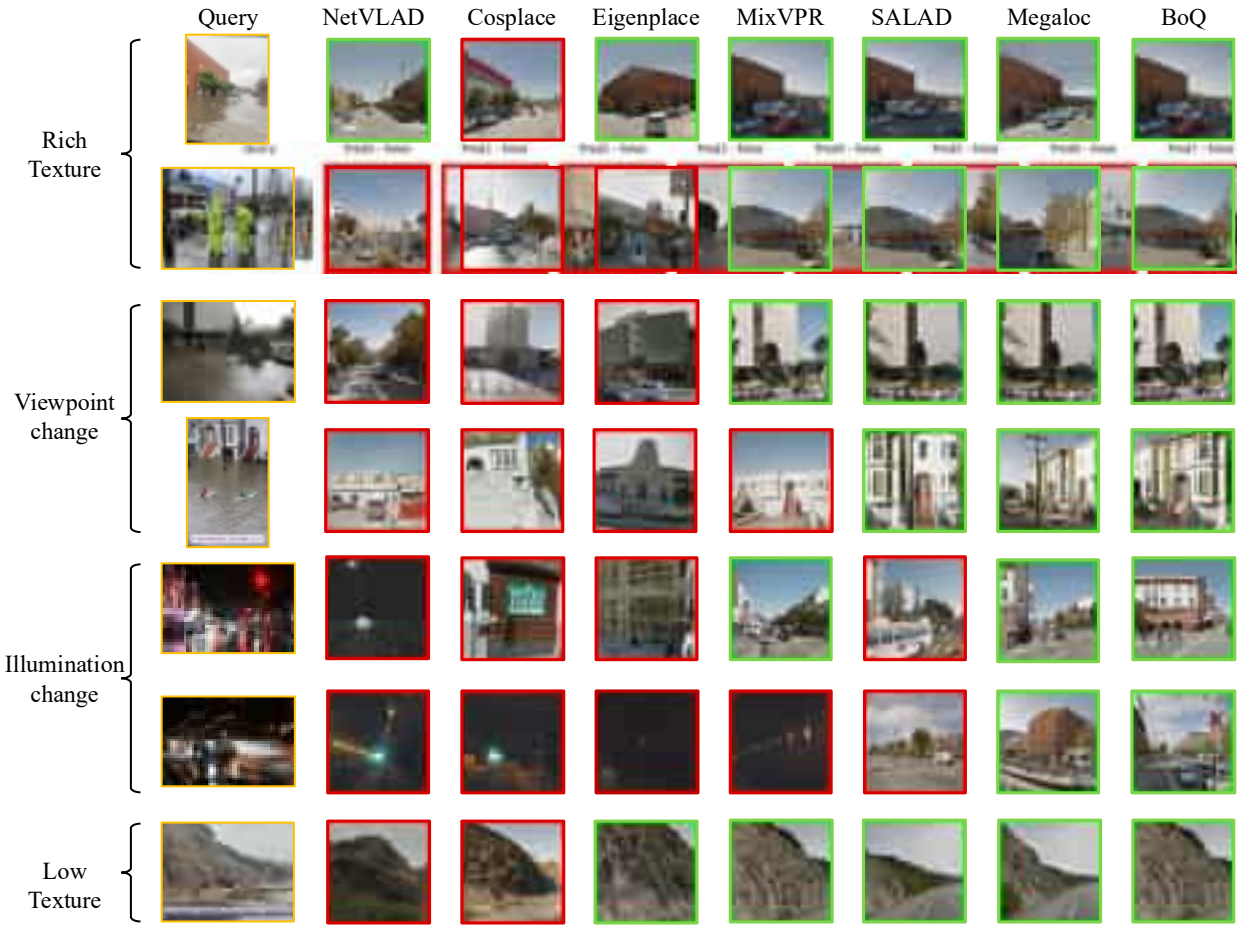


Figure 3. Qualitative retrieval comparison on the SF-Flood dataset under challenging conditions. Green borders indicate correct localizations within 50 meters of the ground truth, while red borders denote failures, illustrating representative mislocalized cases and highlighting failure modes under severe visual ambiguity.

Figure 3 provides qualitative examples under several difficult conditions, including viewpoint changes, nighttime scenes, rich-texture settings, and low-texture settings. In many of these cases, conventional methods retrieve visually similar but incorrect locations, whereas the stronger foundation-model-based methods more often recognize the right place or keep it within the top-ranked list. At the same time, the red-box examples in Figure 3 show that failure cases still remain, especially when visual ambiguity is high. Overall, the results indicate that flood localization is harder than standard urban geo-localization and that foundation-model-based descriptors provide the most reliable retrieval backbone among the methods tested.

3.3 Visual Evidence Validation

Recall metrics summarize retrieval performance at the dataset level, but practical use also requires evidence for whether a prediction is trustworthy. We therefore apply a post-retrieval geometric verification step using SIFT-LightGlue to assess the top-ranked match.

Representative verification examples are shown in Figure 4. Even under floodwater, reflections, nighttime conditions, and partial occlusion, the verified correspondences are concentrated on stable structural elements, such as façades, poles, and traffic infrastructure, rather than transient flood textures. This provides visual evidence for why some heavily degraded queries can still be localized correctly.



Figure 4. Qualitative examples of visual evidence validation.

Figure 5 further quantifies the role of visual evidence extraction. Figure 5(a) shows the localization accuracy within the 50 m threshold across different inlier ranges. All groups maintain relatively high accuracy, indicating that visual evidence is informative for match reliability. Also, the relationship is not monotonic, especially in a sparsely scattered range, which suggests that inlier count should not be interpreted as a universal hard threshold.

Figure 5(b) shows that successful predictions are not only below the 50 meter threshold, but are often concentrated at very small error distances (average 2.4 m), indicating that correct matches are typically spatially close to the ground truth. By contrast, the failed cases form a smaller group (23 cases) above the 50 m threshold and are more common when inlier support is limited. Still, because some correct cases also appear at relatively low inlier counts, the plot does not support a single clean cutoff. Instead, it supports using inlier count as a practical confidence indicator to prioritize stronger matches and flag uncertain ones for review.

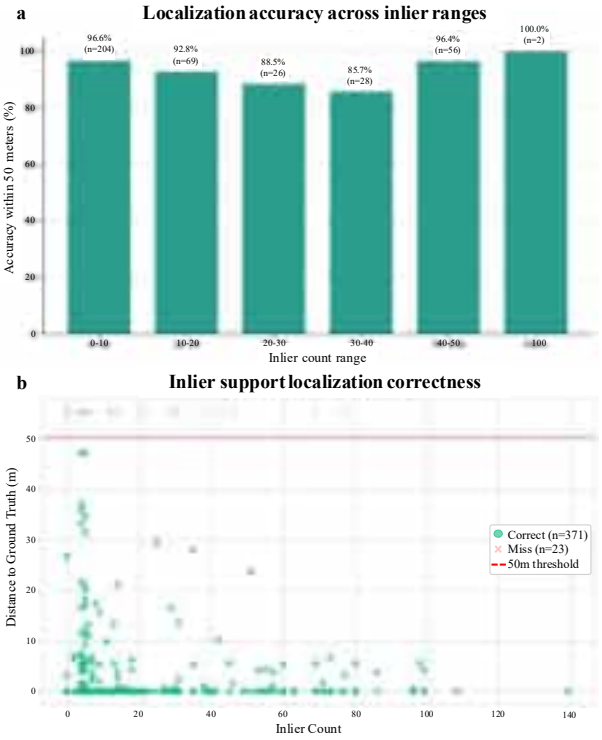


Figure 5. Quantitative analysis of inlier-support visual evidence validation.

This result is consistent with the qualitative examples in Figure 4. Visual evidence verification does not eliminate all failures, but it provides a practical way to distinguish stronger from weaker predictions. In deployment, this makes the retrieved location more interpretable and supports human-in-the-loop review when flood imagery is ambiguous.

4 Discussion and Conclusion

This study presents an end-to-end framework for geo-localizing flood-related social media images, with a focus on reliable sensing for resilient urban infrastructure

operations. Rather than proposing a new VPR backbone, the study addresses a practical question: which existing VPR paradigms remain reliable under severe flood-induced visual disruption, and how can their predictions be verified before entering human decision workflows? The results show that flood scenarios introduce a much harder setting than standard urban benchmarks, and that foundation-model-based descriptors currently provide the most reliable retrieval backbone among the methods tested. More importantly, the added verification stage makes the retrieved result interpretable by attaching explicit visual evidence to the predicted location.

The proposed framework shows how crowdsourced visual data can be transformed into infrastructure-oriented situational awareness, supporting tasks such as rapid identification of affected urban areas and awareness of access routes. In this sense, the study extends social media imagery from online content into an opportunistic sensing layer for urban infrastructure management.

There are several limitations in this study. First, the current evaluation is limited to a single-city case study in San Francisco, and broader validation across cities with different streetscapes, infrastructure conditions, and development contexts remains necessary. Second, the SF-Flood dataset contains 400 query images and, while deliberately designed to include difficult cases such as nighttime scenes, viewpoint changes, and visually degraded flood conditions, it still cannot fully capture the diversity of urban flood imagery. Finally, although the current system is suitable for time-sensitive analysis, lighter verification pipelines and more comprehensive runtime evaluation are still needed for future deployments.

Future work should therefore move in three directions. The first is broader validation, including multi-city testing and more systematic analysis of visually ambiguous urban forms. The second is deployment-oriented refinement, such as lighter matching models and latency tests. The third is multi-modal extension, where visual localization can be combined with textual, temporal, or other contextual signals from social media to improve robustness when image evidence alone is insufficient.

Acknowledgement

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (Grant No. 72404233) and the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2025A1515010190).

References

[1] A. F. Prein, R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland. The future intensification of hourly precipitation extremes. *Nature Climate Change*, 7(1):48–52, 2017.

[2] IPCC. AR6 WGII, Chapter 6: Cities, settlements and key infrastructure. IPCC. On-line: https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_Chapter06.pdf, Accessed: 15/12/2025.

[3] H. Kreibich, A. F. Van Loon, K. Schröter, P. J. Ward, M. Mazzoleni, N. Sairam, et al. The challenge of unprecedented floods and droughts in risk management. *Nature*, 608(7921):80–86, 2022.

[4] C. Mydlarz, P. Sai Venkat Challagonda, B. Steers, J. Rucker, T. Brain, B. Branco, et al. FloodNet: Low-cost ultrasonic sensors for real-time measurement of hyperlocal, street-level floods in New York City. *Water Resources Research*, 60(5):e2023WR036806, 2024.

[5] E. Portalés-Julià, G. Mateo-García, C. Purcell, and L. Gómez-Chova. Global flood extent segmentation in optical satellite images. *Scientific Reports*, 13(1):20316, 2023.

[6] M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

[7] Y. Chen, M. Hu, X. Chen, F. Wang, B. Liu, and Z. Huo. An approach of using social media data to detect the real time spatio-temporal variations of urban waterlogging. *Journal of Hydrology*, 625:130128, 2023.

[8] J. Li, R. Cai, Y. Tan, H. Zhou, A. M. Sadick, W. Shou, and X. Wang. Automatic detection of actual water depth of urban floods from social media images. *Measurement*, 216:112891, 2023.

[9] J. K. Kavota, J. R. K. Kamdjoug, and S. F. Wamba. Social media and disaster management: Case of the north and south Kivu regions in the Democratic Republic of the Congo. *International Journal of Information Management*, 52:102068, 2020.

[10] X. X. Zhu, Y. Wang, M. Kochupillai, M. Werner, M. Häberle, E. J. Hoffmann, et al. Geoinformation harvesting from social media data: A community remote sensing approach. *IEEE Geoscience and Remote Sensing Magazine*, 10(4):150–180, 2023.

[11] S. Ying, Y. Huang, L. Qian, and J. Song. Privacy paradox for location tracking in mobile social networking apps: The perspectives of behavioral reasoning and regulatory focus. *Technological Forecasting and Social Change*, 190:122412, 2023.

[12] Y. Feng, X. Huang, and M. Sester. Extraction and analysis of natural disaster-related VGI from social media: review, opportunities and challenges. *International Journal of Geographical Information*

Science, 36(7):1275–1316, 2022.

[13] S. Xiao, H. Gu, D. Shen, Z. Niu, J. Xiao, and F. Yu. A systematic review of social media-enabled flood disaster informatics: Method, technology and application. *International Journal of Applied Earth Observation and Geoinformation*, 144:104935, 2025.

[14] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford. Visual place recognition: A survey. *IEEE Transactions on Robotics*, 32(1):1–19, 2015.

[15] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5297–5307, 2016.

[16] G. Berton, C. Masone, and B. Caputo. Rethinking visual geo-localization for large-scale applications. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4878–4888, 2022.

[17] A. Torii, R. Arandjelovic, J. Sivic, M. Okutomi, and T. Pajdla. 24/7 place recognition by view synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1808–1817, 2015.

[18] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access*, 9:89644–89654, 2021.

[19] M. Zaffar, L. Nan, and J. F. Kooij. On the estimation of image-matching uncertainty in visual place recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 17743–17753, 2024.

[20] D. M. Chen, G. Baatz, K. Koser, S. S. Tsai, R. Vedantham, ¨ T. Pylvan¨ ainen, K. Roimela, X. Chen, J. Bach, M. Pollefeys, ¨ B. Girod, and R. Grzeszczuk. City-scale landmark identification on mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 737–744, 2011.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.

[22] P. Lindenberger, P. E. Sarlin, and M. Pollefeys. Lightglue: Local feature matching at light speed. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 17627–17638, 2023.

[23] D. Sferrazza, G. Berton, G. Trivigno, and C. Masone. To match or not to match: Revisiting image matching for reliable visual place recognition. In Proceedings of the Computer Vision and Pattern Recognition Conference, pages 2849–2860, 2025.

[24] A. Gordo, J. Almazan, J. Revaud, and D. Larlus. End-to-end learning of deep visual representations for image retrieval. International Journal of Computer Vision, 124(2):237–254, 2017.

[25] Y. Ge, H. Wang, F. Zhu, R. Zhao, and H. Li. Self-supervising fine-grained region similarities for large-scale image localization. In Proceedings of the European Conference on Computer Vision, pages 369–386, 2020.

[26] G. Berton, G. Trivigno, B. Caputo, and C. Masone. Eigenplaces: Training viewpoint robust models for visual place recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 11080–11090, 2023.

[27] A. Ali-bey, B. Chaib-Draa, and P. Giguere. Gsv-cities: Toward appropriate supervised visual place recognition. Neurocomputing, 513:194–203, 2022.

[28] A. Ali-Bey, B. Chaib-Draa, and P. Giguere. Mixvpr: Feature mixing for visual place recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 2998–3007, 2023.

[29] J. Hu, C. Mao, C. Tan, H. Li, H. Liu, and M. Zheng. Progeo: Generating prompts through image-text contrastive learning for visual geo-localization. In Proceedings of the International Conference on Artificial Neural Networks, pages 448–462, 2024.

[30] S. Izquierdo and J. Civera. Optimal transport aggregation for visual place recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17658–17668, 2024.

[31] S. Izquierdo and J. Civera. Close, but not there: Boosting geographic distance sensitivity in visual place recognition. In Proceedings of the European Conference on Computer Vision, pages 240–257, 2024.

[32] G. Huang, Y. Zhou, X. Hu, C. Zhang, L. Zhao, and W. Gan. DINO-Mix enhancing visual place recognition with foundational vision model and feature mixing. Scientific Reports, 14(1):22100, 2024.

[33] A. Ali-Bey, B. Chaib-draa, and P. Giguere. BoQ: A place is worth a bag of learnable queries. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17794–17803, 2024.

[34] G. Berton and C. Masone. Megaloc: One retrieval to place them all. In Proceedings of the Computer Vision and Pattern Recognition Conference, pages 2861–2867, 2025.

Mapping Digital Solutions for Productivity and Disruption Monitoring in Construction

Zhengyang Ling [1] Danny Murguia [1] Ashan Senel Asmone [1] Sam Brooks [1] Campbell Middleton [1]

[1]Department of Engineering, University of Cambridge, United Kingdom

zl461(corresponding author), dem52, asa79, sjb351, crm11@cam.ac.uk

Abstract -

Productivity and disruption monitoring remains a persistent challenge in construction projects. Although digital technologies such as Internet of Things (IoT) sensors, Building Information Modeling (BIM), computer vision, and AI-driven analytics have been introduced, practitioners often lack systematic guidance on aligning these tools with specific monitoring objectives to improve productivity and reduce waste and delays. This study first identifies key productivity measures and disruption indicators. Current digital approaches for capturing these are then clustered into six classes: sensor-based, vision-based, location-based, apps, digitised legacy documents, and audio-based systems. A review of the existing literature is conducted to map the metrics and indicators with suitable digital solutions and assess their maturity. A preliminary case study demonstrates the applicability of the proposed digital solutions in response to monitoring needs. Future work includes developing a decision-support system to select appropriate technologies tailored to monitoring requirements.

Keywords -

Digital solution; Productivity; Disruption; Monitoring

1 Introduction

The construction industry suffers from low productivity growth and persistent schedule and cost overruns that degrade project performance worldwide. Numerous studies have identified that supply chain issues, rework, and disruptions are dominant contributors to poor outcomes [1]. A recent systematic review of construction productivity found that it is shaped by interactions among organisational, technical, and managerial factors, making productivity improvement a persistent industry challenge [2].

Accurate productivity measurement is essential for identifying performance gaps and guiding improvements. Without measurement, inefficiencies remain hidden, causing resource waste, delays, and higher costs [3]. Monitoring productivity enables informed decision-making, helping managers allocate labour and equipment effectively and assess the impact of interventions [2, 4].

Disruption events directly affect construction flow and project performance [5]. Unlike delays, which are typically identified after schedule deviations occur, disruption manifests during execution as interruptions to workflow continuity, resource utilization, and productivity [4]. As such, disruption functions as an early-warning signal; systematic monitoring enables project teams to detect performance degradation before inefficiencies accumulate into delays or cost overruns [6]. However, organisations frequently struggle to establish the structural mechanisms needed to systematically capture, interpret, and act on disruption data, allowing disturbances to propagate into cascading delays [5].



Figure 1. Illustrating the research gap: stakeholders need a tool to support decision-making in selecting appropriate digital solutions for productivity and disruption monitoring across different construction sites.

Monitoring productivity and disruption plays a key role in enabling automation in construction. Real-time measurement provides the basis for identifying opportunities for automation and robotic applications to improve productivity and mitigate disruptions. Building on these data-driven insights, digital solutions can support automated execution and adaptive construction processes. Linking

measurement, decision-making, and automation can significantly enhance project performance.

However, digital monitoring of productivity and disruption faces challenges. Construction firms vary widely in digital maturity, making it difficult to align solutions with capabilities and goals [7]. Unstructured monitoring requirements complicate translation into actionable designs [8]. With the proliferation of ConTech solutions, adoption is often siloed and ad-hoc, lacking systematic evaluation and integration, which limits their ability to generate actionable insights [9]. Therefore, this study addresses the question: How can organisations systematically integrate monitoring requirements with digital solutions to generate data-driven insights for productivity improvement? It proposes a structured approach to help construction companies efficiently select suitable digital solutions, as shown in Fig. 1.

2 Literature Review

2.1 Construction Productivity

Construction productivity refers to the efficiency with which inputs (labour, materials, equipment) are transformed into outputs, commonly expressed as output-to-input ratio and measured as work quantity per unit of labour, time, or cost [10]. Beyond this quantitative ratio, productivity is increasingly understood as dynamic and context-dependent, influenced by work conditions, management practices, and disruptions. Empirical studies have shown that productivity fluctuates in response to changes, interruptions, and cumulative disruption impacts [11]. Contemporary research defines construction productivity as the ability to achieve planned output efficiently whilst minimising productivity losses from disruptions and waste [2, 12]. This definition emphasises real-time measurement of outputs and inputs to accurately reflect execution efficiency under site conditions and inform production planning and control. However, existing research largely focuses on vision-based approaches (i.e., outputs)[13], with limited attention to how other digital technologies can be selected and aligned to simultaneously measure inputs.

2.2 Construction Disruption

Disruption is conceptualised as an unplanned event or condition that adversely affects work execution, resulting in reduced productivity, increased costs, and potential schedule impacts [6]. Disruption is conceptually distinct from delay and waste. Delay refers to time overrun relative to the planned schedule [1]. Whilst disruptions can cause delays, productivity losses may occur without immediately affecting critical path activities. Waste refers to non-value-adding activities consuming resources without contributing to progress [6]. Both delay and waste are often consequences rather than disruptions themselves [2]. As such, a timely detection of disruptions can explain productivity variability and provide site managers with additional insights into performance. Crucially, this is relevant to multi-tiered construction, where agents in the upper tiers (i.e., general contractors) have less visibility into data and outputs from lower tiers (i.e., trade contractors). Disruption is thus a multidimensional performance anomaly manifested through adverse impacts on schedule, productivity, and quality. By identifying and addressing disruptions early, teams can proactively prevent downstream effects. However, disruption monitoring has been less widely addressed than productivity monitoring, partly due to a lack of systematic definitions of disruption causes and associated digital measurement solutions.

2.3 Digital Solutions for Construction Monitoring

Traditional manual monitoring is fragmented and reactive, hindering rapid detection of productivity changes and disruptions. Digital solutions provide real-time, objective, and automated observation via sensors, cameras, and mobile applications. Vision-based approaches using laser scanning and computer vision automate progress monitoring by comparing as-built conditions with planned or BIM-based models [14]. Advances in object recognition and multi-object tracking enable detailed activity-level assessments [13, 15], providing a basis for the project status information digital twin construction [16]. Real-time locating systems (RTLS) using UWB and BLE monitor movement and interaction of workers, equipment, and materials, enabling quantitative analysis of workflow, space utilisation, and non-productive time [17]. Such approaches provide critical inputs for understanding disruption mechanisms related to congestion, waiting, and out-of-sequence work. In addition, mobile devices facilitate on-site data collection and decision support [18]. IoT frameworks enable the integration of heterogeneous data sources for real-time project performance analysis [19]. Audio-based monitoring has emerged as a viable alternative when visual sensing is constrained. Research has investigated the use of ambient sounds and machine learning to identify activities, equipment operation, and safety-critical events [20]. Despite the wide range of options for capturing real-time data, selecting suitable digital solutions tailored to specific requirements remains challenging for stakeholders lacking sufficient technology knowledge.

As the industry moves toward real-time monitoring of productivity and disruptions, the lack of unified disruption indicators creates uncertainty in selecting appropriate digital solutions [8, 13]. In this study, disruption indicators are defined as quantifiable measures captured and analysed

in real-time to support data-driven decision-making, and digital solutions as purpose-built systems integrating one or more technologies for tasks such as data collection, processing, storage, and performance visualisation. Despite extensive research on individual technologies, no framework systematically links productivity measures and disruption indicators to multiple digital solutions; this study addresses that gap.

3 Methodology

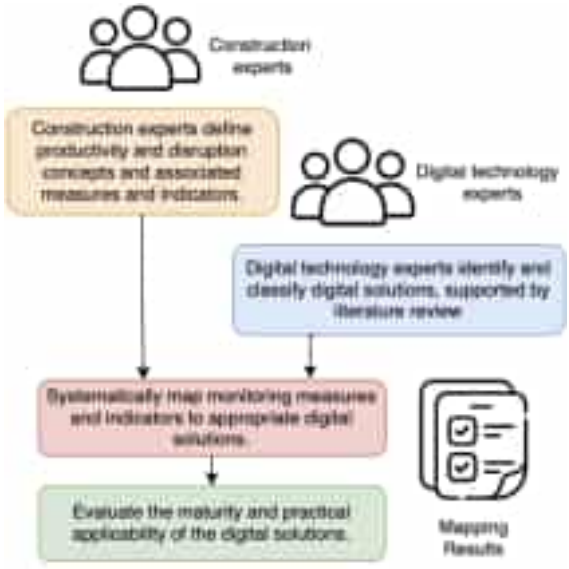


Figure 2. Overview of the proposed methodology.

The overall methodology comprises four key steps, integrating inputs from construction experts and digital technology experts, as illustrated in Fig. 2.

1. *Definition and identification of productivity measures and disruption indicators:* This step involves construction experts defining productivity and disruption concepts, along with associated measures and indicators, supported by evidence from multiple case studies in UK-based projects [4]. To explain variability in production and productivity, it was necessary to standardise data collection on blockers to productivity, which are here defined as disruptions. An initial list of disruptions was derived from Rathnayake and Middleton [2], piloted across three projects, and refined through multiple cycles of industry engagement. Three construction engineering experts were involved in this process.

2. *Identification and classification of digital solutions for data collection:* This step involves digital technology experts identifying and classifying available digital solutions for construction monitoring, supported by a literature review of major construction journals. The literature search was conducted using databases including Scopus, Web of Science, and Google Scholar, with keywords such as "construction monitoring", "productivity monitoring", "disruption", "digital construction", and "computer vision". Approximately 100 key papers were identified through database searches and subsequently selected based on relevance through expert screening (see Section 2.3). These technologies include those capable of capturing visual, spatial, acoustic, and other physical data. Based on the type of data collected and their applicability to monitoring productivity and disruptions, the solutions were grouped into categories to maintain a manageable level of abstraction. This classification enables users to first identify suitable solution types and then refine application details.

3. *Mapping productivity measures and disruption indicators to digital solutions:* This step systematically maps productivity measures and disruption indicators to the most appropriate digital solution types for monitoring. The mapping was conducted through close collaboration between construction professionals and digital technology experts, ensuring alignment between monitoring objectives and technological capabilities. The resulting mapping provides a structured framework for evaluating the suitability of digital solutions for specific monitoring goals.

4. *Evaluation of maturity and practical applicability of digital solutions:* The final step evaluates the maturity and practical applicability of the identified digital solutions. This includes assessing technological readiness, accuracy, scalability, cost, and operational feasibility. Understanding the strengths and limitations of each solution type provides a foundation for selecting appropriate monitoring strategies and supports the effective deployment of digital solutions in practice.

Overall, this study integrates knowledge from construction and digital technology experts to bridge the gap between the requirements for real-time monitoring of productivity and disruptions and the widespread availability of digital solutions that are often underutilised or ineffectively selected in construction projects. It provides a foundation for developing a structured approach that links monitoring requirements to the development and deployment of digital solutions in construction.

Table 1. Types of digital solutions for productivity and disruption monitoring.

Solution Type	Description	Examples
Sensor-based Monitoring Systems	Use of IoT-enabled sensors to collect real-time data related to equipment condition, environmental parameters, and worker activities.	Equipment condition monitoring; process monitoring; energy usage monitoring
Vision-based Monitoring Systems	Use of cameras or laser scanners to observe site activities, detect events, and track construction progress.	Progress monitoring; working conditions monitoring; automated quality inspection
Location-based Tracking Systems	Technologies such as GPS, RFID, or BLE used for real-time tracking of workers, equipment, and materials on construction sites.	Inventory tracking and status updates; tool and equipment tracking and assignment
Web-based / Mobile Applications	Web- or mobile-accessible user interfaces for data entry, storage, analysis, and visualisation.	Visualisation of financial KPIs; sales tracking; progress payment tracking; integrated change management and issue reporting between design and construction
Digitised Legacy Documents	Conversion of legacy documents (e.g., handwritten notes, PDFs, word documents, spreadsheets) into structured digital formats.	Digitalisation of printed labour task allocation sheets into structured digital records
Audio-based Monitoring Systems	Use of microphones and acoustic analysis techniques to capture and interpret sound patterns from machines, environments, or workflows for monitoring and anomaly detection.	Detection of abnormal machine noises; noise-level compliance monitoring; identification of process issues through sound patterns; recording managerial instructions or voice logs

4 Results

4.1 Productivity Measures

Three productivity metrics were defined: production rates (output/duration) and labour productivity (output/worker-hour), and cost-effectiveness (output/cost) for both planned and actual values. The output is measured as progress, which refers to the amount of work completed over time, such as the number of square meters constructed each day [21, 15]. Inputs are measured using labour data, which can be monitored by recording labour hours or workforce utilisation on each activity or work package [22]. Cost can be tracked by estimating expenditures and resource consumption over time [10, 6].

Two additional measures are derived from the programme and BIM models. The programme provides the planned duration for activities and work packages per location, while BIM models provide the planned quantities over time, and serves as a baseline for monitoring progress and productivity [23]. Together, these measures enable real-time monitoring of productivity metrics and allow comparison with planned values.

4.2 Disruption Indicators

Compared to productivity, construction disruption lacks consistent measurement metrics. Unlike traditional disruption measures used in claims analysis or post-event evaluation [24], our indicators are designed to support continuous monitoring through digital solutions, enabling early detection of emerging disruptions.

The resulting indicator set is organized into ten major categories: labour, materials, equipment, planning, logistics, health and safety, design and constructability, quality, weather, and external factors. These indicators span multiple dimensions of project performance. Labour-related indicators capture workforce availability and efficiency losses [22], while material and equipment indicators reflect interruptions in resource flow [10]. Planning and logistics indicators represent workflow reliability and spatial constraints [21], and design and quality indicators capture upstream information gaps and rework-related disturbances [6]. Weather and external indicators account for exogenous shocks affecting site operations [1]. This multidimensional structure supports a system-level understanding of disruption as a dynamic production anomaly rather than a single, isolated event.

We further divided them into 33 monitoring indicators, as shown in Section 4.5. These indicators specify information that enables digital solutions to directly detect and monitor events during execution, as detailed in the subsequent mapping sections.

4.3 Classification of Digital Solutions

In Section 2.3, we reviewed digital solutions for construction monitoring and further classified these techniques into six solution categories, as shown in Table 1. These solution types include sensor-based monitoring systems, vision-based systems, location-based tracking systems, web-based and mobile applications, digitised legacy documents [25], and audio-based monitoring systems. A

Table 2. Mapping of productivity measures to digital solution types (H = High maturity, M = Medium maturity, L = Low maturity).

Productivity Measure	Sensor	Vision	Location	Apps	Digitised Documents	Audio
Progress	M	H		H	M	L
Labour	H	M	M	H	M	L
Cost				H		
Planned duration (from programme)				H		
Planned quantity (from BIM model)				H		

detailed description of each solution type is provided in Table 1. This classification also builds on the work of Yilmaz et al. [26], who identified 58 digital solution areas for construction SMEs through eight interactive workshops with 70 UK SMEs. By organising these solutions into the six categories presented here, we cover most of the 58 solution areas identified by Yilmaz et al. [26], some of which are provided as examples in Table 1.

To further evaluate the applicability of these digital solutions, we define three maturity levels as follows: High maturity technologies are well-established, commercially available, and easily integrated into existing workflows. Medium maturity technologies are technically feasible and validated in pilot studies but require configuration, integration, or specialist expertise and are not yet widely adopted. Low maturity technologies are experimental, primarily at the prototype or conceptual stage, and require further development and validation before reliable application in construction projects.

4.4 Mapping of Digital Solution to Productivity Monitoring

Table 2 presents the mapping of construction productivity measures to different categories of digital solutions, along with their corresponding maturity level. Progress data are most effectively captured using vision-based and application-based solutions (H), whereas audio-based monitoring remains at a low maturity level (L). However, voice logging is an emerging area of development. Labour data can be monitored efficiently with sensor- and application-based systems (H), while location- and vision-based systems provide medium-level support. Cost, programme, and BIM model data are primarily supported by high-maturity, application-based solutions, reflecting the widespread use of integrated project management software in industry practice.

4.5 Mapping of Digital Solution to Disruption Monitoring

Table 3 presents a detailed mapping of 33 disruption indicators to appropriate digital solution types, classified by maturity levels. Compared to productivity measures, disruption indicators capture deviations in workflow execution, requiring targeted monitoring approaches for early detection.

Software applications represent high-maturity solutions widely used in construction, from which most disruption indicators can be derived. However, their reliance on manual input may delay early warnings, and AI-driven analysis is needed to extract critical information from large datasets. Sensor-based systems are mature for monitoring equipment, materials, and weather-related disruptions, while wearable sensors such as IMUs are emerging for labour activity monitoring. Tracking technologies are well established for logistics disruptions and show further potential in material, equipment, and safety tracking (e.g., detecting workers in hazardous zones). Vision-based systems are widely applied to site monitoring but depend heavily on advanced AI and are sensitive to occlusion and lighting conditions; their role in disruption monitoring is expected to grow as reliability improves. Application-based solutions provide broad coverage but are limited by data latency from manual input, while digitisation and AI-based document analysis can unlock valuable information from legacy data. Audio-based monitoring is less common and primarily used for equipment anomaly detection; however, advances in voice logging may allow abnormal work patterns to be recorded efficiently.

5 Preliminary Study

A preliminary study was conducted with a construction company on a recent project. Productivity had previously been recorded manually and calculated monthly or over longer periods using spreadsheets. Although the company sought more real-time monitoring, it lacked awareness of suitable digital solutions, making technology selection difficult. Using the mapping table developed in this study, the company's monitoring needs were systematically linked to relevant digital solution types, enabling a more structured and objective selection process.

After further criteria-based discussions, several solutions were selected and implemented. Two CCTV systems were deployed to capture image data, and progress information was extracted using an advanced segmentation approach (SAM2 [27]). This selection was guided by the mapping framework, which identified vision-based so-

Table 3. Mapping of disruption indicators to digital solution types (H = High maturity, M = Medium maturity, L = Low maturity).

Category	ID	Disruption Indicator	Sensor	Vision	Location	Apps	Digitised Docs	Audio
Labour	1	Worker attendance variance			M	H	M	M
	2	Labour idle time	L	L		M		L
	3	Reduced working speed	L	M		M		L
	4	Labour output rate below plan		M		H		L
Materials	5	On-site material availability	H	L	M	H	M	
	6	Material delivery delay	M		M	H	M	
Equipment	7	Equipment availability rate	H	M	M	H	H	M
	8	Equipment downtime	H	M		H		M
	9	Equipment maintenance duration	H	M		H		M
Planning	10	Out-of-sequence work		M	M	M	H	
	11	Trade handover delay		M	H	M	H	
	12	Work area access delay		M	H	M	H	
Logistics	13	Crane utilisation conflict	H	M		H		
	14	Loading/storage area congestion		M	H	M	H	
	15	Material transport time		M	H	M	H	
	16	Material unavailability at work-face		M	H	M		
	17	Worker travel time to welfare			H	M		
Health & Safety	18	Safety induction time				H		
	19	Safety non-compliance events		M	M	H	H	
	20	Safety incidents / near misses		L		H		
	21	Safety stand-down duration		M		H		
Design & Constructability	22	Missing design information				H	M	
	23	Design errors / clarifications				H	M	
	24	Change order frequency				H	M	
	25	Design clash occurrence				H	M	
Quality	26	Quality non-conformance		M		H		
	27	Rework duration		M	M	H		L
	28	In-process corrective work		M		H		L
Weather	29	High wind events	H			H		M
	30	Heavy rainfall events	H			H		M
	31	Extreme temperature exposure	H			H		M
	32	Snow / ice conditions	H	M		H		M
External	33	External operational interruptions				H	H	

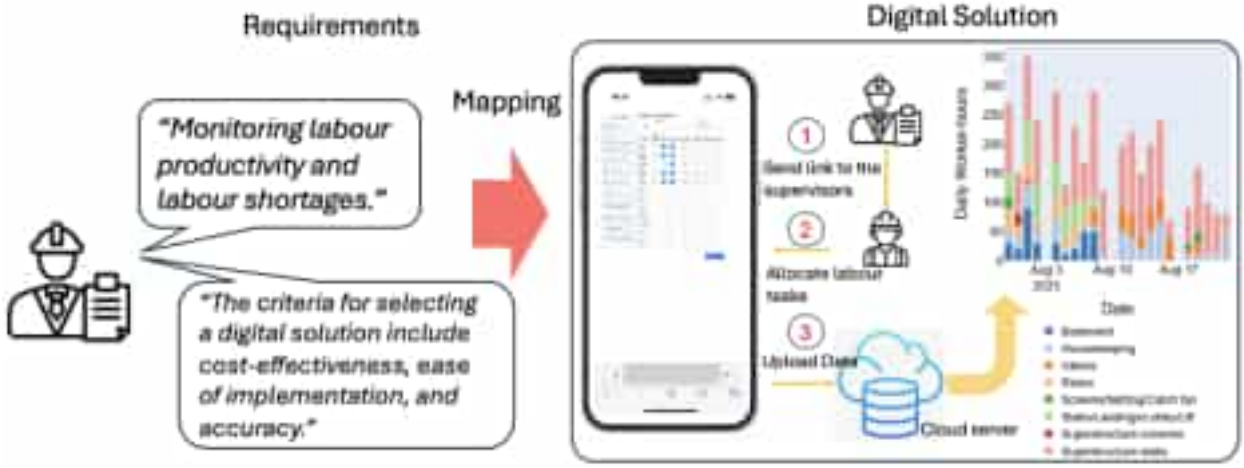


Figure 3. Case study of labour productivity and disruption monitoring.

lutions as suitable for progress monitoring due to their high maturity and ability to capture output-related data. Due to space constraints, detailed methodology and results will be reported in a future publication.

Labour data was initially captured through the turnstile system, which only records site entry and exit and does not capture task-level activity information for labour productivity measurement. A web application was developed and deployed on a cloud server (see Fig. 3). Each working day, site managers assigned activities to construction crews, while supervisors used the system to allocate tasks to workers. The application allowed supervisors to monitor labour input for each activity and to send messages reporting labour shortages for specific activities or task re-allocations to managers. This approach addressed a key disruption indicator, labour shortage, by enabling daily visibility of workforce allocation and deviations from planned activities.

The web application has been in use on-site for over two months. Data collected through the system has been successfully used to monitor daily labour productivity. Alerts can be generated by supervisors to identify potential disruptions caused by labour shortages, enabling site managers to reallocate labour accordingly. During the preliminary study, positive feedback from the company confirmed that the initial requirements of productivity monitoring had been met. Furthermore, the use case demonstrates how the proposed mapping framework supports construction managers in selecting and deploying digital solutions by linking monitoring requirements with appropriate technologies. The daily collected data also enabled timely decision-making, such as reallocating labour tasks.

However, disruption measurement of labour shortage using the web application relies on manual input and does not support fully real-time or automated monitoring. Location-based tracking systems may enable more automated and real-time disruption detection in the future, although their maturity remains limited. The mapping also highlights potential research directions for industry implementation.

6 Discussion and Conclusion

This paper proposes a structured mapping approach of digital solutions for productivity measurement. To achieve this, we first identified the key productivity measures and disruption indicators. The various digital solutions were then clustered into six groups to facilitate straightforward mapping. Based on this framework, two mapping tables were created, and the maturity of the available digital solutions was evaluated. A preliminary study was conducted to gather user requirements, apply the mapping tables, and select and deploy the most suitable solution. This approach provides a practical guideline for construction companies to efficiently deploy digital solutions for monitoring productivity and disruptions.

However, the mapping tables require further refinement through additional pilot studies and by linking to a broader range of off-the-shelf solutions. Further work is also needed to deepen disruption-focused analysis and validate the effectiveness of these solutions through additional case studies. The selection criteria can be enhanced by incorporating additional factors to support improved decision-making.

Finally, the proposed framework can be developed into a decision support tool that leverages the mapping tables and selection criteria to make solution selection more objective and automated. By enabling the systematic selection of sensing and monitoring solutions, the framework supports reliable data collection and digitalisation, providing a foundation for construction automation, including data-driven decision-making and robotic applications.

Acknowledgments

We would like to thank Trimble for funding this research and our industry partners for their continued engagement.

References

[1] S. A. Assaf and S. Al-Hejji. Causes of delay in large construction projects. *International journal of project management*, 24(4):349–357, 2006.

[2] A. Rathnayake and C. Middleton. Systematic review of the literature on construction productivity. *Journal of Construction Engineering and Management*, 149 (6):03123005, 2023.

[3] A. L. Huang, R. E. Chapman, and D. T. Butry. Metrics and tools for measuring construction productivity: Technical and empirical considerations. 2009.

[4] D. Murguia, A. Rathnayake, T. Jansen van Vuuren, and C. Middleton. Measuring construction productivity across projects: multilevel three-dimensional framework. *Journal of Construction Engineering and Management*, 150(11):04024151, 2024.

[5] R. Sacks, O. Seppänen, V. Priven, and J. Savosnick. Construction flow index: a metric of production flow quality in construction. *Construction management and economics*, 35(1-2):45–63, 2017.

[6] S. Besklubova and X. Zhang. Improving construction productivity by integrating the lean concept and the clancey heuristic model. *Sustainability*, 11(17): 4535, 2019.

[7] J. Whyte. How digital information transforms project delivery models. *Project management journal*, 50 (2):177–194, 2019.

[8] Z. Ling, D. McFarlane, S. Brooks, L. de Silva, G. Hawkridge, and A. Thorne. Automated configuration for cost-effective digital solutions. *Computers in Industry*, 173:104397, 2025.

[9] D. Murguia, R. Soetanto, M. Szczygiel, C. I. Goodier, and A. Kavuri. Construction 4.0 implementation for performance improvement: an innovation management perspective. *Construction Innovation*, 2024.

[10] B. Hwang, S. R. Thomas, C. T. Haas, and C. H. Caldas. Measuring the impact of rework on construction cost performance. *Journal of construction engineering and management*, 135(3):187–198, 2009.

[11] M. Arashpour, R. Wakefield, N. Blismas, and E.W.M. Lee. Analysis of disruptions caused by construction field rework on productivity in residential projects. *Journal of construction engineering and management*, 140(2):04013053, 2014.

[12] C. Görsch, O. Seppänen, A. Peltokorpi, and R. Lavikka. Unlocking productivity: revealing waste and hidden disturbances impacting mep workers. *Journal of Construction Engineering and Management*, 150(9):04024108, 2024.

[13] B. Ekanayake, J. K. Wong, A. A. Fini, and P. Smith. Computer vision-based interior construction progress monitoring: A literature review and future research directions. *Automation in construction*, 127:103705, 2021.

[14] C. Zhang and D. Arditi. Automated progress control using laser scanning technology. *Automation in construction*, 36:108–116, 2013.

[15] A. Pal, J. Lin, S. Hsieh, and M. Golparvar-Fard. Automated vision-based construction progress monitoring in built environment through digital twin. *Developments in the Built Environment*, 16:100247, 2023.

[16] R. Sacks, I. Brilakis, E. Pikas, and M Girolami. Digital twin construction. *Data-Centric Engineering*, 1: 1–26, 2020.

[17] N. Soltanmohammadlou, S. Sadeghi, C. Hon, and F. Mokhtarpour-Khanghah. Real-time locating systems and safety in construction sites: A literature review. *Safety science*, 117:229–242, 2019.

[18] S. O. Bonney, J. E. Yankah, K. O. Adjei, and C. K. Tieru. Software and mobile apps as a strategy for productivity improvement in the construction industry. *Journal of Building Construction and Planning Research*, 12(1):1–35, 2024.

[19] T. S. Mohamed and S. Aydin. Iot-based intrusion detection systems: a review. *Smart Science*, 10(4): 265–282, 2022.

[20] I. Kim, Y. Kim, and S. Chin. Sound localization framework for construction site monitoring. *Applied Sciences*, 12(21):10783, 2022.

[21] Z. Pučko, N. Šuman, and D. Rebolj. Automated continuous construction progress monitoring using multiple workplace real time 3d scans. *Advanced Engineering Informatics*, 38:27–40, 2018.

[22] A. M. Al-Janabi, M. S. Abdel-Monem, and K. M. El-Dash. Factors causing rework and their impact on projects' performance in egypt. *Journal of civil engineering and management*, 26(7):666–689, 2020.

[23] F. Mostofi, O. B. Tokdemir, and V. Toğan. A decision-support productive resource recommendation system for enhanced construction project management. *Advanced Engineering Informatics*, 62: 102793, 2024.

[24] M. Cevikbas, O. Okudan, and Z. Işık. Identification and assessment of disruption claim management risks in construction projects: a life cycle-based approach. *Engineering, Construction and Architectural Management*, 31(1):1–27, 2024.

[25] Z. Ling, D. Murguia, A.S. Asmone, and C. Middleton. Automated task-based labour allocation extraction from scanned tables to estimate productivity. In *IET Conference Proceedings LODISA*, volume 2025, pages 147–153. IET, 2025.

[26] G. Yilmaz, L. Salter, D. McFarlane, and B. Schönfuß. Low-cost (shoestring) digital solution areas for enabling digitalisation in construction smes. *Computers in Industry*, 150:103941, 2023.

[27] Y.-T. Hu R. Hu C. Ryali T. Ma H. Khedr R. Rädle C. Rolland L. Gustafson et al. N. Ravi, V. Gabeur. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.

BIM-Integrated UAV Workflow for Automatic Facade Temperature Field Reconstruction Using IR Imagery and Component-Level Thermal Analysis

Enrique Aldao[1], Gálata Martínez-Alonso[1], Gabriel Fontenla-Carrera[1], Higinio González-Jorge[1]

[1]Aerolab, Instituto de Física Computatión e Ciencias Aeroespaciais (IFCAE), Universidade de Vigo, Ourense, 32004, Galicia, Spain

enrique.aldao.pensado@uvigo.gal, galata.martinez@uvigo.gal gabriel.fontenla@uvigo.gal higiniog@uvigo.gal

Abstract -

Infrared thermography is a widely adopted non-destructive technique for infrastructure inspection, enabling the detection of defects such as thermal bridges, insulation deficiencies, moisture accumulation, and material degradation. In recent years, the use of this technique alongside Unmanned Aerial Vehicles (UAVs) has attracted growing interest due to their ability to access hard-to-reach areas such as facades and bridges. This work presents an automatised workflow for the efficient acquisition of facade temperature fields using UAVs. RGB imagery is employed to perform Structure-from-Motion and camera pose estimation, which is then used to reconstruct radiance fields of the facade surface. Radiometric calibration is performed using reference temperature probes, allowing the reconstruction of absolute temperature fields from raw infrared imagery. The proposed method enables the direct integration of calibrated thermal fields into Building Information Models (BIM), providing per-component thermal information for detailed inspection and data analysis. A practical case study on a multi-material facade demonstrates the workflow, highlighting its potential to enhance preventive maintenance and support energy-efficiency evaluations.

Keywords -

IR Imagery; Perspective-n-Point (PnP); Building Information Model (BIM); Unmanned Aerial Vehicle (UAV)

1 Introduction

Infrared thermography has become a widely used non-destructive technique for the condition assessment of building facades, as it enables the detection of surface-temperature anomalies without physical contact [1, 2]. By measuring the spatial distribution of thermal radiation emitted by facade materials, the method can reveal a wide range of defects that may go unnoticed during visual inspections, including thermal bridges, insulation deficiencies, moisture accumulation, air leakage, and areas of material degradation [3, 4]. These anomalies are closely related to energy efficiency and structural integrity, making thermal imaging an effective tool for the preventive maintenance of built infrastructure [5].

Traditionally, facade inspections have been carried out by trained personnel using cranes, scaffolding, or specialised elevated platforms to access different areas of the structure and collect imagery. However, these methods are not only costly and time-consuming but also expose workers to significant risks associated with working at height [6]. To overcome these limitations, the use of Unmanned Aerial Vehicles (UAVs) has increased significantly in recent years. These platforms can be equipped with infrared cameras and complementary sensors, enabling the rapid, and efficient acquisition of data in hard-to-reach locations [7, 8, 9]. Furthermore, through the use of photogrammetric techniques, UAV imagery can be processed to generate high-resolution orthomosaics and detailed 3D reconstructions, thereby facilitating the integration of thermal and visual data within Building Information Models (BIM).

Although photogrammetry is well established for visual imagery, its direct application to infrared data remains challenging due to the lower spatial resolution of thermal cameras and the difficulty of identifying Ground Control Points (GCPs) in thermal images [10, 11]. To address this issue, several studies employ targets made of materials with contrasting emissivity or reflectivity as GCPs, which are easily detectable in infrared imagery. These targets are typically combined with high-accuracy positioning systems such as RTK-GNSS or terrestrial laser scanning to obtain precise reference coordinates [12]. While effective, this approach can be impractical for facade inspection because placing and surveying the targets at significant heights is often difficult and unsafe.

Conversely, other methods reconstruct the scene by combining RGB and infrared imagery. The RGB images are used to detect reference features such as corners or edges, which are then associated with their corresponding 3D coordinates [13, 14]. A Perspective-n-Point (PnP) approach is employed to estimate the camera poses, and the known relative transformation between the RGB and infrared cameras is used to project thermal textures onto

the 3D model. Commercial software packages such as Agisoft Metashape and Pix4Dmapper currently support this workflow [15, 16]. However, their integration with BIM environments remains limited. Although they produce 3D reconstructions with mapped thermal textures, these results do not allow for the direct analysis of individual facade components. In addition, the thermal textures correspond to raw radiance values measured by the sensor, which require radiometric calibration to recover the actual temperature field. In light of these limitations, this work presents an IR inspection pipeline for reconstructing facade temperature fields using UAVs. The main contributions are as follows:

1. **Radiometric calibration:** Using a temperature probe and a set of radiometric control points, the proposed method reconstructs absolute temperature fields from raw infrared imagery.

2. **Direct integration with BIM:** The workflow enables the direct incorporation of calibrated temperature fields into BIM models, providing per-component thermal information that facilitates detailed inspection and data analysis.

The remainder of this article is structured as follows. Section 2 outlines the methodology used to reconstruct the facade temperature field, Section 3 presents and analyses the results, and Section 4 summarises the main findings.

2 Methodology

Figure 1 presents the workflow diagram of this work. First, experimental data are acquired, including UAV imagery, ground control point (GCP) data, and thermal reference point data of the building facade (Section 2.1). Next, this information is processed to perform structure-from-motion reconstruction and compute all camera poses (Section 2.2). Finally, the thermal field is calculated and integrated into the BIM reference frame (Section 2.3).



Figure 1. Thermal field reconstruction workflow.

2.1 Experimental data acquisition

As a practical case study, the facade shown in Figure 2 is selected for the IR inspection. It is a flat wall with different surface finishes and materials, as well as windows and a roof with varying reflectivity properties, which makes it particularly suitable for IR inspections. The facility is located in Ourense and belongs to Misturas SA, a company with which the University of Vigo collaborates through the RADIANCE project to develop intelligent robotic solutions in construction.



Figure 2. Facade chosen for the study.

To relate the camera poses to the BIM model, the positions of the Ground Control Points (GCPs) were first surveyed using a Leica DISTO X6 laser rangefinder [17] mounted on a Leica DST 360-X station. The DST 360-X includes a calibrated inclinometer and a horizontal encoder, allowing the direct measurement of 3D points with centimetre-level accuracy. Using this setup, the 3D coordinates of all window corners were recorded, providing a total of 20 GCPs distributed across the facade surface.

For the RGB and thermal data acquisition procedure, a UAV-based approach was chosen for practicality given the height of the building, for which a DJI Matrice 300 was employed. This UAV is equipped with a Zenmuse H20T hybrid camera system, which enables the synchronised acquisition of both RGB and infrared (IR) imagery. Table 1 summarises the main specifications of this imaging system. As shown, the visible-light sensor provides substantially higher resolution and a wider field of view, which facilitates the identification of GCPs and improves the accuracy of the camera pose estimation.

Table 1. Zenmuse H20T hybrid camera specifications according to the manufacturer [18].

RGB Wide Camera	
Sensor	1/2.3” CMOS, 12 MP
Resolution	4056×3040
Spectral Band	Visible spectrum
DFOV	82.9°
Focal Length	4.5 mm (24 mm eq.)
Aperture	f/2.8
Thermal IR Camera	
Sensor	Uncooled VOx Microbolometer
Resolution	640×512 (16 bit, R-JPEG)
Spectral Band	8–14 μm
DFOV	40.6°
Focal Length	13.5 mm (58 mm eq.)
Aperture	f/1.0
Temperature Range	-40°C–150°C (HG)

With this set-up, a survey of the facade was conducted. Photogrammetry on vertical structures presents several

challenges for path planning, particularly due to the absence of a reference plane for defining parallel trajectories. For this reason, the survey was performed manually by following a zigzag flight pattern at a stand-off distance of 5 m from the building. A total of 1,300 images were captured at 0.5 m intervals to ensure sufficiently high overlap for computing the Structure from Motion (SfM) of the scene. Image acquisition was synchronised, allowing the IR and RGB images to be correlated, and their timestamps were recorded in a .txt file.

2.2 Camera Pose estimation

For the camera pose estimation, Agisoft Metashape software is used. First, Ground Control Points (GCPs) are manually annotated within the images, and their corresponding 3D coordinates, measured independently, are input into the software. Using this information, the Structure from Motion (SfM) of the images is computed. The software applies feature extraction algorithms to detect distinctive points in the images, which are then matched and correlated across overlapping photos to compute the 3D camera poses (position and orientation).

All data are expressed in the reference frame of the DISTO X6 rangefinder, whose origin and orientation are defined by the base reference station. To integrate this information within the BIM model, the 3D point cloud derived from Metashape is used to compute the coordinate transformation. First, a coarse registration is performed by aligning corresponding point pairs from the BIM geometry and the point cloud. Reference points, such as corners and window locations, are manually selected in both geometries, and a rigid transformation is computed to align the two datasets. To further improve the registration, the Iterative Closest Point (ICP) algorithm is applied to locally refine the alignment and maximise the overlap between the two geometries. Once these transformations are computed, the camera poses in the BIM reference frame are obtained, as illustrated in Figure 3. Note that the displayed geometry corresponds to the BIM 3D model rather than the photogrammetry point cloud; the point clouds are used solely to align the camera poses within the BIM reference frame.

To verify the reconstruction accuracy, the Root Mean Square Error (RMSE) in the fitting of GCPs was measured in Metashape, yielding a value of 2.8 cm. Overall, all camera poses could be computed with high precision, which minimises projection errors in the IR imagery. As both cameras are mounted in a rigid dual-camera setup, the resulting pixel errors in the IR projections are expected to be on the same order of magnitude, which is highly accurate given the resolution of the thermal imagery. This result was achieved thanks to the precise measurement of GCPs with a laser system and the presence of textured, recognisable patterns on the building facade, which facilitated the SfM calculation.

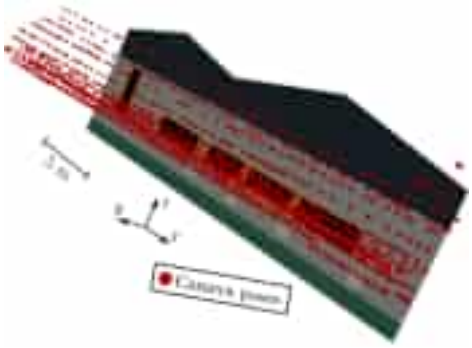


Figure 3. 3D camera-pose estimation using Agisoft Metashape. Red dots indicate the camera poses next to the BIM geometry.

Although this approach requires photogrammetry post-processing, its main advantage is the high accuracy of camera pose estimation, which minimises projection errors in the reconstructed temperature field. Furthermore, because the sensors are synchronised, the pose of the IR camera can be computed automatically from the RGB camera, streamlining the reconstruction workflow.

Alternative approaches, such as LiDAR, do not allow for direct pose calculation because they do not acquire data in a shutter-like manner, which complicates sensor synchronisation. Instead, LiDAR sensors steer infrared beams continuously, recording returns over multiple timestamps during each scan sweep. Another limitation is that the geometric accuracy of LiDAR strongly depends on precise navigation or SLAM systems for point cloud motion correction. To prevent potential drifts that could cause geometric distortions, GNSS RTK systems are typically required. However, these systems are highly susceptible to GNSS signal occlusions and multipath effects, which can be particularly limiting during close-range inspections where accurate positioning is critical. Furthermore, the geometric accuracy of lightweight LiDAR sensors—commonly mounted on UAVs—is generally comparable to, or even lower than, that of photogrammetry, with RMSE values typically on the order of a few centimetres [19].

2.3 Thermal field reconstruction

For the thermal field reconstruction, the surface of the BIM model is rasterised at a 3 cm resolution. For each point, denoted as $\mathbf{X}^i_{\mathrm{BIM}}$, the corresponding radiance values from the IR images are extracted using a pinhole camera model. First, the coordinates of each point are expressed in the camera reference frame using Equation (1):

$$\mathbf{X_c}^{i,j} = \begin{bmatrix} X_c^{i,j} \\ Y_c^{i,j} \\ Z_c^{i,j} \end{bmatrix} = \mathbf{R}^j \, \mathbf{X}^i_{\mathrm{BIM}} + \mathbf{t}^j . + \begin{bmatrix} d \\ 0 \\ 0 \end{bmatrix} \qquad (1)$$

where:

- $\mathbf{X_c}^{i,j}$ are the coordinates of $\mathbf{X}^i_{\text{BIM}}$ in the reference frame of camera pose j.
- $\mathbf{R}^j$ and $\mathbf{t}^j$ are the rotation matrix and translation vector associated with camera pose j.
- d is a translation constant equal to 4.5 cm. Note that the camera poses computed in Section 2.2 correspond to the RGB sensor. In the Zenmuse H20T, the IR sensor is aligned with the RGB camera but is spatially separated, as shown in Figure 4.

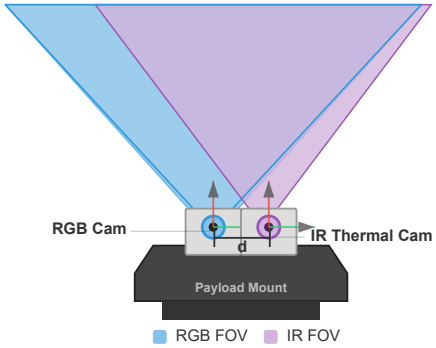


Figure 4. Layout of the Zenmuse H20T. The IR and RGB sensors are aligned along the camera's x-axis but are separated by a distance d.

The corresponding pixel coordinates $u^{i,j}$ and $v^{i,j}$ are then obtained using Equation (2):

$$u^{i,j} = f\frac{X_c^{i,j}}{Z_c^{i,j}} + c_x, \qquad v^{i,j} = f\frac{Y_c^{i,j}}{Z_c^{i,j}} + c_y, \tag{2}$$

where f is the focal length of the camera in pixel units, and c_x and c_y are the coordinates of the principal point (optical centre) on the image plane. All these quantities are derived from the specifications listed in Table 1.

For each pixel coordinate, the following verifications are performed to ensure projection robustness:

1. **Incidence angle filter.** Using the facade STL mesh derived from the Metashape photogrammetry reconstruction, the surface normal vector $\mathbf{n}$ at each facade point $\mathbf{X}_{\text{STL}}$ is obtained. For each camera position $\mathbf{t}^j$, the viewing direction vector is defined as

$$\mathbf{d} = \mathbf{t}^j - \mathbf{X}_{\text{STL}}. \tag{3}$$

The incidence angle θ is computed as

$$\cos\theta = \frac{\mathbf{n}\cdot\mathbf{d}}{\|\mathbf{d}\|}. \tag{4}$$

To ensure robustness with respect to normal orientation, the absolute value is considered. A sample is accepted only if

$$|\cos\theta| \geq \cos(\theta_{\min}) = 0.5, \tag{5}$$

which corresponds to a maximum allowable incidence angle of $\theta_{\max} = 60°$. This condition rejects grazing views, which are more sensitive to emissivity effects, ensuring that camera viewing directions are approximately perpendicular to the facade surface.

2. **Visibility / self-occlusion test.** For each valid projection, defined by the direction vector $\mathbf{d}$ and camera pose $\mathbf{t}^j$, a visibility check is performed using the Möller-Trumbore algorithm. Ray-triangle intersection tests are carried out with the STL mesh obtained from the photogrammetric reconstruction, removing projections that fall on occluded regions.

Only the values of $u^{i,j}$ and $v^{i,j}$ that satisfy these conditions are retained, and the corresponding radiance values are extracted from the infrared images. The average radiance at each facade point is then computed. Note that only camera poses for which the point lies within the camera field of view are considered in this calculation.

To obtain absolute temperature fields, reference temperature values are sampled at the facade points shown in Figure 5 using a contact probe Testo 720. The measurements are performed manually by directly recording the in-situ temperature at each point (Figure 6). Thermal paste is applied to ensure good thermal contact with the surface, thereby improving the accuracy of the probe readings. Measured temperatures are correlated with IR camera radiance values at the same locations, and linear regression is used to obtain the absolute temperature fields of the facade surface.



Figure 5. Probe contact point location across the facade surface

Note that this approach assumes uniform thermal emissivity across the facade surface. While this simplification is convenient, it may introduce temperature measurement errors in cases where the facade exhibits significant variations in emissivity, as is typically the case for glass surfaces such as windows. However, since the majority of the surface consists of concrete—with relatively uniform optical properties—and given that the primary focus of the inspection is on this material (as windows provide limited relevant information), a simple linear calibration was deemed appropriate for this case study. Nevertheless, for buildings with more heterogeneous material compositions, the proposed methodology could be extended by

incorporating material-specific calibration procedures for different facade elements.



Figure 6. Reference temperature sampling using a Testo 720 probe.

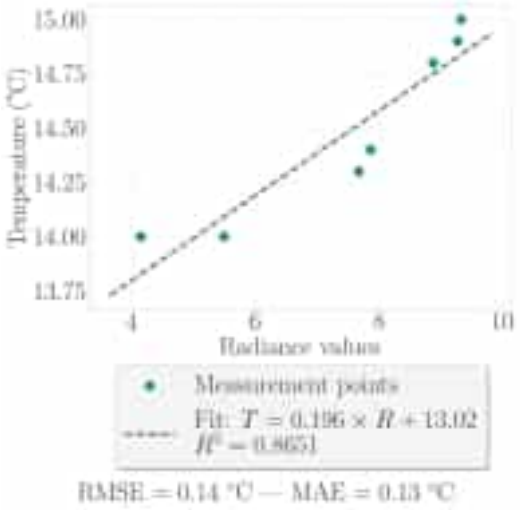


Figure 7. IR camera calibration. A linear regression is employed to correlate radiance values with absolute temperatures measured with the probe.

3 Results

Figure 7 shows the results of the IR-camera calibration (Section 2.3), obtained using the absolute temperature values measured by the contact probe. The probe measurements and the average radiance recorded by the IR camera over the rasterised facade surface deviate to some extent from the fitted line. However, this scatter is expected and consistent with typical error sources in IR imagery [20, 21]. Differences in emissivity among facade materials, sensor noise, and temperature variations during the acquisition can fully explain the observed deviations, which are on the order of several tenths of a degree. Nevertheless, the data still exhibits a reasonable linear correlation, with an R^2 coefficient of 0.8651. The fitted line indicates that temperature changes only slightly with radiance, with a slope of 0.196 and an intercept of 13.02. A Root Mean Square Error (RMSE) of 0.14ºC between the adjusted line is obtained, a fairly accurate value. These differences can be attributed to several factors, including sensor noise, temperature variations between probe sampling and image acquisition, or measurement errors.

Figure 8 shows the reconstructed temperature field of the facade overlaid with the semantic surface segmentation of the BIM model. Overall, the temperature remains nearly constant across the facade wall, with only minor variations of a couple of degrees near the upper left and right corners. These differences are likely due to measurement errors from the IR camera. In these corners, the camera was pointing in the direction of the sun, and this radiation may have affected the readings.

Although the different facade components are made of materials with varying absorption coefficients, the facade remained in the shade during the inspection and in the preceding hours of the morning, which explains the observed uniformity in the temperature field.

For the integration of inspection data within standardised BIM formats such as IFC, the semantically segmented facade components are mapped to BIM-native entities, creating a structured, object-based representation of the building envelope. In this framework, thermographic analysis can be used to detect potential deficiencies, such as localised thermal anomalies, which are then associated with the corresponding BIM elements. When identified, these deficiencies can be incorporated as property sets linked to the affected objects, optionally including their spatial coordinates. This enriched, object-level representation enables the integration of inspection results into a unified BIM model, supporting condition assessment and maintenance workflows while ensuring interoperability.

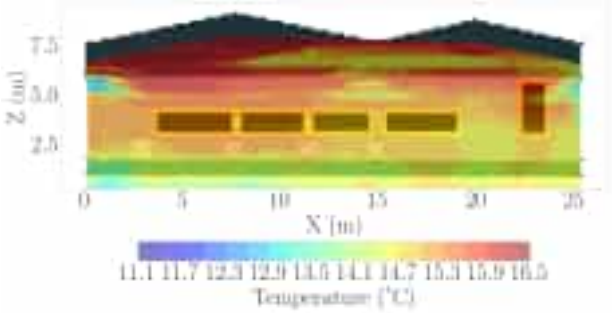


Figure 8. Overlay of the facade temperature field on the components of the BIM model.

Figure 9 displays a histogram of the temperature distribution across the raster points of the facade. As shown, most points remain at a temperature of around 15 °C, with

only a few outliers reaching values as low as 12 °C. However, the frequency of these outliers is minimal and, as discussed previously, they can likely be attributed to IR measurement errors. Overall, the temperature field remains uniform, and no insulation deficiencies or moisture accumulation issues are observed.

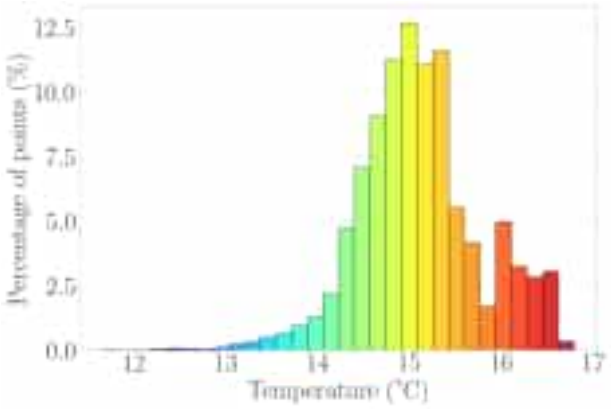


Figure 9. Histogram of the facade temperature distribution.

To verify the consistency of the results, the reconstructed temperature field is compared with the air temperature at the time of the inspection. Table 2 shows the air temperature during the inspection, retrieved from a nearby meteorological station managed by MeteoGalicia in Ourense [22], the region's main meteorological agency. The station is situated within the University of Vigo campus, at a distance of approximately 1 km from the inspected building facade. As shown, the air temperature remained within 15–16 ºC during the inspection, which is consistent with the measured temperature field.

Table 2. Temperature measurements from MeteoGalicia's station on 06/11/2025 [22].

Time	**Temperature (°C)**
12:00:00	14.92
12:10:00	15.00
12:20:00	15.20
12:30:00	16.20
12:40:00	16.66
12:50:00	15.86
13:00:00	15.62

Figure 10 shows the number of viewing cameras at each point of the projected raster model. As observed, the distribution of viewing cameras is not uniform, since the facade inspection was conducted by manually piloting the UAV. To improve inspection uniformity, automatic path planning tools could be employed. In general, however, almost all raster points exhibit a high camera count. During the facade survey, a total of 1,433 images of the wall were captured (Figure 3), resulting in a high level of overlap between photographs.

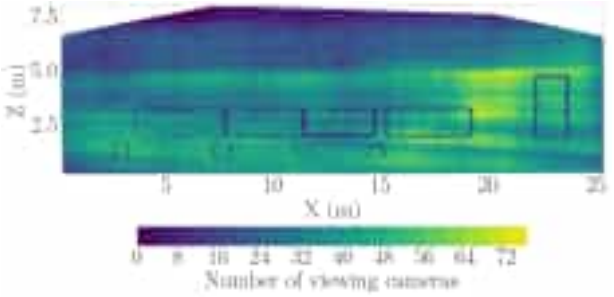


Figure 10. 2D map of the number of cameras viewing the facade surface.

Figure 11 presents a histogram of the distribution of viewing cameras. Most points are observed by 10 or more cameras, with an average of 36.76 viewpoints across the facade surface. Only in isolated cases does the camera count fall below 10, primarily at points near the building roof where fewer images of the facade were acquired. This highlights one of the main advantages of UAV imagery over traditional methods: up to thousands of images of the entire facade can be captured in just a few minutes, significantly reducing both inspection time and the associated operational risk.

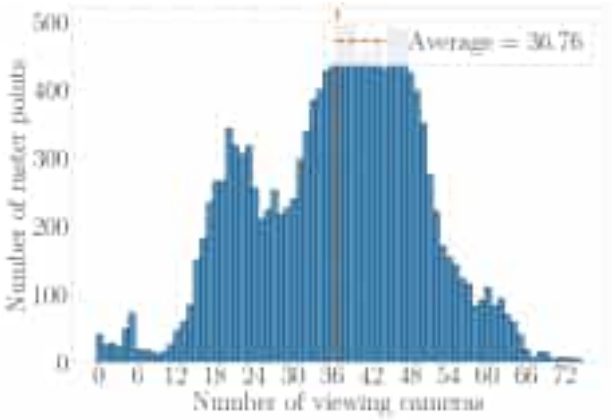


Figure 11. Histogram of the number of cameras viewing the facade.

4 Conclusions

This work presents an automated workflow for the efficient reconstruction of facade temperature fields using UAVs. Visual imagery is used to perform Structure-from-Motion and estimate camera poses, and this information is then employed to project IR radiance fields onto a BIM model of the structure. Absolute temperature values are

obtained through a radiometric calibration using contact probes. To validate the proposed approach, a case study was conducted in Ourense, where a temperature field of the facade was reconstructed and segmented according to the BIM model surfaces. Overall, a uniform temperature field was observed, as the inspected wall was in the shade. The results were also consistent with the air temperature values recorded at a nearby meteorological station, further supporting the validity of the method.

In future work, the methodology will be integrated with path planning and autonomous navigation systems to enable more efficient and fully automated facade inspections. Additionally, artificial intelligence algorithms will be incorporated for automatic defect detection, and the approach will be validated on facades exhibiting moisture, thermal bridges, and insulation deficiencies to quantify the accuracy of the proposed pipeline. Inspections of facades under sunlight exposure will also be conducted to assess the impact of solar radiation on the measured temperature values and the accuracy of insulation defect detection, thereby further validating the methodology under varied environmental conditions.

Acknowledgements

This work was funded by the European Union - Horizon Europe, Agencia Estatal de Investigación, Ministerio de Ciencia, Innovación y Universidades, and Xunta de Galicia through the following grants: SARIL A302, STORCITO 101182153, RADIANCE 101235536, PID2021-125060OB-I00 A202, PID2024-156356OB-I00 A507, CPP2022-009679 A308, FPU21/01176, FEADER2024/052B A406, Ciencias Mariñas-MRR C286, GPC-ED431B 2025/31 A505.

References

[1] Elton Bauer, Elier Pavón, Eva Barreira, and Eliane Kraus De Castro. Analysis of building facade defects using infrared thermography: Laboratory studies. *Journal of Building Engineering*, 6, 2016. ISSN 23527102. doi:10.1016/j.jobe.2016.02.012.

[2] Jun Chen, Shenghan Zhang, Gang Zhao, Dhanada Mishra, and Matthew Ming Fai Yuen. Infrared thermography of façade delamination: Quantifying environmental variability and heat transfer mechanisms. *Construction and Building Materials*, 494, 2025. ISSN 09500618. doi:10.1016/j.conbuildmat.2025.143177.

[3] Ali Waqas, Mohamad T. Araji, and Sherif S. Sherif. A novel method for monitoring material behaviour in urban buildings using aerial multi-modality imaging. *Building and Environment*, 282, 2025. ISSN 03601323. doi:10.1016/j.buildenv.2025.113300.

[4] Marta Videras Rodríguez, Sergio Gómez Melgar, and José Manuel Andújar Márquez. Evaluation of aerial thermography for measuring the thermal transmittance (U-value) of a building façade. *Energy and Buildings*, 324, 2024. ISSN 03787788. doi:10.1016/j.enbuild.2024.114874.

[5] Yinglong He, Jiayu Pan, Ramit Debnath, Ronita Bardhan, Luke Cullen, Marco Gomez Jenkins, Erik Mackie, George Hawker, and Ian Parry. Energy mapping of existing building stock in cambridge using energy performance certificates and thermal infrared imagery. *Environmental Data Science*, 3, 2025. ISSN 26344602. doi:10.1017/eds.2024.26.

[6] Alberto Villarino, Hugo Valenzuela, Natividad Antón, Manuel Domínguez, and Ximena Celia Méndez Cubillos. Uav applications for monitoring and management of civil infrastructures. *Infrastructures*, 10(5), 2025. doi:10.3390/infrastructures10050106.

[7] Mariluz Gil-Docampo, Juan Ortiz Sanz, Ignacio Cañas Guerrero, and Manés Fernández Cabanas. Uas ir-thermograms processing and photogrammetry of thermal images for the inspection of building envelopes. *Applied Sciences*, 13(6):3948, 2023.

[8] Albert Ficapal and Ivan Mutis. Framework for the detection, diagnosis, and evaluation of thermal bridges using infrared thermography and unmanned aerial vehicles. *Buildings*, 9, 2019. ISSN 20755309. doi:10.3390/buildings9080179.

[9] Tarek Rakha and Alice Gorodetsky. Review of Unmanned Aerial System (UAS) applications in the built environment: Towards automated building inspection procedures using drones. *Automation in Construction*, 93:252–264, 2018. ISSN 0926-5805. doi:10.1016/j.autcon.2018.05.002.

[10] Adam Dlesk, Karel Vach, and Karel Pavelka. Transformations in the Photogrammetric Co-Processing of Thermal Infrared Images and RGB Images. *Sensors*, 21(15), 2021. doi:10.3390/s21155061.

[11] Yingying Kong, Henry Leung, Bowen Zhang, Shiyu Xing, and Yingying Chen. 3-D Point Cloud Reconstruction of Infrared Images Based on Improved Structure from Motion. In *2018 2nd European Conference on Electrical Engineering and Computer Science (EECS)*, pages 307–310, 2018. doi:10.1109/EECS.2018.00063.

[12] Krishna Ribeiro-Gomes, David Hernández-López, José F. Ortega, Rocío Ballesteros, Tomás Poblete, and Miguel A. Moreno. Uncooled Thermal Camera Calibration and Optimization of the Photogrammetry Process for UAV Applications in Agriculture. *Sensors*, 17(10), 2017. doi:10.3390/s17102173.

[13] Adam Dlesk, Karel Vach, and Karel Pavelka. Photogrammetric Co-Processing of Thermal Infrared Images and RGB Images. *Sensors*, 22(4), 2022. doi:10.3390/s22041655.

[14] Rudraksh Kapil, Guillermo Castilla, Seyed Mojtaba Marvasti-Zadeh, Devin Goodsman, Nadir Erbilgin, and Nilanjan Ray. Orthomosaicking Thermal Drone Images of Forests via Simultaneously Acquired RGB Images. *Remote Sensing*, 15(10), 2023. doi:10.3390/rs15102653.

[15] Agisoft LLC. Agisoft metashape. https://www.agisoft.com/, 2025.

[16] Pix4D. Pix4dmapper. https://www.pix4d.com/, 2025.

[17] Leica Geosystems. Leica disto x6 laser distance meter. Product page, 2025. URL https://shop.leica-geosystems.com/global/measurement-tools/disto/leica-disto-x6/buy. Accessed: 2025-12-11.

[18] DJI. Zenmuse h20 series specifications. https://enterprise.dji.com/zenmuse-h20-series/specs, 2025. Accessed: 2025-12-11.

[19] J. D. D. Casisirano and A. R. C. Claridades. Comparative analysis of uav photogrammetry and terrestrial laser scanning for 3d building reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-5/W4-2025:143–150, 2026. doi:10.5194/isprs-annals-X-5-W4-2025-143-2026. URL https://isprs-annals.copernicus.org/articles/X-5-W4-2025/143/2026/.

[20] MG Acorsi, LM Gimenez, and M Martello. Assessing the performance of a low-cost thermal camera in proximal and aerial conditions. remote sens 12: 3591, 2020.

[21] Krishna Ribeiro-Gomes, David Hernández-López, José F Ortega, Rocío Ballesteros, Tomás Poblete, and Miguel A Moreno. Uncooled thermal camera calibration and optimization of the photogrammetry process for uav applications in agriculture. *Sensors*, 17(10):2173, 2017.

[22] MeteoGalicia. MeteoVisor: observación meteorolóxica en Galicia. https://www.meteogalicia.gal/web/observacion/meteovisor, 2025. Accessed: 2025-12-11.

Hyperledger Fabric-Based Blockchain Application in Modular Construction Manufacturing

Angat Bhatia[1,3], Vafa Rostamiasl[1,2] and Osama Moselhi[1,2]

[1]Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada
[2]Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Concordia University, Montreal, QC, Canada
[3]Risk Management Department, Hatch, Gurugram, India
angat.bhatia@hatch.com, vafa.rostamiasl@concordia.ca, moselhi@encs.concordia.ca

Abstract –

Modular Construction Manufacturing (MCM) has attracted considerable attention in recent years due to its potential to increase productivity, minimize waste, and improve the predictability of building projects. Despite these benefits, effective coordination and real-time information exchange among multiple stakeholders in the delivery of modular construction remains a significant challenge, particularly in managing production line data for module components. Current data-sharing practices in MCM are often fragmented and rely on manual reporting, leading to inconsistencies, limited traceability, and reduced trust among stakeholders. To address these challenges, this paper introduces a permissioned blockchain method for MCM. The proposed method includes: *(i)* network design, *(ii)* smart contract development, and *(iii)* transaction execution and evaluation. A smart contract (chaincode) is developed to automate the creation, updating, and querying of production data, while enforcing validation rules and access control exclusively among participating stakeholders. The blockchain network is established with two peer organizations representing key stakeholders to facilitate decentralized transaction endorsement. This method ensures consistent ledger synchronization, strengthens data integrity, and provides tamper-resistant and traceable documentation of module component records throughout the production line. The implementation of the proposed method utilizing the wall panel production line data confirms the feasibility of permissioned blockchain-enabled information management in MCM and highlight its potential to support more transparent, reliable, and collaborative production workflows. In addition, it protects data privacy by allowing only authorized project members to participate in and maintain the network.

Keywords – Modular Construction; Blockchain; Hyperledger Fabric; Smart Contracts; Production Data Management

1 Introduction

Modular construction has emerged as a transformative approach within the construction industry, offering significant advantages over conventional construction. By shifting major construction activities to a controlled factory environment, modular construction enhances production efficiency and reduces production line waste (e.g., worker waiting time) [1-2]. In particular, Modular Construction Manufacturing (MCM) enables standardized fabrication of module components (e.g., wall panels) in the production line, supporting higher levels of automation and repeatability [3]. This construction process has become increasingly relevant as governments and industry pursue scalable and sustainable solutions to address housing shortages [4]. Despite its benefits, the MCM process is characterized by fragmented information flows and limited interoperability between stakeholders [5].

Production line managers and on-site supervisors often maintain independent systems to record and exchange information, resulting in inconsistencies and duplicate records. These inefficiencies become more pronounced when stakeholders participate in multiple projects. Considering the dynamic nature of the production line, which requires frequent changes in production line planning (i.e., change orders), and delays/inaccuracies in sharing these module components production information (e.g., estimated process time of module components) propagate downstream, affecting production line planning, on-site assembly sequencing, and project makespan [6]. Moreover, the success of MCM relies heavily on the accuracy, transparency, and coordination of production line information (e.g., start and finish time of module components) shared among multiple stakeholders (e.g., production line manager and on-site supervisor).

The modular construction industry requires a reliable, tamper-proof mechanism to capture and share production data in real time across all participating stakeholders (e.g., production line managers and on-site supervisors). Such an approach must ensure trust, traceability, and data

integrity without relying on a centralized system. In this regard, Blockchain has been proposed as a promising solution due to its decentralized architecture, immutability, and built-in consensus mechanisms [7]. Studies have examined blockchain applications in supply chain management, smart contracts for construction payments, and document management for construction applications [8-10]. However, most prior works focus on conceptual frameworks and overall supply chain-level applications, with limited research demonstrating practical blockchain deployment specifically for production-level applications in modular construction. In addition, there is a notable gap in operational prototypes that integrate blockchain with actual production data, smart contract logic, and multi-stakeholders within an MCM environment. Considering the broader adoption of blockchain in the MCM production line, it is necessary to test its potential benefits with a case study. In particular, a smart contract should be designed and implemented to demonstrate its potential to efficiently manage and share production line data to improve the MCM production line process. To address these gaps, this paper proposed a Hyperledger Fabric-based blockchain method for MCM. The proposed method consists of the following procedures: (i) creating a blockchain architecture tailored for MCM, (ii) developing a smart contract for managing module component data, and (iii) evaluating the proposed blockchain method through transaction execution and ledger validation using a MCM case study. The contributions of this study include: (i) a permissioned blockchain network with peer organizations representing key stakeholders (i.e., production line manager and on-site supervisor) in the MCM; and (ii) a custom smart contract (chaincode) is developed to record, query, and manage module components information, ensuring authorized project members to participate in the network and tamper-resistant information management and exchange. It is worth mentioning that Hyperledger Fabric is chosen due to the following reasons: (i) it protects data privacy by allowing only authorized project members to participate in and maintain the network; and (ii) it is an extensible open-source platform that users can customize according to their needs [8].

2 Literature Review

2.1 Blockchain Technology in Construction

Blockchain has recently emerged as a transformative technology that can enhance transparency, trust, and data integrity in construction projects. As a decentralized and immutable ledger system, blockchain enables distributed parties to record transactions securely without relying on centralized authorities [11]. Early construction-oriented studies highlight blockchain's potential benefits in supply chain, contract administration, and automated payment mechanisms [12]. Moreover, blockchain's ability to ensure tamper-proof and auditable records aligns with the construction industry's need for verifiable information exchanges among fragmented stakeholders [13]. However, most blockchain research in construction focuses on supply chains and payment automation, with limited work at the production-line level in modular construction. Moreover, many studies propose conceptual frameworks rather than operational prototypes, and research on permissioned blockchains, such as Hyperledger Fabric in a modular construction production line environment, is limited.

2.2 Digital Technologies for Modular Construction

Modular construction relies heavily on digital technologies to support precision, standardization, and coordination. Building Information Modeling (BIM) enhances design integration and clash detection, enabling improved off-site fabrication planning [14]. Likewise, cloud-based platforms also facilitate data sharing and decision-making across the modular construction lifecycle [15]. Despite these advances, digital systems in MCM often operate as isolated platforms controlled by single organizations, resulting in fragmented and inconsistent production data. Studies note persistent issues, including a lack of interoperability, limited trust in shared information, and difficulty maintaining accurate records across stakeholders [16]. These challenges highlight the need for decentralized solutions that can complement BIM and cloud systems by ensuring secure, tamper-resistant, multi-party data integration.

2.3 Blockchain Adoption for Manufacturing and Modular Construction

Manufacturing research highlights blockchain's role in enabling transparent production monitoring, machine-to-machine communication, and component traceability [17]. Smart contracts have been used to automate quality checks, material verification, and workflow approvals, demonstrating the potential for blockchain to enhance cyber–physical coordination in factory environments [8-10]. Modular construction shares similar challenges with general manufacturing, such as multi-stakeholder interactions, frequent design changes, and strict quality control requirements. In this regard, Blockchain has been proposed as a suitable enabler for secure data exchanges between manufacturers, suppliers, and contractors [9]. However, most existing studies remain conceptual, and few have developed and tested blockchain solutions integrated with real modular production data. In addition, there is a gap in end-to-end blockchain prototypes demonstrating live transactions for module components (e.g., wall panels) production line data.

3 Proposed Method

This paper adopts a structured method to design, implement, and validate a blockchain-enabled information management method for modular construction manufacturing (MCM). As illustrated in Figure 1, the proposed method consists of three main steps: (i) network design; (ii) smart contract (chaincode) development and deployment; and (iii) transaction execution and evaluation. The proposed method begins by designing a permissioned blockchain network (i.e., Hyperledger Fabric) tailored to the MCM process. Subsequently, smart contracts are developed to automate module components (e.g., wall panel) production data transactions (e.g., estimated process time). Finally, the proposed system is evaluated through transaction execution and ledger evaluation.

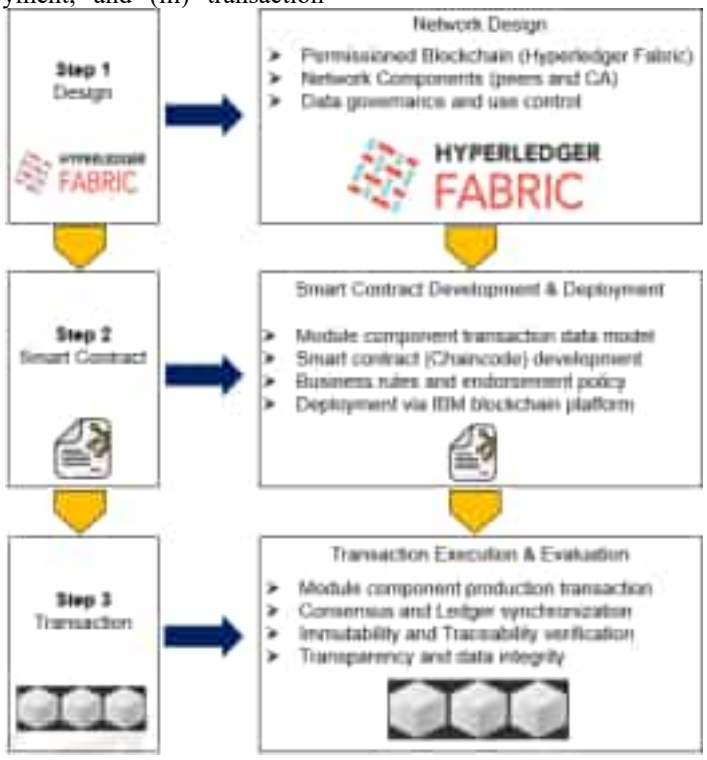


Fig. 1: Overview of the proposed Hyperledger Fabric-based blockchain method for MCM

3.1 Network Design

The first step focuses on designing a blockchain network that supports a secure, transparent, and multi-stakeholder information exchange in MCM. The prefabrication process of module components (e.g., wall panels) was selected as the target use case due to its repetitive nature and reliance on accurate production data. A permissioned blockchain architecture was adopted to ensure controlled access and data privacy among participating stakeholders (e.g., production line manager and on-site supervisor). Core network components, such as peer nodes, channels, ordering services, and certificate authorities (CA), are defined to enable secure identity management, transaction ordering, and ledger synchronization. In addition, data governance rules and access permissions were established to govern the recording, sharing, and querying of production information within the network. This design ensures that production data remain tamper-proof, traceable, and accessible only to authorized participants; thereby addressing common information fragmentation challenges in current MCM.

3.2 Smart Contract Development

The proposed blockchain method incorporates a set of smart contract components to regulate information exchange and data governance in MCM, as illustrated in Figure 2**.** These components are implemented as executable smart contracts (chaincode) on a permissioned blockchain network and designed to manage participant access, record production line transactions, and support secure data querying. Authorized stakeholders, including the production line manager and on-site supervisor, are registered through a certificate authority (CA) and assigned predefined roles to ensure controlled access to production data. Production-related transactions for module components,

such as process time, are validated against predefined endorsement policies and recorded on the distributed ledger following consensus. In addition, query functions are provided to enable authorized participants to retrieve real-time and historical production information without permitting data modification. By formalizing access rules within smart contracts, the proposed components enhance transparency, traceability, and data integrity while eliminating reliance on centralized information management systems.

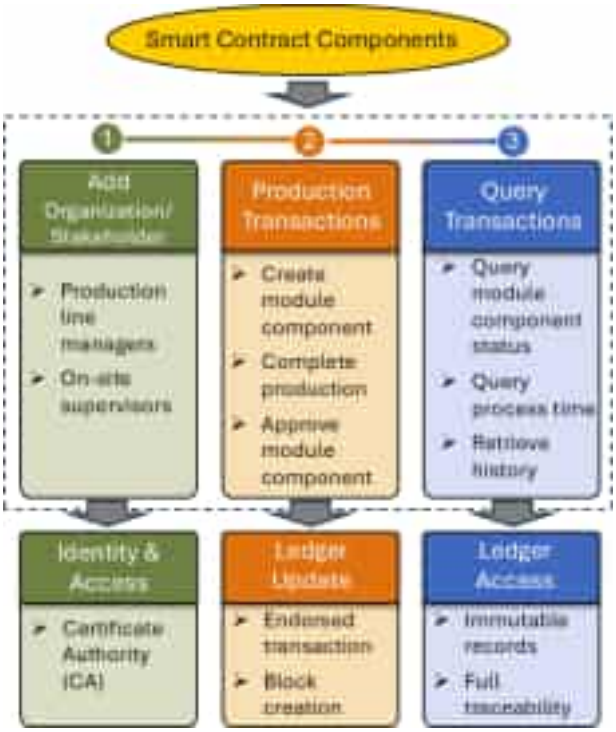


Fig. 2: Proposed smart contract components for MCM

As shown in Figure 3, smart contracts were developed to encode manufacturing business rules governing the creation, update, and querying of module components (e.g., wall panel ID). These contracts ensure that production information is recorded in a consistent and standardized manner while enforcing predefined validation rules. Endorsement policies were specified to require transaction approval from relevant stakeholders (i.e., production line manager and on-site supervisor) before records are committed to the distributed ledger. The blockchain network and smart contract were deployed using the IBM Blockchain platform, which provides a graphical interface for configuring Hyperledger Fabric infrastructure. The deployment process included: (i) installing the smart contract on both organization peers; (ii) approving the chaincode definition to reach endorsement consensus; (iii) committing the chaincode to the channel, making it operational; and (iv) instantiating the ledger to begin recording production transactions. Additionally, the module component data (e.g., number of studs, windows, start and finish times of module components at the workstation) used in the blockchain environment was collected utilizing an RFID system. The process times (i.e., the time required to complete one module component at a given workstation) are extracted using Equation 1:

$$PT\,m,w = IRT\,m,w+1 - IRT\,m,w \qquad (1)$$

Where PT is the process time of module component ID “m” at current workstation “w”, “w+1” is the next workstation, and IRT is the initial read time of the module component at a given workstation of the production line. The process times are initiated as transactions in the smart contract using a deep neural network (DNN) based prediction model [18]. Hyperparameter tuning—the process of identifying optimum parameters (i.e., number of hidden layers, number of neurons, learning rate and momentum) is performed by employing a genetic algorithm (GA). The optimization objective is to minimize the model's prediction error (i.e., MAE), as expressed in Equation 2:

$$MAE = \frac{\Sigma^{n}_{i=1} |A_i - P_i|}{n} \qquad (2)$$

Where Ai is the actual process time of a module component, and Pi is the predicted process time of a module component. A detailed discussion of the developed prediction model is beyond the scope of this paper and is reported in a paper by Bhatia et al. [18].

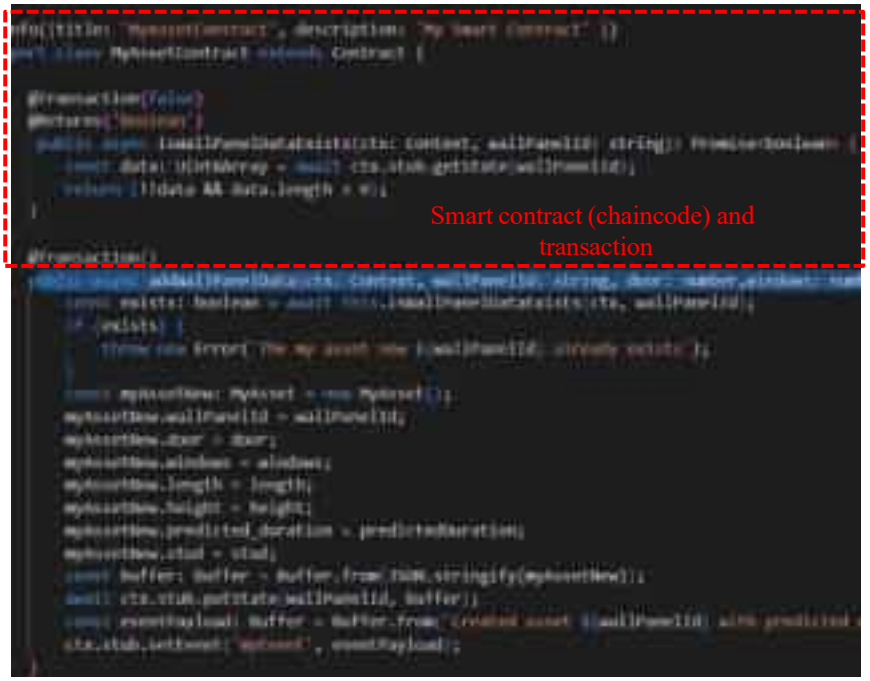


Fig. 3: Snippet of the Smart Contract

3.3 Transaction Execution and Evaluation

The final step evaluates the blockchain network through transaction execution utilizing a case study. Blockchain transactions corresponding to module components information (e.g., number of doors and windows) were executed to simulate real-time manufacturing operations. These transactions were endorsed by peer nodes, ordered into blocks, and committed to the distributed ledger in accordance with the defined consensus mechanism. When transactions were submitted, peers endorsed the transaction according to the network policy, and the ordering service generated a new block. Ledger synchronization across peer nodes was examined to verify the consistency and immutability of recorded production data. The traceability of wall panel records was assessed by querying historical transactions and validating their resistance to unauthorized modification. The evaluation focuses on the system's ability to enhance transparency, ensure data integrity, and support trustworthy information sharing among stakeholders. This demonstrates the feasibility of applying blockchain technology to MCM and highlights its potential to improve information reliability and coordination across production environment.

4 Case Study

The proposed method utilized data from a wood-frame wall panel production line in Edmonton, Canada. The production line consists of framing station, sheathing station, multifunction bridge, window/door installation lines, where windows/doors are installed on the wall panels and transferred to the wall magazine line (i.e., storage area). Process times at workstations were collected by the modular fabricator company using an RFID system. The data for the workstations includes the time between July 2015 and August 2018. As such, it contains 416,950 timestamps (i.e., start and finish times) for wall panels at various workstations and design attributes (e.g., number of studs and number of windows), for a total of 93,660 wall panels. A detailed discussion of the RFID data and the development of predictive model is beyond the scope of this paper and is reported in papers by Altaf et al. [1] and Bhatia et al. [18]. The objective of the case study is to demonstrate the application of the proposed method by deploying smart contracts, transaction execution and validation through data querying to retrieve immutable production line data of a wall panel production line.

Figure 4 illustrates the deployment of the proposed smart contract within a Hyperledger Fabric network using the IBM Blockchain platform. The smart contract definition, including its name and version, was deployed to a dedicated blockchain channel (i.e., Blockchain-4) and approved by two peer nodes (e.g., Org 1 Peer representing production line manager) operated by separate blockchain organizations. Deploying the smart contract across multiple peers ensures that transaction execution and ledger updates are jointly validated in accordance with predefined endorsement policies. This configuration establishes decentralized trust, prevents unilateral data modification, and enables consistent synchronization of wall panel production data across participating stakeholders (i.e., production line manager and on-site supervisor).

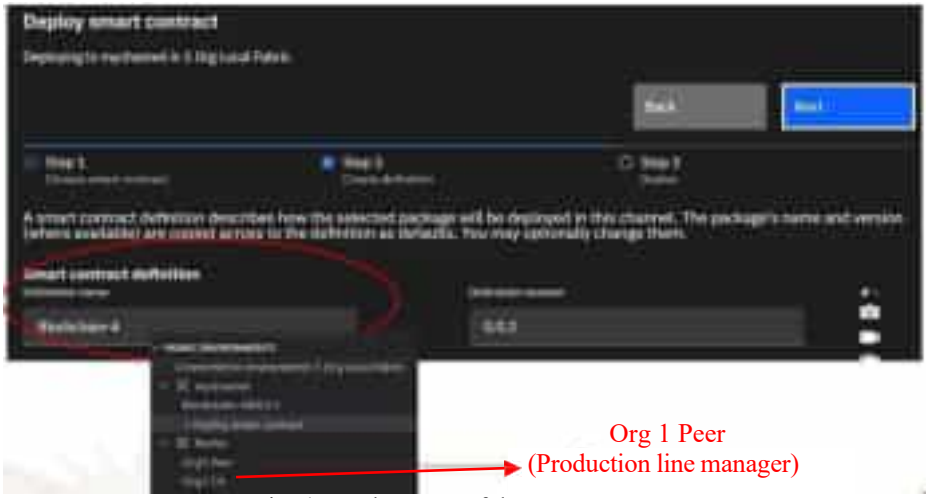


Fig. 4: Deployment of the smart contract

Following smart contract deployment, blockchain transactions were executed in response to wall panel production events. Figure 5 presents the invocation of a smart contract transaction, where a specific 'Wall Panel ID' is submitted as an input parameter. For example, the user submitted 'Wall Panel ID:7' to generate the wall panel information. The submission of this identifier represents the occurrence of a manufacturing event associated with the corresponding wall panel. Upon submission, the transaction is endorsed by peer nodes and committed to the distributed ledger following the defined consensus mechanism. This process demonstrates that physical production events in MCM are translated into blockchain transactions, enabling real-time and tamper-resistant recording of manufacturing activities.

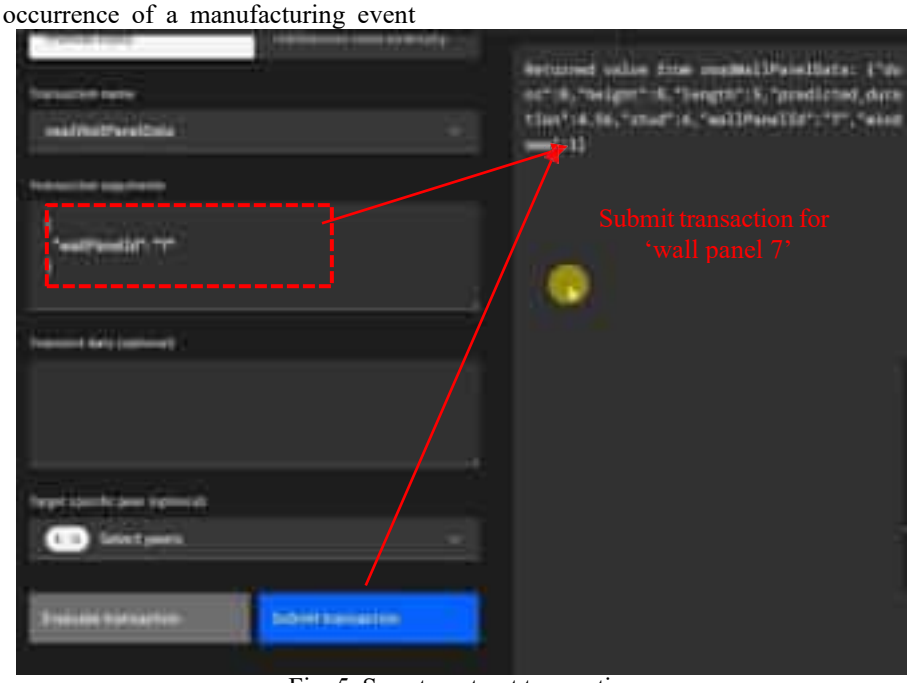


Fig. 5: Smart contract transaction

To ensure data integrity and consistency, manufacturing workflows were formalized within the smart contract logic. Figure 6 shows an excerpt of the implemented smart contract (chaincode) for the case study. The smart contract includes validation mechanisms that verify the existence of wall panel records (e.g., new window value, new height value and new predicted duration) before allowing updates to be committed to the ledger. This logic prevents duplicate entries and enforces predefined business rules governing production data submission. Invalid transactions are automatically rejected, ensuring that only verified manufacturing events are immutably recorded. This implementation confirms that data validation is enforced at the blockchain layer without reliance on centralized control.

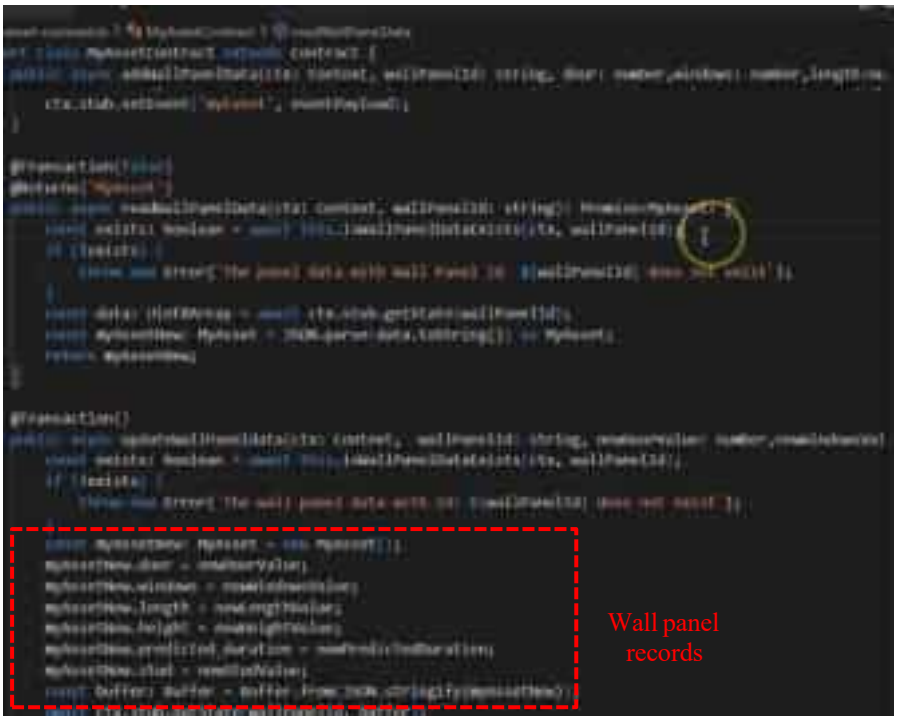


Fig. 6: Implemented smart contract

In addition to transaction execution, this paper evaluated the system's ability to retrieve and verify stored production data. Figure 7 illustrates the execution of a read-only query transaction used to retrieve wall panel production information from the blockchain ledger. For example, the 'wall panel ID 2' has: 3 doors, 11ft height, 12 studs and an estimated process time of 14.12 minutes. The query output confirms that production data recorded during earlier manufacturing events remain persistently stored and accessible in a structured format. Authorized participants can retrieve wall panel status, timestamps, and historical records without modifying ledger contents. This capability demonstrates the traceability, transparency, and immutability of the proposed method, supporting post-event verification and informed decision-making in MCM.

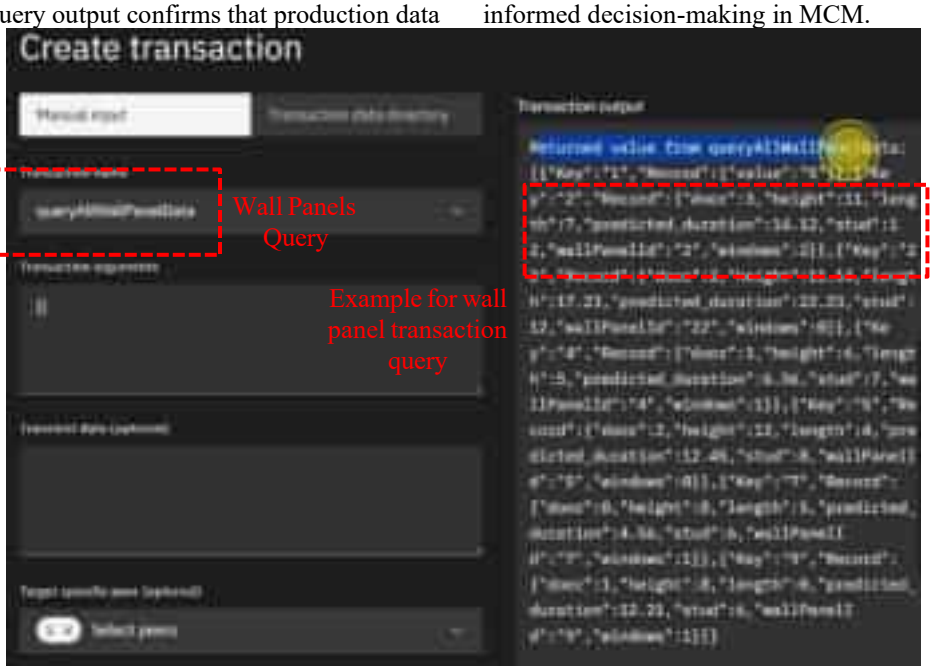


Fig. 7: Execution and query of a transaction

Collectively, this demonstrates the end-to-end application of the proposed method, from smart contract deployment and event-driven transaction execution to validation and data retrieval. The successful implementation enhances data integrity, transparency, and stakeholder trust. In addition, it supports the feasibility of integrating blockchain as a digital infrastructure for managing production information in modular construction.

5 Conclusion and Future Works

This paper presented a newly developed permissioned blockchain-based method using Hyperledger Fabric to enhance data integrity, transparency, and multi-stakeholder coordination (i.e., production line manager and on-site supervisor) in MCM. The method integrates a multi-peer blockchain network, customized smart contracts for managing wall panel production information, and a distributed ledger structure capable of securely recording and synchronizing manufacturing events. Through a case study implementation, the system successfully executed event-driven transactions, enforced smart-contract-based validation rules, and enabled the immutable storage and retrieval of production data. This demonstrates that wall panel information from the production line can be systematically translated into blockchain transactions and subsequently queried for verification and traceability.

However, the implementation involved only two peer organizations (i.e., production line manager and on-site supervisor), which limits the evaluation of scalability and interoperability across a broader modular construction supply chain. Future research should therefore focus on expanding the blockchain network to additional stakeholders and evaluating system performance in full-scale manufacturing environments.

References

[1] Altaf, M.S., Bouferguene, A., Liu, H., Al-Hussein, M. and Yu, H., 2018. Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID. Automation in Construction, 85, 369-383.

[2] Salama, T., Salah, A. and Moselhi, O., 2021. Integrating critical chain project management with last planner system for linear scheduling of modular construction. Construction Innovation, 21(4), 525-554.

[3] Bhatia, A.P.S., 2024. Automated Planning and Scheduling Method for Modular Construction Manufacturing (Doctoral dissertation, Concordia University).

[4] MBI (Modular Building Institute). 2023. "Permanent modular construction report." Accessed on 02 December, 2025. https://offsiteconstructionnetwork.com/2023-modular-construction-annual-reports/

[5] Wu, L., Lu, W., Zhao, R., Xu, J., Li, X. and Xue, F., 2022. Using Blockchain to Improve Information Sharing Accuracy in the Onsite Assembly of Modular Construction. Journal of Management in Engineering, 38(3), 04022014.

[6] Bhatia, A., Moselhi, O. and Han, S., 2025a. Hybrid Optimization Approach for Scheduling in Panelized Construction Manufacturing. Journal of Construction Engineering and Management, 151(8), 04025107.

[7] Wood, G., 2014. Ethereum: A secure decentralised generalised transaction ledger. Ethereum project yellow paper, 151, 1-32.

[8] Elghaish, F., Pour Rahimian, F., Hosseini, M.R., Edwards, D. and Shelbourn, M., 2022. Financial management of construction projects: Hyperledger fabric and chaincode solutions. Automation in Construction, 137, 104185.

[9] Ding, S., Hu, H., Xu, F., Chai, Z. and Wang, W., 2024. Blockchain-based security-minded information-sharing in precast construction supply chain management with scalability, efficiency and privacy improvements. Automation in Construction, 168, 105698.

[10] Das, M., Tao, X., Liu, Y. and Cheng, J.C., 2022. A blockchain-based integrated document management framework for construction applications. Automation in Construction, 133, 104001.

[11] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

[12] Li, J., Greenwood, D. and Kassem, M., 2019. Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases. Automation in Construction, 102, 288-307.

[13] Turk, Ž. and Klinc, R., 2017. Potentials of blockchain technology for construction management. Procedia Engineering, 196, 638-645.

[14] Sacks, R., Eastman, C., Lee, G. and Teicholz, P., 2018. BIM handbook: A guide to building information modelling for owners, designers, engineers, contractors, and facility managers. John Wiley & Sons.

[15] Dave, B., Boddy, S. and Koskela, L., 2013. Challenges and opportunities in implementing lean and BIM on an infrastructure project. In: Annual Conference of the International Group for Lean Construction, 21.

[16] Pan, W. and Sidwell, R., 2011. Demystifying the cost barriers to offsite construction in the UK. Construction Management and Economics, 29(11), 1081-1099.

[17] Wu, L., Zhong, Y., Zhu, X., Zhou, X. and Leng, J., 2025. A Review of Blockchained Product Quality Management Towards Smart Manufacturing. Processes, 13(8), 2622.

[18] Bhatia, A.P.S., Moselhi, O. and Han, S., 2025b. Process time estimation for workstations in modular construction production line. Journal of Information Technology in Construction (ITcon), 30(44), 1080-1098.

RepCA-GCN: A Lightweight and Efficient Method for Real-Time Construction Worker Action Recognition via Structural Re-parameterization

Jinhui Liang[1], Hui Deng[1], Zhou Zhang[2], Wenhao Li[1], Yichuan Deng[1,*]

[1] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China.
[2] School of Management Science and Engineering, Guizhou University of Finance and Economics, Guizhou, China.
jhliang29@gmail.com, hdeng@scut.edu.cn, zhangzhou@mail.gufe.edu.cn, 19098092885@163.com, ctycdeng@scut.edu.cn

Abstract –

Accurate and real-time recognition of construction workers' behaviors is critical to intelligent safety supervision and productivity assessment. However, existing skeleton-based methods struggle to balance recognition accuracy with inference latency, while lacking robustness against noisy keypoints caused by frequent occlusions. This study proposes RepCA-GCN, a high-efficiency architecture specifically tailored for construction site action recognition. The method integrates Structural Re-parameterization to decouple training capacity from inference latency, Coordinate Attention to capture long-range spatial dependencies, and a Confidence Gating mechanism to suppress noise from unreliable pose estimations extracted via YOLOv11-Pose. Validated on our constructed SiteWorker Action Recognition (SWAR) dataset of 7,424 video clips, we systematically explored skeletal configurations, finding that a streamlined 14-keypoint scheme helps to mitigate facial keypoint noise. Experimental results demonstrate that RepCA-GCN achieves 95.35% accuracy, matching state-of-the-art complex models (e.g., CTR-GCN) while significantly reducing FLOPs to 0.413G and achieving a low latency of 1.36 ms. Furthermore, a deployed monitoring system utilizing a dual sliding-window strategy achieves over 60 FPS in multi-person scenarios on a workstation equipped with an RTX 4090 GPU. This framework offers a superior speed-accuracy trade-off, making it well-suited for edge deployment in dynamic construction environments.

Keywords –

Construction Safety; Action Recognition; Pose Estimation; Structural Re-parameterization; Graph Convolutional Network; Real-time Monitoring

1 Introduction

The construction industry remains a high-risk sector, where unsafe behaviors of workers are a primary cause of accidents [1,2]. While traditional manual inspections suffer from insufficient coverage and high costs [3], the advent of computer vision has enabled automated monitoring using standard RGB cameras, which are cost-effective and non-intrusive compared to wearable sensors [4,5] or depth cameras [6].

In recent years, skeleton-based action recognition has emerged as a promising direction. By extracting human keypoints using pose estimation models (e.g., OpenPose [7], YOLO-Pose [8]) and processing them with deep learning networks, researchers can effectively classify worker actions [9,10]. Compared with Long Short-Term Memory (LSTM) based methods [10], the Spatio-Temporal Graph Convolutional Network (ST-GCN) [11] serves as a more advanced foundational framework with superior performance. However, applying ST-GCN directly to construction sites faces two critical bottlenecks:

Noise Sensitivity: Construction sites are cluttered, leading to frequent occlusions. Standard GCNs treat all keypoints equally, allowing low-confidence (occluded) keypoints to introduce noise and degrade accuracy.

Computational Efficiency vs. Accuracy: Advanced GCN variants (e.g., 2s-AGCN [12], CTR-GCN [13]) improve accuracy but introduce heavy computational burdens (high FLOPs and latency), making them unsuitable for real-time edge deployment on site. Conversely, lightweight models often sacrifice recognition performance.

To bridge this gap, this paper proposes RepCA-GCN, a lightweight yet powerful network tailored for real-time construction monitoring. We leverage Structural Re-parameterization technology [14], which allows the model to utilize a complex multi-branch architecture for feature extraction during training but collapses into a

single, efficient convolution layer during inference. This effectively decouples training capacity from inference speed. Furthermore, we integrate Coordinate Attention (CA) to enhance spatial feature representation and a Confidence Gating mechanism to filter noise based on keypoint reliability.

Our main contributions are:

We propose RepCA-GCN, incorporating Structural Re-parameterization modules, Coordinate Attention, and Confidence Gating. This architecture achieves SOTA accuracy (95.35%) with significantly lower latency (1.36ms) compared to existing methods.

We construct the SiteWorker Action Recognition (SWAR) dataset. To enhance robustness, we employ a rigorous occlusion-aware data augmentation strategy (including scaling, flipping, and random occlusion) on the training set, expanding the data scale to four times its original size to ensure generalization.

We validate the system's effectiveness through comprehensive experiments, demonstrating its capability for both safety monitoring (e.g., fall detection) and efficiency analysis in real-world scenarios.

2 Related Work

2.1 Vision-based Construction Safety Monitoring

Computer vision has been widely adopted for site safety. Early works focused on object detection, such as detecting Personal Protective Equipment (PPE) using YOLO or Faster R-CNN [15–17]. While effective for static compliance checking, these methods fail to capture temporal dynamics required for identifying complex actions (e.g., working vs. idling). Video-based approaches like 3D-CNNs [18] and Two-Stream networks [19] can capture motion but suffer from high computational costs, limiting real-time application [20].

2.2 Skeleton-based Action Recognition

Skeleton-based methods offer a balance between performance and efficiency. Yan et al. [11] proposed ST-GCN to model skeletal data as graphs. In construction, Wu et al. [9] and Zhou et al. [10] applied pose-based methods for behavior analysis. Zhou et al. [10] utilized a lightweight LSTM model; however, RNN-based structures lack the ability to explicitly model spatial structural dependencies compared to GCNs.

Recent advancements in GCNs focus on adaptive graphs (e.g., 2s-AGCN, CTR-GCN) to improve accuracy, but their complex topology computations increase inference latency. To address the efficiency issue, Structural Re-parameterization (e.g., RepVGG [14]) has gained attention in CNNs for speeding up inference without accuracy loss. In this study, we adapt this technique to the temporal modeling of GCNs, combined with YOLOv11-Pose, to achieve a high-performance, real-time solution for construction sites.

3 Methodology

To address the challenges of occlusion and high computational latency in construction site monitoring, we propose a high-efficiency framework for worker action recognition. As illustrated in Figure 1, the system operates in a two-stage pipeline. First, the YOLOv11-Pose model is employed to extract spatiotemporal human skeletal keypoint sequences from raw video streams. These sequences, which contain both coordinate data and confidence scores, serve as the input for the subsequent stage. Second, the proposed RepCA-GCN model processes these skeletal sequences to classify worker behaviors. This decoupled design allows for the independent optimization of pose estimation and action recognition, ensuring both precision and system flexibility.

3.1 RepCA-GCN Model Architecture

The core of our methodology is the RepCA-GCN (Reparameterized Coordinate-Attention Graph Convolutional Network). As shown in Figure 2, the network processes a skeletal sequence with dimensions (B, T, N, C). The backbone comprises four stacked RepBlock stages. To primarily resolve the conflict between model capacity and real-time inference speed, we employ Structural Re-parameterization in the temporal modeling layers. During the training phase, the temporal block adopts a multi-branch topology consisting of parallel 5×1, 3×1, 1×1 convolutions, and an identity mapping. This design enables the model to learn multi-scale temporal features. However, for deployment, these branches are mathematically fused into a single 9×1 convolutional layer via linear transformations of kernels and biases. This decoupling ensures the model retains the representational power of a complex topology while achieving the low memory access cost and latency of a single-path network during inference.

Complementing this efficient architecture, we integrate Coordinate Attention to refine features. Unlike standard channel attention, Coordinate Attention aggregates the $C \times T \times N$ skeletal graph features along distinct temporal (T) and spatial (N) directions, capturing long-range dependencies with minimal overhead. Additionally, a Confidence Gating mechanism is applied. By using the input confidence score c_n to modulate features via $X_{\text{out}} = X_{\text{gcn}} \odot \sigma(W \cdot c_n + b)$, the model effectively suppresses noise from occluded keypoints.

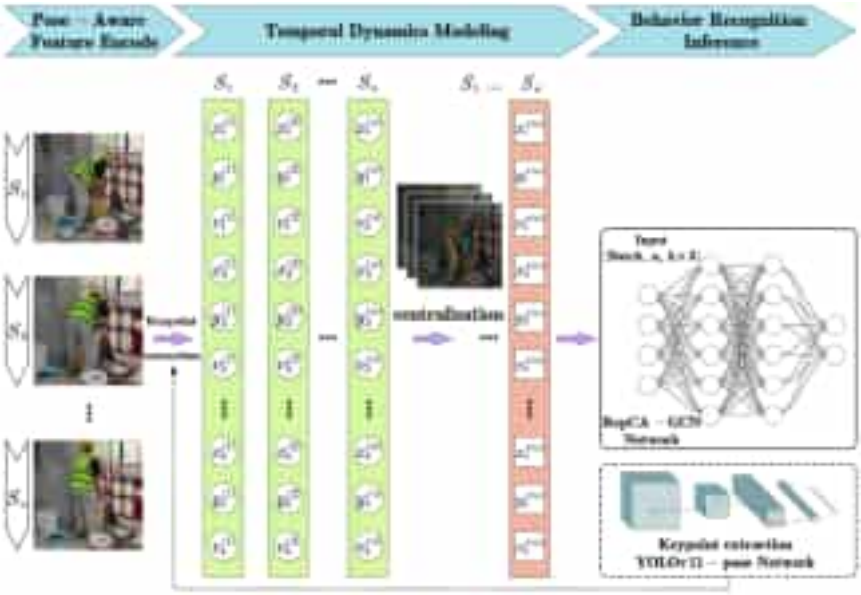


Figure 1. The Proposed Methodological Framework for Skeleton-Based Action Recognition

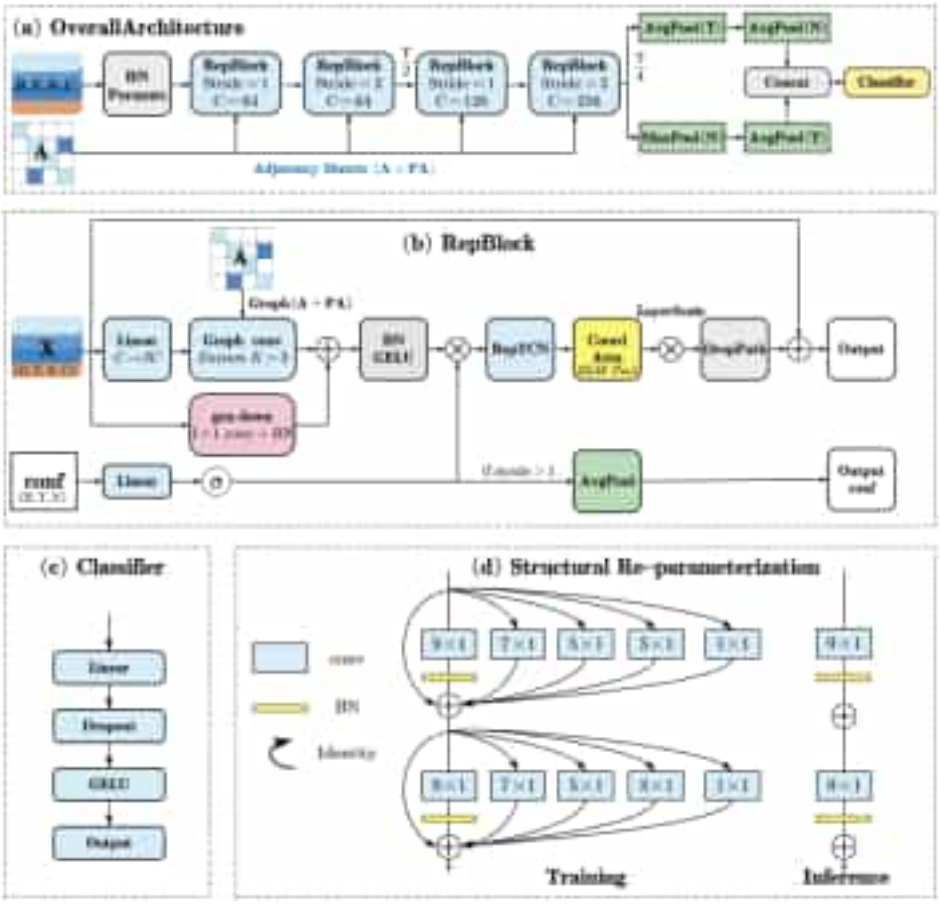


Figure 2. RepCA-GCN Model Architecture

3.2 Data Preparation and Workflow

The comprehensive workflow, ranging from data preparation to model training and practical recognition, is depicted in Figure 3. The process begins with the construction of the SiteWorker Action Recognition (SWAR) dataset. We collected video data from diverse sources, including real-world construction site surveillance, drone footage, and simulated environments, to ensure scenario diversity.

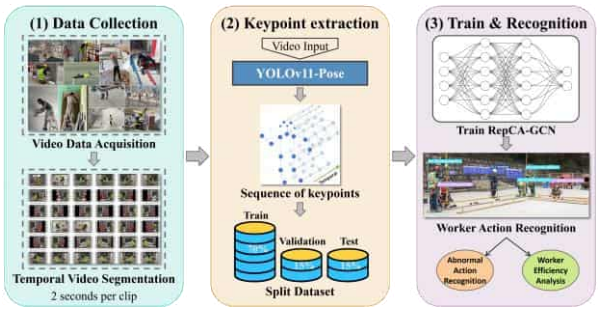


Figure 3. Workflow of data preparation, model training and recognition for the RepCA-GCN

As shown in Figure 4, the dataset is collected from actual construction sites, covering various environmental conditions and camera angles. It includes multiple categories of behaviors recorded under different lighting levels to reflect the complexity of real-world scenes.



Figure 4. SWAR dataset

The composition of the dataset is analyzed in Figure 5. The dataset exhibits a relatively balanced distribution across the six behavioral categories. 'Squatting operation' constitutes the largest portion (24.2%), while 'fall' represents the smallest category (13.1%), thereby mitigating the risk of severe class imbalance. The 'squatting operation' and 'standing operation' categories specifically denote the active construction states of workers in their respective postures. Most videos in the SWAR dataset are in 1080p resolution (1920 × 1080 or 1080 × 1920), with a small portion recorded at 720p. The specific number of video clips for each category is summarized in Table 1.

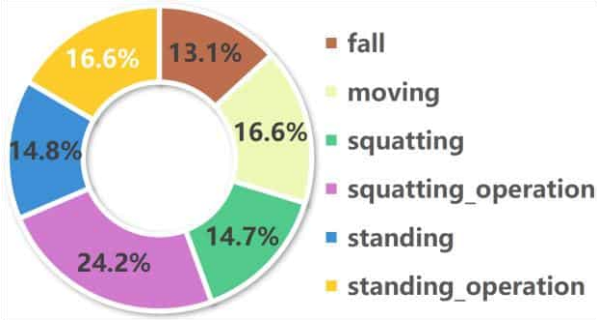


Figure 5. Composition of the SWAR Dataset

Table 1. Number of Video Clips in the SWAR Dataset

Behavior	Count
Fall	976
Moving	1230
Squatting	1091
Squatting operation	1794
Standing	1100
Standing operation	1233
Total	7424

For skeletal representation, we adopt the standard 17-keypoint configuration defined by the COCO format (Figure 6), with all coordinates undergoing normalization and centralization to ensure spatial consistency. Additionally, alternative 13-keypoint and 14-keypoint schemes (Figure 7) were explored and evaluated in the subsequent experiments.

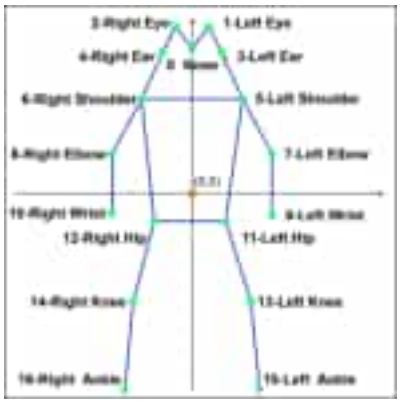


Figure 6. COCO 17 Keypoints Layout

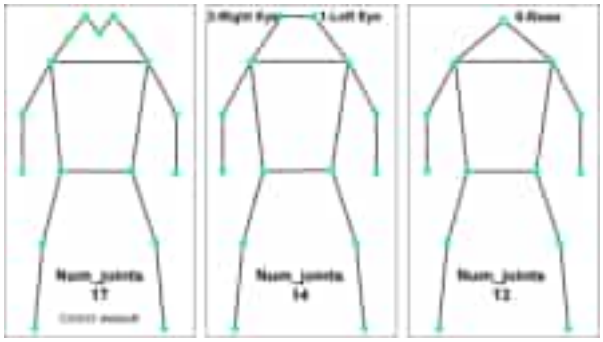


Figure 7. Three Keypoint Selection Schemes

4 Experiments and Results

4.1 Experimental Setup

The dataset partitioning and augmentation strategy are detailed in Table 2. The original SWAR dataset was split into training, validation, and testing sets with a ratio of 70:15:15. To strictly prevent data leakage and ensure realistic evaluation, we implemented a hybrid splitting strategy. Specifically, 30% of the validation and testing samples were randomly selected to ensure global data representation, while the remaining data were allocated sequentially based on video indices. Because raw videos capture continuous actions of identical subjects, this sequential allocation guarantees that frames from the same subject do not overlap across the training and testing sets.

Table 2. Dataset Partition

Subset	Original Samples	Augmented Samples
Train	5195	20780
Validation	1111	-
Test	1118	-
Total	7424	-

To enhance robustness against scale variations and frequent occlusions on construction sites, we exclusively applied three distinct data augmentation pipelines to the training set: (1) adaptive spatial scaling combined with a 50% probability of horizontal flipping; (2) scaling and flipping coupled with random occlusion; and (3) pure random occlusion. The occlusion operation intentionally masks 1 to 2 skeletal joints to simulate the noisy keypoints commonly generated by pose estimators in cluttered environments. This process expanded the training set to four times its original size.

All experiments were conducted on an Ubuntu workstation equipped with an NVIDIA RTX 4090 GPU using PyTorch 2.5.1. The models were trained for 50 epochs with a batch size of 128, optimized via AdamW with a cosine annealing learning rate schedule (decaying from 1e-3 to 1e-5) and gradient clipping for training stabilization.

4.2 Performance Analysis

The training dynamics of the proposed model are illustrated in Figure 8, demonstrating stable convergence without severe overfitting, owing to the rigorous anti-leakage data splitting and occlusion-aware augmentation. As presented in Table 3, the SWAR dataset under the new strict partition proves to be a challenging benchmark. RepCA-GCN achieves a highly competitive accuracy of 92.58%, outperforming the baseline ST-GCN (91.32%) and performing on par with the computationally heavy 2s-AGCN (92.47%).

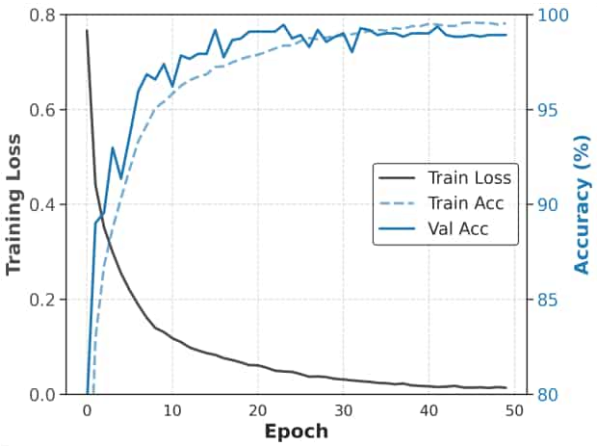


Figure 8. Training and validation results of RepCA-GCN

Table 3. Comparative performance of different models on the SWAR dataset (17 Keypoints)

Model	Acc(%)	Params(M)	FLOPs(G)	latency(ms)
ST-GCN	91.32	2.07	0.795	1.76
2s-AGCN	92.47	3.45	1.305	3.41
CTR-GCN	**93.11**	1.42	0.595	5.53
RepCA-GCN	92.58	**1.19**	**0.413**	**1.36**

Although slightly lower than CTR-GCN (93.11%), RepCA-GCN demonstrates a vastly superior speed-accuracy trade-off. By leveraging structural re-parameterization, it requires only 0.413 GFLOPs and achieves an ultra-low inference latency of 1.36ms, making it a well-suited choice for resource-constrained edge deployment.

Table 4. Accuracy (%) comparison across different models and keypoint configurations

Model	13 Points	14 Points	17 Points
ST-GCN	92.13	90.16	91.32
2s-AGCN	90.52	95.08	92.47
CTR-GCN	**92.31**	94.54	**93.11**
RepCA-GCN	92.22	**95.35**	92.58

Furthermore, we investigated the influence of skeletal representations on model performance (Table 4). Interestingly, RepCA-GCN achieves its peak accuracy of 95.35% when utilizing the 14-keypoint scheme, significantly surpassing the standard 17-keypoint COCO format (92.58%). This confirms that standard pose layouts may not be optimal for construction action recognition. The 17-keypoint format contains dense, highly localized facial keypoints that contribute little to macroscopic body actions but introduce significant noise when partially occluded by hardhats. The streamlined 14-keypoint scheme effectively filters out this redundant facial noise while preserving essential structural integrity, thereby maximizing the efficacy of our Confidence Gating mechanism and yielding the best recognition performance.

4.3 Ablation Study and Efficiency Analysis

As shown in Table 5, activating the structural re-parameterization deployment mode (ON) losslessly reduces parameters and FLOPs by ~40%, lowering latency to 1.36 ms without accuracy degradation.

Table 5. Efficiency comparison of RepCA-GCN before and after structural re-parameterization

Rep deploy	Params(M)	FLOPs(G)	latency(ms)
OFF	2.01	0.697	1.80
ON	**1.19**	**0.413**	**1.36**

This efficiency gain is mathematically guaranteed by the additivity of linear transformations. During training, the multi-branch temporal block outputs $Y = \sum_{i \in S} \mathrm{BN}_i\left(\mathrm{Conv}_i(X)\right)$, where $S = \{9 \times 1, 5 \times 1, 3 \times 1, 1 \times 1, \text{identity}\}$. For inference, each Conv-BN sequence is linearly fused into a single weight W_i' and bias b_i'. By zero-padding smaller kernels and the identity matrix (formulated as a 1×1 Dirac delta) to match the maximal 9×1 kernel (denoted by a padding operator $\mathcal{P}$), all branches equivalently collapse into a single generic convolution:

$$Y = \left(\sum_{i \in S} \mathcal{P}(W_i')\right) X + \sum_{i \in S} b_i' = W_{\text{fused}} X + b_{\text{fused}} \quad (1)$$

This structural decoupling ensures high training capacity while yielding an ultra-fast, single-path network for edge deployment.

5 Application and Evaluation

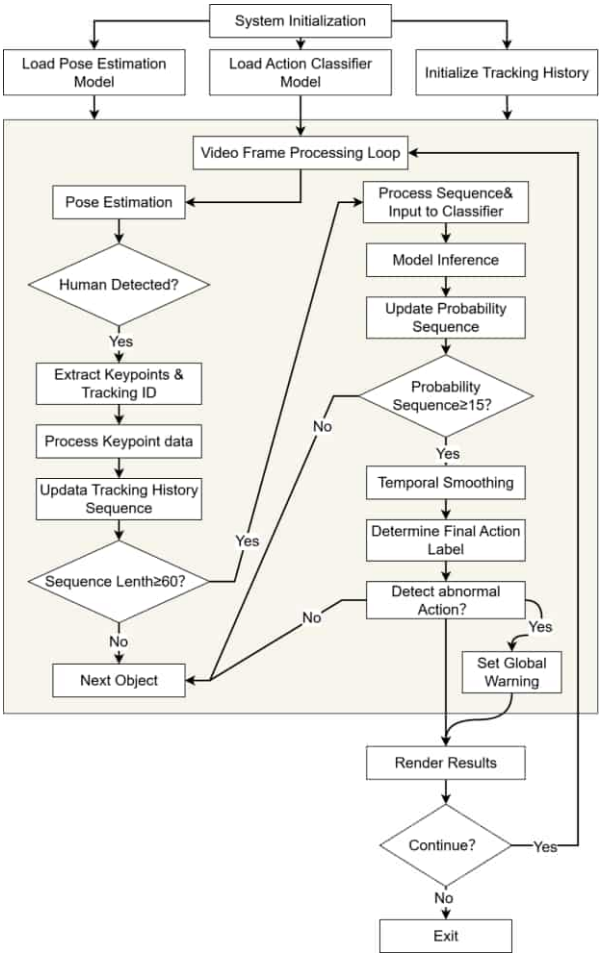


Figure 9. Flowchart of the proposed worker action recognition system for construction sites

To verify its practical utility, we developed a real-time monitoring system (Figure 9) that integrates YOLOv11-Pose for multi-object tracking with RepCA-GCN for action classification. The pipeline assigns unique IDs to detected individuals to extract and preprocess continuous skeletal keypoint streams.

5.1 Real-time Monitoring System Design and Visualization

For robust monitoring, a dual sliding-window mechanism—a 60-frame Temporal Buffer for context and a 15-frame Probabilistic Smoothing Window ($\alpha = 0.96$)—is implemented to filter transient noise and classification jitter. As visualized in Figure 10, the system accurately identifies diverse actions (e.g., "falls") and provides real-time visual feedback for immediate safety alerts.

5.2 System Performance Evaluation

Table 6 details the end-to-end throughput in multi-person scenarios, achieving 60.5 FPS on an RTX 4090 and 22.6 FPS on an RTX 4060. These baseline PyTorch metrics already satisfy real-time constraints, with further acceleration possible via TensorRT, proving RepCA-GCN an efficient and scalable solution for safety supervision.

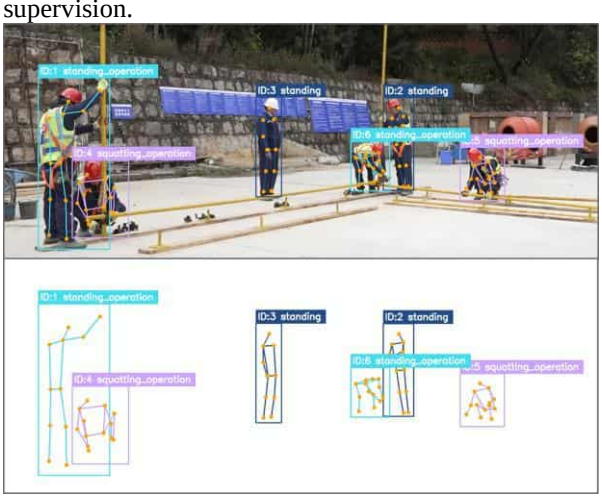


Figure 10. Visualization of Recognition Results

Table 6. End-to-End System Throughput (FPS) on Different Hardware Platforms

Model	Server (RTX 4090)			PC (RTX 4060)		
	Single-person	Multi-person	Avg	Single-person	Multi-person	Avg
ST-GCN	65.1	41.5	53.3	**24.9**	18.9	21.9
2s-AGCN	52.1	27.7	39.9	23.9	15.2	19.6
CTR-GCN	47.6	24.0	35.8	18.9	10.9	14.9
RepCA-GCN	**70.6**	**50.3**	**60.5**	24.1	**21.0**	**22.6**

6 Conclusion and Limitations

This study proposes RepCA-GCN, a lightweight framework for construction worker action recognition. By integrating Structural Re-parameterization and reliability-aware feature modulation, the model addresses computational latency and noisy keypoints caused by occlusions. We constructed the SWAR dataset with extensive augmentation to facilitate robust training. Results show RepCA-GCN achieves state-of-the-art accuracy (95.35%) with reduced computational costs (0.413 GFLOPs) and low latency (1.36 ms). The monitoring system demonstrates robust real-time performance for intelligent safety supervision. However, skeleton-based methods completely discard background information, making them unable to recognize environment-interactive actions.

Acknowledgments

The authors would like to acknowledge the support by National Science Foundation of China (52308314), the support by Guangdong Basic and Applied Basic Research Foundation (2023A1515030169) and the support by Guangzhou Science and Technology Project (2025A04J5216).

References

[1] B. Choi, S. Ahn, S. Lee, Construction Workers' Group Norms and Personal Standards Regarding Safety Behavior: Social Identity Theory Perspective, J. Manage. Eng. 33 (2017) 04017001. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000511.

[2] M.A. Nabi, I.H. El-adaway, C. Dagli, A System Dynamics Model for Construction Safety Behavior, Procedia Computer Science 168 (2020) 249–256. https://doi.org/10.1016/j.procs.2020.02.254.

[3] S. Chi, C.H. Caldas, Automated Object Identification Using Optical Video Cameras on Construction Sites, Comput.-Aided Civil Infrastruct. Eng. 26 (2011) 368–380. https://doi.org/10.1111/j.1467-8667.2010.00690.x.

[4] A. Kelm, L. Laussat, A. Meins-Becker, D. Platz, M.J. Khazaee, A.M. Costin, M. Helmus, J. Teizer,

Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites, Autom. Constr. 36 (2013) 38–52. https://doi.org/10.1016/j.autcon.2013.08.009.

[5] I. Awolusi, E. Marks, M. Hallowell, Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices, Autom. Constr. 85 (2018) 96–106. https://doi.org/10.1016/j.autcon.2017.10.010.

[6] S. Subedi, N. Pradhananga, Sensor-based computational approach to preventing back injuries in construction workers, Autom. Constr. 131 (2021) 103920. https://doi.org/10.1016/j.autcon.2021.103920.

[7] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. Sheikh, OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 172–186. https://doi.org/10.1109/TPAMI.2019.2929257.

[8] D. Maji, S. Nagori, M. Mathew, D. Poddar, YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2022, IEEE, New York, 2022: pp. 2636–2645. https://doi.org/10.1109/CVPRW56347.2022.00297.

[9] H. Wu, Y. Han, M. Zhang, B.D. Abebe, M.B. Legesse, R. Jin, Identifying Unsafe Behavior of Construction Workers: A Dynamic Approach Combining Skeleton Information and Spatiotemporal Features, J. Constr. Eng. Manage. 149 (2023) 04023115. https://doi.org/10.1061/JCEMD4.COENG-13616.

[10] X. Zhou, S. Li, J. Liu, Z. Wu, Y.F. Chen, Construction Activity Analysis of Workers Based on Human Posture Estimation Information, Engineering 33 (2024) 225–236. https://doi.org/10.1016/j.eng.2023.10.004.

[11] S. Yan, Y. Xiong, D. Lin, Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition, in: Thirty-Second AAAI Conference on Artificial Intelligence / Thirtieth Innovative Applications of Artificial Intelligence Conference / Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, Assoc Advancement Artificial Intelligence, Palo Alto, 2018: pp. 7444–7452.

[12] L. Shi, Y. Zhang, J. Cheng, H. Lu, Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), IEEE, New York, 2019: pp. 12018–12027. https://doi.org/10.1109/CVPR.2019.01230.

[13] Y. Chen, Z. Zhang, C. Yuan, B. Li, Y. Deng, W. Hu, Channel-wise Topology Refinement Graph Convolution for Skeleton-Based Action Recognition, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV 2021), IEEE, New York, 2021: pp. 13339–13348. https://doi.org/10.1109/ICCV48922.2021.01311.

[14] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, J. Sun, RepVGG: Making VGG-style ConvNets Great Again, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021, IEEE Computer Soc, Los Alamitos, 2021: pp. 13728–13737. https://doi.org/10.1109/CVPR46437.2021.01352.

[15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New York, 2016: pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

[16] N.D. Nath, A.H. Behzadan, S.G. Paal, Deep learning for site safety: Real-time detection of personal protective equipment, Autom. Constr. 112 (2020) 103085. https://doi.org/10.1016/j.autcon.2020.103085.

[17] L. Lopez, J. Suarez-Ramirez, M. Aleman-Flores, N. Monzon, Automated PPE compliance monitoring in industrial environments using deep learning-based detection and pose estimation, Autom. Constr. 176 (2025) 106231. https://doi.org/10.1016/j.autcon.2025.106231.

[18] J. Carreira, A. Zisserman, Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset, in: 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), IEEE, New York, 2017: pp. 4724–4733. https://doi.org/10.1109/CVPR.2017.502.

[19] C. Feichtenhofer, A. Pinz, A. Zisserman, Convolutional Two-Stream Network Fusion for Video Action Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New York, 2016: pp. 1933–1941. https://doi.org/10.1109/CVPR.2016.213.

[20] C. Feichtenhofer, H. Fan, J. Malik, K. He, SlowFast Networks for Video Recognition, in: 2019 IEEE/CCF International Conference on Computer Vision (ICCV 2019), IEEE, New York, 2019: pp. 6201–6210. https://doi.org/10.1109/ICCV.2019.00630.

Automatic Concrete Pipe Material Detection from Mobile Laser Scanning Data Using Transfer Learning

Kartika Nur Rahma Putri[1,2], Khalegh Barati[2], Xuesong Shen[2], James Linke[3]

[1]Department of Civil and Environmental Engineering, Universitas Gadjah Mada, Indonesia
[2]School of Civil and Environmental Engineering, University of New South Wales, Australia
[3]GeoAI, Sydney, NSW 2019, Australia

kartikanurrahmaputri@ugm.ac.id, khalegh.barati@unsw.edu.au, x.shen@unsw.edu.au, j.linke@geoai.au

Abstract

The current prefabricated material tracking method in construction projects usually uses RFID or GPS technology. This technique requires tagging materials, which can be labour-intensive for large volumes of materials on a construction site. Deep learning-based object detection can provide a solution for material tracking in construction projects without the need for material tagging. To obtain this result, a deep learning architecture must be trained to automatically detect construction materials. However, there is a limited dataset of construction material objects in outdoor environments for training. Therefore, this study proposes a framework for material detection at construction sites using a modified dataset and transfer learning from other LiDAR data mainly used for autonomous driving (e.g., cars, pedestrians, and cyclists). The method includes workflows for data acquisition using a Mobile Laser Scanner (MLS) system, data preprocessing, training, and testing. A site-specific case study using concrete pipe material was used to validate the proposed approach. The technical performance of the material object detection model trained with transfer learning increased from 35.40% F1 score to 80.56% F1 score. Furthermore, with an additional modified dataset, the F1 score increased to 87.18%.

Keywords

Construction material detection; 3D point cloud; mobile laser scanning; transfer learning; deep learning

1 Introduction

Tag-based material-tracking technologies, such as Radio Frequency Identification (RFID), require significant manual effort and are often prone to errors and inefficiencies. For instance, workers must manually attach tags to materials and traverse construction sites to scan and extract information at close range. On large construction sites, this process becomes time-consuming and error-prone because of the widespread distribution of materials [1]. Furthermore, positioning technologies such as RFID, ultra-wideband (UWB), infrared, and Global Positioning System (GPS) often depend on additional infrastructure such as Wi-Fi, cellular networks, or Bluetooth to function effectively. This dedicated infrastructure requirement increases the complexity and cost of tag-based methods.

Considering the current challenges of using RFID, UWB, and GPS, which require manual tagging and dedicated infrastructure for data transfer, the application of visual perception technology has gained significant interest. Some technologies, such as image-based technology and point cloud processing, do not require dedicated infrastructure [2]. Although image-based methods for object detection are promising, they only capture 2-dimensional (2D) data, making it difficult to obtain information about the material's 3-dimensional (3D) size. To address this, point cloud data is a possible alternative, as it provides detailed 3D information about the material at a construction site. However, processing large point cloud datasets often requires significant computational power, which may delay real-time applications of the technology.

The current point cloud data used in construction projects is mostly generated by Terrestrial Laser Scanners (TLS), which produce highly accurate, dense point clouds. The TLS method is considered more suitable for scanning small and dense areas. However, for an infrastructure project with a large area, such as a toll road or airport, TLS data generation can be too slow. On the other hand, Mobile Laser Scanning (MLS) provides faster data acquisition for large areas [3], but faces challenges with occlusion and sparse point clouds.

To address the challenges of detecting construction materials from mobile laser scanning (MLS) data, this study proposes a transfer-learning-based framework. A real dataset from a large infrastructure project was used

to evaluate the proposed approach's feasibility. The data were collected using an MLS system and pre-processed through subsampling, segmentation, and data labelling.

A Point Voxel Region-based Convolutional Neural Network (PV-RCNN) [4] was used for 3D object detection. Transfer learning was applied by initialising the model with weights pre-trained on an autonomous driving dataset and then fine-tuning it using the construction MLS data. The results were compared with a model trained from scratch using the same dataset to evaluate the benefit of transfer learning.

2 Related Work

This section provides an overview of current material object detection in construction projects. It highlights relevant studies on the processing of 3D point cloud data and on the application of deep learning methods for the detection of construction materials.

2.1 Construction Material Recognition Method

To support material recognition in construction, several technologies, including 3D object detection and vision-based scene understanding, have been applied in construction projects. One study demonstrated a vision-based method for material recognition to automate construction progress monitoring and generate Building Information Models (BIM) from unordered site image collections [5]. This approach utilised 3D point cloud models created from site photo logs using Structure-from-Motion (SfM) techniques. The primary aim was to classify material types based on appearance to facilitate progress tracking. This approach primarily targets materials that have already been installed on site, which limits its applicability for detecting and managing on-site material inventory.

In terms of material quantification, one recent research developed a method for automatic measurement of corrugated pipes and rebars for large precast concrete beams from a TLS [6]. The segmentation framework was based on the application of radius nearest-neighbours, covariance, and a density-based spatial clustering algorithm to detect pipes and rebar and to extract their sizes, including radius and length. Although the algorithm accurately extracted the object size with 98% accuracy, the method still required a manual clutter-removal stage and relied on dense TLS data.

Previous work introduced a vision-based machine learning method for counting steel pipes on construction sites [7]. This method utilised a Convolutional Neural Network (CNN) combined with image processing technologies to enhance counting accuracy and reduce labour intensity. The primary aim was to automate the counting process to improve efficiency and accuracy. Despite its effectiveness, the method may face challenges in highly cluttered environments where steel pipes overlap significantly.

3D point cloud object detection methods in the construction industry, including vision-based recognition, semantic segmentation, and volumetric measurement, have significantly enhanced monitoring and material management. However, challenges such as occlusion, noise, and data completeness, especially for the MLS point cloud, remain unresolved. Therefore, further research and development to improve the robustness of these techniques are needed to ensure their applicability across diverse construction environments.

2.2 3D Point Cloud Processing

In automatic 3D object detection for point cloud processing, three methods are generally used. The methods are simple geometric models, feature descriptors, and deep learning [8]. The first method assumes that objects can be detected by utilising simple geometric models, such as planes and cylinders, from point cloud data. This method usually uses random sample consensus (RANSAC), which computes the least-squares error between the known object and the points in the point cloud, or the Hough transform, which searches for the object using its parametric shapes, such as the radius or slope-intercept.

The second method uses feature descriptors divided into local and global categories. The local descriptor tries to summarise statistics of the object's features, such as curvature, density, and normal, to describe the neighbourhoods around interest points. Meanwhile, the global descriptors define objects using their lengths, areas, and angles in a single feature vector. Examples of research using global descriptors include [9] that uses the Principal Axes Descriptor to automatically detect construction equipment from a TLS point cloud. In the context of material detection, recent studies have explored the use of geometric feature extraction from mobile laser scanning point clouds to enable tracking and volume estimation of precast concrete pipes [10]. While this method achieved 78.5% accuracy, its performance indicates room for improvement in material object detection.

The last method of object detection uses deep learning to find a direct mapping from the 3D point cloud input to its object labels. The deep learning method requires extensive training on a large number of objects before it can correctly predict the type of object to be detected. Compared to traditional geometric feature extraction, deep learning offers advantages in robustness. One key benefit of deep learning is that it minimises the need for manually engineered feature descriptors and employs more flexible yet robust classification rules [8].

However, there are some concerns about the applicability of machine learning data, especially related to the amount of data required to train the model.

In the construction industry, the most publicly available point cloud dataset with its labelling is an indoor dataset, such as the Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS) [11] and the ScanNet [12] dataset. Meanwhile, another point cloud dataset in outdoor settings is usually used for the autonomous driving process, such as the Karlsruhe Institute of Technology and the Toyota Technological Institute (KITTI) [13]. This dataset class usually contains cars, pedestrians, trees, and cyclists.

Despite growing interest in point cloud technology for construction applications, its use for automated material tracking remains limited. Most existing studies focus on progress monitoring using TLS. While MLS, which is faster and more practical for large infrastructure, generates sparse and occluded point clouds that are more challenging to process. Furthermore, most existing 3D object detection algorithms and publicly available datasets are developed for autonomous driving, covering common classes such as cars and pedestrians, rather than construction-specific materials such as pipes, concrete, and steel. Existing MLS-based approaches for pipe detection have also shown limited accuracy. As a pilot project, this study aims to develop and test a deep learning framework trained on custom MLS data for automated concrete pipe detection under the challenging conditions of single-pass MLS scanning.

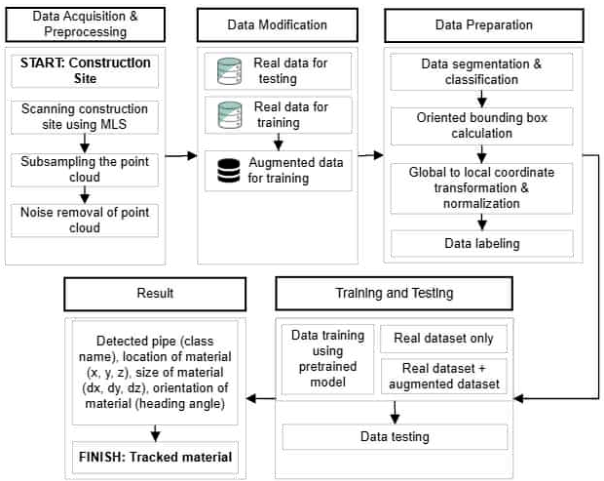


Figure 1. A step-by-step method for 3D point cloud object detection in a construction project.

3 Methods

The proposed methodology aims to detect prefabricated concrete pipes from MLS point cloud data. Figure 1 illustrates the 3D material object detection step-by-step method used in this paper. Point cloud data from a real construction site was used as a case study. From this dataset, only 8 areas can be used. To increase the amount of training data, a preprocessing stage was introduced. Then, a transfer learning approach for 3D object detection was used to detect the pipe.

3.1 Data Acquisition

The 3D point cloud data was obtained using MLS in a single pass with the specifications mentioned in Table 1. MLS was selected in this study due to its fast data acquisition capability, which allows large areas to be scanned in a single continuous pass. Since the MLS system performs only one scan along the road, many of the data is dense on one side and sparse on the other. The resulting data was large, noisy, and mostly occluded, as can be seen in Figure 2. Although multi-pass scanning could improve point density, it was not possible due to this constraint. Furthermore, single-pass data collection reflects how MLS is used in practice, where fast data acquisition is one of its main advantages. Testing our method under these conditions shows that it can work well in real inspection scenarios. To obtain data for labelling and training, the area containing prefabricated concrete pipes was segmented into distinct regions. The initial data for each area includes concrete pipes, ground points, other materials around the pipes, and the scene.

Table 1. Mobile laser scanner specification.

Criteria	Specification
Laser Sensor	XT32
Range Accuracy	±3cm
Vertical Field of View	-16° ~ 15°
Horizontal Field of View	0° – 360°
Maximum Range	120 m



Figure 2. 3D point cloud data.

Point cloud subsampling was performed to reduce the number of points and reduce computational requirements. The subsampling method was obtained using open-source cloud-compare software. To ensure that point features are not removed, the subsampling method used a spatial mode with a minimum distance of 0.02 meters between points. This subsampling method reduced up to 50% of the points for each area.

3.2 Data Preprocessing and Modification

The dataset preparation was obtained to increase the number of areas for data training. From the data-acquisition stage, 8 areas containing prefabricated concrete pipes formed the real dataset for this case study as can be seen in Figure 3. The real dataset was partitioned into training and test sets, with six areas allocated to training and two to testing. To ensure the integrity of the evaluation process, the testing data were entirely excluded from both the training phase and the generation of modified data. This approach prevents the model from acquiring any prior knowledge of the testing dataset's properties.

The additional dataset was obtained by modifying the current dataset. The ground points of each dataset were segmented and deleted to get the modified dataset (Figure 4). Next, only the good pipe from each dataset is segmented without the ground points and scene (Figure 5). Both modifications to the datasets aim to train the model to recognise the correct pipe in a clean environment. From this stage, the training data increased from 6 to 18 areas.



Figure 3. Example of the current dataset.



Figure 4. Current dataset without ground point.



Figure 5. Good pipes only dataset.

3.3 Dataset Preparation

All areas were prepared to match the algorithm. The data was segmented into pipes and scenes. A bounding box was generated for each pipe and the scene. An oriented bounding box (OBB) was used to capture the size and orientation of each pipe. Since the concrete pipes are located on the ground, only the heading angle (also known as the yaw angle) was calculated. The pitch and roll angles were assumed to be zero [13]. The heading angle describes the rotation of the object around the vertical (z) axis and indicates the direction the object faces horizontally. The illustration of the object bounding box, coordinates and heading angle can be seen in Figure 6. Each object was labelled with its centre coordinate (x, y, z), 3D size generated from its bounding box (length: dx, width: dy, height: dz), heading angle, and class name.

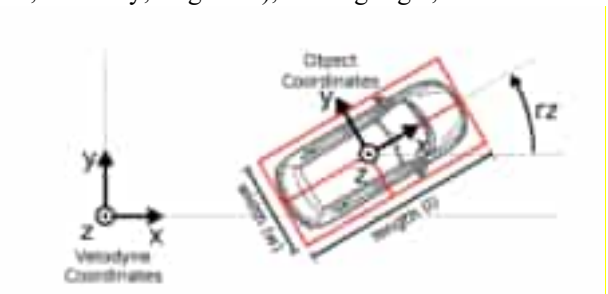


Figure 6. Object coordinates illustration.

Given that the PV-RCNN algorithm utilises both point and voxel features, the intensity data must also be included in the point cloud dataset. Additionally, the data needs to be verified for compliance with the point cloud range and voxel size specified in the OpenPCDET custom dataset information. Each point from the laser scanner data contains information about its global coordinates. Since the area is divided into regions containing pipes, the coordinate information was transformed to local coordinates. To ensure the point cloud range was not too large, each data point was transformed again so that the first point in each point cloud was located around the 0, 0, 0 coordinate (normalisation).

3.4 Dataset Training and Testing

The methodology described was implemented in a Python environment on a Personal Computer with an Intel(R) Core (TM) i7-6700 CPU@ 3.4 GHz processor and installed RAM of 64 GB. The model training time ranged from 20 minutes to 1 hour depending on the dataset size and configuration and to evaluate a single scan took approximately 30 to 40 seconds to process. The primary library used for 3D processing was Open3D. From 8 real datasets, areas 1 and 5 were selected as the test datasets. The baseline for object detection was calculated by training only on a real dataset (6 areas) and testing on 2 areas. The next scenario involved adding a pre-trained model to the dataset and an additional modified dataset for training.

The model's performance was evaluated using precision, recall, and F1 scores as defined in equations 1, 2, and 3. Precision measures the proportion of detected objects that are correct, calculated as the ratio of true positives (TP) to the sum of true positives (TP) and false positives (FP). Recall measures how many of the actual objects are correctly detected, defined as the ratio of true positives to the sum of true positives and false negatives (FN). In this context, TP represents correctly detected objects, FP refers to incorrect detections, and FN indicates objects that were present but missed by the model. The F1 score combines precision and recall into a single metric by computing their harmonic mean.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

4 Results and Discussion

Each dataset configuration was trained separately, and performance was measured at each epoch. This analysis highlights variations in results across different data types and demonstrates the impact of pre-trained models on 3D object detection performance.

4.1 Real Dataset

From 8 real datasets, Areas 1 and 5 were selected as the test datasets. Area 1 represents a region with high-quality pipe data and minimal noise, while Area 5 contains more occlusions and noise. As a baseline, the model was trained on only six real datasets and evaluated on the designated test data. The Ground Truth (GT) bounding box, shown in Figure 7 indicates that Area 1 contains 14 pipes and Figure 8 indicates that Area 5 contains 19 pipes.



Figure 7. Ground truth for Area 1

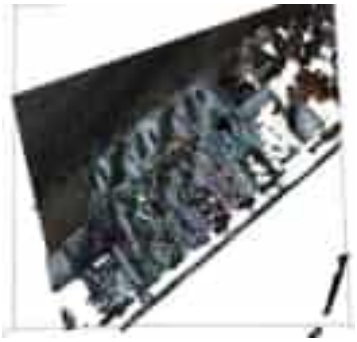


Figure 8. Ground truth for Area 5

In the first experiment, PV-RCNN was trained using six real MLS datasets without pretraining. The baseline detection results, shown in Figure 9 (a) and Figure 9 (b), indicate poor performance due to the limited number of training samples. The model failed to accurately detect pipe objects, with predicted bounding box centres and orientations misaligned from the ground truth. In addition, clutter was frequently misclassified as pipes, and several actual pipes were missed, leading to unreliable detections.

Intersection over Union (IoU) measures the overlap between a predicted bounding box and the ground truth to evaluate object detection accuracy. The IoU 50% means detection is considered correct if at least 50% of the predicted box overlaps with the ground truth. In terms of quantitative performance, object detection using only real data resulted in a precision of 25% and a recall of 60.6% for IoU 50%. Although no pipes were correctly detected according to the visualisation criteria, the precision, recall, and F1 scores at the 50% threshold were non-zero. This is because some predicted bounding boxes partially overlapped with ground-truth pipes and were counted as true positives under the relatively loose 50% IoU criterion, despite their inaccurate geometry.

After incorporating a pre-trained model, detection performance improved substantially. In Area 1, the model correctly detected 13 out of 14 pipes, with bounding box locations, dimensions, and orientations closely aligned with the ground truth, as shown in Figure Figure 9 (c). In the more challenging Area 5, which contains higher noise and clutter, the model detected 17 out of 19 pipes, with two clutter objects misclassified as pipes as can be seen in Figure 9 (d). When the pre-trained model was introduced, precision improved to 72.73%, while recall increased to 90.28%. The recall of 90.28% indicates that the model correctly identified most of the actual pipes, though some remained undetected. These results highlight the effectiveness of integrating a pre-trained model with real data to improve object detection performance compared to the baseline approach.

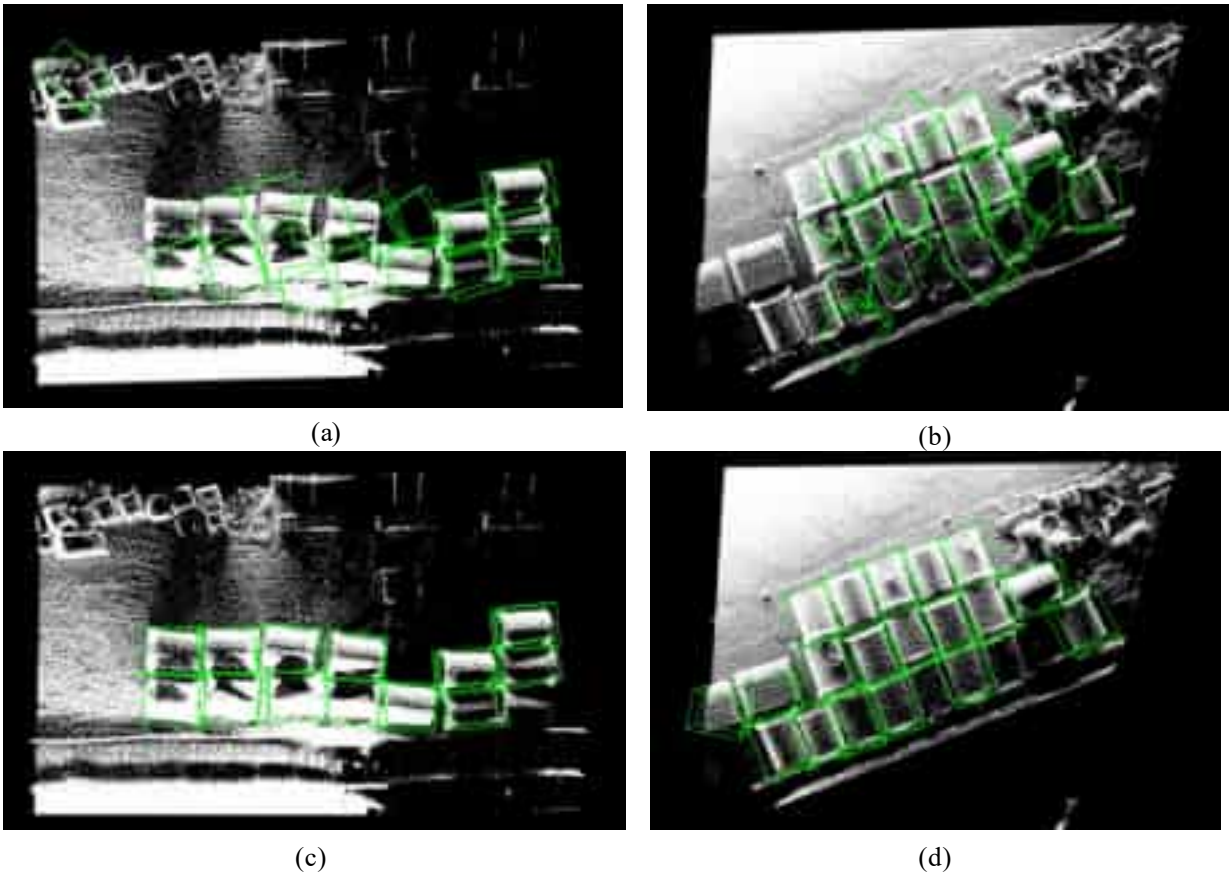


Figure 9. Predicted bounding box from 6 real data: (a) Area 1 without pre-trained model; (b) Area 5 without pre-trained model; (c) Area 1 with pre-trained model; and (d) Area 5 with pre-trained model.

4.2 Real Dataset plus Modified Dataset

The modified datasets were generated to increase the amount of training data. As a result, an additional 12 data points were added to the real data. The remaining areas were then trained using a pre-trained PV-RCNN model for 80 epochs. As shown in Figure 10, the introduction of additional modified datasets significantly improved the results. The algorithm successfully detected 100% of pipes in both Area 1 (14 out of 14 pipes) and Area 5 (19 out of 19 pipes), with no misclassifications. The incorporation of modified datasets into the object detection model increased the F1 score from 80.56% to 87.18%. While the results for Area 1 showed minimal differences between using only real datasets and real plus modified datasets in terms of the number of detected pipes and bounding box size, the improvements are more evident in Area 5. The bounding box locations and orientations in Area 5 became more aligned with the ground truth, and fewer objects were misclassified. The result of object detection performance for all datasets can be seen in Table 23.

The results at IoU 70% are slightly lower than those at IoU 50%, which is expected since a stricter threshold requires more precise bounding box localization. Nevertheless, both thresholds show a consistent improvement in detection performance. At IoU 50%, the F1 score improves from 35.40% in the baseline to 87.18% in the best configuration. Similarly, at IoU 70%, the F1 score increases from 0.00% to 81.97%. This result demonstrates the effectiveness of integrating modified data in refining the model's performance. The poor performance of the baseline model highlights the sensitivity of the proposed methods to the availability of labelled training data. Detection performance also varied across test areas, depending on scene complexity.

Table 2. 3D object detection performance for all datasets.

Real Data	Modified Data	Training type	Precision (50%)	Recall (50%)	F1 (50%)	Precision (50%)	Recall (50%)	F1 (50%)
6 real data	-	-	25.00%	60.60%	35.40%	0.00%	3.03%	0.00%

Real Data	Modified Data	Training type	Precision (50%)	Recall (50%)	F1 (50%)	Precision (50%)	Recall (50%)	F1 (50%)
(baseline)								
6 real data	-	Pretrained model	72.73%	90.28%	80.56%	50.17%	81.81%	62.20%
6 real data	6 modified data	Pretrained model	72.42%	93.93%	81.78%	36.02%	69.69%	47.49%
6 real data	12 modified data	Pretrained model	77.27%	100.00%	87.18%	77.27%	87.27%	81.97%

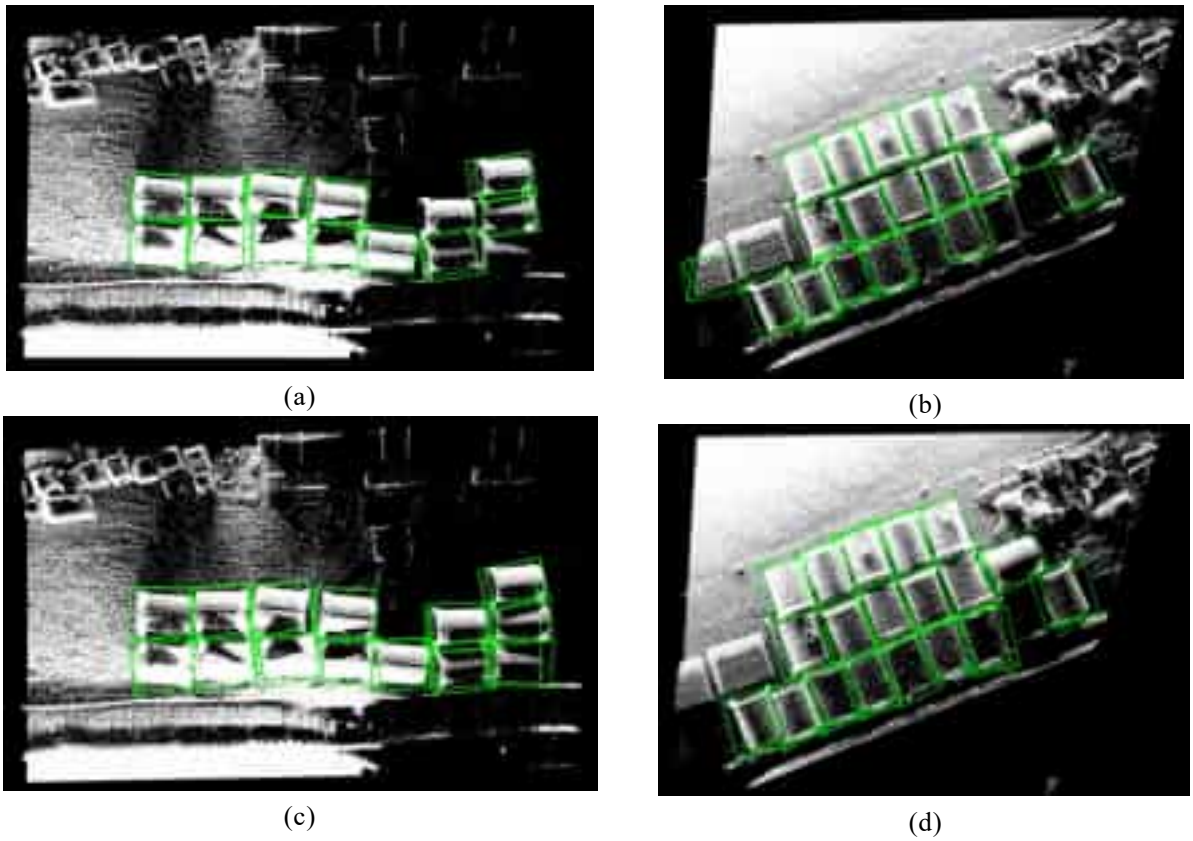


Figure 10. Predicted bounding box from real and modified data with pre-trained model: (a) Area 1 from real and 6 modified data; (b) Area 5 from real and 6 modified data; (c) Area 1 from real and 12 modified data; and (d) Area 5 from real and 12 modified data.

5 Conclusion and Future Work

This research focuses on the application of 3D point clouds for on-site detection of construction materials. The main challenge was the limited actual data and occluded point cloud generated from MLS. This paper addresses the gap by proposing a material-detection method based on transfer learning and modified datasets.

A case study was conducted to verify the proposed method. Two datasets were used: the real dataset and the modified dataset. From the case study results, 3D point cloud object detection improves from 35.40% F1 score on the real dataset without a pre-trained model to 80.56% with a pre-trained model. The results demonstrate that transfer learning combined with dataset preparation can substantially improve pipe detection performance when only a limited amount of real MLS data is available. This improvement is largely due to the use of pre-trained models trained on autonomous driving datasets, which provide strong geometric priors due to their relatively similar scanning conditions. Additionally, the introduction of modified datasets further enhances the F1 score to 87.18%. This result suggests that using additional modified datasets can improve the model's performance.

Although the method shows promising results for material object detection, this study has several limitations. First, the number of testing areas is limited. To comprehensively assess the model's accuracy, additional areas, including those without pipes, should be tested. Second, the detected material is restricted to a single type and size. Future research should explore different material types to evaluate the algorithm's performance across a wider range of conditions. Third, the parameters used in this study, such as subsampling threshold, model hyperparameters, and other preprocessing settings, were not fully optimized. Only a single configuration was tested, and future work should explore the sensitivity of the method to different parameter settings to better understand their impact on detection performance. Additionally, the data used in this study is limited to 3D point clouds from MLS only. Further research could incorporate other data sources, such as images, videos, and 3D point clouds from airborne laser scanners, to assess whether integrating multiple data types improves detection accuracy.

Acknowledgement

The authors are grateful to BPPT (Center for Higher Education Funding), Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, LPDP (Indonesia Endowment Fund for Education) [Grant No. 1253/J5/KM.01.00/2021]. This research is also supported by funding from the Australian Research Council (ARC) Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), and the ARC Research Hub for Human-Robot Teaming for Sustainable and Resilient Construction (IH240100016). The laser-scanning data provided by Linke & Linke Surveys is greatly appreciated.

References

[1] S. Kang, M. W. Park, and W. Suh, "Feasibility study of the unmanned-aerial-vehicle radio-frequency identification system for localizing construction materials on large-scale open sites," *Sensors Mater.*, vol. 31, no. 5, pp. 1449–1465, 2019, doi: 10.18494/SAM.2019.2266.

[2] A. Alarifi *et al.*, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors (Switzerland)*, vol. 16, no. 5, pp. 1–36, 2016, doi: 10.3390/s16050707.

[3] B. Yang, Y. Zang, Z. Dong, and R. Huang, "An automated method to register airborne and terrestrial laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 109, pp. 62–76, 2015, doi: 10.1016/j.isprsjprs.2015.08.006.

[4] S. Shi *et al.*, "PV-RCNN: Point-voxel feature set abstraction for 3D object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10526–10535. doi: 10.1109/CVPR42600.2020.01054.

[5] A. Dimitrov and M. Golparvar-Fard, "Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections," *Adv. Eng. Informatics*, vol. 28, no. 1, pp. 37–49, 2014, doi: 10.1016/j.aei.2013.11.002.

[6] J. Shu, X. Zhang, W. Li, Z. Zeng, H. Zhang, and Y. Duan, "Point cloud and machine learning-based automated recognition and measurement of corrugated pipes and rebars for large precast concrete beams," *Autom. Constr.*, vol. 165, no. June, 2024, doi: 10.1016/j.autcon.2024.105493.

[7] Y. Li and J. Chen, "Computer Vision–Based Counting Model for Dense Steel Pipe on Construction Sites," *J. Constr. Eng. Manag.*, vol. 148, no. 1, pp. 1–16, 2022, doi: 10.1061/(asce)co.1943-7862.0002217.

[8] J. Chen, Z. Kira, and Y. K. Cho, "Deep Learning Approach to Point Cloud Scene Understanding for Automated Scan to 3D Reconstruction," *J. Comput. Civ. Eng.*, vol. 33, no. 4, pp. 1–10, 2019, doi: 10.1061/(asce)cp.1943-5487.0000842.

[9] J. Chen, Y. Fang, Y. K. Cho, and C. Kim, "Principal Axes Descriptor for Automated Construction-Equipment Classification from Point Clouds," *J. Comput. Civ. Eng.*, vol. 31, no. 2, p. 04016058, 2017, doi: 10.1061/(asce)cp.1943-5487.0000628.

[10] K. N. R. Putri, X. Shen, M. Aleksandrov, K. Barati, and J. Linke, "Acquiring Semantic Information of Precast Concrete Pipe Using Geometric Feature Extraction from Mobile Laser Scanning Point Cloud," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2024, vol. X–4, no. October, pp. 263–270.

[11] I. Armeni *et al.*, "3D Semantic Parsing of Large-Scale Indoor Spaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1534–1543. doi: 10.1109/CVPR.2016.170.

[12] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 2432–2443. doi: 10.1109/CVPR.2017.261.

[13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research," *Int. J. Rob. Res.*, no. October, pp. 1–6, 2013.

A Multi-Camera Human Tracking Framework for Safe Human-Robot Collaboration in Construction-Site Robotic Arm Operations

Ren-Jie Li [1], Meng-Han Tsai [1], Liang-Ting Tsai [2], Yuxiang Chen [2] and Ci-Jyun Liang [3]

[1] National Taiwan University of Science and Technology, Taipei, Taiwan
[2]University of Alberta, Edmonton, Alberta, Canada
[3]Stony Brook University, Stony Brook, New York, USA
d11105003@mail.ntust.edu.tw, menghan@mail.ntust.edu.tw, liangtin@ualberta.ca, yuxiang.chen@ualberta.ca, cjliang@umich.edu

Abstract

The construction industry is characterized by a high degree of structural complexity, making it difficult to achieve a high level of modularization at the current stage. During operations involving robotic arms, human operators are often required to enter the robot's workspace at certain stages to perform fine adjustments, which may expose them to safety risks. Therefore, this study aims to develop a human-robot collaboration safety system framework that utilizes multi-view cameras to monitor, in real time, the positions of both the robotic arm and the operator. Object detection techniques are applied to streaming video for human tracking, while human pose estimation is used to analyze body posture. In addition, multi-view trajectory integration is employed to fuse information from different camera perspectives, enabling the system to determine whether the operator has entered a hazardous operating zone of the robotic arm. Ultimately, the proposed system provides real-time risk alerts while minimizing unnecessary interruptions to normal construction workflows, with emergency stop triggered when safety is compromised. This design reduces the likelihood of accidents caused by unintended entry into dangerous areas and enhances overall safety in human-robot collaborative operations.

Keywords -

Human-robot collaborative; Safety monitoring; Object detection; Human Pose Estimation

1 Introduction

The use of robotic arms has been widely adopted in many highly modular industries, such as automotive manufacturing, electronics assembly, and warehouse logistics. However, the introduction of robotic arms into the construction industry faces far greater challenges. In addition to the inherent complexity of construction projects, even components of the same type may vary in size, specifications, and construction requirements due to differences in design conditions, site constraints, and client demands. Moreover, to accommodate on-site conditions, design modifications are often made during the construction process. These high levels of uncertainty and environmental variability pose significant challenges to the integration of robotic arms into prefabrication and assembly workflows in construction.

In the early stages of design and planning, operators must repeatedly test processing workflows and related parameters to ensure production efficiency and product quality. At the same time, robotic arms used in construction can weigh several tons, and their end-effectors are often equipped with high-risk tools, such as nail guns and cutting machines, for timber cutting and assembly tasks. To prevent irreversible accidents caused by human error, this study proposes a safety framework for construction-site robotic arm operations. By integrating multi-camera human tracking and safety-zone assessment, the proposed framework not only improves the efficiency of human-robot collaboration but also provides enhanced safety supervision in robotic construction environments.

2 Literature Review

Human-robot collaboration (HRC) has gradually become a key trend in the development of smart manufacturing industries [1][2]. However, when operators and robotic arms work within the same workspace, potential safety hazards may arise, including collision, crushing, pinching, and puncture injuries. To prevent such safety incidents, a common protective strategy is to establish a protective field , typically implemented through the Safety-Rated Monitored Stop (SRMS) mode, in which a robot and its human partner

are not allowed to operate simultaneously within the same workspace [3]. Under SRMS, personnel are restricted from entering the operational area while the equipment is active. This can be achieved using physical barriers or sensor-based solutions such as light detection devices. Once human presence is detected within the operational zones during robot operation, the system automatically stops the equipment. However, this approach inevitably reduces the effective shared workspace between the robot and the human operator, potentially constraining task flexibility and productivity. To overcome this limitation, an alternative collaboration strategy is the Speed and Separation Monitoring (SSM) approach, which allows humans and robots to move simultaneously within the shared workspace while maintaining a minimum protective distance at all times [3]. Such safety strategies rely on reliable perception of worker locations in real time. To capture worker locations in complex construction environments, multi-camera system plays an important role in a wide range of domains, including motion capture, surveillance, robotics [4]. Multi-camera people tracking (MCPT) is a challenging task, particularly in complex real-world environments, and is essential for the development of intelligent surveillance and monitoring applications [5]. This task also involves solving multiple computer vision challenges, such as object detection and multiple object tracking (MOT) [6]. However, most existing MCPT research has focused on general surveillance or crowd monitoring scenarios, and is often conducted in controlled or simulated environments. In contrast, HRC settings present additional real-world challenges such as frequent occlusions, limited camera viewpoints, real-time safety requirements, and the need to associate tracking results with robot operating zones. Therefore, a reliable multi-camera tracking framework tailored for HRC scenarios is required to accurately monitor human positions and safety-critical movements in real time. In this study, we develop a multi-camera human tracking and safety monitoring system that integrates object detection, MOT, and human-robot spatial interaction analysis to support SSM-based collaborative operations.

3 Method

This study develops a Multi-camera human tracking and safety-warning system framework designed for construction-site robotic arm operations (Figure 1). The system consists of three main modules: a tracking module, a projection module, and a safety-zone decision module. Two cameras are installed around the robotic arm workspace to capture on-site operation videos from multiple viewpoints. The tracking results are then projected onto a 2D ground-plane representation, enabling analysis of the relative spatial positions between workers and the robotic arm during operation. Based on the proposed three-stage safety-zone design, the system further determines the robot operating speed in a sequential manner, including full speed, reduced speed, and emergency stop.

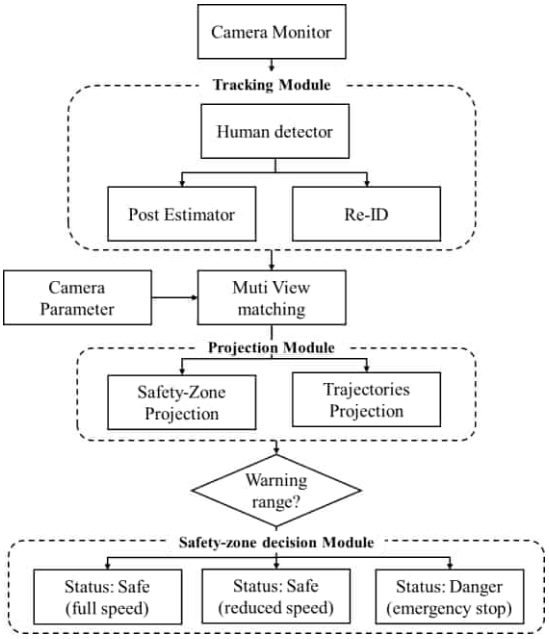


Figure 1. Multi-camera human tracking and safety-warning framework.

3.1 Tracking Module

The tracking module extracts identity-consistent worker trajectories from multi-view videos. First, each camera frame is processed using a YOLO-based detector to localize workers with bounding boxes. The model is fine-tuned using project-specific construction-site videos to improve robustness under real-world conditions. Next, a pose estimation model predicts 2D skeletal keypoints for each detection, providing structural cues that support cross-view association. Finally, person re-identification (ReID) features are extracted to match workers across cameras and time. By integrating appearance similarity, pose consistency, and temporal proximity, fragmented single-view tracks are fused into unified multi-camera trajectories.

3.2 Projection Module

In the projection module, fused multi-camera trajectories are mapped onto a 2D ground-plane representation of the workspace. Camera views are geometrically aligned with the site layout, and predefined safety-zone boundaries are projected onto the same ground-plane coordinate system. The system then computes worker-robot proximity by analyzing distances

between workers and predefined hazardous regions within the robotic arm's working envelope. This projection supports real-time proximity monitoring and safety-aware decision-making under SSM-based human-robot collaboration scenarios.

3.3 Safety-Zone Decision Module

The safety-zone decision module evaluates operator proximity based on the projected 2D worker positions and predefined safety zones associated with the robotic arm's working envelope. Following the Speed and Separation Monitoring (SSM) principle, the workspace is divided into three staged zones reflecting different safety risk levels. The system continuously computes the relative distance between each worker and the hazardous operational region of the robotic arm. According to the zone in which the worker is located, the module determines the robotic arm operating mode sequentially:

1. Full speed operation when no workers are within the warning zones.
2. Speed reduction when a worker enters the intermediate zone.
3. Emergency stop when a worker enters the closest/high-risk zone.

This design provides an explicit safety supervision mechanism for real-time human-robot collaboration and reduces the likelihood of accidents caused by human error.

4 System prototype

To validate the feasibility of the proposed framework, a proof-of-concept prototype was developed with a focus on the tracking module. Since accurate human localization and identity-consistent trajectory generation provide the essential inputs for both ground-plane projection and safety-zone decision-making, tracking performance largely determines the effectiveness of the overall pipeline. To verify whether the estimated movement paths are consistent with real operator trajectories, a Unity-based simulation environment was constructed to replicate the robotic arm workspace (Figure 2)

The prototype implements multi-view tracking using three components: human detection, pose estimation, and person re-identification (ReID). Multi-view MP4 videos are decoded frame by frame, and a YOLO-based detector is applied to obtain worker bounding boxes. A top-down pose estimator predicts 2D keypoints for each detection, while a ResNet-50-based ReID model extracts appearance embeddings to maintain identity consistency across frames and camera views. The system then performs tracking integration by associating detections over time using fused cues from pose similarity, geometric constraints, bounding-box overlap (IoU), and ReID similarity.

Finally, for each confirmed track, the system projects the person's position onto a common 2D ground plane and compares the worker's location with the predefined safety-zone boundaries (Figure 3). When the operator is outside the robot work area, the robotic arm operates normally. Once the operator enters the fenced work zone, the robot is required to slow down. If the operator further enters the red danger zone, the robotic arm must stop immediately. This process is used to verify positional accuracy and assess whether the tracking result is consistent with the expected safety-zone relationship.

To quantitatively evaluate tracking accuracy, we measure the spatial error between predicted trajectories and ground-truth paths under different distance thresholds. The accuracy is defined as the percentage of frames where the localization error is within a given threshold. The results are summarized as follows: 0.1 m threshold: 26.3%, 0.3 m threshold: 80.4%, 0.5 m threshold: 88.9%, 1.0 m threshold: 95.3%. These results indicate that the proposed system achieves high localization reliability under practical safety margins (0.3-1.0 m), which are consistent with safety requirements in human-robot collaboration scenarios.

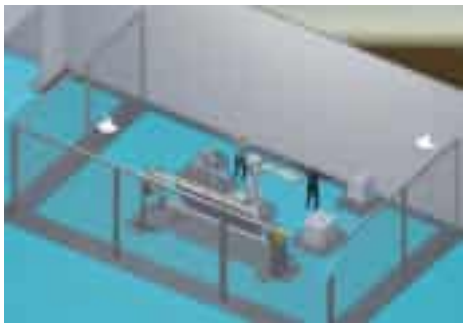


Figure 2. 3D on-site environment simulated in Unity.

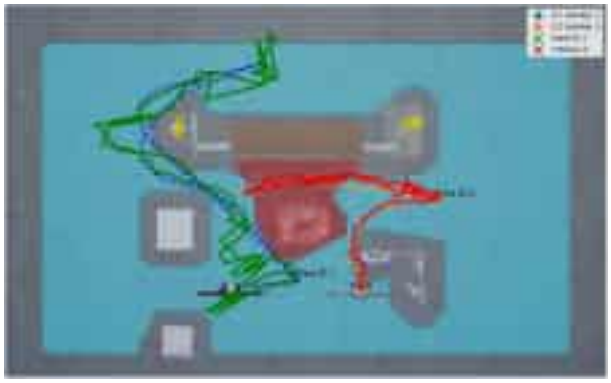


Figure 3. Human movement trajectories.

5 Future works

The proposed framework provides a safety supervision mechanism for human-robot collaboration in complex construction-site robotic arm workspaces. Compared with the traditional Safety-Rated Monitored Stop (SRMS) mode, which often reduces collaboration efficiency, this framework follows the Speed and Separation Monitoring (SSM) principle to gradually adjust the robotic arm operation state according to the actual safety risk. However, the current implementation adopts predefined safety zones for simplicity and real-time feasibility, which may not fully capture task-dependent hazards in dynamic construction environments. In practical scenarios, safety risks are influenced by multiple factors, including the robotic arm's reachability, tool path, and tool geometry. Therefore, future work will focus on integrating robotic arm motion planning and task-level information to dynamically model the working envelope of the robotic system. By incorporating task scripts and kinematic constraints, the system can generate adaptive safety zones that reflect real-time operational conditions. This enhancement will enable more accurate and context-aware safety assessment, improving the applicability of the proposed framework in real-world construction environments.

6 Conclusions

This study develops a Multi-camera human tracking and safety-warning system framework designed for application in robotic-arm operation environments within construction and factory settings. By deploying multiple video cameras around the robotic workspace, the system enables synchronized monitoring of operator activities and tracks human movement trajectories in relation to the robot's operating zone. The proposed framework is composed of three main modules: the tracking module, the projection module, and the safety-zone decision module. Through this pipeline, the system converts multi-camera visual information into a unified 2D spatial representation of worker positions, allowing the analysis of proximity and potential safety-risk interactions between operators and the robotic arm.

The proposed framework demonstrates the feasibility of using multi-camera human tracking as a safety-oriented perception layer for human-robot collaboration in construction-site robotic arm operations. By transforming tracking results into a spatial representation of worker-robot proximity, the system enhances situational awareness and supports collaborative operation strategies such as SSM, thereby contributing to safer interactions between workers and robotic systems. The system is designed to balance operational efficiency and safety by minimizing unnecessary interruptions while ensuring immediate emergency stop when hazardous conditions are detected, aligning with safety requirements in SSM-based human-robot collaboration.

Acknowledgements

The authors would like to thank Prof. Jacob J. Lin and Mr. Cheng-Yun Tsai (Department of Civil Engineering, National Taiwan University) for their support in prototype development, including coding assistance and constructive feedback. The authors also acknowledge RoBIM Technologies Inc. for sharing practical knowledge and providing support for this study.

References

[1] Occupational Safety and Health Administration, Ministry of Labor. Collaborative Robot Operation Safety Assessment Report (original in Mandarin). April 2019. Available at: https://www.osha.gov.tw/. Accessed on 2026/01/01.

[2] M. Valori, A. Scibilia, I. Fassi, J. Saenz, R. Behrens, S. Herbster, C. Bidard, E. Lucet, A. Magisson, L. Schaake, J. Bessler, G. B. Prange-Lasonder, M. Kühnrich, A. B. Lassen, and K. Nielsen. Validating Safety in Human-Robot Collaboration: Standards and New Perspectives. *Robotics*, 10(2):65, 2021. https://doi.org/10.3390/robotics10020065.

[3] W. Li, Y. Hu, Y. Zhou, and D. T. Pham. Safe Human-Robot Collaboration for Industrial Settings: A Survey. *Journal of Intelligent Manufacturing*, 35:2235-2261, 2024.

[4] S.-E. Lee, K. Nishino, and S. Nobuhara. *Spatiotemporal Multi-Camera Calibration Using Freely Moving People*. *IEEE Robotics and Automation Letters*, 10(5):4818-4825, May 2025. https://doi.org/10.1109/LRA.2025.3554102

[5] J. Kim, W. Shin, H. Park, and J. Baek. Addressing the Occlusion Problem in Multi-Camera People Tracking with Human Pose Estimation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 5463-5469, 2023.

[6] S. Hassan, G. Mujtaba, A. Rajput, and N. Fatima. Multi-Object Tracking: A Systematic Literature Review. *Multimedia Tools and Applications*, 83:43439-43492, 2024.

Modern Methods of Construction and IoT Adoption in New Zealand School Buildings: A Descriptive Multiple-Case Study

Odinaka Chukwu[1], Tochukwu Nnaji[2] and Lilian Obi-George[1]

[1]School of Built Environment, Massey University, New Zealand
[2]Faculty of Architecture and Design Innovation, Victoria University of Wellington, New Zealand

ochukwu@massey.ac.nz, tochukwu.nnaji@vuw.ac.nz, l.obigeorge@massey.ac.nz

Abstract -

New Zealand's school building programme is deploying modern methods of construction (MMC) and nationwide IoT sensors faster than almost any other public sector, yet this descriptive multi-case analysis of 17 schools shows that this is disconnected from net-zero goals. Government data indicate that 60% of new classrooms now use MMC, but only 18% of audited schools published any details about their modular systems, and none disclosed embodied-carbon targets or energy-performance benchmarks. Similarly, while every school received IoT sensors for temperature, humidity and CO_2 under the Ministry's Te Haratau rollout, no project documents reveal data-privacy protocols, cybersecurity standards or performance dashboards. The descriptive, multi-case analysis demonstrates that MMC and IoT are being adopted to solve capacity, cost and resilience pressures, not to deliver measurable carbon reductions. Mandatory public reporting of construction typology, lifecycle carbon metrics and sensor governance would align New Zealand's most standardised building portfolio with its climate commitments and create a replicable model for other public infrastructure.

Keywords -

Modern Methods of Construction (MMC); Internet of Things (IoT); School buildings; New Zealand; Net-Zero

1 Introduction

New Zealand (NZ) has committed to net-zero greenhouse-gas emissions by 2050, placing the building and construction sector under increasing pressure to deliver both capacity and carbon reductions simultaneously [1]. School buildings are central to this transition: they represent one of the country's largest, most standardised, publicly funded building portfolios, governed centrally by the Ministry of Education (MoE), and are already the focus of policy initiatives to improve performance and resilience [1, 2].

Recent policy pushes have accelerated two major forms of innovation in the school property programme. First, a 2024 government directive instructed the MoE to prioritise MMC, contributing to a rapid increase in modular and prefabricated classrooms: over 60% of new classroom builds used MMC in Q1 2024, up from under 20% in Q4 2023 [2, 3, 4]. Second, the MoE's Te Haratau programme (TeH) is deploying IoT-based Internal Environment Monitoring (IEM) devices across more than 2,500 state schools to capture indoor air-quality and comfort data at scale [5, 6]. Together, these initiatives create an unprecedented opportunity to link construction innovation and smart monitoring with national climate objectives.

However, existing research and reports on zero-carbon buildings in New Zealand highlights institutional, regulatory and knowledge gaps that hinder the translation of high-level climate targets into project-level decisions [7, 8, 9]. Little empirical evidence currently shows how MMC and IoT are combined in completed school projects, how they are framed in publicly available documentation, or whether they are explicitly aligned with net-zero, embodied-carbon or energy-performance targets.

In this context, MMC is understood to include prefabrication, modularisation, volumetric off-site manufacturing and digitally integrated design-to-production workflows [2], while IoT encompasses sensor networks and building-management systems capable of streaming data on energy use, environmental quality, occupancy and equipment performance [10, 11].

This paper addresses these gaps through a descriptive multiple-case study of 17 New Zealand school projects with identifiable MMC typologies and confirmed IoT sensor deployment. The objective is to map how MMC and IoT are documented and justified in these projects, and to assess the extent to which sustainability and carbon-reduction narratives are integrated into their public reporting.

2 Methods and Materials

2.1 Study Design

The study adopts a descriptive multiple-case research design following Yin's approach to case study research [12], focusing on mapping patterns of adoption and framing rather than testing causal relationships. The unit of

analysis is the individual school campus (primary, intermediate, secondary and kura kaupapa Māori).

Purposive, multi-stage sampling was used to identify and select cases from the national Directory of Educational Institutions and associated MoE property information [13]. Eligible schools were state or state-integrated institutions that:

- completed significant capital works between 2015 and 2024
- demonstrated some use of MMC
- deployed IoT monitoring systems, and
- showed at least some net-zero or sustainability framing at policy or design level.

Initial screening of MoE and public documentation yielded approximately 35–40 schools with recent capital activity and potential MMC involvement. A more detailed documentary review then confirmed the presence of (1) MMC components, (2) IoT systems and (3) sustainability framing. From these, 17 schools were selected to ensure variation in region, school type and project type (new builds, major redevelopments, roll-growth additions and resilience/recovery works). Schools are anonymised as S1–S17.

2.2 Data Sources and Coding Framework

Data were collected from publicly accessible sources including GETS tender documents, MoE property data, council consent records, architect and contractor case studies, Education Review Office (ERO) reports, media coverage and vendor listings. For each school, a within-case summary was compiled, and key information was extracted into structured coding tables.

A multi-dimensional coding framework operationalised three primary domains:

- **MMC typologies** – manufacturing location (off-site/hybrid/traditional), component scale (volumetric/panelised/structural) and deployment intent (permanent/temporary/adaptive), drawing on MMC literature and UK off-site classifications [2, 14].
- **IoT functions** – confirmation of hardware deployment; intended function (environmental monitoring, building management system, occupancy detection); platform governance (stand-alone vs. integrated, cloud management, open vs. proprietary protocols); and occupant-facing transparency (real-time dashboards vs. opaque backend systems) [6, 10, 11, 15].
- **Carbon reduction claims** – presence and strength of sustainability framing, ranging from explicit building-level net-zero targets to broader programme-level or values-only statements, covering both operational and embodied emissions where stated [7].

Modular classroom blocks were coded as OP-V (off-site, prefabricated, volumetric, permanent) and OP-V-R (off-site, prefabricated, volumetric, relocatable) for analysis. The multi-stage case selection and analysis process is shown in Figure 1.

3 Results

3.1 Audit Group Overview

The audit examined 17 NZ school campuses across eight regions (Auckland, Wellington, Waikato, Bay of Plenty, Hawke's Bay, Christchurch, Palmerston North, Lower Hutt). Group composition: primary schools (8), intermediate (4), secondary (4), kura kaupapa Māori (1). Projects ranged from new builds (3), major redevelopments (7), roll-growth additions (4), and resilience/recovery initiatives (3). For each school, approximately 6–28 distinct documents were available, providing sufficient material to reach theoretical saturation for the descriptive aims of the study.

3.2 MMC Adoption

Only 18% of schools (3 of 17) demonstrated MMC. Despite government policy accelerating MMC adoption to 60% of new classrooms in Q1 2024, only a minority had publicly documented evidence of MMC, reflecting a public-sector documentation gap.

3.2.1 Case 1: TKKM Te Ara Wahanui (S8)

Minister Erica Stanford announced OP-V construction for 8 modular classrooms [16], emphasizing operational efficiency: "By using modular construction, we can deliver classrooms more quickly." Classrooms designed for 20-to-50-year lifespan. Public narrative focused on safety and warmth; no explicit sustainability or carbon-reduction language despite government climate commitments.

3.2.2 Case 2: Addington Primary (S9)

CEBUS relocatable block (OP-V-R, 192 m^2, $235,594) deployed May 2024. Previously relocated; designed for future relocation. Part of post-earthquake rebuild (2015–2020), exceeding Building Code standards. Framed as "stronger, safer" rebuild without net-zero claims.

3.2.3 Case 3: Ōtūmoetai College (S14)

$47M redevelopment addressing weathertightness and capacity [17]. Multiple prefab classrooms on field to be removed [18]. Existing OP-V-R stock; unclear if new modular units will replace structures. Carbon narrative driven by "leaky buildings" rather than sustainability.



Figure 1. Multi-stage case selection and analytic workflow of MMC and IoT-enabled school buildings.

3.3 IoT and Smart Building Systems

All 17 schools had confirmed IoT infrastructure via MoE TeH programme documentation and IoT provider lists. Zero schools publicly documented BMS, data governance, or performance expectations, creating a transparency gap. Without school-level documentation, system functionality cannot be classified.

The MoE TeH programme plans national rollout to all 2,500 state schools by 2030. Investment: $12.5–25M ($5,000–10,000 per installation). All 17 schools are part of a managed, standardised IoT rollout. Heterogeneity in outcomes (energy performance, occupant comfort) is unlikely to reflect system design but rather building typology, envelope performance, and occupancy patterns.

3.4 Carbon Reduction Claims and Sustainability Narratives

Overall, the carbon claim distribution for the 17 schools showed that none used explicit terms like "Net-Zero," "Zero Carbon," or "Carbon Neutral." Only 12% of 17 schools demonstrated any form of implicit sustainability framing through formal environmental certifications or embedded programmes.

Sustainability framing dissociated from procurement. Capacity, cost, and schedule are primary drivers, environmental performance secondary or absent.

3.4.1 Case 4: Waterloo School (S16)

S16 demonstrates institutional commitment to sustainability as a whole-school ethos through Enviroschools Bronze certification active since 2014, with evidence from ERO reports stating that ecological sustainability is highly evident in the school's philosophy and values, although there are no explicit net-zero claims or building-specific carbon targets in its substantial property redevelopments [19].

3.4.2 Case 5: Te Kauwhata Primary School (S17)

Strongest implicit sustainability commitment: Enviroschools Green-Gold certification (2024, highest level). Formal programmes include Community Compost Hub and solar-heated pool for GHG reduction. However, capital activity tenders lack construction method or embodied carbon documentation. Focus on operational metrics (waste, solar) without building-level net-zero strategies.

3.5 Co-Adoption Patterns: MMC, IoT, and Sustainability Alignment

Fusing the three domains reveals three main co-adoption patterns:

- **Full convergence (MMC + IoT + explicit net-zero claim)** – No school in the sample combined documented MMC use, IoT monitoring and explicit building-level carbon-reduction targets.

- **Hardware-only adoption (all 17 schools)** – Every school had IoT infrastructure through TeH, but almost none linked this hardware to measurable energy or carbon-performance expectations, and only a small minority documented MMC. In these cases, MMC and IoT appear as independent responses to capacity and operational pressures rather than components of an integrated net-zero strategy.
- **Split adoption (~82%)** – Most schools combined IoT deployment with traditionally documented or unspecified construction typologies. In other words, smart-monitoring infrastructure is being rolled out regardless of whether buildings use MMC or conventional construction, with little evidence that the two are co-designed for carbon outcomes.

These patterns point to a structural gap: technology adoption is not currently being harnessed to deliver or verify progress towards net-zero commitments at the level of individual school projects.

As shown in Table 1, only 3 schools (18%) had documented MMC components, despite national policy pushing for 60% adoption. All 17 received IoT sensors, yet none reported data usage or governance publicly. Similarly, no schools set explicit carbon targets.

4 Discussion

4.1 The Documentation Paradox: MMC Adoption versus Visibility

The gap between 60% national MMC adoption (Q1 2024, up from 20% in Q4 2023) and only 18% documented MMC use in this cohort signals systematic transparency deficits in public-sector reporting [4]. This discrepancy does not indicate low adoption; rather, it reveals that construction innovation proceeds without public accountability mechanisms.

Documented cases (S8, S9, S14) consistently emphasized operational efficiency, speed, cost, resilience, over embodied carbon reduction, despite NZ's legally binding net-zero by 2050 target. S8's ministerial announcement framed modular construction as delivering "classrooms more quickly" without referencing lifecycle performance. S14's decommissioning of existing prefab classrooms in favour of permanent structures suggests institutional preferences for traditional construction despite MMC availability. This pattern reflects immediate capacity pressures overriding environmental considerations in procurement decisions.

Addressing this gap requires standardized reporting protocols that mandate disclosure of MMC typology, embodied carbon estimates, and operational performance expectations in all publicly funded school projects. Without such mechanisms, policy acceleration proceeds independently of climate accountability.

4.2 Universal IoT Deployment, Opaque Governance: The Te Haratau Programme

The TeH programme's three pillars: condition assessment, fitness for purpose and operational efficiency (energy monitoring), establish a data infrastructure unprecedented in the sector.

Table 1 also shows that 17 schools (100%) confirmed IoT deployment measuring CO_2, temperature, and humidity, yet zero schools publicly documented BMS configurations, data governance protocols, or environmental performance dashboards. This opacity raises critical questions about transparency and occupant agency, particularly in educational settings where children constitute the monitored population. While the MoE centrally manages TeH data via cloud platform, lack of school-level documentation on algorithms or system interoperability prevents independent evaluation of effectiveness.

Aligning TeH with net-zero ambitions would therefore require not only technical rollout but also governance frameworks that specify responsibilities, transparency requirements and mechanisms for linking IoT-derived insights to measurable energy and carbon-performance indicators at school level. The national IoT infrastructure therefore represents a latent platform for net-zero implementation, but only if governance, transparency and integration with BMS are strengthened.

4.3 Decoupled Narratives: Construction Innovation without Sustainability Targets

The most consequential finding is systematic decoupling of MMC and IoT adoption from explicit sustainability performance targets. No schools articulated net-zero, zero-carbon, or carbon-neutral claims. As summarised in Table 1, only 12% (n=2) demonstrated implicit sustainability commitment through formal Enviroschools certification, while 41% (n=7) framed capital projects through resilience and post-disaster recovery rather than climate mitigation.

This decoupling from the Carbon Neutral Government Programme [8] and 2022 Emissions Reduction Plan [9] suggests procurement narratives prioritize operational drivers capacity, cost, speed, resilience over environmental performance. The gap reflects immediate capacity pressures, absence of mandatory embodied carbon caps in Building Code, and persistent institutional knowledge gaps identified in zero-carbon building research [7]. Without enforceable mechanisms linking MMC and IoT adoption to carbon performance targets, technological innovation proceeds independently of climate commitments, under-

Table 1. MMC, IoT, and Carbon Framing Across 17 NZ Schools.

Dimension	Category	Count (n=17)	%	Notes
MMC Documentation	Documented MMC Use	3	18%	All were off-site volumetric; 2 temporary, 1 permanent
	Unknown / No Documentation	14	82%	Indicates a significant public documentation gap
IoT Deployment	Sensors Installed	17	100%	Standard MoE rollout (Temp, CO_2, Humidity)
	Public Data / Dashboards	0	0%	No BMS, transparency, or governance disclosed
Carbon Framing	Explicit Net-Zero Targets	0	0%	No formal carbon or energy benchmarks published
	Implicit Framing (e.g. Enviroschools)	2	12%	Green-Gold and Bronze certifications
	Resilience-Focused Framing	7	41%	Projects linked to earthquakes / recovery
	No Sustainability Framing	8	47%	No carbon / environmental goals evident

mining policy coherence between infrastructure delivery and emissions reduction objectives.

Given their scale, standardisation and central governance, New Zealand schools represent an ideal scene for developing structured net-zero procurement and reporting practices that could subsequently be extended to broader public infrastructure.

5 Conclusion

This case study reveals a paradox in New Zealand's school-building programme: despite accelerated MMC deployment (60% of new classrooms, 2024) and universal IoT rollout, these innovations remain disconnected from net-zero-by-2050 commitments. Across 17 schools, only 18% published construction documentation; none disclosed carbon targets or data-governance protocols. The gap is institutional: procurement rewards speed and cost while Building Code imposes no embodied carbon limits. As New Zealand's most standardized public infrastructure, schools offer strategic opportunity to develop reporting protocols linking construction innovation to measurable emissions reduction, providing a replicable framework for broader public infrastructure decarbonization.

References

[1] Ministry of Business, Innovation and Employment. Building for climate change: Transforming the building and construction sector to reduce emissions and improve climate resilience. Technical report, Ministry of Business, Innovation and Employment, Wellington, New Zealand, 2020.

[2] D. Bain and A. Wilen. Continued growth for MMC and off-site construction, despite setbacks in modular housing. On-line: https://manufacturers.thenbs.com/resources/knowledge/continued-growth-for-mmc-and-off-site-construction Accessed: 04/08/2025.

[3] C. Bishop and E. Stanford. Response to ministerial inquiry into school property. On-line: https://www.beehive.govt.nz/release/response-ministerial-inquiry-school-property, Accessed: 22/08/2025.

[4] Ministry of Education. Standardising school property. On-line: https://www.education.govt.nz/our-work/strategies-policies-and-programmes/property-and-infrastructure/standardising-school-property, Accessed: 29/08/2025.

[5] Ministry of Education. Te haratau: Lifting the quality of New Zealand's physical learning environments. Technical report, Ministry of Education, Wellington, New Zealand, 2024. On-line: https://www.education.govt.nz/our-work/strategies-policies-and-programmes/property-and-infrastructure/te-haratau, Accessed: 29/08/2025.

[6] Ministry of Education. Internal environment monitoring devices. On-line: https://www.education.govt.nz/our-work/strategies-policies-and-programmes/property-and-infrastructure/te-haratau/internal-environment-monitoring-devices, Accessed: 29/07/2025.

[7] T. T. P. Bui, C. MacGregor, S. Wilkinson, and N. Domingo. Towards zero carbon buildings: issues and challenges in the New Zealand construction sector. *International Journal of Construction Management*, 23(15):2709–2716, 2023. doi:10.1080/15623599.2022.2110642.

[8] Ministry for the Environment. Carbon neutral government programme: A guide to managing your greenhouse gas emissions – measuring, reporting, target-setting and reduction planning. Technical report, Ministry for the Environment, Wellington, New Zealand, 2023. On-line: https://www.sustainability.gov/pdfs/new-zealand-nzgi-roadmap.pdf, Accessed: 09/08/2025.

[9] Ministry for the Environment Manatū Mō Te Taiao. Te hau mārohi ki anamata: towards a productive, sustainable and inclusive economy: Aotearoa New Zealand's first emissions reduction plan. Technical report, Ministry for the Environment Manatū Mō Te Taiao, Wellington, New Zealand, 2022.

[10] A. Taboada-Orozco, K. Yetongnon, and C. Nicolle. Smart buildings: a comprehensive systematic literature review on data-driven building management systems. *Sensors*, 24(13):4405, 2024. doi:10.3390/s24134405.

[11] M. A. Ahmed, S. A. Chavez, A. M. Eltamaly, H. O. Garces, A. J. Rojas, and Y.-C. Kim. Toward an intelligent campus: IoT platform for remote monitoring and control of smart buildings. *Sensors*, 22(23): 9045, 2022. doi:10.3390/s22239045.

[12] R. K. Yin. *Case Study Research and Applications*. Sage, Thousand Oaks, CA, USA, 6 edition, 2018.

[13] New Zealand Ministry of Education. Directory of educational institutions. data.govt.nz dataset. On-line: https://catalogue.data.govt.nz/dataset/directory-of-educational-institutions/resource/4b292323-9fcc-41f8-814b-3c7b19cf14b3, Accessed: 14/11/2025.

[14] C. Goodier and A. Gibb. Future opportunities for offsite in the UK. *Construction Management and Economics*, 25(6):585–595, 2007. doi:10.1080/01446190601071821.

[15] D. Amaxilatis, O. Akrivopoulos, G. Mylonas, and I. Chatzigiannakis. An IoT-based solution for monitoring a fleet of educational buildings focusing on energy efficiency. *Sensors*, 17(10):2296, 2017. doi:10.3390/s17102296.

[16] T. Natanahira. Watch: Minister erica stanford makes education announcement. Radio New Zealand, news article. On-line: https://www.rnz.co.nz/news/political/567061/watch-minister-erica-stanford-makes-education-announcement, Accessed: 13/11/2025.

[17] New Zealand Herald. $47 million to be spent repairing Otumoetai college buildings. Bay of Plenty Times, news article. On-line: https://www.nzherald.co.nz/bay-of-plenty-times/news/47-million-to-be-spent-repairing-otumoetai-college-bu C5MCXBKYGIA3SCUKJ4AMQEJ3BI/, Accessed: 08/11/2025.

[18] R. Crawford. Transformative renovation at ōtūmoetai college. SunLive, news article. On-line: https://www.sunlive.co.nz/news/364783-transformative-renovation-at---t--moetai-colle html, Accessed: 11/08/2025.

[19] S. Boyd, M. Denzler, M. Li, and R. Bolstad. School board leadership: Key issues 2025. Technical report, New Zealand Council for Educational Research, Wellington, New Zealand, 2025.

Identifying and Classifying Residents' demand for Smart Community Development: An Empirical Analysis from China

Yujia Shan[1] and Tiantian Gu[1]

[1] School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou, China

ts24030295a31tm@cumt.edu.cn, 6072@cumt.edu.cn

Abstract

Smart community development is pivotal for modernizing grassroots governance. However, few studies have established a systematic classification model for residents' multi-dimensional needs. To fill this research gap, this paper constructs a comprehensive demand analysis framework covering three core dimensions: community safety, livability services and community governance. Based on the survey data of 191 valid samples from 5 typical smart communities in Huizhou, this study uses hierarchical clustering and K-means clustering to identify and classify residents' demand types. The results divide residents into four clusters: weak service-oriented demand group, safety and governance priority demand group, strong service demand group and comprehensive demand group. This taxonomy framework offers a replicable methodology for smart community planning with diverse resident demand priorities. Additionally, grounded in empirical analysis of Huizhou, China, it provides smart community development strategies for the Chinese context to enhance smart community development efficacy.

Keywords –

Smart community; Residents' demand; Community governance; Cluster analysis

1 Introduction

Smart community development plays a pivotal role in driving sustainable urbanization, serving as the operational backbone of broader smart city frameworks [1]. Generally, such initiatives adopt digital technologies to enhance urban quality of life, operational efficiency, and governance structures [2]. During public health emergencies, smart communities also showed obvious advantages in rapid early warning, closed management and resource allocation [3]. Yet the effectiveness of these efforts hinges on a systematic grasp of what residents actually need [4]. Given that residents hold diverse and often competing priorities, failing to quantitatively capture this demand landscape risks misallocated investments, poor uptake, and ultimately project failure [5]. Therefore, a deep understanding of residents' needs is essential to shift the focus of smart community development from a technology-first approach to one that puts people at the center [6].

In existing literature, the governance functions of smart communities have been extensively analyzed, but discussions on the actual demands of residents remain fragmented. For instance, the roles of various stakeholders have been clarified, and the development orientations of community property service enterprises, government agencies, and other participants have been further analyzed in the process [7]. From the property management side, researchers have examined how firms adapt their operations and business models under digital transformation [4], while public administration scholars have focused on policy execution and governance outcomes [8, 9]. However, in existing research, residents typically occupy a passive position. Consequently, another strand of literature has concentrated on residents' needs and experiences to evaluate specific smart community functions and adjust development strategies. Scholars have investigated satisfaction and supply-demand matching in specific functional domains, including older adult care services [10], residential energy scheduling [11], and water demand management [12]. Additionally, few studies have explored residents' demands for smart community development [13]. Taken together, this work points to two related shortcomings. On the one hand, most studies examine isolated service functions rather than constructing a comprehensive demand indicator framework. On the other hand, rigorous quantitative tools for mapping multi-dimensional demand profiles across resident groups remain absent.

To address these gaps, this study developed a resident demand indicator system covering three dimensions of smart community development, then built a classification model by combining hierarchical

clustering with K-means analysis. Drawing on 191 valid survey responses, the approach identifies empirically grounded demand clusters among residents. Theoretically, this research filled the theoretical gap of lacking a systematic indicator system that integrates multi-dimensional residents' demands. It also introduces a data-driven taxonomy that moves the field beyond isolated functional analyses toward an integrated view of demand patterns. Practically, suggestions were designed for the case communities directly and offered a transferable reference for smart community projects with comparable demand profiles, supporting more effective project implementation in similar contexts.

2 Methodology

2.1 Establishment of the residents' demand indicator system

The selection of resident demand indicators drew on two complementary theoretical foundations: Maslow's Hierarchy of Needs and the Kano Model [14]. These frameworks were chosen because their respective strengths align well with the dual task of ordering and categorizing demands in a smart community context [15]. Maslow's hierarchy traces how human needs progress from basic survival concerns to higher-order aspirations [16], while the Kano Model offers a mechanism for classifying demands based on their impact on satisfaction [17]. Used in combination, the two models support a structured analysis of resident demands that spans from foundational safety concerns to active participation in community governance.

The resulting integrated indicator system (see Figure 1) organizes resident demands into three core dimensions. The first, Community Safety, corresponds to Maslow's safety needs level and captures residents' baseline expectations for protecting life and property [18-22]. Under the Kano Model, these function as Must-be Quality attributes: their absence triggers strong dissatisfaction, yet their presence alone does not generate positive responses. The second dimension, Community Services, maps onto the love and belongingness and esteem needs in Maslow's framework, covering services that raise living standards and reinforce social identity [23-27]. The Kano Model treats these largely as One-dimensional Quality demands, where satisfaction climbs in step with service performance. The third dimension, Community Governance, reflects Maslow's self-actualization level and concerns residents' involvement in community decision-making processes [8, 28-30]. In Kano terms, these are Attractive Quality attributes. Their presence can substantially lift satisfaction, but their absence does not provoke dissatisfaction.



Figure 1. The residents' demand indicator system

Demand indicator selection relied on an expert scoring exercise drawing on 38 valid responses from a panel of 50 invited participants, spanning both academic researchers and seasoned practitioners. The expert panel composition is shown in Table 1. This process resulted in the final demand indicator system, consisting of 3 primary and 25 secondary indicators, as summarized in Figure 1.

Table 1. Information of the 38 experts surveyed

Variable	Items	Percentage (%)
Gender	Male	47.37
	Female	52.63
Years of Experience	Over 5 years	18.42
	4-5 years	31.58
	1-3 years	23.68
	Within one year	26.32
Professional Field	Academia	39.47
	Enterprise	60.53

2.2 Development of the residents' demand classification model

Building on this indicator system, a demand classification model was constructed by combining hierarchical cluster analysis (HCA) with k-means clustering to categorize resident demand profiles in a systematic way. The two methods were paired for their complementary roles. HCA was used first to determine the appropriate number of clusters by interrogating the underlying data structure, after which k-means clustering was applied to partition samples into stable, well-

separated groups [31]. Cluster analysis is a multivariate statistical technique that groups samples according to similarity, making it well suited to uncovering typologies within complex resident demand data [32].

(a). Calculation of initial distances

$$d_{ij} = \sum_{k=1}^{m} (x_{it} - x_{jt})^2 \quad (1)$$

$$S = \sum_{i=1}^{n} (x_{i1} - \frac{\sum_{i=1}^{n} x_{i1}}{n})^2 + \ldots + \sum_{i=1}^{n} (x_{im} - \frac{\sum_{i=1}^{n} x_{im}}{n})^2 \quad (2)$$

where x_{it} and x_{jt} represent the data of the t-th indicator in the i-th sample and j-th sample, respectively. d_{ij} denotes the distance between the i-th sample and the j-th sample in terms of three dimensions. S denotes the sum of squared deviations.

(b). Cluster merging based on minimum distance

If clusters p and q are merged into a new cluster l, the sum of squared deviations of cluster l was:

$$S_l = S_p + S_q + \frac{n_p n_q}{n_p + n_q} (\overline{x_p} - \overline{x_q})^2 \quad (3)$$

$$d_{pq} = \triangle S_{pq} = \frac{n_p n_q}{n_p + n_q} (\overline{x_p} - \overline{x_q})^2 \quad (4)$$

The distance between the new cluster l and any arbitrary cluster r was calculated using the following formula:

$$d_{lr} = \frac{n_p + n_r}{n_l + n_r} d_{pr} + \frac{n_q + n_r}{n_l + n_r} d_{qr} - \frac{n_r}{n_l + n_r} d_{pq} \quad (5)$$

where p and q contain n_p and n_q samples and their corresponding sums of squared deviations are denoted as S_p and S_q.

(c). Determination of the optimal number of clusters (k).

Step (b) is repeated iteratively, merging the two nearest clusters at each pass until all samples converge into a single cluster. This process generates agglomeration coefficients for successive cluster solutions, and the relationship between these coefficients and the number of clusters is then examined to identify the optimal cluster count (k).

(d). Calculation of the new distance matrix

K-means clustering was applied with the optimal cluster number (k=4) determined from the hierarchical clustering analysis. The algorithm employed k-means++ initialization, which selects initial cluster centers based on distance. With k centers established ($c_{j\prime}$, where j'=1, 2, …, k), the Euclidean distance $D(x_i, c_{j\prime})$ between each sample and every center was computed, and the resulting distance matrix was constructed as follows:

$$D = (d_{ij}) = \begin{bmatrix} d_{11\prime} & d_{12\prime} & \ldots & d_{1k} \\ d_{21\prime} & d_{22\prime} & \ldots & d_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ d_{n1\prime} & d_{n2\prime} & \ldots & d_{nk} \end{bmatrix} \quad (6)$$

Each sample is then assigned to the cluster j' with the min $D(x_i, c_{j\prime})$, which satisfied: $D(x_i, c_{j'}) = \min\{D(x_i, c_{j'})\}$ (i=1, 2, …, n; j'= 1, 2, …, k), the sample belongs to the j'-th cluster.

(f). Calculation of the k new cluster centers:

$$C_{j\prime}(I+1) = \frac{1}{n_{j\prime}} \sum_{x' \in C_{j\prime}} x' \quad (7)$$

Where $n_{j\prime}$ denotes the number of samples in cluster $C_{j\prime}$; x' denotes the samples classified into cluster $C_{j\prime}$.

This process continued until cluster assignments no longer changed and the sum of squared errors reached convergence. The procedure yielded k distinct clusters, each carrying a characteristic demand profile across the three core dimensions. Together, these clusters constitute an empirically grounded typology of smart community development patterns, organized around the underlying structure of resident demand.

3 Study area and data collection

3.1 Study area

Huizhou, located in Guangdong Province, China, was among the first batch of 20 pilot smart cities approved by the Chinese government in 2013. In the field of smart community, the Huizhou Municipal Government has been committed to advancing community digital governance, promoting age-friendly and barrier-free renovations, constructing community digital facilities, and establishing information management platforms. These persistent efforts have yielded remarkable achievements in optimizing community services and enhancing governance efficiency. Given that Huizhou is a representative city with a long-standing development history, rich practical experience, and outstanding accomplishments in China's smart community development, it serves as a typical study area for analyzing the residents' demand for smart community development.

To ensure the validity and representativeness of the sample, five communities with smart community outcomes were selected as the study cases, including

Banzhangling, Qingxi, Siqian, Tangjing and Tongxin. These communities all have smart upgrades in community safety and governance dimensions, and are equipped with functions such as resident demand management, grid-based governance, management of the floating population and intelligent patrol systems. Additionally, the five communities have integrated community service functions, such as home-based eldercare, smart healthcare services, intelligent charging facilities and community service centers. This selection ensures that the sample represents actual smart community development while capturing variation in residents' demand.



Figure 2. Study area

3.2 Data collection

Survey data were collected through 'Wenjuanxing', a professional online platform for questionnaire distribution and data collection. A total of 230 questionnaires were sent to residents across the 5 selected smart communities (see Figure 2), and after standard validation procedures, 191 responses were retained as complete and usable, corresponding to a response rate of 83.04%.

4 Results

4.1 Quantification of residents' demands for smart community development

Based on the 191 valid resident questionnaires, the demand of residents for smart community development was assessed. The results revealed significant disparities in demand level across the three dimensions of community safety, community services, and community governance. Overall, residents demonstrated their highest prioritization for community safety (avg. 3.85), followed by community services (avg. 3.77) and community governance (avg. 3.66), revealing that current security concerns take precedence over broader issues of service quality and governance participation. A detailed examination of secondary indicators provides further insights. The average demand scores for these indicators are summarized in Figure 3. Within community safety, residents exhibited particularly strong demand for D1-1 (Digital emergency plan) and D1-3 (Propaganda and education of emergency safety), while showing relatively lower demand for D1-8 (Public facilities monitoring). In community services, D2-6 (Community medical services) received the highest average score, while D2-3 (Smart childcare) scored the lowest. The seven indicators of community governance dimension all showed relatively low demand scores, with D3-1 (Integrated government service system) and D3-7 (Multi-sectoral linkage) recording the lowest scores. These patterns indicated that most residents place greater emphasis on safety outcomes in smart community development, particularly in emergency preparedness and safety education, while expressing lower demand for infrastructure monitoring, smart childcare, and community governance functions.

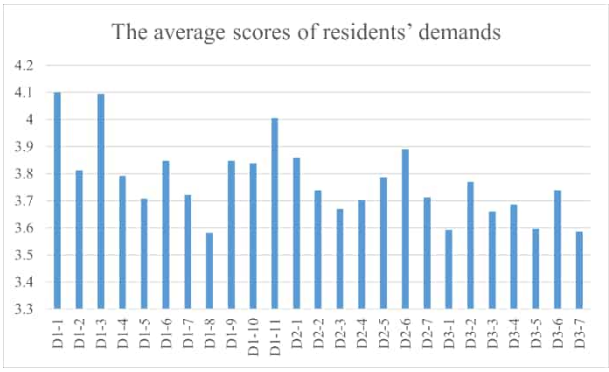


Figure 3. The average demand scores for all indicators

4.2 Clusters of residents' demands for smart community development

To further identify the diverse patterns of residents' demands, a classification model was employed on the demand scores of the 191 respondents. The dataset demonstrated high reliability and validity, with a Cronbach's Alpha of 0.833 and a KMO measure of 0.83 (Bartlett's test $p<0.001$).

The optimal cluster number was subsequently determined using hierarchical clustering with Ward's method. Based on the agglomeration schedule, the optimal cluster number was identified at $k = 4$. Specifically, the growth rate of the agglomeration coefficient exhibited a marked deceleration after the cluster number reached 4, decreasing from 21.07% ($k = 3$ to $k = 4$) to 7.65% ($k = 4$ to $k = 5$). This indicates that

more clusters would gradually diminish the differentiation between resident demand clusters, thus validating k = 4 as the optimal number. Following this, resident demands were partitioned into four clusters using k-means clustering. The clustering process regarded as convergent when the relative change in cluster centers fell below the tolerance threshold of 10^{-4}, indicating that the resulting cluster assignments had stabilized.

The final cluster centers resulting from this analysis are illustrated in Figure 4. The result showed a Silhouette Score of 0.283 and a Calinski-Harabasz index of 106.1, supporting the validity of the cluster structure. Cluster 1 comprised 14 respondents, Cluster 2 comprised 55, Cluster 3 comprised 56, and Cluster 4 comprised 66.

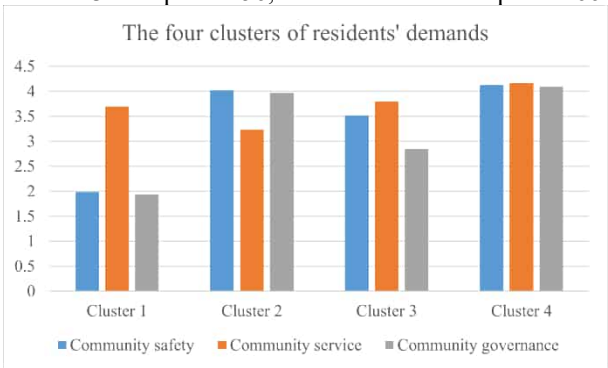


Figure 4. The four clusters of residents' demands

The four clusters showed markedly different demand profiles. Cluster 1 recorded uniformly low scores across all three dimensions (community safety, community services, and community governance). Cluster 2 stand out for its low community services demand, alongside high scores in safety and governance. Cluster 3 posted a high average in community services while falling below average in both safety and governance. Cluster 4 exceeded the overall average across all three dimensions, reflecting a broadly elevated demand profile.

Drawing on these profiles, the four clusters were defined as follows. Cluster 1 (weak service-focused demand) recorded low scores across all three dimensions with a marginal tilt toward basic livability, accounting for the smallest subset of only 14 respondents. Cluster 2 (prioritized safety and governance demand) concentrated demand on safety and governance while showing limited interest in community services, a pattern consistent with the sample-level finding that community safety achieved the highest overall average score. Cluster 3 (strong community services demand) was marked by elevated demand for community services, particularly medical services, alongside lower scores in safety and governance. Cluster 4 (comprehensive demand across all dimensions) maintained high scores across all three domains and represents the largest group at 66 respondents. This empirically grounded grouping captures the demand heterogeneity among residents and offers a practical foundation for formulating differentiated smart community development strategies.

4.3 Strategies to Satisfy residents' Demands for the Smart Community Development

Drawing from quantified demand profiles that identify four distinct resident clusters, effective smart community development requires moving beyond one-size-fits-all solutions toward tailored strategies that address specific resident needs.

For the communities where the residents' demand fell under cluster 1, priority could be given to establish fundamental smart community capabilities that improve daily living. The initial stage focuses on developing digital infrastructure for daily services to bring convenience to residents' everyday lives, including basic medical care, elderly care, community service centers, and other essential services [33]. This objective can be accomplished through the deployment of professional operators, who can enter into performance-based contracts with explicit service-level agreements. These contracts ensure that residents receive reliable and high-quality smart services [34].

For Cluster 2 communities, where residents demonstrate strong demand for safety and governance, targeted public investment is required to address functional deficiencies that affect resident security and civic participation. Emphasis could be placed on establishing centralised community data platforms to improve the efficiency of community safety response and enable active resident participation in governance [35]. Financial sustainability could be supported through instruments such as municipal bonds earmarked for smart city projects [36]. In addition, parallel administrative reforms clarifying jurisdictional boundaries and streamlining service workflows are essential to build digital governance platforms [37], and the clear and streamlined service workflows can enhance public willingness to participate in community governance and improve service experiences.

For communities characterized by Cluster 3, where service demand is high but safety and governance expectations are comparatively low, policy efforts should target market-driven innovation in sectors that directly shape residents' daily lives. A regulatory environment that incentivises private investment in domains residents prioritise, such as community healthcare and smart recycling, can help to achieve this [38]. At the same time, minimum service standards for smart community security [39] and operational oversight arrangements [40] should be put in place to guard against the under-provision of safety and governance functions.

Finally, for communities exhibiting the demand profile of Cluster 4, residents present strong and comprehensive demands across all dimensions. In addition to the three strategies mentioned above, such communities need to address stable funding chains and sustainable business environments. Consequently, the strategic focus should shift toward establishing structured public-private partnerships that draw on combined strengths to directly serve resident needs directly [41]. Delivered services include intelligent safety systems with real-time alerts, digital medical services with seamless access, and governance platforms that support direct resident participation [42]. To achieve stable cooperation, the governments are suggested to develop standardized contractual frameworks that clarify responsibilities, risk sharing and service delivery standards, which ensure that residents receive consistent and high-quality integrated services in safety, healthcare, and governance [43]. This clarity is critical for attracting investment and technical expertise that can be translated into tangible benefits for residents, and practical service outcomes can guarantee residents' willingness to use community services.

5 Conclusion

This study proposed and applied a multi-dimensional classification model to map the demand landscape of residents in smart community development. Drawing on a hybrid clustering approach applied to 191 survey respondents from China, four resident demand clusters were identified: Cluster 1 (weak, service-focused demand), Cluster 2 (prioritized safety and governance demand), Cluster 3 (strong community services demand), and Cluster 4 (comprehensive demand).

This study presented valuable insights in theoretical, and practically terms. On the theoretical side, it offers a data-driven taxonomy that quantitatively organizes resident demands across multiple dimensions, moving the field away from isolated and function-specific analyses toward a more integrated interpretive framework. On the practical side, the findings give policymakers and planners a concrete classification tool. Matching a community to its demand profile allows for the design of targeted interventions, ranging from building foundational digital capabilities in Cluster 1 communities to instituting collaborative governance structures in Cluster 4, thereby improving resource allocation and strengthening the strategic impact of smart city programs.

Despite these contributions, two limitations should be acknowledged. On the one hand, the model's application has been within the specific socio-political context of Chinese cities, which may affect the generalizability of the demand cluster definitions. Future research could validate and refine this taxonomy in diverse international settings. On the other hand, it is necessary to conduct a further analysis of the evolutionary pathway among the clusters of residents' demands, thereby generating more precise strategy framework for smart community development.

References

[1] N. Zhang, X. J. Zhao, and X. P. He, Understanding the relationships between information architectures and business models: An empirical study on the success configurations of smart communities, *Government Information Quarterly,* 37(2):101439, 2020.

[2] K. Jakobsen, M. Mikalsen, and G. Lilleng, A literature review of smart technology domains with implications for research on smart rural communities, *Technology in Society,* 75:102397, 2023.

[3] A. Asadzadeh, S. Pakkhoo, M. M. Saeidabad, H. Khezri, and R. Ferdousi, Information technology in emergency management of COVID-19 outbreak, *Informatics in Medicine Unlocked,* 21:100475, 2020.

[4] T. T. Gu, Q. Z. Xu, X. Y. Song, E. Y. Hao, P. Cui, and M. H. Xie, Analysis of influencing factors and their inner mechanism of the market participation in the smart community construction of China, *Ain Shams Engineering Journal,* 15(7):102761, 2024.

[5] Y. Lee, S. Han, and Y. Cho, Navigating the Path to Smart and Sustainable Cities: Insights from South Korea's National Strategic Smart City Program, *Land,* 14(5):928, 2025.

[6] N. Streitz, Beyond "smart-only' cities: redefining the "smart-everything' paradigm, *Journal of Ambient Intelligence and Humanized Computing,* 10(2):791-812, 2019.

[7] C. Gao and X. Zhao, Concept of a demand-response model for smart community construction: a case study in Ningbo, China, *Scientific Reports,* 15(1):6893, 2025.

[8] J. Guo and W. H. Ling, Impact of Smart City Planning and Construction on Community Governance under Dynamic Game, *Complexity,* 2021:6690648, 2021.

[9] S. L. Miao and H. Tan, Models and Logic of Smart Community Governance in a Centralised System, *China-an International Journal,* 23(3):45-63, 2025.

[10] T. T. Gu, D. Z. Li, and L. Z. Li, The Elderly's Demand for Community-Based Care Services and Its Determinants: A Comparison of the Elderly in the Affordable Housing Community and Commercial Housing Community of China, *Journal of Healthcare Engineering,* 2020:1840543,

2020.
[11] C. D. Vizia, E. Patti, E. Macii, and L. Bottaccioli, A User-Centric View of a Demand Side Management Program: From Surveys to Simulation and Analysis, *IEEE Systems Journal,* 16(2):1885-1896, 2022.
[12] X. Hu, Y. Li, Y. Sun, C. Zhang, W. Ding, and J. Deng, Water permits trading framework for urban water demand management based on smart metering, *Journal of Environmental Management,* 304:114208, 2022.
[13] W. Zhang, B. Guo, W. Zhao, Y. He, and X. Wang, Making Smart Cities Human-Centric: A Framework for Dynamic Resident Demand Identification and Forecasting, *Sustainability,* 17(21):9423, 2025.
[14] Y. F. Jiang, H. L. Long, Y. T. Tang, and W. Deng, Measuring the role of land consolidation to community revitalization in rapidly urbanizing rural China: A perspective of functional supply-demand, *Habitat International,* 155:103237, 2025.
[15] T. T. Gu, E. Y. Hao, C. Y. Wang, S. Y. Zhu, and Y. C. Wang, CRITIC-PROMETHEE II-Based Evaluation of Smart Community Services: A Case Study of Shenzhen, China, *Journal of the Knowledge Economy,* 16:3286–3320, 2024.
[16] P. Jinqing, L. Zhengyi, T. Yutong, J. Haihao, Y. Rongxin, and Y. Jinyue, Balancing stakeholder benefits: A many-objective optimal dispatch framework for home energy systems inspired by Maslow's Hierarchy of needs, *Advances in Applied Energy,* 13:100160, 2024.
[17] D. Zhang, Z. Li, J. Zhong, and J. Yang, A framework for prioritizing urban ecological infrastructure (UEI) implementation tasks based on residents' ecological demands and government policies, *Journal of Environmental Management,* 354:120369, 2024.
[18] X. Wang, X. X. Zhang, and J. J. He, Challenges to the system of reserve medical supplies for public health emergencies: reflections on the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic in China, *Bioscience Trends,* 14(1):3-8, 2020.
[19] C. L. Chen, Z. Y. Lim, and H. C. Liao, Blockchain-Based Community Safety Security System with IoT Secure Devices, *Sustainability,* 13(24):13994, 2021.
[20] Y. Sha, M. Li, H. Xu, S. Zhang, and T. Feng, Smart City Public Safety Intelligent Early Warning and Detection, *Scientific Programming,* 2022(1):7552601, 2022.
[21] C. Wang, L. Wang, T. Gu, J. Yin, and E. Hao, CRITIC-TOPSIS-Based Evaluation of Smart Community Safety: A Case Study of Shenzhen, China, *Buildings,* 13(2):476, 2023.
[22] P. Sun, G. Draughon, R. Hou, and J. P. Lynch, Automated Human Use Mapping of Social Infrastructure by Deep Learning Methods Applied to Smart City Camera Systems, *Journal of Computing in Civil Engineering,* 36(4):04022011, 2022.
[23] S. Barns, E. Cosgrave, M. Acuto, and D. McNeill, Digital Infrastructures and Urban Governance, *Urban Policy and Research,* 35(1):20-31, 2017.
[24] J. Tan, J. Leng, X. D. Zeng, D. Feng, and P. L. Yu, Digital Twin for Xiegong's Architectural Archaeological Research: A Case Study of Xuanluo Hall, Sichuan, China, *Buildings,* 12(7):1053, 2022.
[25] F. K. Wang, J. Z. Zhang, and P. K. Zhang, Influencing Factors of Smart Community Service Quality: Evidence from China, *Tehnicki Vjesnik-Technical Gazette,* 28(4):1187-1196, 2021.
[26] M. Li, J. Shen, X. X. Wang, Q. Chen, X. Y. Liao, and L. Ren, A theoretical framework based on the needs of smart aged care for Chinese community-dwelling older adults: A grounded theory study, *International Journal of Nursing Knowledge,* 35(1):13-20, 2024.
[27] S. Sun, X. Y. Lyu, and J. Chen, Technology readiness and smart healthcare device usage intentions among Chinese elderly: A moderated mediation model of technology interactivity and subjective norms, *Health Informatics Journal,* 31(2):14604582251344828, 2025.
[28] D. Gagliardi, L. Schina, M. L. Sarcinella, G. Mangialardi, F. Niglia, and A. Corallo, Information and communication technologies and public participation: interactive maps and value added for citizens, *Government Information Quarterly,* 34(1):153-166, 2017.
[29] L. J. Wan and S. Q. Jiang, Research on the Influencing Factors of Sustainable Development of Smart Community, *Mathematical Problems in Engineering,* 2022:8420851, 2022.
[30] J. Y. Yin, J. Q. Wang, C. Y. Wang, L. X. Wang, and Z. Y. Chang, CRITIC-TOPSIS Based Evaluation of Smart Community Governance: A Case Study in China, *Sustainability,* 15(3):1923, 2023.
[31] L. Ivan, J. Dikken, and J. van Hoof, Unveiling the experienced age-friendliness of older people in Bucharest: A comprehensive study using the validated Romanian age-friendly cities and communities questionnaire and cluster analysis, *Habitat International,* 143:102973, 2024.
[32] R. Inbakaran and M. Jackson, Resident Attitudes Inside Victoria's Tourism Product Regions: A Cluster Analysis, *Journal of Hospitality and Tourism Management,* 13(1):59-74, 2006.
[33] F. Yang, X. G. Wen, A. Aziz, and A. K. Luhach, The need for local adaptation of smart infrastructure for sustainable economic management,

Environmental Impact Assessment Review, 88:106565, 2021.
[34] H. Guo, C. C. Yang, B. B. Liu, and F. Yang, Performance-based contracts in the sharing economy: A supply chain framework with application of Internet of Things, *Annals of Operations Research,* 326(SUPPL 1):1, 2023.
[35] B. Choudhuri, P. R. Srivastava, S. K. Mangla, and Y. Kazancoglu, Enterprise architecture as a responsible data driven urban digitization framework: enabling circular cities in India, *Annals of Operations Research,* 354(1):193-221, 2025.
[36] G. Esposito, A. Terlizzi, M. Guarino, and N. Crutzen, Interpreting digital governance at the municipal level: Evidence from smart city projects in Belgium, *International Review of Administrative Sciences,* 90(2):301-317, 2024.
[37] H. M. S. Badar, K. Mahmood, W. Akram, Z. Ghaffar, M. Umar, and A. K. Das, Secure authentication protocol for home area network in smart grid-based smart cities, *Computers & Electrical Engineering,* 108:108721, 2023.
[38] Z. J. Yan, L. Jiang, X. L. Huang, L. F. Zhang, and X. X. Zhou, Intelligent urbanism with artificial intelligence in shaping tomorrow's smart cities: current developments, trends, and future directions, *Journal of Cloud Computing-Advances Systems and Applications,* 12(1):179, 2023.
[39] A. S. Sani, E. Bertino, D. Yuan, K. Meng, and Z. Y. Dong, SPrivAD: A secure and privacy-preserving mutually dependent authentication and data access scheme for smart communities, *Computers & Security,* 115:102610, 2022.
[40] J. Y. Cho and S. K. Park, Key factors for sustainable operation of smart rural communities in aging societies: Voices of Korean community leaders, *Technology in Society,* 74:102306, 2023.
[41] G. Siokas, V. Kelaidi, and A. Tsakanikas, The smart city as a hub for nourishing public-private partnerships, *Sustainable Cities and Society,* 76:103466, 2022.
[42] D. Ali *et al.*, Integrating private health facilities in government-led health systems: a case study of the public-private mix approach in Ethiopia, *Bmc Health Services Research,* 22(1):1477, 2022.
[43] C. Benitez-Avila, A. Hartmann, and G. Dewulf, Contractual and Relational Governance as Positioned-Practices in Ongoing Public-Private Partnership Projects, *Project Management Journal,* 50(6):716-733, 2019.

An LLM-Integrated BIM Workflow for Rapid Early-Stage Layout Generation of Worksite Trailers

Chao-Hsu Pan[1], Ren-Jie Li[1], Liang-Ting Tsai[2], Cheng-Hsuan Yang[3], and Meng-Han Tsai[1]

[1] National Taiwan University of Science and Technology, Taipei, Taiwan
[2] University of Alberta, Edmonton, Alberta, Canada
[3] RoBIM Technologies Inc., Edmonton, Alberta, Canada
menghan@mail.ntust.edu.tw

Abstract

Worksite trailers are widely used as rapidly deployable prefabricated units, yet their design process must balance standardization for efficiency with customization for diverse functional needs. This study proposes a proof-of-concept early-stage design workflow that integrates a Large Language Model (LLM) with Building Information Modeling (BIM) to support requirement interpretation, historical case reuse, and preliminary layout generation for worksite trailers. The system accepts natural-language descriptions and dimensional constraints, converts them into structured JSON-based parameters, and connects them to BIM generation in Autodesk Revit.

To reduce repetitive modeling effort and improve design reuse, the workflow includes three modules: Historical Case Retrieval, Unit-Based Spatial Assembly, and Conversational Parametric BIM Modeling. Historical cases are matched through a weighted multi-criteria process based on function, spatial composition, occupancy, dimensions, and special-space requirements. Retrieved cases may be reused directly or decomposed into reusable spatial units for recombination. When no suitable precedent is available, or when further refinement is needed, users can generate or modify layouts through natural-language commands that are translated into executable BIM operations.

Prototype demonstrations show that the workflow can generate editable preliminary configurations, support iterative refinement, and improve early-stage design reuse. However, the system remains a proof-of-concept prototype, and formal time-efficiency testing, user studies, and large-scale retrieval benchmarks have not yet been conducted.

Keywords –

off-site manufacturing, Large Language Model (LLM), human–machine interface (HMI)

1 Introduction

Worksite trailers are a prevalent form of prefabricated modular construction, widely deployed as temporary accommodation units, offices, and command centers in construction sites and remote working environments due to their rapid deployment and transportability [1]. In principle, modularity should benefit from standardization to enable efficient delivery and time savings [2]. In practice, however, diverse usage scenarios create a persistent tension between "standardization for efficiency" and "customization for flexibility," where customer-specific requirements often challenge standardized production logic [3]. Even seemingly minor changes in spatial proportions, functional zoning, or dimensional adjustments can still require substantial modeling effort, partly because current BIM tools often provide limited user-centered support for rapid early-stage adaptation [4]. Meanwhile, high investment costs in digital hardware technologies may further discourage stakeholders from adopting advanced digital workflows at scale [5].

Current workflows typically address this tension through manual model modification or template-based extension, yet a lack of project-to-project knowledge transfer remains a persistent barrier [6]. Past cases are commonly scattered across drawings, model files, and individual experience, and a large portion of construction data remains unstructured and difficult to retrieve and reuse [7]. Consequently, case reuse still depends heavily on manual interpretation and rework, limiting the potential for digital data to optimize assembly processes and reduce waste in offsite contexts [8]. In early-stage design, stakeholders often need not a fully detailed model, but a usable preliminary model that can be generated quickly and refined iteratively; language-guided early layout generation has therefore become an emerging direction [9]. Prior work also indicates that integrating BIM with DfMA-oriented parametric and algorithmic collaboration can strengthen client engagement during

offsite configuration, but practical workflows for fast customization are still developing [10].

Building on these challenges, this study integrates Large Language Models (LLMs) with BIM-based modeling to support requirement understanding and early-stage spatial configuration generation for worksite trailers. The system accepts natural-language descriptions and dimensional constraints, translates them into structured and executable parameters, and uses these parameters to drive BIM modeling operations, enabling iterative refinement toward a usable configuration [11]. To address fragmented knowledge organization and limited reuse, the framework further incorporates a structured representation and matching mechanism for historical cases to support retrieval and reuse. Unlike one-shot generative layout methods, the proposed approach emphasizes stepwise refinement, reusable case knowledge, and direct editability in the BIM environment. Accordingly, the objectives of this study are to: (1) develop an LLM-centric requirement interpretation workflow that converts natural-language inputs into executable design parameters; (2) establish a reusable design knowledge workflow by structuring historical configuration cases for retrieval and adjustment; and (3) implement a BIM-based prototype to demonstrate an end-to-end pipeline—from requirement input to the generation of a usable early-stage configuration—thereby validating the feasibility of language-driven BIM for worksite trailer design.

2 Methodology

2.1 System Overview

This study proposes an early-stage worksite trailer design workflow that integrates Large Language Models (LLMs) with Building Information Modeling (BIM). The primary goal is to rapidly generate a usable initial configuration during the early design phase while supporting iterative adjustment and convergence. As shown in Figure 1, the proposed workflow comprises four stages: requirement input and interpretation, historical case retrieval and matching, parameterized model generation and update, and data saving and output. Through this structure, case reuse and parameterized BIM modeling are unified into a single coherent process, enabling consistent model creation whether users start from a retrieved seed layout or directly specified parameters.

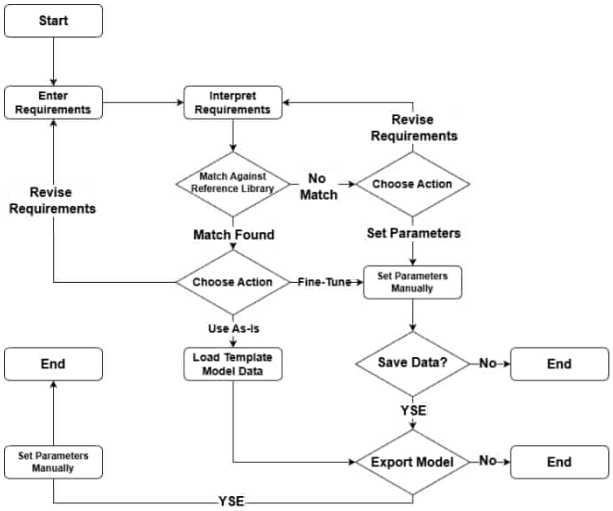


Figure 1. Framework of the proposed approach.

2.2 System Architecture

Figure 2 illustrates the system architecture of the proposed workflow. The framework is composed of three core modules: Historical Case Retrieval, Unit-Based Spatial Assembly, and Conversational Parametric Modeling. These modules do not function independently; instead, they are coordinated through a shared data mechanism based on JSON representations and a centralized ProjectStore for global state management.

Within this architecture, the Historical Case Retrieval module provides reusable reference cases based on user-specified requirements. The Unit-Based Spatial Assembly module enables the decomposition and recombination of reusable spatial units from retrieved or predefined cases. The Conversational Parametric Modeling module interprets user instructions and translates them into executable BIM parameters for model generation and update in Autodesk Revit.

A key feature of the proposed system is its common data mechanism. Rather than maintaining isolated intermediate results in each module, the system uses structured JSON data and ProjectStore to record global design states, shared schema definitions, reusable unit information, and BIM-related parameter mappings. This design supports multi-turn interaction, traceable parameter updates, and consistent synchronization between semantic interpretation and geometric model execution. The resulting BIM parameters are then exported to the Revit environment for model creation and modification.



Figure 2. System architecture of the proposed workflow

2.3 Historical Case Retrieval and Reuse

To improve retrieval consistency, each historical case is stored in a structured JSON-based representation containing key attributes such as function type, spatial composition, occupancy, approximate dimensions, and special-space requirements. Candidate cases are then scored through a weighted multi-criteria matching process based on these attributes. To reduce semantic inconsistency, room names are normalized into canonical categories, and partial similarity is allowed for related room types. The final ranking is generated from the aggregated similarity score, and the top-ranked cases are presented as candidate seed layouts for further reuse and refinement.

When a similar case is identified, the retrieved configuration is treated as a seed layout. The selected case is loaded as an editable set of design parameters, allowing users to adopt it as a starting point and refine it according to project-specific needs. This mechanism reduces modeling ambiguity, shortens the exploration cycle, and enables structured reuse across recurring worksite trailer scenarios.

If the retrieved seed layout still deviates from the user's requirements, the system supports iterative customization through parameter editing. Such parameterization not only supports geometric adjustment but also ensures that subsequent BIM updates remain consistent with the retrieved case context. The system does not require users to provide a complete specification at the beginning; requirements may be supplied progressively. In general, richer and more explicit inputs yield more complete retrieval conditions, which improves similarity ranking and increases the stability of case matching.

In the current prototype, retrieval effectiveness is examined mainly through qualitative inspection of returned cases rather than a formal large-scale benchmark, because the present case library is relatively small and manually curated.

2.4 Unit-Based Spatial Assembly

In addition to whole-case retrieval, the proposed workflow supports unit-based spatial assembly. This module allows users to decompose historical cases into reusable spatial units, such as bedrooms, washrooms, kitchens, office areas, or other functional spaces, and recombine them into a new configuration. As a result, the system supports not only case retrieval, but also case decomposition and case recombination.

The extracted spatial units preserve not only geometric properties but also basic semantic information, including space type, unit dimensions, functional role, and relative layout relationships with adjacent spaces. This enables the system to perform semantically informed spatial recomposition rather than simple geometric block placement.

The system also supports case decomposition preview. After a case is retrieved, users may inspect its internal composition and review which spatial units it contains before deciding which parts should be preserved or discarded. This function helps users better understand the internal structure of a precedent case and provides a basis for subsequent recombination.

A further advantage of this module is that spatial units may be selected across different cases. For example, users may choose a bedroom from one case, a washroom from another, and a kitchen from a third, and then reorganize these units into a new solution. This extends design knowledge reuse beyond rigid template-based reuse and allows a finer-grained design recombination process.

After the desired units are selected, the assembled result is converted into a new design basis for downstream modeling. The recombined configuration can then be connected to subsequent processes such as outer frame generation, room allocation, interior wall generation, opening placement, and space labeling. In this sense, Unit-Based Spatial Assembly serves as an intermediate design-restructuring mechanism between historical case knowledge and parameterized BIM generation.

2.5 Parameterized Layout Generation

When case matching is insufficient or no suitable case exists, users may directly specify model parameters, and the system generates an initial configuration through the same parameterized workflow. This design ensures that a usable early-stage model can still be produced even when the case library provides limited coverage.

In the parameterized modeling stage, the system drives BIM generation and updates using an editable set of model data. The process sequentially performs: (1) overall envelope dimension setting, (2) room space partitioning, (3) interior wall layout, (4) door/window

and other opening placement, and (5) room labeling. User instructions are interpreted through task-specific prompts and converted into structured JSON outputs, which are then mapped to internal data objects and BIM parameters.

After each parameter change, the system updates the BIM model and provides visual feedback, enabling low-cost iterative refinement and rapid convergence of the initial configuration. This approach allows the generated model to remain editable and traceable throughout the design process.

Finally, the workflow includes two post-processing options. First, users may store the generated result as a new case to enrich the case library. Second, they may export the BIM model as a baseline for subsequent detailing and engineering coordination.

3 Implementation

The proposed workflow has been implemented as a functional proof-of-concept prototype within a Building Information Modeling (BIM) environment. Autodesk Revit is adopted as the modeling platform, and the system operates as a Revit add-in that uses the Revit API (Application Programming Interface) to transform structured design data into corresponding BIM elements. The current prototype uses an API-based Large Language Model, GPT-4o mini, for requirement parsing and command interpretation. During development, another lightweight model was also tested; however, GPT-4o mini was selected because it provided sufficiently stable parameter extraction performance at a lower cost. This implementation enables an end-to-end pipeline from natural-language requirement input to BIM model generation.

To connect semantic interpretation and BIM execution, the prototype employs a unified structured data format as an intermediate layer. Rather than relying on a single universal prompt, the system adopts a task-specific prompt design strategy. Different prompts are used for subtasks such as requirement extraction, room partitioning, wall operations, opening placement, dimension modification, and intent routing. Each prompt explicitly requires a single valid JSON (JavaScript Object Notation) object without additional explanatory text, and the returned content is parsed into predefined internal data structures for storage and subsequent BIM execution.

In the implemented system, Historical Case Retrieval serves as an important starting point for early-stage design. When a user enters a request such as “I need a dorm unit for 3 people, including a kitchen, bedroom, and bathroom,” the system first analyzes the request and extracts key conditions, such as function, capacity, required spaces, and required dimensions. Figure 3 presents the case retrieval and selection interface, where the system displays candidate historical cases, matching levels, and selected projects based on the parsed conditions for further comparison and confirmation. This mechanism allows users to quickly identify potentially relevant configurations from previous cases, reducing the burden of manually searching and reviewing existing designs, while enabling early-stage design to begin from precedent knowledge.



Figure 3. Case retrieval and selection interface

In addition to presenting candidate cases, the system also provides more detailed matching information to help users understand why each project is recommended. Figure 4 shows the matched project results together with their associated attributes, including project function, spatial composition, capacity, dimensional conditions, and their degree of correspondence with the user’s requirements. Through this information, users can not only review the recommended cases, but also understand how each candidate aligns with the intended design conditions, thereby improving the transparency and interpretability of case comparison and selection.



Figure 4. Matched project results and associated attribute information

After candidate cases are obtained, the system further supports Unit-Based Spatial Assembly. Instead of only allowing whole-case reuse, the system can decompose

selected cases into smaller reusable spatial units, such as bedrooms, washrooms, kitchens, and service spaces. Figure 5 illustrates the preview of decomposed units from historical cases. This allows users to inspect the internal composition of each case, determine which spatial units should be retained, and identify which parts may be replaced. In this way, historical cases are no longer treated merely as complete templates, but are transformed into finer-grained design resources.



Figure 5. Preview of decomposed reusable spatial units from selected historical cases

Furthermore, the system supports selecting and recombining spatial units across different historical cases. Users may choose rooms or functional units from multiple projects and reorganize them into a new customized configuration. Figure 6 shows a recombined layout assembled from reusable spatial units originating from different precedent cases. This demonstrates that the system supports not only case reuse, but also flexible spatial recombination for early-stage customization.

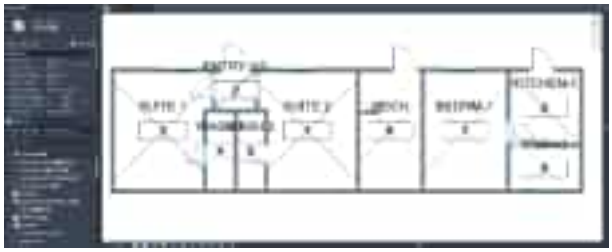


Figure 6. Recombined layout assembled from reusable spatial units

When Historical Case Retrieval and Unit-Based Spatial Assembly cannot provide a satisfactory solution, or when users need to further refine the configurations produced by the previous modules, the system can activate the third module, namely Conversational Parametric BIM Modeling. This module allows users not only to move beyond existing precedent cases, but also to directly describe spatial requirements in natural language or issue localized modification commands for an existing configuration. The system then progressively translates these inputs into executable modeling parameters and corresponding BIM operations, thereby supporting model creation, modification, and iterative refinement.

To achieve this, the system does not rely on a single universal prompt. Instead, it adopts a task-specific prompt design strategy. Dedicated prompts are defined for different subtasks, including outer-frame extraction, room-segmentation extraction, interior-wall command interpretation, wall trimming, opening placement, opening resizing, wall-thickness modification, and intent routing. These prompts explicitly require the model to output only a valid JSON object without any additional explanatory text. The returned content is then immediately parsed into predefined internal data structures for storage and subsequent BIM execution. If the returned output is empty, malformed, or violates predefined geometric or modeling constraints, the system does not execute the corresponding modeling action and instead requests a revised instruction. After user input is processed, the extracted information is written into the shared data structure, and corresponding modeling actions are triggered based on the interpreted parameters. At present, the prototype supports parameterized operations covering overall building dimensions, room count and target sizes, internal spatial subdivision through wall placement, opening placement (e.g., doors and windows), and room label generation for identification and model management. Figure 7 presents the input interface of the third module, where users issue natural-language instructions for direct generation or further refinement, while Figure 6 shows the recombined layout that may serve as the basis for subsequent conversational refinement.



Figure 7. Input interface of the conversational parametric BIM modeling module

Representative examples further illustrate how natural-language commands are translated into executable modeling actions. For overall dimensions, an input such as “Create a building with a total length of 526.875 inches and a width of 159.5 inches” is parsed and stored as envelope parameters in the structured data,

forming the basis for subsequent modeling. For room specification, an instruction such as “Divide it into three rooms with lengths of 211.625, 103.75, and 211.5 inches” updates the room count and assigns target dimensions to each room for later spatial partitioning.

For interior subdivision, the system supports wall placement commands to control room separation. For example, a vertical wall instruction such as “In Room 1, add a vertical wall 98.5 inches from the right side” is interpreted as a vertical partition wall with a specified offset relative to the room boundary, whereas a horizontal wall instruction such as “In Room 3, place a horizontal wall 60.5 inches above the bottom edge” is converted into a corresponding horizontal wall parameter. These wall definitions are recorded in the structured data together with orientation and relative offset information, enabling subsequent BIM generation.

For opening placement, the system allows users to specify doors, windows, or other openings through natural-language descriptions. For instance, a command such as “On the top wall of Room 1, put a door in the center of the right half” is parsed to determine the target room, host wall, relative placement region, and opening type. The system then generates an opening record containing wall references, positional information, and size parameters for BIM execution.

The system also supports label and identification updates. For example, an instruction such as “Rename the label of Room 1 to BEDRM-1” updates the corresponding room identification fields in order to maintain naming consistency within the BIM model. Such label adjustments not only improve model management, but also help align the generated result with subsequent design communication and engineering coordination.

All parsed design information is persistently stored and reused during subsequent modeling steps. The BIM modeling module sequentially reads these structured data and uses the Revit API to generate or update the building envelope, space partitions, walls, openings, and labels. Through this step-by-step update process, the model provides immediate visual feedback after each parameter change, thereby enabling low-cost iterative refinement and rapid convergence. Overall, the third module is capable not only of generating a new model directly from textual requirements when no suitable precedent exists, but also of serving as an effective mechanism for detailed refinement after the first two modules have produced an initial configuration. This prototype implementation demonstrates the feasibility and repeatability of natural-language-driven BIM configuration generation in an operational modeling environment.

4 Discussion and Future Work

This study investigates the integration of Large Language Models (LLMs) with Building Information Modeling (BIM) to support early-stage layout design for worksite trailers. The proposed workflow demonstrates that natural-language interaction can be effectively linked to structured parameter generation and BIM-based modeling, enabling preliminary configurations to be developed more directly than in conventional manual workflows. For worksite trailer projects, where repeated customization is often required under dimensional and transport constraints, this approach offers a practical means of reducing the need to begin each design task from scratch.

A key strength of the proposed workflow lies in its integration of Historical Case Retrieval, Unit-Based Spatial Assembly, and Conversational Parametric BIM Modeling within a unified data mechanism. Through this structure, prior design cases are not treated merely as static references, but can instead be retrieved, examined, reorganized, and further refined according to project-specific needs. This makes the workflow particularly valuable in situations where both design reuse and flexibility are required, especially when user requirements do not fully correspond to any single precedent case.

At the same time, several limitations remain in the current study. The existing case library is still limited in scale and diversity, which affects the breadth of retrieval and the stability of case-based reuse. In addition, the present validation is based mainly on prototype implementation and representative demonstrations. Formal time-efficiency testing, user studies, and quantitative comparisons with conventional workflows have not yet been conducted. Although task-specific prompts and validation rules are applied to improve execution reliability, ambiguous or incomplete user input may still lead to inconsistent interpretation results. Furthermore, the current prototype focuses primarily on orthogonal layouts for container-type modular units and does not yet incorporate code compliance checking or deeper integration with downstream construction processes.

These observations suggest that the system should currently be understood as a proof-of-concept prototype rather than a fully deployable practical platform. Its main contribution lies in demonstrating that LLM-based requirement interpretation, reusable design knowledge, and BIM-based model generation can be organized into a coherent early-stage design workflow. This provides a meaningful foundation for subsequent development toward more robust, scalable, and practically applicable modular design support systems.

Future work should therefore focus on strengthening both the technical robustness and practical applicability

of the workflow. One important direction is the expansion and systematic organization of the historical case library. A broader and more diverse collection of cases would improve retrieval quality, enhance recommendation stability, and support more reliable reuse of spatial units across different design scenarios. Another important direction is the adoption of more rigorous evaluation methods, including time-efficiency analysis, controlled user experiments, and comparisons with conventional design workflows, in order to better assess the actual benefits of the system in practice. Such studies would also help clarify its effectiveness for users with different levels of design expertise. In addition, the interpretation of natural-language input can be further improved through stronger rule-based validation, better handling of vague or approximate expressions, and more advanced semantic reasoning mechanisms. In the longer term, the workflow may be extended beyond single-unit layouts to support larger-scale planning scenarios, such as multi-unit trailer arrangements and mixed-function camp configurations, together with building regulation checking, structural considerations, and mechanical-electrical-plumbing (MEP) integration.

5 Conclusion

This study proposed an LLM-integrated BIM workflow for rapid early-stage layout generation of worksite trailers. By connecting natural-language requirement interpretation with structured parameter processing and BIM-based modeling, the workflow supports the creation of editable preliminary configurations while also enabling the reuse and adjustment of prior design knowledge.

The prototype implementation confirms the feasibility of this approach and suggests that language-driven BIM workflows can provide a useful foundation for improving accessibility, efficiency, and design reuse in modular construction. Overall, this research offers an initial step toward more responsive and knowledge-supported design methods for worksite trailer planning.

References

[1] Smith, R. E. (2010). *Prefab Architecture: A Guide to Modular Design and Construction*. John Wiley & Sons.

[2] Gutiérrez, N., Negrão, J., Dias, A., and Guindos, P. (2024). Bibliometric Review of Prefabricated and Modular Timber Construction from 1990 to 2023: Evolution, Trends, and Current Challenges. *Sustainability*, 16(5), 2134.

[3] Marchesi, M., and Matt, D. T. (2017). Design for Mass Customization: Rethinking Prefabricated Housing Using Axiomatic Design. *Journal of Architectural Engineering*, 23(3), 05017004.

[4] Park, D. Y., Choi, J., Ryu, S., and Kim, M. J. (2022). A User-Centered Approach to the Application of BIM in Smart Working Environments. *Sensors*, 22(8), 2871.

[5] Hedayati, E., Kolaei, A. Z., Khanzadi, M., and Amiri, G. G. (2025). Implementation of hardware technologies in offsite construction (2014–2023). *Automation in Construction*, 170, 105948.

[6] World Economic Forum. (2016). *Shaping the Future of Construction: A Breakthrough in Mindset and Technology*. World Economic Forum.

[7] Wu, C., Li, X., Guo, Y., Wang, J., Ren, Z., Wang, M., and Yang, Z. (2022). Natural language processing for smart construction: Current status and future directions. *Automation in Construction*, 134, 104059.

[8] Gbadamosi, A.-Q., Mahamadu, A.-M., Oyedele, L. O., Akinade, O. O., Manu, P., Mahdjoubi, L., and Aigbavboa, C. (2019). Offsite construction: Developing a BIM-Based optimizer for assembly. *Journal of Cleaner Production*, 215, 1180–1190.

[9] Leng, S., Zhou, Y., Dupty, M. H., Lee, W. S., Joyce, S. C., and Lu, W. (2023). Tell2Design: A Dataset for Language-Guided Floor Plan Generation. *arXiv* preprint arXiv:2311.15941.

[10] Bakhshi, S., Chenaghlou, M. R., Rahimian, F. P., Edwards, D. J., and Dawood, N. (2022). Integrated BIM and DfMA parametric and algorithmic design based collaboration for supporting client engagement within offsite construction. *Automation in Construction*, 133, 104015.

[11] Qiu, Z., Liu, J., Wu, Y., Liu, P., Qi, H., Liang, H., and Xia, Y. (2025). LLM-based framework for automated and customized floor plan design. *Automation in Construction*, 180, 106512.

Prompt Engineering and Video Segmentation Strategies for VLM-Based Action Recognition in Prefabrication Workflows

Chien-Wen Chen[1], Ren-Jie Li[1], Liang-Ting Tsai[2], Cheng-Hsuan Yang[3], Meng-Han Tsai [1]

[1] National Taiwan University of Science and Technology, Taipei, Taiwan
[2] University of Alberta, Edmonton, Alberta, Canada
[3] RoBIM Technologies Inc., Edmonton, Alberta, Canada
menghan@mail.ntust.edu.tw

Abstract –

Automated productivity monitoring and robotic learning in prefabrication are hindered by the scarcity of structured demonstration data. To address this, this study investigates a Vision-Language Model (VLM) workflow for extracting engineering logs from wall panel assembly videos. We conducted a comparative empirical evaluation using a 2 × 3 factorial design to evaluate the interaction between video processing strategies (Continuous vs. Segmented) and prompt complexity. Results indicate that relying on internal attention mechanisms for long videos leads to significant temporal hallucinations. In contrast, the proposed "Segment-and-Caption" framework, combining physical temporal segmentation (visual constraint) with domain-specific prompting (semantic constraint), achieved the highest reliability, yielding a Duration-weighted Accuracy of 65.44% and a Macro-F1 score of 0.547 to account for class imbalance. Crucially, we identify a fundamental trade-off: while this "stateless" segmentation eliminates attention drift, it introduces an "Identity Gap" that precludes consistent worker tracking. This work establishes that decoupling visual context from temporal search is a critical strategy for reliable VLM-based monitoring, providing a validated baseline for converting unstructured footage into machine-readable data for downstream robotic planning.

Keywords –

Prefabrication; Prompt Engineering; Vision-Language Models; Action Recognition

1 Introduction

1.1 Industrial Background and Practical Motivation

Prefabrication relies heavily on synchronized assembly-line workflows to enhance efficiency [1, 2]. Despite this industrial shift, core tasks like wall panel assembly remain labor-intensive, with essential actions (e.g., fastening, alignment) performed manually due to automation limitations [3, 4]. Consequently, accurate monitoring is essential. However, current practices relying on manual supervision or passive CCTV are labor-intensive and fail to generate the structured, machine-readable data required for productivity profiling and human-robot collaboration [5].

1.2 Limitations of Existing Automation and Vision-Based Approaches

While sensor-based technologies like RFID track location effectively [6], they require intrusive instrumentation and lack the semantic context of specific assembly actions. Computer vision offers a non-intrusive alternative, yet conventional task-specific models require extensive training data and struggle to generalize across diverse site conditions [7]. Even emerging open-vocabulary models often underperform specialized detectors in identifying domain-specific elements [8]. Moreover, scaling these tools to real-world scenarios remains challenging due to robustness issues [9] and the difficulty of transforming raw visuals into digital twin-ready data [10].

1.3 Vision-Language Models for Construction Video Understanding: Opportunities and Challenges

Vision-Language Models (VLMs) offer a promising alternative by jointly reasoning over visual and natural language inputs. Unlike task-specific models, VLMs facilitate flexible, zero-shot activity understanding through text descriptions, eliminating the need for massive annotated datasets [11, 12]. Recent studies demonstrate that ontology-based prompting can guide VLMs to interpret construction images [13]. Theoretically, this allows for generating detailed activity logs directly from visual data, offering a pathway to automated productivity analysis.

1.4 Research Gap and Study Objectives

Despite their potential, applying VLMs to long, untrimmed prefabrication videos presents challenges. Models relying on internal attention often suffer from temporal hallucinations, inaccurately timestamping actions [14]. Furthermore, VLM performance is highly sensitive to prompt engineering [13]. While Video-LLMs are evolving, the specific interaction between prompt formulation and video segmentation strategies remains underexplored. To bridge this gap, this study provides initial empirical evidence by employing a 2 × 3 factorial design to evaluate how physical segmentation and domain-informed prompting jointly impact VLM performance and mitigate temporal hallucinations.

2 Methodology

2.1 Proposed Approach

We propose the "Segment-and-Caption" framework to strictly constrain VLM inference (**Figure 1**). As illustrated, the **baseline (top path)** applies temporal prompting directly to the raw, continuous video, relying on internal attention mechanisms, a process often leading to contextual hallucinations due to attention drift. In contrast, our **proposed method (bottom path)** introduces a modular decomposition to isolate temporal constraints. Specifically, the workflow enforces two primary mechanisms:

Physical Temporal Segmentation slices the footage into discrete clips based on ground truth timestamps. This "hard constraint" effectively renders the input stateless, limiting the model's focus to visual evidence within a specific time window. Complementing this visual isolation, Domain-Informed Prompting acts as a "semantic constraint." By injecting a construction knowledge into the prompt logic, we prevent generic descriptions and ensure outputs remain structurally consistent with engineering standards.

2.2 Experimental Design

To evaluate reliability, we adopted a 2 × 3 factorial design intersecting Input Strategy and Prompt Complexity. The Input Strategy compares the full-length Continuous baseline against the Segmented Input strategy, where action events are physically isolated to remove the burden of temporal search. Superimposed on these are three prompt levels: (1) P-Natural (naive instructions), (2) P-Syntactic (employing structural delimiters like XML to optimize parsing and enforce CSV outputs), and (3) P-Domain (integrating the construction knowledge). This progression allows us to determine whether improvements stem from formatting rules or domain-specific knowledge.

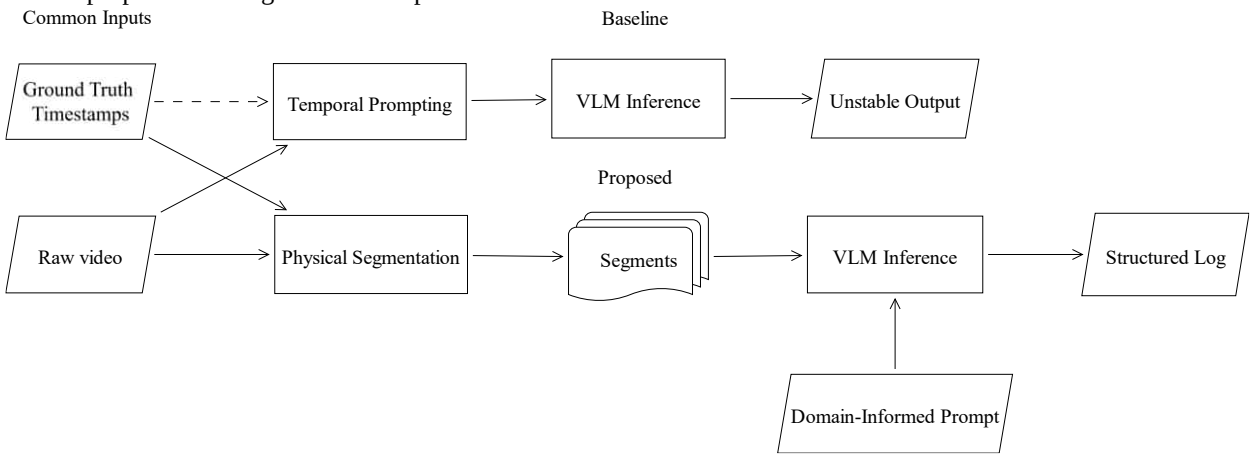


Figure 1. Workflow comparison between the Continuous Baseline and the Proposed Framework

3 Experimental Results and Analysis

3.1 Experimental Setup

The experiment utilized a high-density collaboration sequence (Duration: 6:57, 720p, 60fps) captured from a Canadian prefabrication factory, characterized by frequent occlusion and complex multi-worker interactions. The selection of this specific duration was governed by the native technical constraints of the Gemini 3 Pro web-based interface, which imposes a 2GB file upload limit. At approximately 1.87GB, this sequence represents the maximum continuous temporal context currently supportable by the platform's native multimodal pipeline, providing rigorous evaluation context for temporal stability without exceeding ingestion boundaries.

To simulate an open-world industrial monitoring scenario and ensure dataset transparency, the VLM was prompted with a construction knowledge comprising 8 task categories: *prepping*, *waiting*, *handling*, *fastening*, *process*, *measuring*, *fixing*, and *unsure*. While the ground-truth (GT) sequence contained 27 instances distributed across 4 active classes (*fastening*, *measuring*, *prepping*, *waiting*), the remaining 4 categories were treated as distractor classes to evaluate the model's robustness against over-classification.

All evaluations were performed using the Gemini 3 Pro (Web Version) with system-level decoding parameters (e.g., Temperature, Top-K, and Top-P) maintained at their default platform settings to ensure a standardized baseline and eliminate manual tuning bias. For the Continuous Input baseline, we fed the maximum possible context length directly into the Gemini 3 Pro web-based interface. Unlike API-based approaches that allow for programmatic chunking, this setup meant we were bounded by the platform's native file upload limits. Conversely, the Segmented Input strategy involved slicing this sequence into 27 discrete clips based on ground truth timestamps, isolating individual action events to test the efficacy of our "hard constraint" approach.

3.2 Quantitative Results

To provide a comprehensive evaluation and clarify the performance metrics against inherent class imbalance, we report three indicators in Table 1: Duration-weighted Accuracy, which reflects the proportion of total video time correctly classified; Clip-level Accuracy (Top-1), representing the standard discrete instance-based classification rate; and the Macro-F1 score, calculated exclusively over the active ground-truth categories to account for the recall of minority tasks.

As detailed in Table 1, the experimental data confirms that the proposed Segmented P-Domain strategy yielded the most reliable performance across all metrics, achieving a Duration-weighted Accuracy of 65.44%, a Clip-level Accuracy of 55.56%, and a Macro-F1 score of 0.547. This result validates the hypothesis that layering physical segmentation as a visual constraint serves as an effective method for stabilizing VLM outputs.

It is worth noting, however, that these metrics require careful interpretation, particularly regarding the Segmented P-Syntactic model. While this configuration achieved a seemingly competitive accuracy of 58.17%, a deeper inspection and the resulting lower Macro-F1 score (0.419) reveals this score was inflated by frequency bias (class imbalance). Given that the dataset is naturally dominated by repetitive "Fastening" tasks, the P-Syntactic model adopted a heuristic strategy of predicting the majority class for nearly every segment. Consequently, it failed to register minority but critical tasks such as "Checking" or "Alignment". In contrast, the P-Domain strategy demonstrated a genuine sensitivity to action variance as evidenced by its superior Macro-F1 performance, making it far more practically viable for industrial monitoring despite the modest numerical gap in raw accuracy.

Table 1. Comparative Performance of VLM Workflows

Input Strategy	Prompt Level	Duration-weighted Acc	Clip-level Acc (Top-1)	Macro-F1
Segmented	P-Domain (Proposed)	65.44%	55.56%	0.547
Segmented	P-Syntactic	58.17%	48.15%	0.419
Continuous	P-Syntactic	47.54%	37.04%	0.383
Continuous	P-Domain	47.09%	33.33%	0.257
Continuous	P-Natural	42.62%	37.04%	0.246
Segmented	P-Natural	42.06%	33.33%	0.167

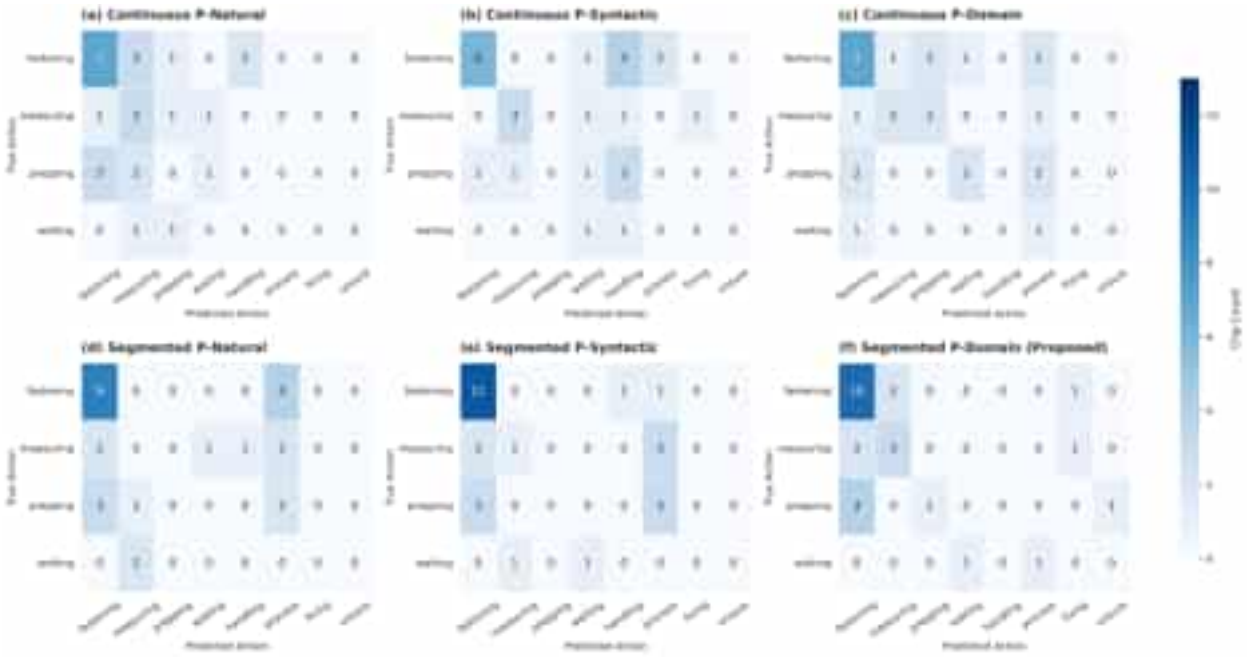


Figure 2. Comparison of action recognition performance across six experimental configurations.

3.3 Qualitative Analysis

While the quantitative metrics demonstrate the stability of the proposed method, a granular analysis utilizing the comprehensive 6-panel Confusion Matrix (Figure 2) reveals distinct operational characteristics and vulnerabilities of VLMs. As illustrated in Figure 2, the Proposed Method (f) achieves the highest concentration of correct predictions along the diagonal. This visually demonstrates the superior reliability of combining physical temporal segmentation with domain-informed prompting, effectively mitigating the attention drift and temporal inconsistency observed in continuous baselines.

The most prominent limitation is Small Object Blindness. Across all testing configurations, the models consistently struggled to resolve small tools, such as chalk lines or spray cans. The continuous baseline often compensated for this visual deficit by leveraging the full temporal sequence to infer actions (e.g., deducing "Chalking" from walking motions). However, the segmented approach, lacking this sequential context, frequently misclassified these same motions as generic "Measuring."

This visual limitation was mirrored by a semantic rigidity in the domain-informed models. By enforcing a strict construction knowledge, the VLM exhibited a tendency toward Taxonomy Over-fitting. As illustrated in Figure 2, by evaluating the model against 8 predefined options, we observed that it frequently misclassified ambiguous manual segments into distractor categories like *'fixing'* or *'process'* that were absent in the physical sequence. In its attempt to satisfy the "cognitive constraint," the model often forced nuanced or ambiguous behaviors into the nearest predefined category. This semantic hallucination explains the performance gap between raw accuracy and Macro-F1, as misclassifications into zero-support distractor classes directly penalized the recall of actual minority tasks like *prepping*. This indicates that while domain constraints successfully standardize output formatting, providing an overly broad knowledge without dynamic context-awareness can inadvertently degrade precision.

4 Discussion and Conclusion

4.1 Limitation and Future Work

Our results confirm that physical segmentation effectively eliminates the temporal hallucinations observed in continuous baselines; however, the relatively small sample size derived from this initial high-complexity sequence precludes a rigorous inferential statistical analysis at this stage. Consequently, these findings provide initial empirical evidence supporting the efficacy of the 'Segment-and-Caption' framework. While the current sample size precludes broad statistical claims of universal superiority, the mechanistic advantages of decoupling visual context from temporal search are clearly demonstrated.

Furthermore, this architectural choice creates a fundamental trade-off: by gaining temporal stability, we sacrifice narrative continuity. This results in an "Identity

Gap," where the "stateless" processing prevents the system from maintaining consistent worker identities across discrete clips (e.g., labeling the same individual as "Worker A" then "Worker B"). Consequently, while action recognition is accurate, individual productivity profiling remains challenging. Additionally, the current workflow relies on ground-truth timestamps. Future iterations will address these dual constraints by: (1) developing an Automated Event Detection module to dynamically slice video streams, removing manual dependency; and (2) fusing the VLM workflow with traditional computer vision-based tracking (Re-ID) to link worker identities across disjointed segments, thereby restoring continuity without sacrificing the stability gains of segmentation.

4.2 Conclusion

This study addresses the critical bottleneck of converting unstructured construction videos into machine-readable engineering logs. By validating the "Segment-and-Caption" framework, we demonstrated that decoupling visual context from temporal search is a necessary evolution for industrial monitoring. Our results confirm that applying physical segmentation as a hard constraint significantly outperforms continuous baseline methods in metric stability, achieving an Action Type Accuracy of 65.44%. While trade-offs in semantic depth and identity continuity remain, this framework establishes a viable technical pathway for automating granular productivity tracking in the next generation of prefabrication factories.

References

[1] Y. Fu, J. Chen, and W. Lu. Human-robot collaboration for modular construction manufacturing: Review of academic research. *Automation in Construction*, 158:105196, 2024. Doi: 10.1016/j.autcon.2023.105196

[2] A. Abu Nokta. *A Lean-based Approach to Increase Process Efficiency in Precast Concrete Panel Manufacturing*. Master's thesis, University of Alberta, Edmonton, Canada, 2025. Doi:10.7939/83076

[3] C. P. Chea, Y. Bai, X. Pan, M. Arashpour, and Y. Xie. An integrated review of automation and robotic technologies for structural prefabrication and construction. *Transportation Safety and Environment*, 2(2):81-96, 2020. Doi: 10.1093/tse/tdaa007

[4] X. Yang, F. Amtsberg, M. Sedlmair, and A. Menges. Challenges and potential for human-robot collaboration in timber prefabrication. *Automation in Construction*, 160:105333, 2024. Doi: 10.1016/j.autcon.2024.105333

[5] Z. Ren and J.I. Kim. The role of AI in on-site construction robotics: A state-of-the-art review using the sense-think-act framework. *Buildings*, 15(13):2374, 2025. Doi: 10.3390/buildings15132374

[6] I. Awolusi, E. Marks, and M. Hallowell. Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. Automation in Construction, 85:96-106, 2018. Doi: 10.1016/j.autcon.2017.10.010

[7] R. Xiong, Y. Wang, J. Cai, K. Liu, Y. Zhu, P. Tang, and N. El-Gohary. OpenConstruction: A systematic synthesis of open visual datasets for data-centric artificial intelligence in construction monitoring. *arXiv*, 2025. Doi: 10.48550/arXiv.2508.11482

[8] A. Abdalwhab, A. Imran, S. Heydarian, I. Iordanova, and D. St-Onge. Are open-vocabulary models ready for detection of MEP elements on construction sites. *arXiv*, 2025. Doi: 10.48550/arXiv.2501.09267

[9] M. Wang, J. Cai, D. Hu, Y. Hu, Z. Han, and S. Li. AI-based robots in industrialized building manufacturing. *Frontiers of Engineering Management*, 12(1):59-85, 2025. Doi: 10.1007/s42524-025-4099-x

[10] A. Pal, J. J. Lin, S. H. Hsieh, and M. Golparvar-Fard. Automated vision-based construction progress monitoring in built environment through digital twin. *Developments in the Built Environment*, 16:100247, 2023. Doi: 10.1016/j.dibe.2023.100247

[11] Y. Tang, J. Bi, S. Xu, L. Song, S. Liang, T. Wang, D. Zhang, J. An, J. Lin, R. Zhu, A. Vosoughi, C. Huang, Z. Zhang, P. Liu, M. Feng, F. Zheng, J. Zhang, P. Luo, J. Luo, and C. Xu. Video understanding with large language models: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. Doi: 10.48550/arXiv.2312.17432

[12] S. Yin, C. Fu, S. Zhao, K. Li, X. Sun, T. Xu, and E. Chen. A survey on multimodal large language models. *National Science Review*, 11(12):nwae403, 2024. Doi: 10.1093/nsr/nwae403

[13] C. Zeng, T. Hartmann, and L. Ma. Ontology-based prompting with large language models for inferring construction activities from construction images. *Advanced Engineering Informatics*, 69:103869, 2026. Doi: 10.1016/j.aei.2025.103869

[14] Y. Wang, Y. Wang, P. Wu, J. Liang, D. Zhao, Y. Liu, and Z. Zheng. Efficient temporal extrapolation of multimodal large language models with temporal grounding bridge. *arXiv*, 2024. Doi: 10.48550/arXiv.2402.16050

SAGE: A Knowledge-Grounded Multi-Agent Framework for Automated MEP Drawing Interpretation

Vivek Vishwas Vichare[1], Pratik Khandelwal[1], Aditya Debnath[2], Divya Singh Rathore[3], Gauri Lamb[4], and Paul O'Neill[4]

[1]Crafted Data Labs Private Limited (Encraft AI), India, [2]Indian Institute of Technology Kharagpur, India, [3]Indian Institute of Technology (BHU) Varanasi, India, [4]FHP Group, United Kingdom
vivek@encraft.ai, pratik@bingeclip.ai, aditya0202debnath@kgpian.iitkgp.ac.in, divyasingh.rathore.phy19@itbhu.ac.in, g.lamb@fhpgroup.com, P.O'Neill@fhpgroup.com

Abstract – Mechanical, Electrical, and Plumbing (MEP) engineering drawings serve as the contractual foundation for construction projects, yet verifying their consistency remains manual and error-prone. While Vision-Language Models show promise, they exhibit symbol fragility, spatial reasoning failures, and limited cross-document reasoning on technical drawings. We propose SAGE, a multi-agent framework employing: (1) YOLO11 for region segmentation (0.912 mAP50); (2) Cache-Augmented Generation for project-specific knowledge grounding; (3) a dual-mode adaptive extraction engine with coordinate grid overlay; and (4) hierarchical assembly combining extractions across project drawings. Evaluation across three construction projects (23 drawings, 1043 components) demonstrated 91.9% precision, 98.6% recall, and 97.4% spatial accuracy—substantially outperforming zero-shot VLM baselines (GPT 5.2, Claude Opus 4.6, Gemini 3 Pro) which achieved comparable precision but recall degrading sharply with drawing complexity (20–29% across projects vs. SAGE's consistent 98–100%). Ablation studies confirm the critical role of each architectural component.

Keywords –
MEP Drawings; Vision-Language Models; Multi-Agent Systems; Construction Automation; Drawing Interpretation

1 Introduction

The construction industry struggles to maintain consistency between CAD drawings and specifications, especially in retrofit projects where existing conditions overlap with new designs. Most design errors arise during handovers between architects, engineers, and contractors, causing costly rework, delays, and material conflicts. MEP systems account for over 50% of project investment [1], while rework consumes 4–10% of costs—52% linked to poor data and miscommunication [2] leading to $88.7 billion in global losses and $15.8 billion in annual interoperability costs [3]. Recent Vision-Language Models (VLMs) show promise for document understanding [4] but still face spatial reasoning and symbol interpretation challenges [5]. Fine-tuning improves accuracy—up to 52% F1 gain on GD&T extraction [6]—yet scales poorly with new component types.

In MEP projects, information flows sequentially from architects to MEP engineers to contractors, with each handover introducing potential for specification drift and interpretation errors. Design-intent drawings may reference equipment schedules, legend sheets, and specification documents that reside in separate files, creating cross-document dependencies that manual review struggles to verify consistently. Discrepancies such as mismatched equipment ratings between schedules and plan drawings, or spatial conflicts between HVAC ductwork and electrical conduit routing, often remain undetected until construction, when correction costs escalate significantly.

This paper presents SAGE, a multi-agent framework addressing these limitations through: (1) YOLO11 [7]-based region segmentation isolating drawing blocks, legends, notes, and title blocks; (2) Cache-Augmented Generation for knowledge grounding, enabling project-specific symbol interpretation; (3) a dual-mode adaptive extraction engine with coordinate grid overlay for spatial referencing; (4) hierarchical assembly combining extractions across all project drawings; and (5) industry validation on real construction projects. Unlike prior automated drawing interpretation pipelines that rely on static processing, SAGE introduces four novel mechanisms: (a) an adaptive grid density function that dynamically adjusts tiling granularity based on measured drawing complexity; (b) a Surgical Merge Agent implementing zone-based deduplication for overlapping tile regions; (c) the first application of Cache-Augmented Generation for domain-specific symbol grounding in

technical drawings; and (d) a multi-agent orchestration pattern with hierarchical semantic assembly, representing a principled four-stage decomposition specifically designed for the MEP drawing domain, where each agent targets an identified VLM failure mode through focused task specialisation. Importantly, as the underlying VLM capabilities continue to advance, the agentic architecture is designed to inherit these improvements—stronger base models directly translate into improved extraction quality within the orchestrated pipeline, without requiring architectural changes.

2 Related Work

Engineering drawing digitization has evolved from rule-based systems to deep learning approaches [8]. For P&ID diagrams, deep learning pipelines have achieved 94.5% symbol recognition accuracy [9], while transformer-based approaches demonstrate 83.6% AP for simultaneous node and relationship detection [10]. Hybrid frameworks combining rotation-aware object detection with transformer-based vision-language parsers have demonstrated promising results for mechanical drawings [11].

VLMs face significant challenges with technical drawings, exhibiting 10-30% hallucination rates on complex scenes [12] and 50-60% spatial reasoning accuracy. Recent work demonstrates that fine-tuned open-source VLMs (e.g., Florence-2 with 0.23B parameters) can achieve 52% F1 improvement over zero-shot GPT-4o and Claude-3.5-Sonnet [6]. However, fine-tuning suffers scaling limitations, motivating our agentic approach leveraging zero-shot VLMs with structured knowledge grounding. Set-of-Mark (SoM) prompting enhances VLM visual grounding by overlaying markers on image regions [13], providing foundation for SAGE's grid overlay. Multi-agent architectures like AutoGen enable task decomposition with iterative refinement [14].

Despite these advances, no existing framework addresses the complete MEP drawing interpretation pipeline—from region isolation through cross-document knowledge grounding to spatially-aware component extraction. Rule-based systems lack the semantic flexibility to handle diverse drafting conventions across projects, while fine-tuned VLMs require costly per-project retraining. SAGE bridges this gap through a multi-agent architecture that preserves zero-shot VLM generality while grounding interpretation in project-specific context via CAG, eliminating the need for per-project fine-tuning. Notably, because SAGE treats the VLM as a replaceable inference engine within a structured agentic workflow, improvements in base model capabilities—such as enhanced spatial reasoning or reduced hallucination in future VLM generations—are expected to propagate directly into improved pipeline performance without architectural modification.

3 VLM Limitations on MEP Drawings

Systematic evaluation of zero-shot VLMs (GPT-4V, Claude 3.5, Gemini Pro Vision) on real-world MEP drawings identified four critical failure modes: Symbol Fragility—custom drafting symbols cause systematic misclassification without legend context; Spatial Reasoning Failure—models fail to establish spatial relationships, achieving only 50-60% accuracy [12]; OCR Degradation—performance on rotated, scaled, or densely packed annotations degrades significantly [15]; Limited Cross-Document Reasoning: VLMs struggle to incorporate information from referenced schedules and legend sheets. These findings motivated SAGE's multi-agent architecture, where specialized agents address each limitation through focused task decomposition. Section 5 presents a systematic quantitative baseline study using current-generation models (GPT-5.2, Claude Opus 4.6, Gemini 3 Pro) to validate that these limitations persist and intensify with drawing complexity despite substantial advances in base model capabilities.

4 Methodology

4.1 Framework Overview

SAGE decomposes MEP drawing interpretation into four sequential stages following multi-agent design principles [14]: S1 (Region Isolation) for semantic segmentation into functional regions; S2 (Knowledge Grounding) for pre-loading project-specific context; S3 (Adaptive Extraction) for dual-mode component extraction with coordinate grid overlay; and S4 (Assembly) for synthesis into structured semantic representation as shown in the Figure 1. We utilised Gemini 3 Pro (Vision API) for developing extractor, merging and assembly agents due to its superior visual reasoning capabilities. Each stage introduces a targeted mechanism addressing a specific VLM failure mode identified in Section 3: region isolation reduces visual search space, CAG eliminates symbol ambiguity, the adaptive grid forces exhaustive spatial traversal, and hierarchical assembly enables cross-document reasoning.

4.2 Stage 1: Region Isolation

YOLO11, fine-tuned on 200 annotated MEP drawings, partitions each sheet into four semantic region classes: Drawing Blocks (primary schematic content), Legend Blocks (symbol definitions), Note Blocks (specifications), and Title Blocks (drawing metadata).

The model achieved 0.912 mAP50 and 0.696 mAP50-95 on held-out test data.

4.3 Stage 2: Knowledge Grounding

SAGE constructs project-specific context by aggregating legend symbol-to-specification mappings, note block content, and project documentation. This context is pre-loaded using Cache-Augmented Generation (CAG) [16], enabling interpretation of project-specific symbol conventions.

4.4 Stage 3: Adaptive Extraction Engine

The extraction implements dual-mode processing with adaptive visual grid overlay inspired by Set-of-Mark prompting. Grid density G (Complexity Decision) = f(ρs, ρt, R) is computed from symbol density (ρs), text density (ρt), and resolution (R), estimated via a preliminary VLM assessment call on each Drawing Block. Standard-density drawings are processed by a single VLM instance (Baseline Mode); high-density schematics are partitioned into overlapping sub-regions processed in parallel (Dynamic Tiling Mode). Both modes extract: (1) component identification, (2) metadata (specifications, model numbers, sizing), and (3) grid cell location.

In Dynamic Tiling mode, 10% tile overlap ensures boundary components are fully visible but introduces redundant detections. A "Surgical" Merge Agent accepts components outside the seam zone as "Safe Entities" while forwarding "Conflict Candidates" within overlap boundaries for IoU-based deduplication and attribute union—keeping processing linear to overlap count.

4.5 Stage 4: Semantic Model Assembly

The Aggregation Module combines results from all project drawings into unified JSON representation. Stage 4 performs hierarchical merger across drawing blocks via Global Coordinate Registration: local coordinates transform to global system using block origins from YOLO segmentation. Output encodes component instances, spatial relationships from grid proximity, and system membership classification.

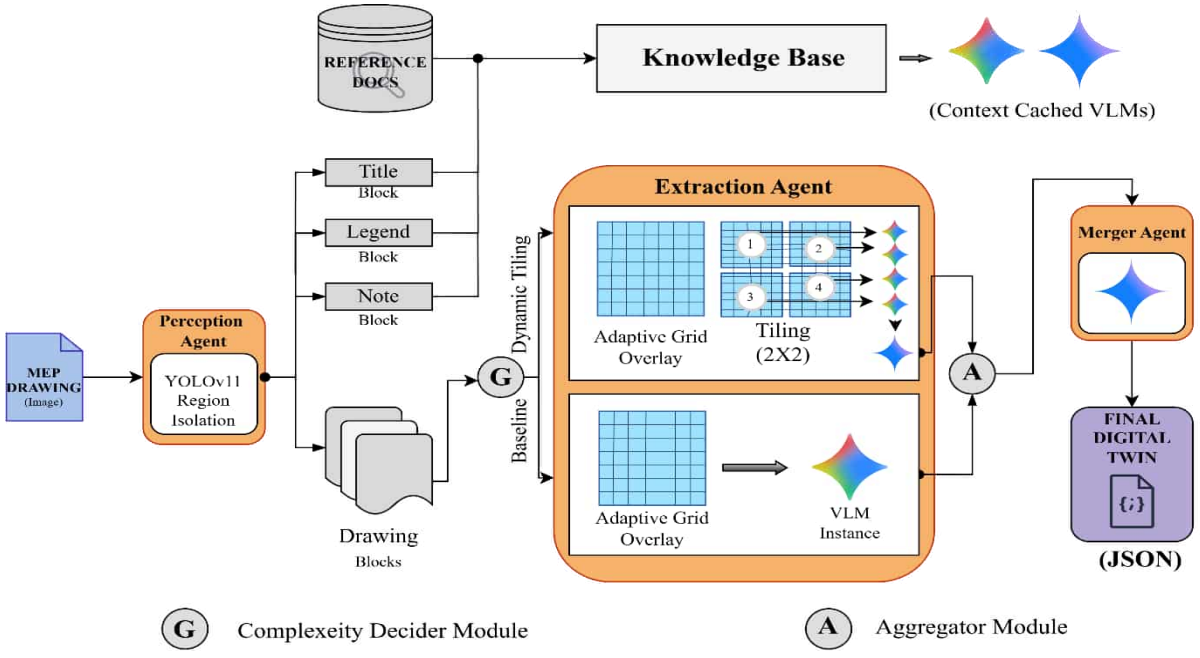


Figure 1. Multi-Agent Architecture (SAGE) for MEP drawing interpretation and Semantic Model Generation

5 Industry Case Study

5.1 Experimental Setup

We evaluated SAGE across three active construction projects provided by FHP Group (UK-based MEP consultancy). The dataset comprised 23 drawing sheets with 1,043 total extracted components spanning diverse MEP systems as detailed in Table 1. All drawings were exported as PDF from the original CAD/DWG files, representing the typical document format used in contractor handover and tendering workflows.

Table 1. Dataset Characteristics

Project	# of Drawings	Components	MEP System
1	5	66	Controls/BMS, Electrical, HVAC
2	6	339	HVAC, Electrical
3	12	638	Electrical, Public Health

5.2 Results

Results were obtained in JSON format as shown in Figure 2 and were accordingly compared with the ground truth validated by the expert design team at FHP Group.

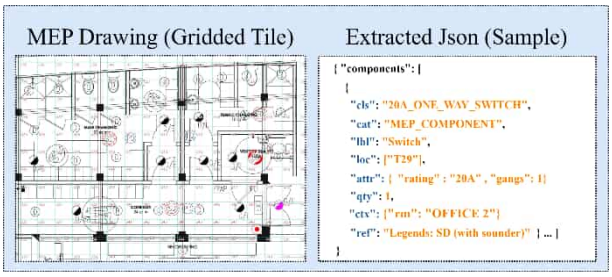


Figure 2. Sample extraction result showing a localized MEP drawing tile and its corresponding semantic JSON representation.

Precision measures the proportion of SAGE's extracted MEP components that match ground truth annotations, while recall captures the proportion of actual drawing components successfully detected. Spatial accuracy evaluates whether components are correctly localized to their grid cell positions.

Table 2. Evaluation Metrics

Project	Precision (%)	Recall (%)	F1-Score (%)	Spatial Accuracy (%)
1	100.0	98.48	99.24	100.0
2	90.16	99.71	94.69	97.05
3	92.05	97.96	94.91	97.44

To systematically quantify the failure modes identified qualitatively on earlier models, we conducted a baseline evaluation using current-generation VLMs. Notably, SAGE itself employs Gemini 3 Pro as its underlying VLM, making the Gemini 3 Pro baseline a direct controlled comparison—isolating the contribution of the agentic pipeline (region isolation, knowledge grounding, adaptive tiling, hierarchical assembly) from the base model's zero-shot capabilities. Spatial accuracy (Table 2) is excluded from the baseline comparison because it is architecturally inseparable from SAGE's coordinate grid overlay: without the grid, zero-shot VLMs produce only coarse qualitative localisations (e.g., "upper-left area") that are not commensurable with the grid-cell-level metric. The baseline comparison therefore focuses on detection quality (Precision, Recall, F1), where the models can be compared on equal terms. The results confirm that the same failure patterns persist even with substantially more capable models—and intensify with drawing complexity. Zero-shot VLMs achieve precision comparable to SAGE (88–91%) but weighted average recall of only 20–29%, with performance degrading sharply on complex drawings: on Project 1 (66 components), recall ranged from 51–92%, while on Project 3 (638 components, dense multi-system layouts), it collapsed to 8–23% across all three baselines. SAGE maintains near-complete recall (98–100%) regardless of drawing complexity. This demonstrates that raw VLM visual reasoning is adequate for sparse, well-separated layouts but fundamentally insufficient for the dense, information-rich drawings typical of real-world MEP projects; a limitation that persists across model generations. It is worth noting that as VLM architectures continue to mature with improved spatial reasoning, reduced hallucination, and stronger OCR on rotated annotations—these baseline capabilities are expected to improve. Because SAGE employs VLMs as inference components within a structured multi-agent workflow rather than as monolithic end-to-end systems, such advances in base model performance will propagate directly into improved pipeline-level metrics without requiring architectural changes to the framework.

Table 3. Zero-Shot VLM Baseline Comparison (Average across 3 Projects)

Model	Precision (%)	Recall (%)	F1-Score (%)
GPT 5.2	88.2	19.5	31.0
Claude Opus 4.6	89.3	29.3	43.3
Gemini 3 Pro	90.8	23.0	32.3
SAGE	91.9	98.6	95.1

As with the baseline comparison, the ablation evaluates detection quality rather than spatial accuracy, which is architecturally inseparable from the grid overlay.

Table 4. Ablation Study (Weighted Average across projects)

Configuration	Precision (%)	Recall (%)	F1-Score (%)
SAGE w/o CAG	50.5	49.7	48.0
SAGE w/o Region Isolation	91.5	79.5	86.2
SAGE w/o Grid Overlay	90.3	75.6	82.3
SAGE w/o Dynamic Tiling	89.9	45.2	59.3
SAGE	91.9	98.6	95.1

Project 1 (66 components, low-density) showed negligible ablation effects, as simpler layouts fall within zero-shot VLM capability; the impact is concentrated on Projects 2 and 3. Removing CAG caused the most severe degradation (F1: 95.1%→48.0%), as the VLM hallucinates without legend-to-specification grounding for visually homogeneous MEP symbols. Removing dynamic tiling caused the largest recall drop (recall: 98.6%→45.2%, F1→59.3%), as single-pass extraction misses dense component clusters. Removing the grid overlay maintained precision (90.3%) but penalised recall (75.6%), confirming its role as a deterministic spatial traversal mechanism. Removing YOLO produced moderate degradation (F1: 86.2%) from increased background noise exposure. The full pipeline achieves the optimal balance: CAG grounds semantics, the grid ensures spatial completeness, YOLO isolates signal from noise, and tiling ensures exhaustive extraction.

6 Discussions

SAGE demonstrates that knowledge-grounded multi-agent architectures effectively overcome zero-shot VLM limitations in technical drawings. The coordinate grid overlay achieved 97.4% spatial accuracy, while the Surgical Merge Agent eliminated duplicate detections in low-to-medium density cases. High-density seam regions remain challenging. The baseline comparison (Table 3) confirms that zero-shot VLMs maintain high precision (88–91%) but only 20–29% weighted recall. This high-precision/low-recall pattern reflects three limitations: VLMs lack mechanisms for exhaustive spatial scanning of dense technical layouts; project-specific drafting symbols fall outside pre-training distributions, causing systematic omission; and attention mechanisms struggle with counting repetitive elements in dense arrays. SAGE targets each limitation through CAG (symbol grounding), the adaptive grid (exhaustive traversal), and YOLO (signal-to-noise isolation). The framework accurately extracted and validated drawing details against specifications, identifying discrepancies in parameters such as pressure and power ratings—highlighting its value for automated compliance verification.

Current limitations include incomplete topological inference, reduced counting accuracy in dense regions, and a 3× computational overhead from dynamic tiling. Average processing time per drawing block is ~20 seconds when tiling is not included and ~120 seconds in Dynamic Tiling Mode. The grid density function $G = f(\rho s, \rho t, R)$ determines mode selection based on symbol density, text density, and resolution; across our dataset, approximately 40% of drawing blocks exceeded the complexity threshold triggering Dynamic Tiling, yielding an effective ~3× computational overhead at the project level. This cost is concentrated on high-density drawings where the accuracy-cost trade-off is most favourable. Early experiments using recursive legend-symbol prompting and integrating non-VLM methods (e.g., YOLO, R-CNN) show promise for improving counting.

Future work will extend region detection to new drawing types, explore graph-based spatial reasoning, adapt tile sizes dynamically, and develop hybrid pipelines combining VLM semantics with deterministic detection for greater robustness.

7 Conclusions

This paper presented SAGE, a multi-agent framework enabling automated semantic model generation from MEP drawings, providing AI-powered quality assurance before contractor handover. Evaluation across three real-world construction projects demonstrated 91.9% precision and 98.6% recall (F1: 95.1%), with 97.4% spatial accuracy. Zero-shot VLM baselines (GPT-5.2, Claude Opus 4.6, Gemini 3 Pro) achieved comparable precision (88–91%) but only 20–29% weighted average recall, confirming that knowledge grounding and structured spatial traversal, not raw visual reasoning alone, are necessary for reliable MEP drawing interpretation. Ablation studies confirmed the complementary roles of CAG (semantic grounding), adaptive grid overlay (spatial completeness), YOLO-based region isolation (signal-to-noise filtering), and dynamic tiling (exhaustive extraction in dense regions). As VLM capabilities continue to advance, SAGE's architecture is designed to inherit these improvements: stronger base models translate directly into enhanced pipeline performance without structural modification, positioning the framework for sustained relevance in construction automation workflows.

Acknowledgments

The authors thank FHP Group UK for providing project drawings and validation expertise.

References

[1] B. Wang et al. *Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes*, 162:105308, 2024.

[2] FMI Corporation. *Harnessing the Data Advantage in Engineering and Construction*, 2024.

[3] M. P. Gallaher et al. *Cost Analysis of Inadequate Interoperability*, NIST GCR 04-867, 2004.

[4] OpenAI. GPT-4 Technical Report. arXiv:2303.08774, 2023.

[5] C. Picard et al. *From Concept to Manufacturing: Evaluating Vision-Language Models for Engineering Design*, Artificial Intelligence Review, 2024.

[6] M. T. Khan et al. *Fine-Tuning VLM for Engineering Drawing Extraction*. arXiv:2411.03707, 2024.

[7] G. Jocher and J. Qiu, Ultralytics YOLO11," Version 11.0.0, Ultralytics, 2024. [Online]

[8] C. F. Moreno-García et al. *Digitisation of complex engineering drawings*, Neural Computing and Applications, 31:1695-1712, 2019.

[9] G. Su et al. *P&ID recognition based on deep learning*, Biomimetic Intelligence and Robotics, 4(1):100142, 2024.

[10] R. Rahul et al. *P&ID Digitization using Transformers*, arXiv:2411.13929, 2024.

[11] M. T. Khan et al. *From Drawings to Decisions*, Robotics and Computer-Integrated Manufacturing, 2025.

[12] Y. Huang et al. *A Survey on Hallucination in Large Vision-Language Models*, arXiv:2402.00253, 2024.

[13] J. Yang et al. *Set-of-Mark Prompting: Improving Visual Grounding in VLMs*, arXiv:2310.11441, 2023.

[14] Q. Wu et al. *AutoGen: Enabling Next-Generation Multi-Agent Collaboration*, Microsoft Research, 2023.

[15] T. Li et al. *Visual Merit or Linguistic Crutch? A Close Look at DeepSeek-OCR*, arXiv:2601.03714, 2026.

[16] Chan et al. *Don't Do RAG: When Cache-Augmented Generation Is All You Need for Knowledge Tasks.*, *arXiv:2412.15605*, Dec 2024.

A Systems Engineering Framework for Adapting General-Purpose Service Robots for Construction Project Operations

Hadrien Roy[1], Rongbo Hu[2], Satwik Arawalli[1], Shubham Singhal[1], Sanjiv B N[1], Zi Jie Tan[2], Hongjie Cai[2], Keiji Tanaka[2] and Soungho Chae[2]

[1]ugo, Inc., Japan
[2]Kajima Technical Research Institute Singapore, Kajima Corporation, Singapore
h-roy@ugo.plus, r.hu@kajima.com.sg, s-arawalli@ugo.plus, s-singhal@ugo.plus, s-bn@ugo.plus, zj.tan@kajima.com.sg, h.cai@kajima.com.sg, k.tanaka@kajima.com.sg, s.chae@kajima.com.sg

Abstract –

Construction project operations remain difficult to automate due to dynamic site conditions, limited workflow standardization, and the absence of adaptable robotic platforms. General-purpose service robots (GPSRs), while mature in navigation, sensing, and data acquisition, lack a structured methodology for deployment in construction environments. This paper proposes a domain-specific systems engineering framework to systematically transform existing GPSRs into operationally viable tools for construction project operations such as monitoring and inspection. Using a Modified V-Model with iterative prototyping loops, the framework integrates requirements engineering, modular system architecture design, and field-based validation. A case study based on a GPSR identifies functional gaps related to environmental variability, collaboration safety, and data interoperability. Implementation workflows demonstrate how GPSRs can be systematically integrated into construction projects. The results establish a reproducible engineering process that reduces development time and enables scalable, practical use of GPSRs on construction sites.

Keywords –

Construction Robotics; Digitalization; Mobile Robots; Monitoring; Requirements Engineering; Service Robots

1 Introduction

Experts believe that robotics has the potential to improve productivity in the construction industry [1]. However, its fragmented, project-based nature, dynamic environments, and reliance on conventional practices have slowed the adoption of robotics [1]. While other industries, such as manufacturing and agriculture, have successfully integrated robots, construction remains largely underserved. Existing solutions struggle with environmental variability, task versatility, safe human-robot collaboration, high cost of operation and maintenance, and integration with project workflows [1], [2]. General-purpose service robots (GPSRs) are general robotic platforms designed for monitoring and operational support in various environments such as factories and industrial facilities [3]. However, these robots are often not originally intended for the construction industry. Their adaptability, modularity, and robust sensing make them a promising foundation for new applications. This paper explores how GPSRs can be transformed to meet the unique demands of construction sites, overcoming current limitations while enhancing functionality, safety, and workflow integration. By leveraging existing capabilities in a new context, this paper aims to demonstrate a systematic pathway for robots to contribute to construction automation for the purpose of operation, monitoring, and maintenance.

1.1 Current Robotics Landscape in the Construction Industry

Most existing solutions are highly specialized for single tasks such as layout marking, welding, or inspection and are called Single Task Construction Robots (STCRs) [4]. While these systems demonstrate technical feasibility, they face persistent challenges:

- **Specialization and limited reuse:** Robots are often built for one task and certain site conditions, making it difficult to reuse them across different projects, phases, or dynamic environments [5]. This lack of generalization restricts long-term value and increases deployment costs.
- **High setup effort:** Many construction robots require extensive calibration, site preparation, or custom workflows, reducing the productivity gains that were intended [4].
- **Fragility in dynamic environments:** Frequent layout changes, unpredictable obstacles, dust, and weather can quickly degrade performance, cause

failures or require frequent human intervention [5].

- **High implementation cost:** High implementation and maintenance costs are often cited as one of the common challenges for construction practitioners to deploy robots [1], [6].
- **Cultural and legal constraints:** Despite technical viability and capability, implementation can also be hindered by site culture, workforce acceptance, and local regulations [1], [2].

In contrast, GPSRs have matured significantly in indoor industrial and commercial environments. The cost and price of general-purpose robots (e.g., service robots, industrial manipulators, humanoids) have continued to decline over the past years [7], largely due to growing scale and intense competition. They provide robust capabilities in:

- **Navigation and mapping:** Autonomous point-to-point and waypoint navigation with obstacle avoidance.
- **Perception and sensing:** Computer vision analysis, high-resolution photo/video capture, thermal imaging, and Bluetooth-enabled sensor data collection.
- **Data and communication:** Live streaming, long-term recording, and two-way communication via network connectivity for remote operations.

This contrast between the narrow, task-specific nature of current construction robots and the reusable capabilities of general-purpose robots creates an opportunity to explore how adaptable service robots can support construction project operations more effectively.

1.2 Problem Statement

To date, construction project operations remain heavily manual and inconsistent across sites. These operational tasks are critical for safety, progress tracking, and decision-making, yet they suffer from limited automation due to constantly changing environments and the lack of adaptable robotic platforms. At the same time, GPSRs have matured in navigation, sensing, and data collection, but are designed for controlled indoor settings such as offices, factories, and hospitals rather than construction sites. Despite advances in general-purpose robotics, there is no established methodology for adapting these systems to construction project operations. Construction sites are highly dynamic, unstructured, and often harsh, presenting challenges in:

- **Environmental variability:** Robots designed for controlled settings struggle with uneven terrain, dust, and constantly changing layouts.
- **Safety and human–robot collaboration**: Effective operation alongside personnel and machinery.
- **Task versatility**: Many operational tasks demand adaptability to irregular objects and variable conditions, yet existing robots are typically single-purpose.

These factors highlight the need for a reproducible engineering framework that systematically translates general-purpose robotic capabilities into safe, reliable, and operationally relevant tools for construction project environments

1.3 Objective

The objective of this study is to develop and (partially) validate a domain-specific systems engineering framework for rapidly transforming GPSRs to support construction project operations with clear requirements from the user. Specifically, the framework addresses inspection, monitoring, and workflow execution without requiring a full redesign.

2 Methods

To accelerate development and introduce more robotics companies to the construction industry, a domain-specific systems engineering framework was created to guide the transformation of GPSRs for on-site applications. Based on a case study, the methodology adapts established systems engineering principles, using a Modified V-Model for lifecycle structure and a structured requirements engineering process for concept development [8], [9].

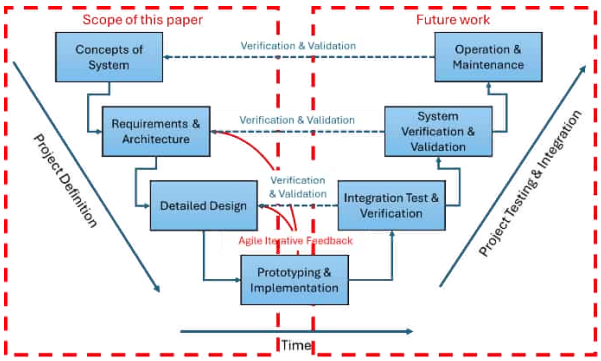


Figure 1. Modified agile V-Model in this research

2.1 Methodological Framework: The Modified V-Model

The V-Model was selected for its structured approach that links development phases with corresponding testing and validation activities. Since the original V-Model was developed for software development, our framework modifies this model by incorporating agile principles

(Figure 1) [8]. While the V-Model provides the overarching structure, small, informal iterative loops are introduced during prototyping to rapidly develop and test Minimum Viable Products (MVPs) [10]. This hybrid approach supports experimental validation of qualitative system features. This paper focuses on V-Model phases from "Concepts of System" through "Prototyping and Implementation," with later phases left for future work.

2.2 Requirements Engineering Process

A structured requirements engineering process is critical for translating multidisciplinary inputs into formal requirements and a coherent System Architecture. The systems engineering process (Figure 2), adapted from [8] and [9], begins with gathering Raw Requirements from domain-specific sources:

- **Key stakeholder interviews:** Structured interviews with key personnel such as site supervisors, engineers, robot operators, and environment safety contractors from the case study site in the target market are essential to understand their needs and pain points in daily construction project operations.
- **On-site observation via Gemba Walks:** Gemba (Japanese term for "real place") Walk is a lean management term referring to physically going to the real site where leaders can identify areas for improvement and cultivate a culture of continuous learning [11]. Direct observation of the construction environment through Gemba Walks at the case study construction site in the target market is essential to understand the dynamic conditions and conventional practices [12].
- **Market and state-of-the-art analysis:** Review of existing robotics solutions is helpful to identify common pitfalls and functional gaps.
- **Initial case study trial:** Baseline field test of GPSRs in the relevant environment is necessary to identify functional deficiencies and collect performance data.

Raw requirements are then converted into Formalized Requirements through (1) categorization, (2) reformulation using standardized syntax, and (3) assignment of reference codes. Requirements were grouped into Functional, Non-functional (i.e., performance), Domain (i.e., the construction sector), and Regulatory (i.e., especially relevant to the target market) categories [8].

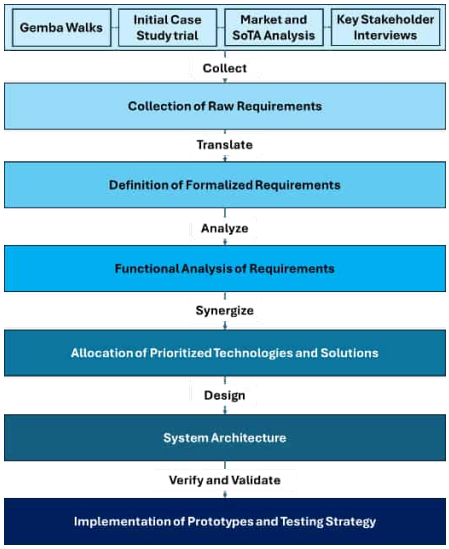


Figure 2. Requirements engineering methodological framework for rapidly transforming GPSRs to construction projects

2.3 System Architecture

Formalized requirements are organized to develop the system architecture, translating the "what" into the "how" [8]. The goal is to define a modular architecture that answers the research question: What system architecture enables modular transformation without full redesign? [9] Requirements are mapped to prioritized solutions—existing and new—to create a development roadmap. The resulting architecture blueprint forms the basis of the "Detailed Design" phase.

2.4 Detailed Design

The "Detailed Design" phase specifies the technical implementations derived from the system architecture. This includes selecting technologies and solutions and prioritizing them according to the requirements engineering outcomes [8]. The phase produces concrete hardware modifications and software integrations required to adapt a GPSR for construction tasks.

2.5 Prototyping and Testing Strategy

The framework then progresses to "Prototyping and Implementation". As outlined in Section 3.1, this phase uses an Agile methodology within the V-Model structure. Iterative prototyping is essential for construction environments where requirements evolve; prototypes validate individual functions quickly, enabling refinement until a stable system is achieved [5].

2.6 Integration Test and Verification

The next phase to be addressed in future work is the

"Integration Test, and Verification" phase. Once an MVP is produced, it undergoes live-environment integration testing. This stage employs the "Early Testing" loops [5] to confirm that the adapted system satisfies the functional requirements from Section 3.2 before progressing to full-scale deployment.

3 Results

As a highly developed tropical city state, Singapore has a robust economy and plays a pivotal role as a regional financial and innovation center in Southeast Asia, linking Eastern and Western markets. Its construction sector is one of the country's most important sectors, employing over 450,000 workers [13]. Like many countries, Singapore's construction sector faces persistent challenges such as stagnant productivity, construction waste, safety risks, labor shortages, and dependence on foreign workers [14]. Against this backdrop, government agencies have introduced a range of incentives to encourage both local and international firms to adopt innovative solutions in the built environment [12]. These conditions make Singapore an especially relevant and suitable setting for a case study, with the potential to scale the outcomes to other markets thereafter. This study introduces a commercialized GPSR platform from Japan to Singapore to explore how such platforms can be adapted to the construction industry in a completely different context (Figure 3).



Figure 3. GPSRs in the case study (credit: ugo, Inc.)

A preliminary field trial was conducted over a 6-month period at an active large-scale manufacturing facility construction site to assess technical and environmental constraints and determine necessary operational requirements. The field trial included participants from the robot manufacturer, construction company, and various project contractors. The robot supported project operations by executing scheduled patrols, monitoring environmental parameters, and providing real-time surveillance. Quantitative and qualitative data were gathered regarding operational metrics, points of failure, and site characteristics. After the conclusion of the field trial, interviews with multiple project managers were conducted, which included feedback from staff members that interacted with the robots directly. The empirical data obtained from this case study, in conjunction with the requirements engineering framework detailed in Section 2.2, were utilized to form both raw and formalized requirements for the construction robot.

3.1 Requirements Engineering Output

Based on the requirements analysis of the key stakeholders, the raw and formalized requirements in Table 1 are split into four categories – Functional Requirements (FR), Non-Functional Requirements (NFR), Domain Requirements (DR), and Regulatory Requirements (RR), representing four key areas that the stakeholders wish to focus on the most.

Table 1. Outputs of requirements engineering process

Requirements No.	Requirements
FR01	**Raw:** Unsafe environment/hazard detection
	Formalized: Capability to detect unsafe environment conditions on construction sites such as holes, steps, and barricades
FR02	**Raw:** Movable live streaming camera
	Formalized: Capability to turn or move livestreaming camera up and down, left and right to monitor more angles (e.g., ceiling, floor)
FR03	**Raw:** Automatic route generation by easy parameter input
	Formalized: Enable user parameter input function to auto-generate robot patrolling route and automatically link routes in different rooms/areas
NFR01	**Raw:** Quick setup and maintenance
	Formalized: Improve and enable rapid system setup (i.e., plug-and-play setup), integration with site management software and smooth maintenance workflows
NFR02	**Raw:** Network connectivity and stability improvements
	Formalized: Improve weak/unstable Wi-Fi/5G signal connectivity issue at remote or underground location
DR01	**Raw:** Sensor protection
	Formalized: Improve design of GPSRs to keep sensors inside free from dust and water from construction sites
DR02	**Raw:** Large language model **(**LLM) functions customized for construction industry
	Formalized: Enable LLM voice

	function for construction specific context
RR01	**Raw:** Comply with Singapore regulations
	Formalized: Adhere to regulatory requirements in the following aspects – Built Environment and Productivity, Workplace and Risk Management, Environmental Health and Pollution, and Cybersecurity and Communication Compliance

3.2 System Architecture and Functional Design

State-of-the-art solutions were surveyed for each requirement to determine the essential technological components. These components were then stratified into their respective requirement categories for the construction sector (e.g., FR, NFR, DR, RR), yielding the system architecture depicted in Figure 4.

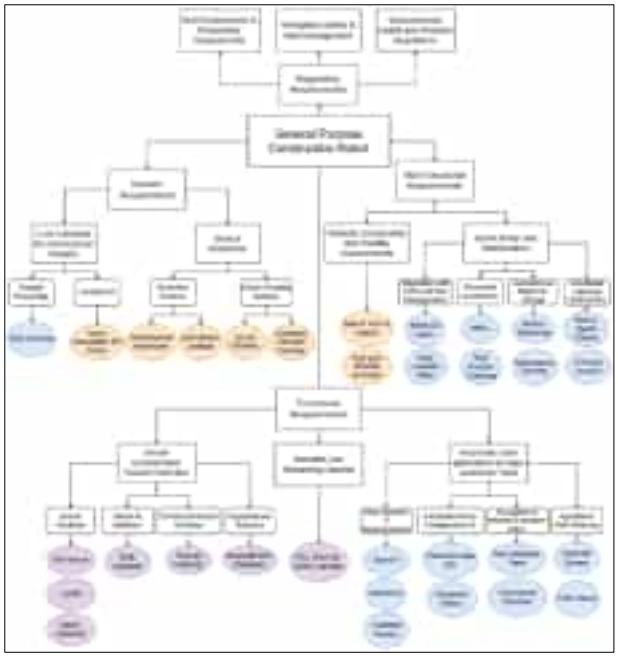


Figure 4. System Architecture of the adapted GPSR derived from requirements engineering

Functionally, it will support multi-modal hazard detection, movable live video streaming, and parameter-driven autonomous navigation. From a performance perspective, it enhances network connectivity and reliability while incorporating weatherproof protection for the sensors. Specifically for the construction sector, it integrates construction-focused LLM capabilities for domain knowledge and on-site interaction and enables quick setup and maintenance. Furthermore, it also complies with regulatory requirements in Singapore related to construction productivity, workplace safety, and environmental protection.

3.3 Development Planning

Based on the system architecture, a development plan was formulated following an evaluation of current systems and an analysis of requisite upgrades across the domains. Subsequent developments were defined based on solution feasibility and priorities, determined in collaboration with construction industry experts.

4 Implementation and Operational Workflow

Following the systems engineering and requirements definition process, the implementation phase aims to operationalize the proposed framework on real construction sites. Rather than focusing solely on technical upgrades, this stage emphasizes how a GPSR can be embedded into real-world construction workflows from preparation and deployment to charging, task execution, and data utilization.

This stage therefore functions as the robot's formal onboarding process. The implementation workflow bridges the gap between technical readiness and operational readiness, ensuring that the robot can function safely, autonomously, and efficiently within real project operations.

4.1 Site Preparation and Integration

Successful deployment begins with a structured preparation and integration process that ensures the robot's capabilities are safely and reliably embedded into site operations. Formalizing these steps makes deployment predictable, scalable, and repeatable across construction projects.

- **Site assessment:** A detailed survey identifies operational zones, access routes, hazard areas, and suitable dock locations. Terrain irregularities, material-strewn paths, temporary installations, and other construction hazards are mapped. Network coverage (Wi-Fi, 5G) is validated to ensure high-bandwidth connectivity across mission zones.
- **Infrastructure setup:** Setup includes installing charging docks, network repeaters, and safety markers along robot pathways. This step ensures stable power and safe dock placement. Physical boundaries are defined to prevent unintended entry into high-risk areas.
- **Software configuration:** Site layouts are imported into the mapping interface, followed by a calibration run to align digital maps with real

conditions. Digital no-go zones, speed-limited corridors, and temporary hazard boundaries are configured.

- **Connectivity integration:** Secure, stable data exchange is established between the robot's management portal and site systems, enabling real-time monitoring of position, mission status, battery level, and system's health.

4.2 Operational Workflow

Daily operation follows modular, repeatable workflows designed for efficiency, safety, and site awareness, balancing autonomy with human oversight.

- **Mission planning:** Operators define patrol, inspection, or monitoring tasks through a graphical user interface (GUI). Routes are generated based on the provided site maps. Typical tasks include safety compliance checks, environmental monitoring, hazard detection, meter readings, and routine visual inspections.
- **Autonomous execution:** The robot executes missions autonomously, navigating slightly uneven flooring, debris, temporary cables, and moving personnel. Onboard perception and hazard-detection algorithms support safe operation.
- **Data capture and reporting:** Images, sensor readings, and hazard detections are continuously recorded and uploaded to the management platform, where they are formatted into actionable reports based on site needs.
- **Human-robot coordination:** Personnel are briefed on robot operations, with clear escalation procedures for unexpected behavior. Supervisors can assign ad hoc tasks or trigger emergency stops without disrupting schedules.

4.3 Charging and Maintenance Procedures

Energy management and maintenance ensure reliable, safe autonomous performance and define the robot's Operational Design Domain (ODD) [15].

- **Autonomous return-to-charge:** After missions or at low battery, the robot automatically docks and begins charging, with status updates displayed in the operator's dashboard.
- **Maintenance scheduling:** Construction environments are particle-rich and harsh, which require proactive maintenance through scheduled tasks including: 1) Sensor integrity checks: Cleaning the LiDAR, cameras, and ultrasonics; running diagnostics, 2) Thermal management: Monitoring temperatures and clearing dust from cooling components, 3) Physical inspection: Checking wheels, actuators, seals, and enclosures for debris or damage, 4) Joint maintenance plan: Jointly defining a coordinated maintenance plan between the robot manufacturer and construction company and executing it by the construction company.

4.4 Workflow Adaptation for Construction Site

As construction environments evolve daily, robotic workflows must remain adaptive to maintain utility and relevance.

- **Periodic map uploads:** Operators upload site maps acquired from the site team, reflecting obstructions, and restricted zones, paralleling the site integration process.
- **Dynamic route regeneration:** The system recalculates routes when encountering new barriers, temporary structures, or shifting pathways, enabling safe navigation without manual reprogramming.
- **Collaborative operation:** Robotic tasks are synchronized with site operations - for example, performing patrols during off-hours, with supervisors reviewing inspection data at shift handover.
- **Data feedback loop:** Collected insights such as hazard detections, environmental metrics, and safety observations are fed back to the site team to facilitate regulatory reporting and proactive risk management.

4.5 Summary of Implementation Steps

The implementation plan follows a structured five-phase approach consistent with systems engineering, incorporating safety, maintenance, and integration practices required for construction sites (Table 2). Overall, implementation focuses on creating a structured yet adaptable operational environment where GPSRs can perform useful functions without a full redesign.

Table 2. Implementation plan derived from the requirements engineering process

Phase	Key Activities	Outputs
i. Preparation	Select pilot site; conduct survey; identify hazards; install dock, connectivity, and safety markers	Deployment-ready infrastructure and safety boundaries
ii. Integration	Import maps; configure no-go zones and tasks; perform calibration runs	Safe, validated operational test environment
iii. Field	Conduct autonomous	Operational

Operation	patrols; verify hazard detection; monitor logs, thermal health, and sensor integrity	data and performance logs
iv. Evaluation	Assess navigation reliability, data sync, detection accuracy, and worker–robot interaction	Pilot performance and safety compliance report
v. Optimization	Refine mission plans, update maps, adjust parameters, improve site workflows	A fully integrated, scalable operational framework

5 Discussion

The following sections highlight the methodology's novelty and effectiveness, the value of the engineering results, and the roadmap for future validation. In addition, the research limitations are outlined.

5.1 Novelty and Implications

The Modified V-Model successfully bridged the historical disconnect between robotic capabilities and construction realities in a rapid manner. By integrating "Gemba Walks", the framework captured dynamic site constraints, such as temporary cabling and dust accumulation, often absent from standard specifications. This observational data was directly formalized into the System Architecture and operational workflows, ensuring that the system was designed around real site operations (e.g., site induction & data handovers) rather than technical autonomy alone. Furthermore, the domain-specific requirements engineering process established a rigorous, traceable blueprint [8]. By linking every System Architecture component to specific "raw" requirements, the framework created a logical audit trail that prevented scope creep and ensured a demand-driven design. This rigor extended to defining a strict ODD [15], which explicitly mapped environmental boundaries such as surface flatness and lighting minimums to clarify the robot's role as a specialized tool for specific zones rather than a universal solution. Moreover, the systems engineering approach rapidly identified and specified all necessary components of the adapted GPSR, enabling early-stage cost-benefit analysis (CBA). By leveraging the established CBA framework for construction robots [16], key financial indicators of the adapted GPSR can be readily estimated prior to on-site deployment, which not only further validates the feasibility of the adapted robot but also strengthens client confidence.

5.2 Limitations

While the framework establishes a robust structural foundation, several requirement categories were excluded from this initial iteration. These limitations fall into three distinct aspects. First, addressable scope limitations such as advanced AI integration (e.g., LLMs) and local 5G infrastructure analysis were deferred to prioritize immediate feasibility but remain part of the future development roadmap. Second, inherent hardware constraints impose fixed operational limits. Specifically, the GPSR's wheel configuration restricts movement on rough terrain, a physical limitation that cannot be resolved through framework adaptation. Furthermore, as the study focuses on framework development, long-term feedback and quantitative metrics have not yet been gathered for verification and validation.

6 Conclusion

This paper presents a domain-specific systems engineering framework for rapidly transforming GPSRs into operational tools for construction project environments. By integrating a Modified V-Model with structured requirements engineering, the proposed framework offers a reproducible methodology that bridges the gap between controlled-indoor robotic platforms and the complex, dynamic conditions of construction sites, laying out the foundation for the next-phase development. The case study demonstrates that, while current GPSRs possess strong baseline capabilities in navigation, sensing, and data acquisition, they require systematic adaptation to address construction-specific challenges such as environmental variability, human–robot collaboration, and workflow integration. With the completion of the remaining work, the study is expected to fully validate the effectiveness and practicality of the proposed methodological framework.

The study mainly benefits construction project teams and robotics providers who want to deploy GPSRs for real site operations without the need for a full robot redesign, further lowering development cost, and accelerating development speed. More broadly, it provides a valuable reference for those who wish to rapidly adapt general-purpose technologies to new application scenarios and industries.

As indicated in Figure 1, future work will advance the project from the "Detailed Design" phase into the "Prototyping and Implementation" phase of the V-Model, eventually verifying and validating the system design, requirements, and project concept. The specific roadmap includes:

- **Prototyping and implementation:** Executing iterative prototyping loops to integrate sensing payloads and software modules.

- **Site testing:** Conducting field trials to validate performance under real conditions and collect quantitative data on localization and docking reliability.
- **Regulatory compliance and expansion:** Addressing deferred scope items, achieving compliance with Singapore's (and potentially other target markets') safety regulations, and integrating capabilities such as LLMs and 5G connectivity.
- **Iterative refinement:** Using site-testing results to drive feedback loops that may return the process to requirements engineering, system architecture, or detailed design for further optimization.

References

[1] J. M. Davila Delgado *et al.*, "Robotics and automated systems in construction: Understanding industry-specific challenges for adoption," *J. Build. Eng.*, vol. 26, p. 100868, 2019.

[2] G. Garcés, "Human-robot synergy in building construction: advances, challenges, and future horizons for construction 5.0," *Discov. Civ. Eng.*, vol. 2, no. 1, p. 135, 2025.

[3] M. Shirasaka *et al.*, "Self-Recovery Prompting: Promptable General Purpose Service Robot System with Foundation Models and Self-Recovery," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 17395–17402.

[4] T. Bock and T. Linner, "Construction Robots: Elementary Technologies and Single-Task Construction Robots," vol. 3. Cambridge University Press, p. 334, 2016.

[5] S. A. Prieto, X. Xu, and B. García de Soto, "A guide for construction practitioners to integrate robotic systems in their construction applications," *Front. Built Environ.*, vol. Volume 10, 2024.

[6] M. S. R. Espinoza, M. A. C. Contreras, K. M. A. Estela, C. J. N. Varillas, and G. F. R. Torpoco, "Robotics in building construction: Advantages and Barriers," in *2023 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, 2023, pp. 1–5.

[7] E. Demaitre, "Robot Prices Drop as AI Investments, Ethics Worries Increase, Finds Stanford HAI AI Index 2022 Report," Robotics 24/7. [Online]. Available: https://www.robotics247.com/article/robot_prices_drop_ai_investments_ethics_worries_increase_finds_stanford_hai_ai_index_2022_report#:~:text=Robot arms are becoming cheaper,top five requested by employers.

[8] T. Linner, R. Hu, K. Pawlitza, and J. Güttler, "Systems engineering for prevention- oriented , assistive technology situated in a multidisciplinary context," *Gerontechnology*, vol. 16, no. 3, pp. 129–138, 2017.

[9] T. Linner, R. Hu, K. Iturralde, and T. Bock, "A Procedure Model for the Development of Construction Robots," in *Innovation in Construction - A Practical Guide to Transforming the Construction Industry*, S. H. Ghaffar, P. Mullett, E. Pei, and J. Roberts, Eds., Cham: Springer International Publishing, 2022, pp. 321–352.

[10] R. Stevenson, D. Burnell, and G. Fisher, "The Minimum Viable Product (MVP): Theory and Practice," *J. Manage.*, vol. 50, no. 8, pp. 3202–3231, Mar. 2024.

[11] G. D. Boca, "The Gemba Walk – A Tool For Management and Leadership," *Ovidius Univ. Ann. Econ. Sci. Ser.*, vol. 0, no. 1, pp. 450–455, 2015, [Online]. Available: https://ideas.repec.org/a/ovi/oviste/vxvy2015i1p450-455.html

[12] R. Hu, J. Mao, S. Chae, and J. K. W. Yeoh, "Benefits of Construction Robots for On-Site Lean Construction: A Comparison and Evaluation Framework," *Eng. Constr. Archit. Manag.*, vol. 33, no. 2, pp. 914–936, 2026.

[13] W. H. Gan and D. Koh, "COVID-19 and Return-To-Work for the Construction Sector: Lessons From Singapore," *Saf. Health Work*, vol. 12, no. 2, pp. 277–281, 2021.

[14] M. F. Dulaimi, F. Y. Y. Ling, and G. Ofori, "Engines for change in Singapore's construction industry: an industry view of Singapore's Construction 21 report," *Build. Environ.*, vol. 39, no. 6, pp. 699–711, 2004.

[15] L. Mendiboure, M. L. Benzagouta, D. Gruyer, T. Sylla, M. Adedjouma, and A. Hedhli, "Operational Design Domain for Automated Driving Systems: Taxonomy Definition and Application," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–6.

[16] R. Hu *et al.*, "A Simple Framework for the Cost–Benefit Analysis of Single-Task Construction Robots Based on a Case Study of a Cable-Driven Facade Installation Robot," *Buildings*, vol. 11, no. 1, p. 8, 2021.

Development of a Prototype AI-Driven Career Guidance System for the Australian Construction Industry

Diya Yan[1], Yi Ding[2], Cynthia Changxin Wang[1], Riza Yosia Sunindijo[1], Xushuo Tang[2], Ziyao Lu[2], Wenqian Zhang[2], and Zhengyi Yang[2]

[1]School of Built Environment, University of New South Wales (UNSW) Sydney, Australia
[2]School of Computer Science and Engineering, University of New South Wales (UNSW) Sydney, Australia
diya.yan@unsw.edu.au, yi.k.ding@unsw.edu.au, cynthia.wang@unsw.edu.au, r.sunindijo@unsw.edu.au, xushuo.tang@unsw.edu.au, ziyao.lu@unsw.edu.au, wenqian.zhang@unsw.edu.au, zhengyi.yang@unsw.edu.au

Abstract –

The construction industry is undergoing rapid transformation driven by digitalization, automation, and changing workforce structures. Despite these shifts, career guidance practices within the sector remain largely manual, fragmented, and poorly aligned with dynamic labor market conditions, limiting their effectiveness for workforce planning and innovation adoption. This study addresses this gap through the design and development of an artificial intelligence (AI)-driven career guidance system tailored to the Australian construction industry. The proposed system architecture comprises four integrated layers: knowledge base, reasoning algorithm, user interface, and cloud-based deployment environment. The knowledge base consolidates theoretical insights from a systematic literature review and occupational data from government and industry sources. A hybrid reasoning approach combining rule-based logic and large language model (LLM) capabilities enables automated, scalable, and personalized career recommendations. The user interface was delivered through a web-based platform to support intuitive design and ease of use. Deployment was implemented through Amazon Web Services (AWS) Elastic Compute Cloud (EC2) to support scalable and reliable system hosting. After the prototype was developed, 32 construction students were invited to pilot the system and complete a questionnaire on perceived usefulness, usability, and impact, while also providing recommendations for improvements. This study contributes to construction technology management and service innovation by demonstrating how AI enabled automation tools can support workforce development, enhance career decision making, and inform strategic talent pipelines for education providers and industry stakeholders.

Keywords –

Artificial Intelligence (AI); Australia; career pathways; construction industry; Large Language Model (LLM)

1 Introduction

Technological advancement and the growing emphasis on digitalization and sustainability have substantially increased the complexity of construction work. Modern construction roles now demand not only technical and project management competencies but also digital literacy, data analysis, and sustainability-related skills [1]. For example, the rise of Building Information Modelling (BIM), digital twin technologies, and green building practices has transformed traditional site-based positions into multidisciplinary roles requiring hybrid technical and analytical capabilities [2]. As a result, the career landscape in construction is rapidly evolving, creating new opportunities while intensifying the ongoing skills shortages that threaten the industry's long-term sustainability and productivity.

Career counselling and guidance have long been recognized in literature as key measures for supporting informed and sustainable career decisions [3]. However, despite its importance, traditional career guidance remains limited in both scale and effectiveness. The advice provided rely heavily on the expertise and personal judgement of individual career advisers, and often lack up-to-date, industry-specific insights [4]. For example, career advisers who are unfamiliar with modern construction technologies or professional roles may unintentionally reinforce outdated stereotypes or overlook emerging career pathways such as digital engineering, sustainability consulting, or project data analysis. These limitations highlight the need for more systematic, scalable, and data-driven approaches to career guidance that can deliver accurate, personalized, and up-to-date information aligned with the realities of the modern construction industry.

The rapid advancement of artificial intelligence (AI)

and Large Language Model (LLM) in recent years has introduced new opportunities to address many of the limitations of traditional career counselling. Previous studies have attempted to use AI as a career advisory tool, demonstrating the potential of AI systems to simulate the functions of a human career coach. For example, Kulugh, et al. [5] developed an AI-based career recommendation chatbot that analyzed students' educational profiles to suggest potential occupational paths, while Cheng and Liang [6] combined natural language processing (NLP) with machine learning classifiers to align user goals with their backgrounds. These developments have marked a shift from static career databases to dynamic, conversational systems capable of analyzing user profiles and generating adaptive career suggestions. However, most existing AI career coaching tools are designed for general audiences and rely on global or non-specialized datasets, lacking the domain-specific depth required to address the distinct structure and workforce characteristics of the Australian construction industry. As noted by Miller, et al. [7], integrating structured industry knowledge into AI systems significantly enhances their precision, contextual relevance, and trustworthiness.

To fill this gap, this study aims to design and develop an AI-powered career guidance system for the Australian construction industry, supporting students, graduates, and career changers in making informed career decisions.

2 System Architecture Design

2.1 Knowledge Base Layer

A knowledge base layer serves as the foundation of the system by providing structured and domain specific information to suit reasoning algorithm. It ensures that the AI operates with targeted knowledge rather than relying solely on general purpose information, thereby reducing the risk of irrelevant outputs and hallucinations [8]. In this study, the knowledge base layer incorporates a career information repository that was developed to provide data-driven and evidence-based insights tailored to the Australian construction industry.

To establish the foundation for developing a comprehensive career information repository, this study first identified key factors influencing career decisions in the construction industry through a systematic literature review. As a widely used research method, a systematic literature review facilitates the identification, evaluation, and synthesis of existing studies on a specific topic [9]. The thematic synthesis approach was applied, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach to ensure transparency, reproducibility, and methodological rigor.

Scopus was selected for its comprehensive coverage of peer-reviewed journals in construction management, engineering, and career development. An initial title search using "construction" and "career" returned 493 papers. A three-stage screening process was applied. First, filters limited results to English-language journal or journal/review articles published after 2015 from comparable labor markets (Australia, US, UK, New Zealand, Canada), reducing the set to 110 papers. Second, titles and abstracts were manually screened to confirm relevance to construction careers and full-text availability, yielding 40 papers. Finally, only studies examining or proposing factors influencing career choice or development in construction were retained, resulting in 14 papers. This number aligns with other systematic reviews in the construction field [10].

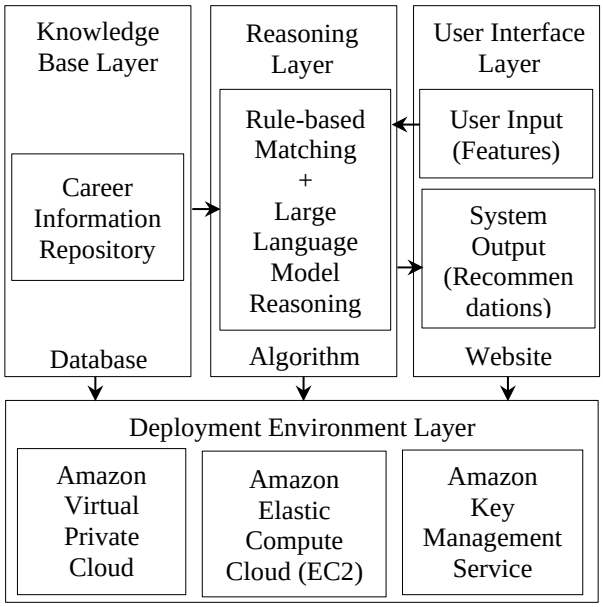


Figure 1. System architecture

The selected papers were analyzed using thematic analysis conducted in Lumivero NVivo 14, which facilitates efficient data management, coding, and theme development [11]. Each paper was reviewed manually, and descriptive codes were assigned to text segments that represented influential factors shaping careers in the construction industry. Codes with overlapping meanings were merged to reduce redundancy. Based on the thematic synthesis, the factors influencing career decisions were grouped into two major categories. Extrinsic factors refer to external conditions and contextual incentives that shape an individual's career choices, including opportunities for future growth, salary expectations, work environment, flexibility, and workplace diversity. Intrinsic factors relate to internal motivations and personal features that influence career preferences, educational background, technical competencies, soft skills, work style, and prior exposure to the industry. Table 1 summarizes the mapping of factors identified across the 14 reviewed studies.

Based on those key factors identified through the systematic literature review, the career information

repository was designed to incorporate 16 features, encompassing general role information, extrinsic factors, and intrinsic factors, as summarized in Table 2.

Table 1. Factors mapping overview

	Factors	Literatures
Extrinsic	Future growth	1, 2, 4, 11, 13,14
	Salary	1, 2, 11
	Work environment	1
	Flexibility	2, 3, 7, 12, 13
	Diversity	1, 4, 7, 8, 9, 13
	Physical demand	1
Intrinsic	Education	5, 9
	Technical skills	1
	Soft skills	8, 9, 10, 13
	Work style	2, 6, 7
	Prior exposure	3, 5, 8, 9, 10

Note: 1 = [12]; 2 = [13]; 3 = [14]; 4 = [15]; 5 = [16]; 6 = [17]; 7 = [18]; 8 = [19]; 9 = [20]; 10 = [21]; 11 = [22]; 12 = [23]; 13 = [24]; 14 = [25].

Table 2. Features included in the knowledge base

	Features	Description
Role	*discipline*	Construction disciplines
	role_description	General scope of daily work
	typical_titles	Variations in job titles
Extrinsic	*future_growth*	Growth outlook for the role
	career_pathway	Example progression
	salary_range	Annual salary range
	work_envir	The primary work location
	outside_hours	Work outside standard hours
	location_flex	Flexibility in work location
	hour_flex	Flexibility in work hours
	gender_ratio	Percentage of women
	physical_dem	Level of physical effort
Intrinsic	*minimum_edu*	Entry qualification required
	technical_skills	Technical competencies
	soft_skills	Interpersonal skills
	work_style	Characteristics of the role

The knowledge base of the prototype was scoped to focus on entry-level roles in the Australian construction industry, reflecting the initial development stage and time and budget constraints, with additional roles to be incorporated in future system iterations. A search was conducted on three commonly used job seeking websites, SEEK, CareerOne, and Workforce Australia, using the keyword “entry level” and the industry classification “construction”. Additional roles were identified through open discussions with industry professionals and members of the research team with backgrounds in the Australian construction industry, informing the selection of roles to be included. As a result, the knowledge base captured 70 commonly found roles across 10 disciplines in the Australian construction industry, as outlined below.

1. Construction Management and Project Management
2. Quantity Surveying and Cost Management
3. Civil Engineering and Structural Engineering
4. Architecture, Design, and Urban Planning
5. Building and Construction Trades
6. Health, Safety, Environment, and Quality
7. Building Surveying and Certification
8. Property, Real Estate, and Facilities Management
9. Mechanical, Electrical, and Services Engineering
10. Surveying, Geospatial, and Geographic Information System (GIS)

A web scraping process was conducted to collect career-related data from a range of credible public sources (n= 83), including government websites, industry associations, educational institutions, and major job-seeking platforms. The scraping was implemented in Python using the BeautifulSoup and Requests libraries to extract structured information. To ensure accuracy and data integrity, the final knowledge base was manually reviewed and validated by the research team, with inconsistencies cross-checked against official government and industry sources.

2.2 Reasoning Layer

The reasoning layer refers to the component responsible for processing user input, matching with the knowledge layer, and generating role recommendations. The core of this layer is the algorithms, which is designed to combine rule-based matching with LLM reasoning, as suggested by recent studies in information systems and artificial intelligence [26]. This hybrid approach allowed the system to balance accuracy with flexibility in handling complex and variable inputs. Among the 16 features, nine were utilized within the reasoning layer, while the remaining features were presented as system outputs.

Categorical features, including *work_envir*, *physical_dem*, *outside_hours*, *location_flex*, *hour_flex*, and *work_style* were designed to be matched using rule-based methods to ensure accuracy and consistency, as these data points in the knowledge base were predefined and discrete. Each feature was assigned a discrete score based on the degree of correspondence between user preferences and role characteristics.

Textual features, namely *minimum_education*, *technical_skills*, and *soft_skills*, were scored using LLM reasoning, as these inputs contain diverse linguistic expressions, synonyms, and contextual nuances that cannot be effectively captured through rule-based matching alone. To guide the LLM’s reasoning process and ensure consistent evaluation across users, prompt engineering techniques were applied to reduce hallucinations. It refers to the process of designing

structured instructions that guide the LLM to generate accurate, consistent, and contextually appropriate responses. The template is presented below.

- System Role: You are a professional career advisor with expertise in the Australian construction industry.
- Task: Analyze the user's background and match it with suitable roles using the knowledge base.
- Scoring Rules: Assign numerical scores for each feature based on the level of alignment between the user's profile and role requirements. Use continuous scoring rather than fixed anchor points to allow finer differentiation. Clearly identify matched and unmatched features, highlighting areas of strength and potential improvement.
- Response Structure: Strengths: Your background aligns well with this role because [analysis]. Areas for Development: Areas that may need further development include [analysis]. Interpersonal Skills: Employers often look for [skills]. Other Recommendations: To expand your career options, [personalized advice, industry associations, mentorship].
- Output Requirements: Return all responses in valid JSON format. Do not include any text outside the JSON structure. Maintain a warm, professional, and easy to understand tone using Australian English. Do not recommend qualifications lower than the user's current education level.

The total score for a given role r, expressed on a 100-point scale, is calculated as:

$$S_{rule}(r) = 5 \times \sum_{i=1}^{6} m_i \quad (1)$$

$$S_{LLM}(r) = (0.4 \times S_{edu} + 0.15 \times S_{tech} + 0.15 \times S_{soft}) \quad (2)$$

$$S_{final}(r) = S_{rule}(r) + S_{LLM}(r) \quad (3)$$

where m_i denotes the discrete score based on the degree of correspondence (0 for mismatch, 0.5 for partial match, and 1 for exact match). The terms S_{edu}, S_{tech}, and S_{soft} represent the feature scores (each on a 100-point scale) assigned by LLM. The weightings used in the formula were determined through research team discussions and remain adjustable through future performance testing. Future versions will also allow users to adjust the weighting scheme according to their preferences. In addition, a comparison was conducted to assess the performance of the system across five LLM backends: GPT-4.1, Claude-sonnet-4, Gemini-2.5-flash, Qwen-plus, and Deepseek-chat. Using ten participant resumes, each model generated career recommendations. The outputs were evaluated by three industry experts within the research team, each with over five years of experience in the Australian construction sector. Scores were assigned on a 1–5 scale, where 1 indicates the lowest level of satisfaction and 5 indicates strong satisfaction. Processing time and estimated computational cost were also recorded. Table 3 summarizes the results.

Table 3. LLMs Comparison Results

Model	Score ($\mu \pm \sigma$)	Time (s) ($\mu \pm \sigma$)	Cost (USD)
GPT-4.1	**4.45 ± 0.37**	10.92 ± 0.94	1.08
claude-sonnet-4	3.84 ± 0.34	11.60 ± 0.25	1.1
gemini-2.5-flash	3.78 ± 0.34	30.63 ± 2.50	0.96
qwen-plus	3.78 ± 0.34	**8.26 ± 0.34**	0.12
deepseek-chat	4.00 ± 0.34	15.65 ± 0.69	**0.05**

Note: μ = the mean value; σ = the standard deviation.

The results revealed performance of trade-offs among the models. GPT-4.1 achieved the highest expert-rated quality, indicating its suitability for high-stakes or decision-critical applications. Deepseek-chat demonstrated the most favorable cost efficiency while maintaining competitive performance, supporting scalable deployment scenarios. Qwen-plus consistently delivered the fastest response times, making it well-suited for interactive applications with modest quality trade-offs. Claude-sonnet-4 showed moderate performance across all metrics without a clear advantage, whereas Gemini-2.5-flash exhibited relatively higher latency without commensurate gains in quality or cost. Future work plans to allow users to select and compare different models within the system.

2.3 User Interface Layer

The user interface layer refers to the component through which users interact with the system. It connects with the reasoning layer to present outputs in clear and readable text. It collects user input, transmits the data to the reasoning layer for analysis, and presents the resulting recommendations in an accessible and user-friendly format. A web-based platform was chosen as the delivery format, as it offers broad accessibility, ease of deployment, and the ability to integrate seamlessly with backend services. The user interface and user experience design followed principles of clarity, simplicity, and accessibility, ensuring that the system is easy to read, navigate, and interact with. This layer was developed using python, Hypertext Markup Language (HTML), and JavaScript connecting to the reasoning layer.

To improve user convenience and minimize manual data entry, a resume importing function was embedded in the interface using spaCy and PyPDF2 for named-entity extraction. This feature allows users to upload their resume in Word or PDF format after consenting to the Privacy Protection Statement. Uploaded resumes are converted into plain text, segmented into key sections

(education, skills, certifications), and analyzed using pattern matching and entity recognition to identify relevant features. The extracted data are automatically mapped to the corresponding input fields in the interface, allowing users to review and edit the pre-filled information before submission. All files are processed temporarily and deleted after use to ensure privacy protection.

Users are required to complete three sections. The first section focuses on educational background, where users list their academic qualifications, certifications, and professional licenses. An example entry is provided to guide users in completing this section. The second section, skills and experience, collects information on users' technical skills, software proficiency, and interpersonal or soft skills. To assist users in formulating their responses, example suggestions are displayed beneath each input field, referencing common competencies in the construction industry. The final section, interest assessment, consists of six multiple choice questions corresponding to the features in the knowledge base, including work environment, physical demands, outside hours, location flexibility, hour flexibility, and work style. Each question is phrased in plain English to ensure comprehension and ease of response. Figure 2 presents the developed prototype.

Upon completion of the three sections, the results page generates a list of five career roles that best align with the user profile based on the provided inputs. Each role includes a brief overview, key characteristics, and AI-generated insights including strengths, areas for improvement, and recommendations for career development. Figure 3 presents an example of a recommended role output screen.

2.4 Deployment Environment Layer

The deployment environment provides the infrastructure necessary for hosting and operating the AI-powered career guidance tool online. It ensures system availability, data security, and scalability for handling multiple user requests simultaneously. Common deployment approaches include cloud-based, on-premises, and hybrid environments. In this study, cloud-based deployment was selected using Amazon Web Services (AWS) due to its scalability, reliability, and cost efficiency [27]. The system was deployed on Amazon Elastic Compute Cloud (EC2), which hosts the application backend, executes rule-based and LLM-driven computations, and manages communication with the knowledge base and user interface. EC2 offers elastic computing resources that can automatically scale to accommodate variable user demand, ensuring consistent system performance. The deployment operates within an Amazon Virtual Private Cloud (VPC), which isolates the system components and controls inbound and outbound network traffic, providing an additional layer of network security. The AWS Key Management Service (KMS) was used to encrypt data stored and processed on the EC2 instance, protecting sensitive information such as user inputs and uploaded resumes. This encryption framework ensures that data confidentiality is maintained in accordance with privacy and security standards. Data exchange between the user interface, backend, and external APIs is managed through AWS Data Transfer, which ensures stable and secure data transmission with low latency.

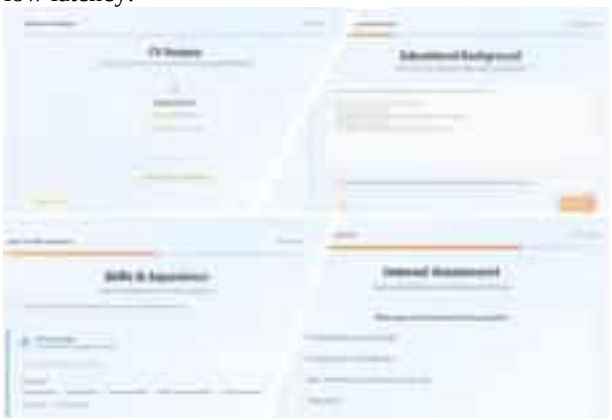


Figure 2. Developed web pages



Figure 3. Results page

The process begins with launching an EC2 instance, followed by configuring the security group to define access permissions. The website code is then deployed to the instance, after which the domain name is configured to enable public access. System monitoring is subsequently set up to track performance and availability. A decision point is used to verify whether the website is running as expected. If issues are identified, troubleshooting is performed before rechecking system status. Once the website is confirmed to be running successfully, the deployment process concludes.

3 Pilot User Experience

After the prototype was developed, users were invited to experience the system and provide feedback to inform

future improvements. A total of 42 students participated in the workshops. The questionnaire responses were screened for completeness and validity prior to analysis. Incomplete submissions (n=2), responses with missing demographic information (n=5), or careless responses (n=3; those in which participants selected the same answer for all questions) were excluded. The final dataset consisted of 32 valid responses. Descriptive analysis was performed using Microsoft Excel to summarize quantitative data, while qualitative feedback was analyzed through manual coding to identify recurring themes and improvement suggestions. The majority were aged between 18-25 years, with 59.3% identifying as female. Most participants (62.5% of total) reported having no prior work experience in the Australian construction industry, reflecting the target user group of the tool.

The quantitative section of the survey consisted of 10 Likert-scale items rated from 1 ("strongly disagree") to 5 ("strongly agree"), measuring three key constructs: usefulness, ease of use, and impact.

Usefulness refers to the extent to which users believed that the tool provided relevant and insightful career guidance. Three items were adapted from Venkatesh and Davis [28] with a Cronbach's alpha of 0.72. *Ease of use* refers to the extent to which users find the system intuitive and user-friendly. Four items were adapted from Venkatesh and Davis [28] with a Cronbach's alpha of 0.88. *Impact* refers to the extent to which users found the tool had an impact on their career and introducing new potential roles within the construction industry. Four items were adapted from Gedrimiene, et al. [29] with a Cronbach's alpha of 0.76. Table 3 presents quantitative results, indicating generally favorable perceptions across all statements.

Two open ended questions were used to capture qualitative insights from participants, focusing on which functions could be improved or made easier to use, and which additional functions users would like to see incorporated into the tool. The findings informed the future development roadmap, which is discussed in the following section.

4 Future Development Roadmap

Regarding existing functions, the most frequently mentioned area for improvement was system response time, noted by more than 40% of users. The commonly requested improvement, supported by over 35% of participants, was the ability to download and share results directly. At present, users can only export their results as a PDF using the browser's print function and share them manually. During the user evaluation workshops, several participants were observed taking photos of their results to keep personal records. Integrating a built-in saving and sharing feature, such as generating downloadable PDF summaries or enabling direct sharing through emails, would further improve the system's usability and accessibility.

Table 4. Descriptive Analysis Results

	Statements	Mean
Usefulness	1. The tool's recommendations were relevant to my skills and experience.	4.09
	2. The tool provided career options that matched my interests.	4.03
	3. The tool offered new career options I had not previously considered.	3.94
Ease of Use	4. The tool was easy to use.	4.44
	5. I felt confident in using the tool.	4.22
	6. The design and layout were clear and intuitive.	4.25
	7. The tool delivered what I expected it to.	3.94
Impact	8. The tool increased my confidence in making career decisions.	3.84
	9. The tool encouraged me to research new career pathways.	4.03
	10. I want to recommend this tool to others.	4.00

When participants suggested additional functions for future development, the most frequently recommended enhancement was career pathway visualization. Currently, the prototype presents future pathways in text only, without visual or interactive elements. Participants suggested that an interactive flow diagram or progression map could help users better understand advancement routes in construction, including transitions from entry-level to middle and senior roles. Such visualization could also enhance user engagement and learning effectiveness, as graphical tools have been shown to enhance users' comprehension of complex information and improve engagement [30].

Another widely supported suggestion was the introduction of an AI-powered mock interview practice module, recommended by approximately half of the participants. This aligns with Shi and Wang [31] emphasizing the value of simulated interview experiences in building communication confidence and employability skills. Such function could offer users a low-pressure environment to practice common industry questions, receive feedback, and prepare for recruitment. Additionally, over 35% of participants proposed the inclusion of an AI-powered chatbot to make the system "more interactive". This chatbot could serve dual purposes, including facilitating mock interview interactions and providing real-time responses to user queries about job recommendations, role descriptions, and next steps. Embedding conversational AI in this way would enhance engagement and deliver a more dynamic, personalized user experience.

In addition, over 40% of users suggested adding a peer stories or role model showcase section. This suggestion aligns with Perrenoud, et al. [32], highlighting the lack of visible role models as a barrier to participation, particularly among women and underrepresented groups in construction. Integrating authentic career stories could humanize the data-driven recommendations, demonstrate diverse pathways to success, and inspire users through relatable real-world examples. Developing this feature would, however, require collaboration with industry partners and professional associations to source and verify participant narratives.

Last but not the least, several participants recommended linking the system directly to current job postings to improve the immediacy and practicality of its recommendations. This could potentially be achieved by connecting the platform with established job-seeking websites such as SEEK or CareerOne. However, implementing such integrations may present challenges related to privacy, data sharing, and potential conflicts of interest. Future research could explore partnership models or API-based data-sharing frameworks that ensure compliance while providing users with up-to-date labor market insights.

5 Conclusion

This study presented the design and development of an AI-powered career guidance system tailored to the Australian construction industry as a technology enabled service innovation. The positive pilot user experience highlights the potential of automation tools and AI to support technology management, workforce planning, and innovative career guidance practices within the construction sector.

This study has several limitations that also present opportunities for future development. First, as the initial stage, the current prototype focuses solely on entry-level positions in the construction industry. Expanding the system to include mid- and senior-level roles would allow users to explore long-term career trajectories and progression pathways within the sector. Second, the features used in the prototype were not exhaustive. Future stages could incorporate additional matching features, such as language proficiency, cultural background, and personal values, to provide a more holistic representation of career decision factors. As with most LLM-based systems, the model remains sensitive to prompt formulation and exhibits potential brittleness, which may affect consistency in outputs. Future work could refine prompt-engineering strategies or adopt hybrid architectures that combine deterministic logic with adaptive AI reasoning to improve reliability. Additionally, future work should include comparison with human advisors to benchmark the system's performance, enabling a systematic evaluation of the relative advantages and limitations of AI- and human-based career guidance. Technical constraints also exist in relation to AWS deployment, including dependence on stable internet connectivity, ongoing server maintenance costs, and token usage fees associated with LLM processing. To enhance scalability and sustainability, future implementations could explore serverless architectures to enable wider accessibility and continuous operation. Last but not least, As the system provides career recommendations, potential bias and fairness issues must be carefully considered. Measures such as bias auditing can help identify whether recommendations systematically differ across demographic groups, such as gender. Providing transparent explanations for recommendations may also help users understand how suggestions are generated. Future work should incorporate bias monitoring and responsible AI practices to support inclusive and equitable career guidance.

References

[1] F. Pariafsai and A. H. Behzadan. Core competencies for construction project management: Literature review and content analysis. *Journal of Civil Engineering Education*, 147(4):04021010, 2021.

[2] G. Bianchi. Sustainability competences. *Publications Office of the European Union*, 1(1):1-73, 2020.

[3] T. Hooley and S. Rice. Ensuring quality in career guidance: A critical review. *British Journal of Guidance & Counselling*, 47(4):472-486, 2019.

[4] N. Thambar. 'I'm just a plain old careers adviser'–recognising the hidden expert. *Journal of the National Institute for Career Education and Counselling*, 37(1):26-32, 2016.

[5] V. E. Kulugh, K. P. Aondover, and A. S. Faki. Artificial intelligence-powered personalised career guidance system. *Dutse Journal of Pure and Applied Sciences*, 11(1c):92-104, 2025.

[6] Y. Cheng and Y. S. Liang. The development of artificial intelligence in career initiation education and implications for china. *European Journal of Artificial Intelligence and Machine Learning*, 2(4):4-10, 2023.

[7] T. Miller, I. Durlik, A. Łobodzińska, L. Dorobczyński, and R. Jasionowski. Ai in context: Harnessing domain knowledge for smarter machine learning. *Applied Sciences*, 14(24):11612, 2024.

[8] Z. Roozbahani. A review of methods for reducing hallucinations in generative artificial intelligence to enhance knowledge economy. *Knowledge Economy Studies*, 2025.

[9] J. Thomas and A. Harden. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC medical research methodology*, 8(1):1-10, 2008.

[10] R. M. Sawsan. A systematic literature review on construction management productivity enhancement by utilizing business information modeling. *Engineering, Technology & Applied Science Research*, 14(2):13702-13705, 2024.

[11] D. B. Allsop, J. M. Chelladurai, E. R. Kimball, L. D. Marks, and J. J. Hendricks. Qualitative methods with nvivo software: A practical guide for analyzing qualitative data. *Psych*, 4(2):142-159, 2022.

[12] B. Bharadwaj, J. Lucas, V. Sharma, D. Gajjar, and P. Wagle. Identifying influencing factors for gen z students to pursue a career in construction. *Journal of Civil Engineering Education*, 151(4):04025013, 2025.

[13] M. Baker, M. Ali, E. Hassett, and R. Jahan Tumpa. Tapping the untapped resource to address construction skills shortages: Perceptions of australian women career changers and construction women. *Journal of Construction Engineering and Management*, 151(4):04025006, 2025.

[14] H. Lingard, M. Turner, and P. Pirzadeh. Is construction my industry of choice? Examining the factors affecting career choice decision-making of young workers. *Engineering, Construction and Architectural Management*, ahead-of-print, 2025.

[15] R. P. Zhang, S. Holdsworth, M. Turner, and M. M. Andamon. Career aspiration and workplace reality–lived experience of early career professional women in construction. *Women's Studies International Forum*, 105:102926, 2024.

[16] E. A. Holt, A. Perrenoud, E. Perkins, and B. F. Bigelow. Recommendations for recruiting and developing early career membership in construction associations. *International journal of construction education and research*, 19(3):242-258, 2023.

[17] E. Bullock-Yowell and R. C. Reardon. Using the self-directed search in the career construction interview. *International Journal for Educational and Vocational Guidance*, 22(1):157-168, 2022.

[18] J. Borg and C. M. Scott-Young. Contributing factors to turnover intentions of early career project management professionals in construction. *Construction Management and Economics*, 40(10):835-853, 2022.

[19] E. O. Rasheed, J. Yu, S. Hale, and N. Booth. The impact of external and internal sources of motivation on young women's interest in construction-related careers: An exploratory study. *International journal of construction education and research*, 18(2):159-178, 2022.

[20] P. Carnemolla and N. Galea. Why australian female high school students do not choose construction as a career: A qualitative investigation into value beliefs about the construction industry. *Journal of Engineering Education*, 110(4):819-839, 2021.

[21] F. Minooei, P. M. Goodrum, and T. R. B. Taylor. Young talent motivations to pursue craft careers in construction: The theory of planned behavior. *Journal of Construction Engineering and Management*, 146(7):04020082, 2020.

[22] B. L. Oo, B. Lim, and S. Feng. Early career women in construction: Are their career expectations being met? *Construction economics and building*, 20(3):1-19, 2020.

[23] T. Bryce, H. Far, and A. Gardner. Barriers to career advancement for female engineers in australia's civil construction industry and recommended solutions. *Australian Journal of Civil Engineering*, 17(1):1-10, 2019.

[24] J. E. Rosa, C. K. H. Hon, B. Xia, and F. Lamari. Challenges, success factors and strategies for women's career development in the australian construction industry. *Construction economics and building*, 17(3):27-46, 2017.

[25] V. Francis. What influences professional women's career advancement in construction? *Construction Management and Economics*, 35(5):254-275, 2017.

[26] X. Li, M. Gao, Z. Zhang, C. Yue, and H. Hu. Rule-based data selection for large language models. *arXiv preprint arXiv:241004715*, 2024.

[27] I. Krissaane *et al.* Scalability and cost-effectiveness analysis of whole genome-wide association studies on google cloud platform and amazon web services. *Journal of the American Medical Informatics Association*, 27(9):1425-1430, 2020.

[28] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186-204, 2000.

[29] E. Gedrimiene, I. Celik, A. Kaasila, K. Mäkitalo, and H. Muukkonen. Artificial intelligence (ai)-enhanced learning analytics (la) for supporting career decisions: Advantages and challenges from user perspective. *Education and Information Technologies*, 29(1):297-322, 2024.

[30] A. V. Moere and H. Purchase. On the role of design in information visualization. *Information Visualization*, 10(4):356-371, 2011.

[31] W. Shi and D. Wang. Empirical research on the application of ai mock interviews in enhancing graduate perceived employability: A case study in hangzhou, china. *Education and Information Technologies*, 30:18461–18484, 2025.

[32] A. J. Perrenoud, B. F. Bigelow, and E. M. Perkins. Advancing women in construction: Gender differences in attraction and retention factors with managers in the electrical construction industry. *Journal of Management in Engineering*, 36(5):04020043, 2020.

Tackling Site-Based Challenges: The Mitigating Role of Construction 4.0 in Advancing Women's Participation

Aba Essanowa Afful[1], Cynthia Changxin Wang[2], Riza Yosia Sunindijo[3]

[1,2,3,]School of Built Environment, University of New South Wales, Australia
aba.afful@unsw.edu.au, cynthia.wang@unsw.edu.au, r.sunindijo@unsw.edu.au

Abstract

This paper examines how Construction 4.0 technologies tackle the challenges of site-based women in construction. The study also examines the role of Construction 4.0 in offering a competitive advantage to women in these site-based roles. The study adopts a qualitative method using semi-structured interviews with 9 women in site-based roles, purposefully selected from a broader study of 29 women across on-site, off-site, and fluid roles. This is followed by a mixed-gender focus group discussion to validate the findings of the interviews. The findings reveal construction technologies used by site-based women include hardware devices, enterprise systems, and design/CAD enabled design tools. Further, key challenges faced by women in these roles include gendered culture that undermines their credibility and voice, capability challenges and progression and career mobility. To address these challenges, study participants indicated that technologies enhance women's performance and credibility in meetings, reduce physical and safety burdens, and enable flexibility, thereby creating role-specific advantages around efficiency, safety, and confidence. However, both interviewees and focus group participants caution that Construction 4.0 offers a critical, but not self-sufficient, platform for change. Technologies can help dismantle some of the gendered challenges, yet their potential depends on being embedded within intentional strategies to challenge masculine industry norms and redesign job roles. This study forms part of a broader research seeking to empower women across on-site, off-site, and fluid roles through Construction 4.0 technologies.

Keywords –

Construction 4.0; women in construction; on-site roles; digital technologies, female participation.

1 Introduction

The construction industry remains one of the most gender-segregated sectors globally, with women significantly underrepresented in site-based and trade-oriented roles despite incremental gains in overall participation. Recent studies report that women are concentrated in office, administrative, and professional roles, while representation in on-site and physically intensive positions often remains below 5–10%, reinforcing a perception of site work as a male domain [1]. This imbalance persists even as labour shortages and ageing workforces intensify pressures on the industry to diversify its talent base and align with global agendas such as the Sustainable Development Goals (SDGs), particularly SDG 5 on gender equality and the empowerment of all women and girls [2].

A growing body of construction and built environment literature highlights that women in site-based roles face persistent challenges, including exclusionary masculine cultures, harassment, limited access to mentoring, inflexible working patterns, and inadequate amenities, which collectively undermine attraction, retention, and progression [3]. Empirical work on early-career women and tradeswomen shows that women on site must continually prove their legitimacy, navigate hostile or dismissive behaviours, and often experience constrained career pathways compared with their male counterparts, even where formal equality policies exist [2,3]. Despite policy interventions, many initiatives have focused on broad demographic targets rather than the role-specific constraints of women in site-based positions, limiting their effectiveness and sustainability.

In parallel, rapid digitalisation and the emergence of Construction 4.0 offer new possibilities to reshape site-based work and reduce reliance on traditional, physically intensive practices. Technologies such as building information modelling (BIM), virtual design and construction (VDC), digital collaboration platforms, automation and robotics, wearable safety devices, and remote monitoring tools are increasingly linked to improvements in productivity, safety, flexibility, and decision-making across project lifecycles [4]. Recent systematic reviews emphasise that these technologies are

also associated with the emergence of new job roles notably "fluid" roles that integrate on-site and off-site responsibilities, potentially re-orienting focus from physical strength and continuous physical presence [2].

However, there is a lack of empirical evidence on whether and how digital tools mitigate challenges such as safety concerns, physical demands, and work–life conflict, for women on site, or whether these tools may inadvertently reproduce existing power imbalances in the industry [2].

Existing studies tend to either map women's roles and challenges in general, or examine Construction 4.0 adoption without a gender lens, resulting in limited understanding of how technologies are experienced by women on site [2, 3]. Again, there is the need to further the argument that construction 4.0 technologies generate competitive advantages for women in construction, e.g., in safety, productivity, visibility, and mobility across job roles. Moreover, relatively few studies have explored women's perspectives using in-depth qualitative methods that capture the socio-material and institutional lens of technology use.

Against this backdrop, this paper focuses specifically on women in site-based roles in the construction industry and examines how construction 4.0 technologies can mitigate challenges and create competitive advantages for these women. Some studies define competitive advantage as the presence of superior characteristics or capabilities that enable sustained outperformance, often operationalised through innovation, differentiation, and operational efficiency [5,6]. This study extends this concept to job roles of WIC, defining competitive advantage as role-specific capabilities and conditions that allow women in on-site construction roles to perform their work more efficiently, safely, and visibly. This is achieved by strategically leveraging Construction 4.0 technologies within male-dominated construction site environments. Aligning with this definition, reductions in physical and safety burdens, enhancements in performance and credibility, and increased flexibility and access are treated as key domains through which Construction 4.0 can generate competitive advantages for women on site.

The study aims to address the research question: *What competitive advantage can construction 4.0 offer to women in on-site job roles, and how might these capabilities address identified challenges?* To answer this question, the study adopts a qualitative research method comprising of 29 semi-structured interviews with women in site-based roles, followed by 3 focus group discussions to validate the emergent themes.

2 Method

Prior to the conduct of the study, ethical clearance was provided from the University of New South Wales Human Research Ethics Committee (approval no. iRECS6102). All participants provided informed consent, were assured of confidentiality and anonymity, and were advised of their right to withdraw at any time.

This study adopted a qualitative, two-stage design using semi-structured interviews and focus group discussions. Participants were recruited via purposive and snowball sampling methods. This approach was employed to identify female participants capable of providing in-depth insights into on-site roles and Construction 4.0 technologies. The inclusion criteria stipulated that participants must be female, aged 18 years or older, currently employed within Australia, possess a minimum of four years of experience in the construction sector, and be actively utilising one or more construction technologies in their present capacity. Recruitment commenced with the researcher's professional network on LinkedIn and engagement with women-centric industry organisations, such as the National Association of Women in Construction (NAWIC), the Architecture, Construction, Engineering, Development (ACED) Women Network, and Women in Building Information Modelling (BIM). Subsequently, recruitment was extended through participant referrals to ensure the inclusion of a diverse array of organisations and projects.

In Stage 1, individual semi-structured interviews were conducted with female construction professionals occupying a range of job roles. It is important to note that for this paper, we rely on a subset of data from the broader qualitative study. In the larger study, 29 female construction personnel were interviewed across on-site, off-site, and fluid roles. Specifically for site based roles, the sample (Table 1) included nine women and it is their interview data that form the basis of the analysis presented here. Example questions included: *What is your current job role and how long have you been in this position? What construction technologies are you currently using or have used in your job roles? How have these technologies impacted your work and responsibilities? In what ways do you believe women can gain a competitive advantage in the construction industry using technology?*

Table 1 Interview Participants information

Code	Profession	Years in Construction	Organisation Type
P1	Site Engineer	6 years	Builder
P5	Plumber	10 years	Sub-contractor
P6	Site Engineer	4 years	Builder
P8	WHSE Coordinator	4 years	Builder
P9	Carpenter	4 years	Builder
P11	WHSE Advisor	4 years	Builder

P12	Site Engineer	5 years	Builder
P24	Site Officer	7 years	Off-site Modular company
P27	Carpenter	8 years	Trade Sub-contractor

In Stage 2, an online focus group was conducted via Microsoft Teams for one hour, to validate the interview findings. The broader project conducted 3 focus groups with 18 participants. However, for this paper, only the 6 participants in on-site roles are included as their findings directly answer the research question raised. To minimise bias, participants were recruited from across different organisations and did not know each other prior to the session. The moderator presented a concise summary of the key interview findings on challenges, technologies, and perceived competitive advantages for women, then used open-ended questions to elicit agreement, disagreement, and additional examples. Structured turn-taking encouraged independent responses and ensured that all participants contributed equally. As seen in Table 2, the group comprised three male and three female participants with closely aligned site-related roles to allow for comparison. The focus group was run for a period of one hour and data was recorded and transcribed automatically on Microsoft Teams.

Table 2 Focus Group Participants

	Code	**Profession**	**Years in Construction**	**Organisation Type**
Males	MSR1	Electrician	4 years	Trade Sub-contractor
	MSR2	Senior site planner	12 years	Builder
	MSR3	Site Engineer	8 years	Builder
Females	FSR1	Site Engineer	6 years	Builder
	FSR2	Principal site engineer	7 years	Builder
	FSR3	Carpenter	5 years	Trade Sub-contractor

Both the interview and focus group transcripts were analysed using Atlas.ti version 2025, following an inductive thematic analysis approach. This software aided in coding, theme generation and analysis of data as well as the production of Sankey visualisation of the data. Analysis began with open coding of the interview transcripts, where segments of data were assigned descriptive codes reflecting participants' accounts of technology use, challenges, and perceived competitive advantages. Atlas.ti assisted the analytical process by searching the transcripts and suggesting inductive codes, as well as visualising how codes interrelated with each other across the dataset. Building on these code suggestions, the research team reviewed, refined, and grouped related codes into broader categories, which were further developed into overarching themes through iterative discussion. For the focus group data, the same coding framework was applied to assess the degree to which focus group participants confirmed, extended, or challenged the themes emerging from the individual interviews. Where focus group responses reinforced interview themes, these were retained as validated findings. Where new perspectives or contradictions arose, the themes were refined accordingly. This two-stage analytical process ensured that the final themes were grounded in both individual and group-level data. Thematic comparison between the interviews and focus group confirmed the main themes from the interview findings and led to minor refinements which are incorporated into the analysis presented in this paper.

3 Findings & Discussion

3.1 On Site Technologies

Figure 1 illustrates the range of Construction 4.0 technologies identified by the 9 site-based participants.



Figure 1- On site technologies identified

Interviewees highlighted hardware devices (e.g., laptops, iPads, smartphones, site cameras) as the most prominent technologies (7 mentions) followed closely by site tools such as hand tools, pliers, saws, hammers, etc. which are used on the sites (6 mentions). The technologies with the next highest mention were included enterprise systems such as (Microsoft suite, Aconex, etc.) which aid in collaboration and documentation and computer-aided design (CAD) tools (5 mentions).

Tradies emphasised that traditional hand tools and site tools remain fundamental, and that for some tasks there is still an expectation of physically handling materials. The carpenter (P27) described construction as "*very hands on*," saying "*you can feel the materials while you use it*…". For her, basic hand and power tools, lifting gear, and other site equipment are still core technologies, even as more advanced digital tools appear around them.

As seen from Figure 1, on-site professionals also relied heavily on enterprise systems to connect office and field work. A WHSE advisor (P11) described her role as "*software heavy*," explaining that while on site, she can "*bring my office with me to site on the iPad*," accessing project documentation while physically walking the project. Another engineer (P1) noted that several colleagues use OneNote or digital to-do lists, showing that email, corporate platforms, and digital note-taking tools are becoming routine parts of site-based work.

Finally, design and modelling software were consistently described as critical construction technologies for site-based roles. P1 stated that "*the modelling software is very useful*," and went on to say that having drawings and models "*consolidated in one*" environment "*definitely helps with the time factor*" when working under programme pressure. In another example she described a meeting where, instead of trying to explain a complex issue verbally, she "*brought up a model*" on the screen so that everyone could see the problem in 3D, which "*takes away a lot of tension and pressure*" and helps the group "*get straight to the solution.*" These technologies were corroborated by the focus group as well. This finding is consistent with research by Hasan et al., [7] who posits that mobile devices such as tablets and smartphones are now embedded in everyday site practice, which improves coordination, shortens information loops, and supports faster decision-making during site inspections and coordination meetings. Similarly, the continued importance of traditional hand tools and site equipment alongside these digital systems is also recognised in the literature. Construction studies note that, even under Construction 4.0, many trade tasks remain highly manual and rely on conventional tools, with digital technologies often augmenting rather than replacing hands-on work [8].

3.2 Challenges Site-Based Women Face

The interview findings revealed *gendered culture* was the most frequently mentioned theme and comprised *sub-codes such as gender bias, negative industry culture and relationships, and negative mindsets and attitudes*. Participants described having to continually prove their competence or manage sexist comments. As a site engineer stated: *"…when I was giving instructions on site about additional reinforcement or what I wanted to change it was questioned a little bit because I was younger and probably because I was a girl talking to 50-year-old men… sometimes [they] don't take you very seriously (P1)."* Capability challenges were the second major theme, with the skills and experience gap emerging as the largest sub-code. The third largest theme, progression and career mobility, centred on organisational culture as the largest sub-theme, with participants highlighting opaque promotion processes, a preference for "traditional" male career paths, and a lack of visible female role models in site leadership roles.

These findings are consistent with literature on women in construction, which identifies gendered industry culture as a primary challenge to women's participation and advancement in site-based roles. Galea et al. [8] demonstrate how informal "rules-in-use" and everyday practices normalise sexism and exclusion, reinforcing gendered bias and hostile or dismissive attitudes. Furthermore, recent work on gender diversity in construction notes that women are often denied substantive on-site tasks and mentoring, leading to fewer opportunities to build technical competence and site experience, which then reinforces perceptions that they "lack" capability [9].

3.3 Technology Impact on Site-based Roles

The highest impact of technology was seen in the efficiency and productivity of on-site tasks. Interviewees emphasised that digital tools and modelling software streamline core site tasks and decision-making. A site engineer noted that consolidating drawings and models "*definitely helps with the time factor*," allowing information to be "*all on the one thing*" instead of cross-referencing separate 2D drawings. Another interviewee reported that software tools "*helped me with my role and helped me be efficient… planning and doing sketches and getting things done," (P1)* and that learning advanced features *"really helps with our job and getting information that we need and processing information (P9)."* As echoed by P12, "*[Without the software] I'd be very slow... it would be a lot of paperwork".* Tools such as Bluebeam and similar platforms were described as *"really important… to be able to measure things at a glance"* and *"streamline things more efficiently (P11)"* when reviewing and checking information before issuing it to clients. A site-based carpenter (P27) explained that supervisors and engineers "*have iPads and they just show me what the building is going to look like,"* adding that seeing the model on the tablet "*puts all the pieces together*" and helps her plan ahead because she can visualise "*what's coming*" rather than relying only on what is half-built on site. The same participant referred to "*cameras*" and other site tools when discussing safety, indicating that digital hardware for monitoring and visual access is now embedded in everyday site practice, alongside tablets and phones as standard devices for drawings and coordination.

Interviewees also highlighted risks associated with over-reliance on construction technologies. One participant cautioned that *"a small thing that's done wrong or entered incorrectly… makes a big difference,"* noting *"a lot of room for human error… the more you digitalise things (P8)".* Others observed that younger

staff can *"rely too much on the technology,"*. P1 vividly captured the dependence on digital tools by stating that *"if I left my laptop/iPad at home, I would die,"* illustrating that over-reliance on these devices can create vulnerabilities when systems fail or are unavailable.

Results from the focus group discussion indicated a general agreement that digital tools, especially 3D modelling and design technologies, support increased participation by opening more non-manual roles where traditional physical challenges are less salient. A male participant (MSR3) noted "*a higher proportion of women on the design side*" and suggested that 3D modelling *"is definitely increasing the female participation"* because those jobs are not *"a manual sort of job."*

Several participants also stressed equal opportunity, arguing that technology "*enhances everyone*" and creates *"a pretty even playing field,"* rather than uniquely privileging women. Relatedly, some were wary that framing technology as something *"to make it easier for women*" could backfire and *"actually do the opposite,"* risking perceptions of special treatment and gender imbalance (MSR2, MSR3). Focus group members also highlighted loss of foundational skills as a risk, echoing interview concerns about overreliance on tools: AI and digital aids were seen as helpful but also could make workers *"a little bit lazy because we're not learning the… foundations of how to actually do something," (FSR3)* especially in trades.

They reinforced that technology cannot by itself fix gendered culture, with MSR2 stating they *"just don't see how"* technology can reduce that challenge, while another argued technology is more likely to shift work structure, well-being, and capability (via online upskilling tools) than culture directly.

3.4 Technologies Addressing the Identified Challenges

Across the interviews and focus group validation, participants identified several advantages that construction technologies offer. The following subsections present the findings organised around three competitive advantage themes emerging from the data: (1) overcoming physical challenges, (2) performance and credibility, and (3) flexibility and access. As summarised in Table 3, each theme directly maps to a key challenge identified by the site-based women in this study, and demonstrate how specific Construction 4.0 technologies create competitive advantages for women in site-based roles.

Table 3 Analytical links between site-based challenges, construction 4.0 technologies, and competitive advantages for women

Challenge	Construction 4.0 technologies	Competitive Advantage (Theme)
Physical demands & high-risk tasks on site	Robotics, drones, assisted lifting aids, prefabrication	**Physical burden reduction;** participation without conforming to hyper-masculine physical norms
Being talked over, competence questioned, credibility undermined	BIM/3D modelling, design software, digital collaboration tools	**Performance & Credibility;** visible, undeniable technical authority in meetings and on site
Always-on site presence, rigid work structures	Mobile devices, iPads, cloud platforms, enterprise systems (MS Teams, Aconex)	**Flexibility & Access;** marginal softening of always-on-site expectations, enabling hybrid work

3.4.1 Overcoming physical challenges

Interviewees and focus group participants highlighted how technologies such as robotics, drones and assisted lifting aids can reduce the need for physically demanding or high-risk tasks on site. For example, P24 described how using robots or drones for external quality checks on high-rise façades means women do not need to *"physically go up and down …*" which was seen as especially helpful for those "*scared of heights*." Focus group comments on assisted aids for handling ceiling sheets similarly emphasised that such equipment allows tasks previously requiring two strong workers to be done more safely thereby reducing the physical strength requirement.

P27 emphasised that construction remains *"very hands on,"* involving heavy lifting, yet acknowledged the value of technologies such as plant and lifting equipment that because they ease some of the physical load. Within the focus group, participants pointed to prefabrication, remote inspection tools, and smart machinery as measures that reduce climbing, manual handling, and exposure to high-risk environments, allowing women to perform core site tasks without conforming to an idea of hyper-masculinity. This is consistent with Construction 4.0 literature which argues that automation, robotics, and digital tools progressively separate productivity from brute strength and shift value toward supervision, and coordination [4,10]. Evidence from other sectors indicates that such shifts can open up new technical roles and reduce gender gaps where physical demands had

previously been used explicitly or implicitly to exclude women [4,10].

3.4.2 Performance and credibility

On-site engineers described digital tools as amplifying their technical performance and making their expertise more visible in high-pressure site environments.

P12 noted that Revit "*helps with understanding consultants' drawings and how they will work together in the physical space, as you can overlay the different layers and easily see where clashes occur",* enabling her to coordinate complex design information under programme pressure. Similarly, P6 explained that having a visual tool, such as a "*Bluebeam markup*", helped her "*communicate more clearly with subcontractors*." She noted that "*you are not only relying on your memory… you have an image… to assist you in communicating what you want"* (P6). This reduced the risk of being misunderstood or dismissed in a male-dominated site environment. Technologies that make expertise visible directly counter these assumptions. These accounts mitigate common challenges such as being talked over, having competence questioned, and struggling to have concerns taken seriously. These challenges are well documented in the women-in-construction literature [10]. Studies of BIM-enabled coordination similarly show that 3D and 4D models support shared understanding, reduce disputes, and shift authority toward those who can control and interpret digital information, thereby strengthening positions of women in project decision-making [11,12].

3.4.3 Flexibility and Access

Site-based women also framed technologies as offering an advantage in navigating rigid, site-centred work. A WHSE advisor (P11) described her role as "quite flexible", explaining that although she "*primarily need[s] to stay on site*," she can "*bring my office with me to site on the iPad*" by using the internal safety system in the field and occasionally working from home "*maybe one day a week*" when necessary (P11). An engineer noted that digital tools "*allow that flexibility… to still be engaged with your team through technology, but work from home,*" even though most responsibilities remain site-based (P6). These examples show how mobile devices, cloud-based documentation, and enterprise platforms marginally soften "*always on site*" expectations. This is corroborated by a plethora of literature that posits digitalisation can support hybrid work arrangements and more sustainable careers for women in construction. However, focus group participants were clear that "*technology alone does not change the nature of the work or the industry" (MSR3)* they emphasise that flexibility is highly dependent on the organisation's willingness to accommodate flexible work patterns.

Other advantages identified by interviewees include skill development, and communication and collaboration which were not validated by the focus group.

3.5 Gender Neutrality of Technologies in Construction

Both interviews and focus groups, however, revealed discomfort with presenting technology as something women uniquely "*need*." FSR1 from the focus group argued that while tools like BIM and project software are crucial, "*it's just a tool*" they "*don't help women any more than… a male"*. She also argues that competitive advantage comes from combining underlying skill with digital knowledge. A focus group participant (MSR1) similarly cautioned that suggesting technology is introduced "*to make it easier for women to do a job in construction"* could be "*misconstrued*" and "*actually do the opposite*" by implying women are less capable. These concerns align with Galea's critique that "*fixing women*" via special initiatives can inadvertently reinforce male advantage and leave underlying structures unchallenged [8].

The findings suggest that construction technologies are formally gender-neutral, but they operate within gendered institutions where men already benefit from norms around physicality, long hours, informal networks, and unexamined authority. In that context, women's strategic use of technologies creates specific, competitive advantages that help them navigate and partially resist those norms.

4 CONCLUSION

This study has revealed that Construction 4.0 technologies offer women in on-site roles concrete, advantages in navigating a workplace historically structured around male norms of physical strength, constant site presence, and informal authority. Across interviews and focus groups, women and men in the industry described how digital tools improved their performance, reduced physical limitations and opened up limited flexibility, in male-dominated site environments. These benefits did not arise from technology alone, but from how women strategically mobilised technologies to overcome challenges such as being talked over in meetings, having their competence questioned, and being constrained by constant site presence.

Literature has argued that digitalisation is reshaping job roles by shifting value toward information handling, coordination, and problem-solving, thereby creating conditions under which women can participate and progress on more equal terms. Wider evidence from research in Australia similarly suggests that digital tools,

automation, and remote-capable platforms can reduce physical demands, enable new hybrid site–office roles, and support more flexible work models that are critical for women's retention [2,10]. At the same time, both this study and the literature caution against assuming that "gender-neutral" technologies will, by themselves, close gender gaps. Without deliberate attention to the deeply embedded rules-in-use that normalise male advantage, there is a risk that digital competencies and high-status technical roles will simply be captured by those already privileged in the system [8].

As part of broader research seeking to empower women across on-site, off-site, and fluid roles through Construction 4.0 technologies, this paper contributes role-specific, empirical insight into how women on-site women actually use technologies to carve out micro-advantages in performance, safety, flexibility, and credibility.

The findings of this study suggests that Construction 4.0 offers a critical, but not self-sufficient, platform for change. Technologies can help dismantle some of the gendered challenges, yet their potential depends on being embedded within intentional strategies to challenge masculine industry norms and redesign job roles. The findings point to a future where construction technologies are not merely tools for efficiency, but central components of an integrated framework to make on-site construction work a viable, attractive, and sustainable career path for women.

Future research should examine the barriers to Construction 4.0 adoption at the organisational and industry level, including cost, digital literacy, and resistance to change. Addressing these barriers is critical to ensuring that the competitive advantages identified in this study are accessible to all women in site-based roles, not only those in well-resourced organisations.

4.1 Practical Implications

Construction 4.0 has often been discussed in terms of broad industry change, but the women in this study mainly experienced it through everyday tools such as laptops, tablets, enterprise systems, and CAD/BIM software. These tools helped them do their work more efficiently, reduce some physical and safety risks, giving them clear advantages in on-site roles. However, the study also shows that technology on its own is not enough to improve gender equality. Simply put, the same tools are introduced into workplaces where long hours, strong site presence, and male-dominated decision-making are still the norm.

This has several practical implications. First, companies should not only provide access to digital tools but also offer targeted training so that women in site roles can move into higher-responsibilities or roles. Second, when organisations roll out new digital systems, they should consider how these changes affect women's workloads, flexibility, and progression, rather than assuming that "technology helps everyone in the same way." Third, clients and policymakers can link digitalisation and gender goals by encouraging or requiring projects to create visible Construction 4.0 pathways for women in on-site roles, for example through social procurement or gender targets.

As part of a broader study on empowering women in on-site, off-site, and fluid roles, these findings suggest that Construction 4.0 is a useful platform for change, but only when combined with deliberate efforts to redesign roles, support skills development, and challenge existing site cultures.

5 ACKNOWLEDGEMENTS

The authors wish to acknowledge the Department of the Prime Minister and Cabinet's Office for Women, under the Australian Government's Women's Leadership and Development Program and UNSW School of Built Environment whose grant has supported this study.

References

[1] R. Edirisinghe, and R. Bimali. Driving Gender Diversity in Construction: Mapping Existing Policies Through the Behavioral Change Wheel Framework. *In CIB Conferences*, 1(1):14, 2025.

[2] A.E. Afful, C.C. Wang, R.Y. Sunindijo, S. Frimpong, E. Boadu, B. Baah. Evolving Job Roles of Women in the Construction Industry. *Journal of Construction Engineering and Management*, *151*(3), p.0312500, 2025

[3] S. Holdsworth, M. Turner, and O. Sandri. Gender bias in the Australian construction industry: Women's experience in trades and semi-skilled roles. *Social Sciences*, *12*(11), p.627. 2023

[4] A. Hajirasouli, A. Assadimoghadam, M.A. Bashir, and S. Banihashemi. Exploring the Impact of Construction 4.0 on Industrial Relations: A Comprehensive Thematic Synthesis of Workforce Transformation in the Digital Era of Construction. *Buildings*, *15*(9), p.1428. 2025

[5] N.W. Njoroge, and G.M. Kinyua.. Strategic orientation as an antecedent of competitive advantage: Evidence from review of literature. *International Journal of Education and Research*, *13*(3), pp.35-68, 2025.

[6] L.W. Keya, and G.M Kinyua. Organizational learning as an antecedent of competitive advantage: evidence from review of literature. *International Journal of Education and Research 13(2)*, pp.13-38, 2025

[7] A. Hasan, A. Elmualim, R. Rameezdeen, B. Baroudi, and A. Marshall. An exploratory study on the impact of mobile ICT on productivity in construction projects. *Built Environment Project and Asset Management*, *8*(3), pp.320-332, 2018

[8] N. Galea, A. Powell, M. Loosemore, and L. Chappell. Demolishing Gender Structures. *UNSW: Sydney*, 2018

[9] S. Holdsworth, M. Turner, C.M. Scott-Young, and K. Sandri. Women in Construction: Exploring the Barriers and Supportive Enablers of Wellbeing in the Workplace. *RMIT University, Melbourne,* 2020

[10] A. Afful, R.Y. Sunindijo, P.A. Tetteh, S. Frimpong. Breaking the concrete wall: The untapped potential of technology to increase the participation of women in the construction industry. In *2nd International Conference on Construction Project Management and Construction Engineering (iCCPMCE-2024)* (pp. 148-153), 2025

[11] A. Waqar, K.A Alrasheed, and H.R. Almujibah.. A BIM-Infused Perspective on Enhanced Visualization and Design Analysis for the Residential Construction. *KSCE Journal of Civil Engineering*, p.100412, 2025

[12] W.U. Yunxia, and L.I. Lei. The impact of digitization in manufacturing on female employment and gender wage gap. *Journal of Asian Economics*, *95*, p.101821. 2024

Strategies to Overcome BIM Barriers in the AEC Industry of Iran

Tahere Asghari* and V. Paul C. Charlesraj

Department of Design and Architecture, School of Arts and Humanities, University of Huddersfield, Huddersfield, United Kingdom.
asghari.ta@gmail.com , p.c.vasantharaj@hud.ac.uk

Abstract

Building Information Modelling (BIM) has emerged as a transformative digital technology that can improve productivity, reduce errors, and enhance collaboration across construction projects worldwide. Despite its well-recognised benefits, the Architecture, Engineering, and Construction (AEC) industry in Iran continues to encounter substantial challenges in BIM adoption. Research exploring effective solutions tailored to the Iranian context remains limited, creating a gap that this study aims to address. The purpose of this research is to develop appropriate strategies to overcome barriers to BIM implementation within Iran's AEC sector. A comprehensive literature review was conducted to examine barriers and existing strategies at global, regional, and local levels. In addition, semi-structured interviews with Iranian industry experts were carried out to capture practical insights specific to Iran's construction industry. The results indicate that although BIM awareness is gradually increasing, its implementation is hindered by inadequate government support, limited professional knowledge and training, and a significant disconnect between academia and industry practice. Based on these findings, this research emphasizes the need for government-led initiatives, including creating a robust legal and technical framework to mandate BIM, implementing incentive-based measures, and integrating BIM-focused curricula into universities. This study contributes original insights and practical recommendations to accelerate BIM adoption in Iran, supporting broader digital transformation in the construction sector.

Keywords

Building Information Modelling (BIM); AEC industry; Iran; BIM Barriers; Adoption Strategies

1 Introduction

The construction industry has historically struggled with low productivity, inefficient planning, and limited collaboration, with traditional methods in the Architecture, Engineering, and Construction (AEC) sector identified as primary contributors to these challenges [1]. Building Information Modelling (BIM) has emerged over the past decade as a transformative digital technology designed to enhance productivity and improve construction performance [2].

BIM represents a fundamental shift from traditional 2D drawings to comprehensive digital representations of buildings' physical and functional characteristics [3]. As a digital model that organizes and visualizes building information throughout a project's lifecycle [4], BIM encompasses data about building components, structures, materials, and lifecycle processes. This technology enables digital management of design information, facilitating more integrated and data-driven project implementation approaches. Research demonstrates that BIM adoption enhances construction productivity, reduces errors and rework, and improves overall project delivery through better coordination, information sharing, and collaborative decision-making [5]. BIM's digital information management capabilities have significantly transformed traditional construction practices [6].

Globally, many countries have embraced BIM and developed adoption strategies for the AEC sector [7]. The United Kingdom experienced dramatic growth in BIM adoption, rising from 13% in 2011 to 74% in 2018, with similar trends observed across developed regions, including the United States, Europe, and Australia [8]. However, BIM adoption remains low in developing countries, including North Africa and the Middle East, where, despite high awareness in Gulf Cooperation Council (GCC) countries, adoption rates remain insufficient [9].

Several barriers hinder widespread BIM acceptance [10], including a lack of motivation, resistance to change, insufficient knowledge, inadequate technology tools, interoperability issues, and undefined processes and standards [11]. In Iran specifically, the AEC industry faces significant adoption challenges, with only 29.5% of construction companies utilizing BIM tools, while 56.8%

have never encountered BIM and 36.4% lack adoption plans [12]. This research aims to explore obstacles hindering BIM usage in Iran's AEC sector and develop strategies based on expert opinions and literature review to address these barriers.

2 Literature Review

2.1 Barriers to BIM Implementation Worldwide

The global adoption of Building Information Modelling (BIM) remains constrained by a complex set of interrelated barriers despite its well-established potential to enhance project integration, productivity, and decision-making [10]. Existing literature emphasizes that BIM implementation is not merely a technological shift but a systemic transformation requiring alignment across organizational structures, workflows, and stakeholder practices [13]. The literature identifies three dominant categories of challenges. First, technical barriers such as system complexity, interoperability issues, and inadequate data standards [17]. Second, organisational and cultural barriers including resistance to change, and shortages of skilled professionals significantly reduce industry readiness [18]. Third, Legal and institutional barriers further complicate BIM adoption. The absence of regulatory frameworks, and unresolved contractual issues related to liability and intellectual property create uncertainty [19]. However, the relative importance of these barriers varies across regions, particularly between developed and developing contexts [20].

2.2 Barriers to BIM Implementation in the Middle East

In the Middle East, BIM implementation remains at an early stage, where adoption is still limited despite increasing awareness [18]. Unlike developed countries, where technical and legal issues such as data security and interoperability are more prominent, the barriers in the Middle East are primarily driven by organizational and human factors, particularly limited awareness, lack of expertise, and insufficient training [20].

In addition, from a technical perspective, challenges such as interoperability issues, inadequate IT infrastructure, and limited access to advanced BIM tools hinder effective implementation [21]. Financial constraints also play a major role, as high initial investment costs and the absence of clear economic incentives discourage organizations from adopting BIM [18]. Furthermore, weak regulatory frameworks, lack of BIM standards, and limited government support create uncertainty and slow down implementation [22]. Overall, BIM barriers in the Middle East are strongly context-dependent, reflecting broader challenges related to human capital, institutional capacity, and economic conditions, which differ significantly from those observed in developed countries [18]. This highlights the importance of examining BIM adoption barriers within specific national contexts such as Iran.

2.3 Barriers to BIM Implementation in Iran

BIM adoption in Iran remains limited despite its potential to enhance productivity and performance in the construction sector, which plays a significant role in the national economy [12], [9]. Existing studies indicate that BIM implementation is still at an early stage, with low awareness and limited practical experience among industry stakeholders [23]. The literature identifies several key barriers, primarily lack of government support, insufficient regulatory frameworks, and weak organizational readiness. In addition, limited BIM knowledge, shortage of skilled professionals, and resistance to change significantly hinder adoption [24]. Financial constraints and unclear return on investment further reduce motivation for implementation [25]. Overall, BIM barriers in Iran reflect a combination of institutional, organizational, and knowledge-related challenges that require targeted, context-specific strategies.

2.4 Overview of Existing Strategies

The implementation of Building Information Modelling (BIM) requires a structured, strategic approach [27]. Research has identified that effective BIM implementation strategies must include changing organizational approaches, providing appropriate incentives, establishing industry partnerships, integrating BIM into existing processes, and promoting collaboration across the AEC sector [28]. Specific strategies that support successful BIM implementation include investing in BIM-related training programs, supporting adoption through adequate financial and technical resources, ensuring appropriate project budgets for BIM integration, developing technical capabilities required for BIM tools and workflows, cultivating organizational commitment to BIM, promoting collaborative work environments, and establishing robust BIM infrastructure and standards [29].

Governments have been identified as the main beneficiaries and key players in BIM implementation, becoming increasingly active in promoting adoption through various policies and initiatives that can be broadly classified into three approaches: government-driven, industry-driven, and mixed approaches combining both governmental and industry initiatives [30]. Government-led strategies have proven particularly effective in accelerating BIM adoption across multiple

countries, with successful implementation measures including establishment of national BIM programs, mandatory BIM requirements, development of BIM standards and protocols, integration of BIM in procurement and tendering processes, awareness and adoption initiatives, development of BIM infrastructure, and the creation of industry-wide BIM classification standards [10].

Research suggests that developing countries need comprehensive BIM implementation models based on theoretical frameworks derived from implementation process theory, incorporating four key strategic elements: top management commitment strategies emphasizing organizational restructuring and cultural change; motivational strategies including regulatory mandates and BIM guidelines development; capacity development strategies focusing on building skills, knowledge, and competencies through education and training programs; and collaborative network relationships creating essential environments for change [18]. BIM implementation in Iran requires a comprehensive approach due to multiple barriers across resource, financial, organisational, regulatory, and market-related dimensions [9]. Efforts toward digitalisation have been initiated by government-related institutions through training programmes, conferences, and the establishment of BIM-focused organisations [26]. These actions reflect increasing recognition of BIM importance within the industry. However, the lack of locally adapted regulations and reliance on international frameworks limit effective implementation [31]. Overall, the need for context-specific strategies tailored to Iran's conditions is highlighted, as limited research has addressed practical solutions to overcome BIM adoption barriers in the Iranian construction sector.

3 Research Methodology

This study adopts a qualitative research design using a two-phase approach to investigate BIM adoption barriers and develop context-specific strategies for the Iranian AEC industry. In the first phase, a structured literature review was conducted to identify key barriers and strategies across global, regional, and Iranian contexts. Relevant studies were collected from academic databases such as Scopus, Web of Science, and Google Scholar using keywords including "BIM adoption barriers" and "BIM implementation strategies." The selected studies were analyzed and grouped into key themes such as technical, organizational, financial, and regulatory factors. The findings from this phase were used to inform the design of the interview questions. Phase Two employed semi-structured interviews with Iranian AEC experts from design firms, contracting organizations, and government bodies, selected purposively for their direct BIM experience (Table 1). This format was chosen for its flexibility to probe unanticipated issues and its capacity to elicit in-depth contextual insights unavailable in the existing literature [32]. While the study is based on three expert interviews, a limitation that restricts generalizability, the focus was on conceptual depth rather than statistical representativeness, with thematic convergence across participants used to assess data sufficiency. Larger-scale validation studies are recommended for future research.

The interview guide, developed from previous studies Marzouk et al. [30], Alreshidi et al. [33], focused on six main themes: existing BIM regulations, government goals and measures, key implementation barriers, facilitators or champions of BIM adoption, recommended strategies, and potential incentives. Initially, eight member companies of the Iranian BIM Association were contacted via email; after no response, participants were approached through WhatsApp and iMessage. Three experts, two managers, and one consultant agreed to participate, representing an acceptable qualitative sample size [34]. Each interview, lasting approximately 40 minutes, was conducted via Google Meet and recorded for analysis.

Table 1. List of interviewees

No	Stakeholder ID	Position	Work Experience	Organisation Type	Area of Expertise
1	Respondent 1	BIM consultant	10 years	Semi-public (Iran BIM Association)	BIM policy development
2	Respondent 2	Director	20 years	Private	Construction management
3	Respondent 3	Manager	12 years	Semi-public (Iran BIM Association)	Training programmes and BIM standard development

3.1 Data Analysis

Interview data were analyses using NVivo software through systematic content analysis. Each participant's responses were coded into themes and subthemes aligned with the study objectives. Following Gray's [30] interpretive approach, the analysis focused on understanding underlying meanings in relation to existing theories rather than simple description. The coding framework identified two main areas: barriers to BIM implementation and recommended strategies for promoting BIM adoption in Iran. Key coded evidence (Table 2) highlighted critical issues such as a lack of government support, the gap between university and industry, and practical strategies, including government incentive schemes and infrastructure development.

Table 2. Summary of Coding in NVivo

Theme	Key Subtheme / Node	Respondent ID	Example of Coded Extract
BIM Implementation Barriers	Lack of government support	Respondent 2	"The main obstacle is the lack of support from government organizations for the implementation of BIM."
BIM Implementation Barriers	The gap between university and industry	Respondent 1	"The lack of university-led applied research to address effective BIM implementation."
Recommended Strategies	Government incentive strategies	Respondent 1	"Reduction of municipal fees in issuing construction approval and permitting."
Recommended Strategies	Infrastructure development	Respondent 1	"Designing and creating urban planning electronic systems with the ability to utilize BIM software infrastructure."

Note: Extracts derived from NVivo qualitative content analysis of expert interviews

4 Results and Discussion

Through NVivo analysis, ten distinct barriers to BIM adoption were identified: (1) lack of government support, (2) lack of awareness among consultants and contractors, (3) limited software infrastructure, (4) the gap between universities and industry, (5) lack of standard frameworks, (6) high implementation costs, (7) resistance from government design and contracting organizations, (8) limited training opportunities, (9) lack of coordination among government bodies, and (10) insufficient communication between academia and industry. Among these, the most critical were the lack of government support, the lack of stakeholder knowledge, and the persistent university–industry gap as shown in Figure 1.

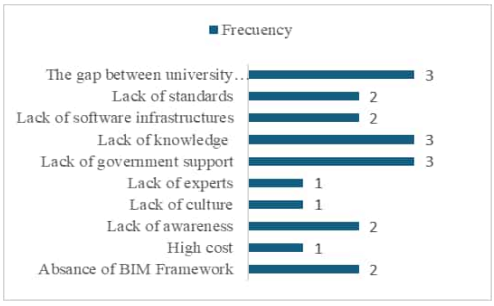


Figure 1. Barriers to BIM adoption

The findings from the semi-structured interviews strongly align with this regional pattern. First, inadequate government support and weak regulatory frameworks emerged as the most critical constraints in the interviews, aligning with previous studies that emphasize the absence of clear BIM standards, mandates, and contractual arrangements in Iran [12], [23]. Second, limited stakeholder knowledge and capacity were widely reported by interviewees, which mirrors global evidence that lack of skills, insufficient training, and resistance to change undermine BIM readiness [11], [14], [17], [27]. Third, the persistent gap between academia and industry, highlighted by respondents and coded as a major barrier in Table 2, echoes findings from both Iranian and international research [9], [18], [12], [23]. Universities in Iran were perceived as insufficiently engaged in applied BIM research and in preparing graduates with practical, implementation-ready skills.

Alongside these core issues, interviewees mentioned high initial costs, fragmented responsibilities among government bodies, limited software infrastructure, and weak inter-organizational coordination. These factors also appear in international and regional studies [10], [14], [18], suggesting that Iran faces many of the same technical and financial challenges documented elsewhere, but under conditions of greater institutional fragility. While the interviews broadly corroborate the literature, they also extend it by providing context-specific nuances

for Iran. International and regional studies frequently stress the role of government as a facilitator of BIM through standards, guidelines, and public-sector procurement [30],[35]. However, the Iranian experts went further, identifying the central government, notably the Office of the Vice President for Technology, as the essential initiator of BIM, with a unique capacity to issue binding executive orders. This emphasis on highly centralized, top-down authority reflects the governance structure of Iran and suggests that models of industry-led diffusion, more common in some Western contexts, are less applicable.

The interviews also revealed a mismatch between policy signals and practical enforcement. For instance, participants referred to Tehran Municipality's 10% discount on construction permits for BIM-based projects and to resolutions that reference BIM in national technical systems. These measures indicate emerging recognition of BIM's importance, yet respondents agreed that they have not translated into systematic adoption, partly due to the absence of a comprehensive national BIM roadmap and partly due to uneven technological maturity across regulatory bodies. This provides an important nuance to the literature on public-sector roles [35]: partial or isolated initiatives, without a robust implementation framework, may have limited impact. Furthermore, whereas global studies often emphasize interoperability issues, and advanced standards [19], [21], the interviews suggest that, in Iran, more basic prerequisites, reliable software access, shared modelling protocols, and digital permitting systems, are still not fully in place. The recommendation to develop "urban planning electronic systems with the ability to utilize BIM software infrastructure" (Table 2) highlights that foundational e-government infrastructure is itself a precondition for BIM, a point less prominent in studies of more digitally mature environments.

Several strategic frameworks for BIM implementation can be found in literature, particularly in developing countries. Olugboyega [27] and Ma et al. [10] emphasize four interrelated elements: (1) top management commitment and organisational change; (2) motivational and regulatory strategies, including mandates and guidelines; (3) capacity development through education and training; and (4) collaborative networks and industry partnerships. The interview findings align closely with this structure and provide concrete, context-specific operationalization terms of top-down leadership and regulation. NVivo node summary for recommended strategies can be found in Figure 2.

All interviewees stressed the need for high-level government directives mandating BIM for selected project types and progressively expanding its scope. This mirrors successful international experiences where mandatory BIM requirements, national BIM programs, and integration into public procurement have accelerated adoption [30], [35], [36]. In the Iranian context, such mandates were seen as necessary to move from pilot initiatives to mainstream practice, while also forcing government agencies themselves to upgrade their digital capabilities.

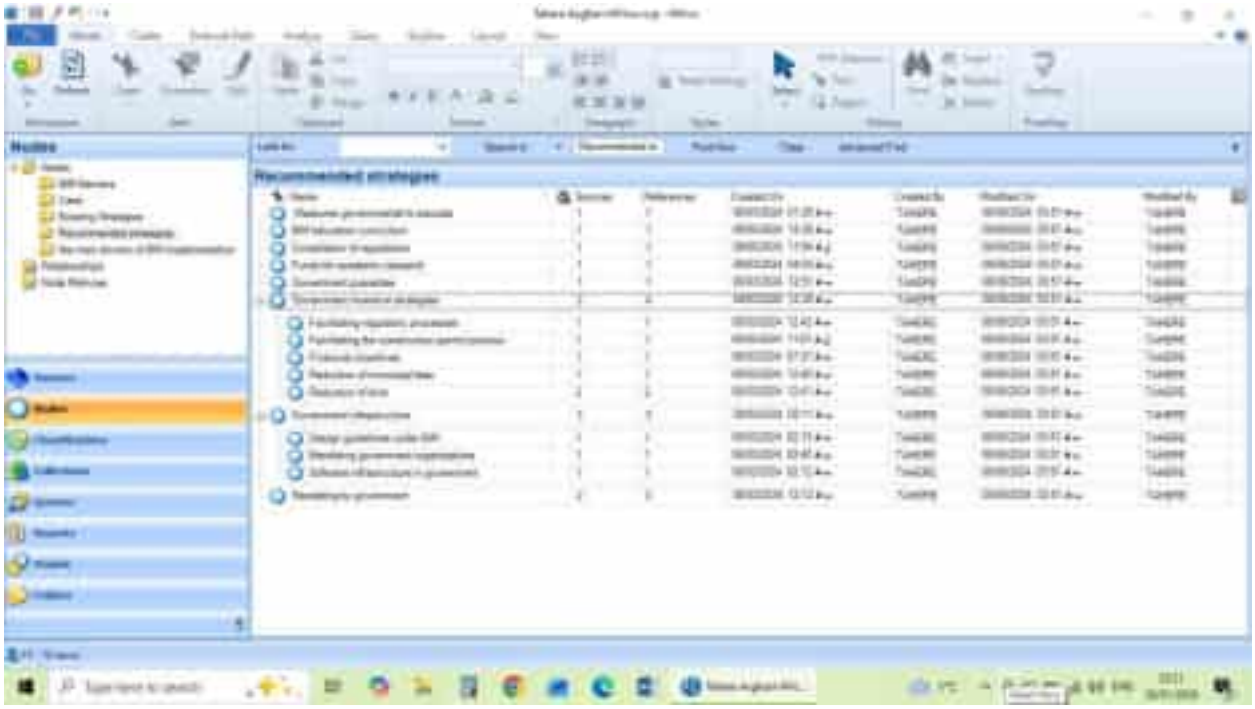


Figure 2. NVivo node summary for recommended strategies

Motivational and incentive-based strategies were the second major cluster of recommendations. Participants proposed streamlined approval processes, reductions in municipal fees, and financial incentives such as discounts on construction and parking permits. These suggestions are consistent with literature that recognizes incentives as effective complements to mandates in mixed policy approaches [35], [36], [38]. The interviews therefore support the view that a combination of regulatory compulsion and economic benefits is likely to be more effective than either approach alone, particularly for resource-constrained firms. Capacity development and education formed the third strategic pillar. Experts called for integrating BIM into university curricula, funding applied research, and expanding professional training opportunities. This directly addresses the university–industry gap identified in both phases of the study and resonates with international evidence that education and continuous professional development are critical success factors for BIM [9], [28], [39]. By explicitly linking curriculum reform and research funding to national BIM objectives, the interviewees extend existing recommendations and emphasize the need for coordinated, state-supported capacity-building.

Taken together, the literature review and interview results indicate that accelerating BIM adoption in Iran requires a dual strategy: strong, centralized government leadership to create a coherent regulatory and technical framework, and sustained investment in human and organizational capacity across the AEC sector. The interviews confirm many barriers previously identified in the literature but also reveal that, in Iran, the absence of an enforceable national BIM framework magnifies other challenges, such as cost, skills, and resistance to change. The proposed strategy therefore comprises three interdependent components: (1) government-led actions, including national BIM standards, clear mandates, and digital infrastructure for BIM-based permitting and procurement; (2) incentive-based policies that reduce financial and administrative burdens on early adopters; and (3) educational and collaborative initiatives to bridge the university–industry gap and develop a skilled BIM workforce.

In addition, continuous professional development and training initiatives for current industry professionals are essential to ensure effective BIM implementation [40]. This emphasis on education and training aligns with global best practices and reinforces the importance of sustained government support for successful BIM adoption. Figure 3 illustrates a visual summary of the recommended strategies for BIM adoption in Iran.

5 Summary and Conclusions

This study examined the barriers to implementing Building Information Modelling (BIM) in Iran's Architecture, Engineering, and Construction (AEC) industry using a comprehensive literature review and semi-structured interviews with experts. Three main barriers were identified: inadequate government support, limited stakeholder knowledge, and a persistent gap between academia and industry. The findings highlight the need for a dual approach combining strong government leadership with targeted capacity-building efforts. To promote BIM adoption, the study proposes three strategic areas: (1) Government-led actions, including legal frameworks and national BIM standards; (2) Incentive-based policies, such as financial benefits and simplified approval processes; and (3) Educational initiatives, emphasizing BIM-focused curricula and collaboration between universities and industry.

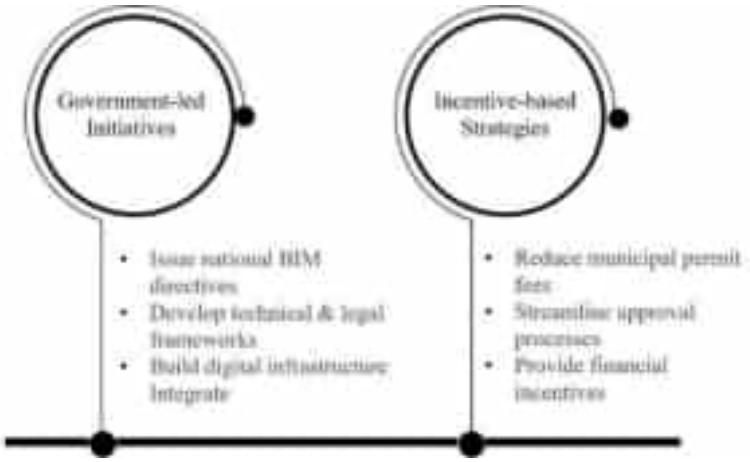


Figure 3. Summary of recommended strategies

Although the study's scope was limited to Iran and based on a small sample, it provides valuable insights into local challenges and offers a foundation for broader, future research incorporating larger datasets and quantitative analysis. However, one of the limitations of this study is its focus on the Iranian context, which may not fully represent BIM implementation challenges in other regions. The small sample size also limits the generalizability of findings, as the study mainly captured perspectives of industry experts without including clients, policymakers, or academics. Additionally, the research relied primarily on qualitative data, suggesting the need for future studies to incorporate quantitative methods. Expanding the scope to include a larger and more diverse group of stakeholders across different regions of Iran would provide a broader and more comprehensive understanding of BIM adoption and its practical implications.

References

[1] R. Fulford and C. Standing, "Construction industry productivity and the potential for collaborative practice," *International journal of project management*, vol. 32, no. 2, pp. 315-326, 2014.

[2] C. M. Eastman, *BIM handbook: a guide to building information modeling for owners, managers, designers, engineers and contractors*, 2e. Hoboken, N.J: Wiley, 2011.

[3] B. Becerik-Gerber and S. Rice, "The value of building information modeling: can we measure the ROI of BIM," *AECbytes Viewpoint*, vol. 47, 2009.

[4] S. Van Nederveen, R. Beheshti, and W. Gielingh, "Modelling concepts for BIM," in *Handbook of research on building information modeling and construction informatics: Concepts and technologies*: IGI Global, 2010, pp. 1-18.

[5] L. Doumbouya, G. Gao, and C. Guan, "Adoption of the Building Information Modeling (BIM) for construction project effectiveness: The review of BIM benefits," *American Journal of Civil Engineering and Architecture*, vol. 4, no. 3, pp. 74-79, 2016.

[6] M. Scheffer, H. Mattern, and M. König, "BIM project management," *Building Information Modeling: Technology Foundations and Industry Practice*, pp. 235-249, 2018.

[7] A. Wong, F. K. Wong, and A. Nadeem, "Comparative roles of major stakeholders for the implementation of BIM in various countries," *Proceeding of Changing Roles 2009*, pp. 23-33, 2009.

[8] J. Wang and W. Lu, "A deployment framework for BIM localization," *Engineering, Construction and Architectural Management*, vol. 29, no. 1, pp. 407-430, 2022.

[9] N. Hatami and A. Rashidi, "Enhancing the adoption of building information modeling in the Iranian AEC sector: insights from a Delphi study," *Engineering, Construction and Architectural Management*, 2023.

[10] L. Ma, R. Lovreglio, W. Yi, T. W. Yiu, and M. Shan, "Barriers and strategies for building information modelling implementation: a comparative study between New Zealand and China," *International Journal of Construction Management*, vol. 23, no. 12, pp. 2067-2076, 2023.

[11] S. Siebelink, H. Voordijk, M. Endedijk, and A. Adriaanse, "Understanding barriers to BIM implementation: Their impact across organizational levels in relation to BIM maturity," *Frontiers of Engineering Management*, vol. 8, pp. 236-257, 2021.

[12] M. R. Hosseini, E. Azari, L. Tivendale, S. Banihashemi, and N. Chileshe, "Building Information Modeling (BIM) in Iran: an exploratory study," 2016.

[13] Q. Phung and N. Tong, "Developing an organizational readiness framework for BIM implementation in large design companies," *International Journal of Sustainable Construction Engineering and Technology*, vol. 12, no. 3, pp. 57-67, 2021.

[14] D. W. Chan, T. O. Olawumi, and A. M. Ho, "Perceived benefits of and barriers to Building Information Modelling (BIM) implementation in construction: The case of Hong Kong," *Journal of Building Engineering*, vol. 25, p. 100764, 2019.

[15] A. Leśniak, M. Górka, and I. Skrzypczak, "Barriers to BIM implementation in architecture, construction, and engineering projects—The polish study," *Energies*, vol. 14, no. 8, p. 2090, 2021.

[16] G. Lee, H. K. Park, and J. Won, "D3 City project—Economic impact of BIM-assisted design validation," *Automation in Construction*, vol. 22, pp. 577-586, 2012.

[17] Z. Sriyolja, N. Harwin, and K. Yahya, "Barriers to implement building information modeling (BIM) in construction industry: A critical review," in *IOP Conference Series: Earth and Environmental Science*, 2021, vol. 738, no. 1: IOP Publishing, p. 012021.

[18] C. El Hajj, G. Martínez Montes, and D. Jawad, "An overview of BIM adoption barriers in the Middle East and North Africa developing countries," *Engineering, Construction and Architectural Management*, vol. 30, no. 2, pp. 889-913, 2023.

[19] Y. Y. Al-Ashmori, I. Othman, and Y. Rahmawati, "Bibliographic analysis of BIM success factors and

other BIM literatures using Vosviewer: A theoretical mapping and discussion," in *Journal of Physics: Conference Series*, 2020, vol. 1529, no. 4: IOP Publishing, p. 042105.
[20] O. I. Olanrewaju, A. F. Kineber, N. Chileshe, and D. J. Edwards, "Modelling the relationship between Building Information Modelling (BIM) implementation barriers, usage and awareness on building project lifecycle," *Building and Environment*, vol. 207, p. 108556, 2022.
[21] S. Tang, D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, and X. Gao, "A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends," *Automation in Construction*, vol. 101, pp. 127-139, 2019.
[22] T. Umar, "Challenges of BIM implementation in GCC construction industry," *Engineering, Construction and Architectural Management*, vol. 29, no. 3, pp. 1139-1168, 2022.
[23] N. Rohani and S. Y. Banihashemi, "Identifying and prioritizing the barriers to BIM implementation in Iran," *Amirkabir Journal of Civil Engineering*, vol. 54, no. 2, pp. 775-792, 2022.
[24] H. Sarvari, P. Asaadsamani, T. O. Olawumi, D. W. Chan, A. Rashidi, and M. Beer, "Perceived barriers to implementing building information modeling in Iranian Small and Medium-Sized Enterprises (SMEs): a Delphi survey of construction experts," *Architectural Engineering and Design Management*, pp. 1-21, 2024.
[25] H. Abbasianjahromi, M. Ahangar, and F. Ghahremani, "A maturity assessment framework for applying BIM in consultant companies," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 43, pp. 637-649, 2019.
[26] S. M. H. M. Zare and A. khosravi Ahangar, "A comparative study on BIM (building information modeling) implementation and maturity across different countries with a review on Iran," in *Proceeding of 3rd International Conference of Building Information Modelling*, 2020, pp. 1-12.
[27] O. Olugboyega, "Differential relationships in the BIM implementation process in a developing country: the role of essential BIM implementation strategies," *Engineering, Construction and Architectural Management*, 2023.
[28] A. Criminale and S. Langar, "Challenges with BIM implementation: a review of literature," in *53rd ASC annual international conference proceedings*, 2017, pp. 329-335.
[29] S. Ahmed, "Barriers to implementation of building information modeling (BIM) to the construction industry: a review," *Journal of civil engineering and construction*, vol. 7, no. 2, pp. 107-113, 2018.
[30] M. Marzouk, H. Elsaay, and A. A. E. Othman, "Analysing BIM implementation in the Egyptian construction industry," *Engineering, Construction and Architectural Management*, vol. 29, no. 10, pp. 4177-4190, 2022.
[31] M. Khanzadi, M. Sheikhkhoshkar, and S. Banihashemi, "BIM applications toward key performance indicators of construction projects in Iran," *International Journal of Construction Management*, vol. 20, no. 4, pp. 305-320, 2020.
[32] B. J. Oates, M. Griffiths, and R. McLean, *Researching information systems and computing*. Sage, 2022.
[33] E. Alreshidi, M. Mourshed, and Y. Rezgui, "Factors for effective BIM governance," *Journal of Building Engineering*, vol. 10, pp. 89-101, 2017.
[34] M. Sandelowski, "Sample size in qualitative research," *Research in nursing & health*, vol. 18, no. 2, pp. 179-183, 1995.
[35] W. Ma, X. Wang, J. Wang, X. Xiang, and J. Sun, "Generative design in building information modelling (BIM): approaches and requirements," *Sensors*, vol. 21, no. 16, p. 5439, 2021.
[36] N. A. H. Hadzaman, R. Takim, and A. H. Nawawi, "BIM roadmap strategic implementation plan: Lesson learnt from Australia, Singapore and Hong Kong," in *Proceedings in 31st annual ARCOM conference*, 2015, pp. 611-620.
[37] A. Porwal and K. N. Hewage, "Building Information Modeling (BIM) partnering framework for public construction projects," *Automation in construction*, vol. 31, pp. 204-214, 2013.
[38] J.-B. Yang and H.-Y. Chou, "Mixed approach to government BIM implementation policy: An empirical study of Taiwan," *Journal of Building Engineering*, vol. 20, pp. 337-343, 2018.
[39] B. Succar and M. Kassem, "Macro-BIM adoption: Conceptual structures," *Automation in construction*, vol. 57, pp. 64-79, 2015.
[40] J. C. Cheng and Q. Lu, "A review of the efforts and roles of the public sector for BIM adoption worldwide," *Journal of Information Technology in Construction (ITcon)*, vol. 20, no. 27, pp. 442-478, 2015.

Decision-Support Framework for Evaluating Steel Connection Systems for Adoption in Construction

Tolulope Oyeyipo[1], Ibukun Awolusi, Ph.D.[2]*, Arturo Schultz, Ph.D.[3], Debra Laefer, Ph.D.[4], and Salam Al-Sabah, Ph.D.[5]

[1]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
[2*]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
[3]School of Civil & Environmental Engineering, and Construction Management, The University of Texas at San Antonio, USA
[4]Center for Urban Science + Progress and Dept. of Civil & Urban Engineering, New York University, USA. Email:
[5]School of Civil Engineering, University College Dublin, Ireland
tolulope.oyeyipo@my.utsa.edu, ibukun.awolusi@utsa.edu, arturo.schultz@utsa.edu, debra.laefer@nyu.edu, salam.al-sabah@ucd.ie

Abstract –

Steel construction currently lacks systematic and validated frameworks that can be used to evaluate innovative systems against established alternatives before field deployment. This contributes to delayed adoption even where there are demonstrated performance advantages. This study presents the development of a conceptual decision-support framework for assessing steel connection systems. It compares an innovative intermeshed steel connection (ISC) with the existing bolted steel connection. The proposed framework integrates building information modeling (BIM) for constructability analysis, computer-vision-based ergonomic assessment for pose estimation, simulation-based performance evaluation, and multi-criteria decision analysis using Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). Using a simulated assembly scenario, the framework systematically evaluates connection systems across the parameters of worker safety (ergonomic risk), assembly efficiency, and construction cost. The framework development process demonstrates how simulated data informs framework parameters using criterion weights from surveys, performance metrics measured from simulation protocols, and decision thresholds. The simulated results indicate that the ISC system reduces ergonomic risks by up to 40% as measured by the Rapid Entire Body Assessment (REBA) mean scores of 4.57 for bolted connection and 2.57 for ISC. The assembly duration of ISC also decreased by approximately 29% and achieved cost savings of up to 10% across geographic locations. This study provides a replicable conceptual methodology for developing decision-support systems for a systematic evaluation of innovative systems. By applying simulated findings to validated decision-support systems, it illustrates how digital technologies such as BIM and computer vision can improve objective technology assessment in construction. The proposed framework offers a basis for future empirical data-based validation studies to guide decision making in steel connection systems adoption.

Keywords –

BIM, Computer vision, Construction, Decision support framework, Simulation, Steel connection, TOPSIS

1 Introduction

The construction industry faces persistent reluctance in adopting innovative technologies even where there is potential to improve safety, efficiency, and cost effectiveness [1]. Construction practitioners often require systematic frameworks to evaluate emerging construction methods and materials against established practices [2]. This oftentimes leads to conservative decision-making and slow technology diffusion [3]. For structural steel construction, where connection technologies have been evolving over time from riveting to welding and high-strength bolting, each transition still required decades of industry acceptance despite empirical performance advantages available at the time [4, 5]. The unavailability of structured decision-support systems integrating multiple performance dimensions, safety, efficiency, and economics, creates a major challenge for innovation adoption in a risk-prone

industry such as the steel construction industry.

Decision-making in construction technology adoption is more complex, sometimes requiring a concession among the competing performance objectives. These objectives may include cost, quality, productivity, worker safety, and sustainability [6]. Traditionally, evaluation methods would focus on a single criterion, mostly cost, and avoid a comprehensive performance assessment. This overlooks the interconnectedness of construction performance metrics, where a change in one dimension can result in cascading benefits in the others [7]. A connection system that can reduce ergonomic strain on workers may also reduce assembly time, reduce injury-related costs on projects, and improve work quality and worker efficiency due to reduced fatigue. Yet these quantitative synergies are vastly unquantified in the traditional decision processes [8,9]. Multi-criteria-decision-making (MCDM) methods such as Analytical Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) have been used as tools to address this complexity. Although interest in MCDM frameworks for innovative construction technology assessments is growing, gaps exist in the literature on conceptual frameworks that systematically integrate digital construction technologies with multi-criteria decision analysis [10]. Most existing studies focus on theoretical MCDM applications while evading data collection limitations in evaluating new innovative technologies [11]. Some studies present case-specific analyses that lack generalizability across various construction contexts, such as steel construction [12, 13]. Conceptual frameworks offer the distinctive advantage of enabling systematic identification of relevant evaluation criteria, establishing relationships between performances, providing a foundation for future empirical validation, and steering the development of measurement protocols ahead of committing major resources to the experimental validation [10].

Integrating digital construction technologies introduces unprecedented opportunities for conceptual framework development. This can be improved by objective, automated performance assessment capabilities. Building information modeling (BIM) is now more than design coordination; it details project constructability analysis, clash detection, 4D scheduling, and assembly sequence optimization [14, 15]. Computer-vision technologies for pose estimation provide a system for objectively assessing worker safety and ergonomics through biomechanical posture analysis. [16, 17]. BIM-based construction sequencing allows for a time-motion analysis used for estimating task duration and productivity rates for comparative assessment [18]. The integration of these tools, alongside expert validation, creates an opportunity to develop a decision framework that bridges theoretical MCDM principles with the practicality of the construction industry.

Steel connection technology is critical to structural integrity, schedule performance, and worker hazard exposure, making it an ideal domain for developing a conceptual decision-support framework. Up to 40% of steel erection labor hours are spent on connection works, ranking the task as high-risk, exposing workers to hazardous risks such as fall from heights and struck-by hazards [6]. An intermeshed steel connection (ISC) is an innovative alternative to bolted steel connections and provides a perfect example for the performance criteria evaluation challenges facing the steel construction industry [19]. ISC uses the interlocking theory to secure steel elements using simple fastening mechanisms that significantly minimize the need for bolts. It potentially reduces assembly duration, minimizing ergonomic hazard exposure, repetitive overhead, and bolt installation work, and decreases material and labor costs [19]. An absence of a standardized evaluation framework, limited performance data, and a level of unfamiliarity among construction professionals pose a challenge, limiting contractors, structural engineers, and project owners in assessing whether the ISC technology is safe for adoption on their specific projects. This uncertainty is aggravated by the interconnectedness of the connection system performance criteria [20].

This study addresses this gap by presenting a comprehensive conceptual decision-support framework for evaluating steel connection systems using ISC and traditional bolted connections on a comparative basis. The conceptual framework will integrate (1) BIM-based constructability analysis, (2) computer-vision-based ergonomic assessment, (3) simulation-based performance evaluation for performance criteria, and (4) TOPSIS-based multi-criteria decision analysis. This study develops a repeatable conceptual methodology for structured decision-support frameworks for both simulation-based estimates and future empirical data. It illustrates the theoretical integration of computer-vision-based ergonomic assessment and BIM constructability, thereby advancing digital tools application beyond design coordination to advance construction method selection. The study also presents an expert validation approach centered on practical applicability, criterion relevance, and implementation of feasibility through professional feedback.

2 Literature Review

2.1 Construction technology adoption and innovation diffusion

The construction sector has a conservative approach to technology adoption. This stems from the industry being driven by project-based production systems,

fragmented supply chains, low profit margins that discourage innovative experimentation, regulatory constraints, and liability from unproven methods [21, 22]. Client demand for innovation, a short-term returns focus, and insufficient collaboration between research institutions and construction practitioners further widen the innovation gap [23, 24]. The study of [25] posited diffusion of innovation theory as a foundational framework for technology adoption patterns. Mitropoulos [26] noted that new technologies must prove compatibility with existing work processes, need minimal learning curves, and provide evidence of advantages with pilot implementations. Where proof of field-tested employment cannot be avoided, it becomes inherently difficult to demonstrate the potential benefits that will justify adoption [27].

Steel connection technology evolution is an example of long adoption curves from riveting to welding and high-strength bolted steel connections. This was despite the clear advantages in connection strength and speed presented at each of these stages [19]. This confirms that the industry requires technical validation and a comprehensive decision-support tool that enables professionals and practitioners to evaluate innovations within their specific context [25, 21]. Intermeshed steel connections are currently facing these same adoption challenges, as they lack an extensive field of performance data and structured evaluation frameworks to provide informed decision-making.

2.2 Multi-criteria decision making in construction management

MCDM enables systematic multi-dimensional evaluation that accommodates diverse stakeholder preferences [10]. It applies across diverse situations, including contractor selection [20], construction method selection [27], sustainable building design [28], safety management strategy evaluation [1], and equipment selection decisions [12]. TOPSIS, an MCDM technique, is particularly suitable for construction applications due to its computational simplicity, large criterion handling, and intuitive logic [29]. The study of [30] successfully applied TOPSIS for green building project delivery methods, [31] utilized TOPSIS for subcontractor selection model, [32] applied TOPSIS for construction optimization decision scheme of urban subway tunnels. However, MCDM applications majorly rely on historical project data, and where innovative technology lacks such data, researchers must rely on expert judgement or develop alternative data generation approaches [1, 24]. These data limitations become even more acute where innovations are evaluated against established alternatives [13]. Recent research in MCDM has proposed framework validation as a means of establishing confidence in decision-support outputs [33].

2.3 Digital technologies in construction assessment

BIM-based constructability analysis evaluates assembly feasibility, detects spatial conflicts, and stimulates installation sequences, providing a significant influence on safety and productivity planning for steel connections [14, 15]. Computer vision pose estimation eliminates subjectivity in ergonomic data collection and enables continuous dynamic monitoring [10]. Its application has been extended to productivity measurement through activity recognition [18] and to simulated ergonomic assessment via virtual reality environments [34]. Despite these advances, systematic integration of computer vision ergonomic assessment into multi-criteria decision-support frameworks for technology evaluation remains unexplored.

Simulation-based research enables controlled investigations of construction processes without full-scale resource requirements, safety, economic concerns, and hidden variables in field experimentation [10, 35]. The extent to which simulation output correctly represents real-world scenarios requires careful consideration of model assumptions, input parameter estimates, and the need for validation against empirical data as soon as it is available [36]. For steel connections, construction process simulation for connection systems remains underdeveloped, with limited research examining the assembly sequences, worker movements, or comparative productivity analysis. Integrating a BIM-based spatial simulation with computer-vision-based ergonomic assessment for connection technology evaluation represents a novel approach to be documented.

Conceptual framework validation organizes challenging dynamics, identifying relationships between variables, and provides step-by-step guides for future empirical designs. The approach varies based on the intent and development stage. Expert panels assess the framework for comprehensiveness, clarity, intended use, and application feasibility [27]. Limited literature guides the development of conceptual decision-making frameworks, especially for incorporating simulation outputs and conducting expert validation in early research stages. This study adopts a simulation-based framework to illustrate how performance data would inform decision processes while awaiting actual empirical data confirmation to address this gap.

3 Framework Development Methodology

This section presents an overview of the process for developing the conceptual decision-support framework. The process follows 4 sequential phases, with each one building on the output of the prior phase. Phase one builds the evaluation criteria by identifying factors that

matter when choosing between steel connection technologies. The factors were identified through a qualitative literature review from academic database searches (Web of Science, Engineering Village, Scopus). Three major performance dimensions of worker safety, assembly efficiency, and construction cost were identified, with a further split into 12 sub-criteria for expert review and feedback. These criteria were confirmed through prior ISC studies [19, 37], which established that up to 40% of steel erection labor hours are devoted to connection work, making ergonomic risks, assembly duration, and cost the most consequential performance dimensions for comparative technology evaluation.

Phase 2 designed the simulation process protocol. The model components (beams, columns, and connection details) were defined (14 ft × 14 ft × 20 ft steel frame comprising four W14×90 columns, four W18×50 beams, and eight connections, modeled in Autodesk Revit and exported to Navisworks for 4D simulation), tools, and equipment set-up positioning. Simulation assembly steps were defined from a work breakdown structure (WBS), and durations for task performance were recorded for analysis. A visualization software, Navisworks, analyzed the 4D timeline, clash detection, worker viewpoints, and time tracking. Next, a computer-vision-based algorithm measured how physically demanding each connection type is for workers. This was achieved through a REBA scoring system that identified worker postures during various stages of steel connections (Openpose algorithm helped to analyze and extract body key points and joint angles), risk levels, and guided required resolution actions. Cost analysis informed by historical cost data of RS Means identified each steel member by line-item description and broke down the assembly costs into material, labor, equipment, and life-cycle costs. A sensitivity analysis across various geographical locations also measured the cost impact expectations.

Phase 3 executed the simulation and data collection process in 3 scenarios: typical assembly process, constrained site conditions, and a detailed production phase. The typical assembly process relied on the ISC dataset from the study of [19] compared against the bolted connection dataset from site recordings. Results were aggregated into a performance data matrix. Weights for each performance criterion and sub-categories were also calculated for a final TOPSIS analysis. A point allocation expert survey was distributed to 8 steel construction professionals, spanning safety managers, general contractors, structural engineers, steel erectors, and project owners. Performance dimensions to be weighed were worker safety, assembly efficiency, and construction costs. Sub-criterion weights within worker safety placed the highest importance on ergonomic risk and fall protection requirements. For assembly efficiency, installation time, and labor requirements were noted as most important, while for construction costs, labor and material costs were both weighted as the highest consideration factors.

Phase 4 is the TOPSIS analysis stage; the two connection technologies are compared mathematically (vector normalization), and the best choice is closest to the ideal solution as much as possible and farthest from the worst solution. This is summarized as presented in Figure 1 above. The TOPSIS model measures the weightings from expert feedback on performance dimensions and sub-criterion factors, and plots against standard ideal ranges of performance. The technology with consistent weighting, closest to the ideal, is selected as a preferred technology.

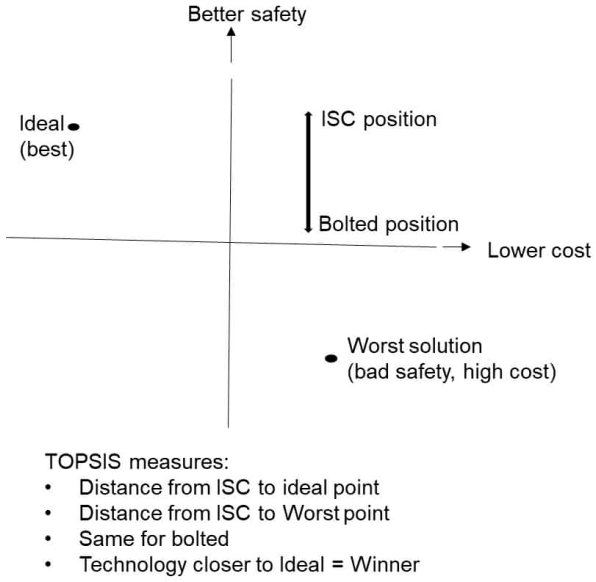


Figure 1. TOPSIS decision support analysis chart

The typical framework outputs and applications can be seen as presented in Table 1 below. The outputs translate analytical evaluation into practical decision support for multiple stakeholders. Overall closeness coefficients provide a single performance score that helps project managers and contractors decide whether the ISC system should be adopted. Criterion-level scores break this result down by technical, cost, or safety dimensions, allowing engineers and safety managers to identify specific strengths and weaknesses. Weight sensitivity results show how stable the recommendation is when priorities change, supporting confidence among risk-averse decision-makers. Cost breakdowns offer location-specific financial insight for owners and estimators, while safety metrics quantify hazard exposure to help safety directors and regulators assess whether the ISC system provides a safer alternative.

Table 1. Proposed Framework Output and Application Process

Output	What it shows	Who uses it	Decisions supported
Closeness coefficients	Overall technology performance score	Project managers, contractors	Should the ISC system be adopted?
Criterion-Level scores	Performance weight on specific dimensions or criteria	Engineers, Safety managers	Where does ISC perform better or worse?
Weight Sensitivity	How robust is the recommendation?	Risk-averse stakeholders	Can this decision be trusted?
Cost Breakdowns	Detailed cost comparison by location	Estimators, Owners	Market cost impact
Safety Metrics	Hazard exposure quantification	Safety directors, OSHA	Is the ISC system safer?

4 Results

The proposed conceptual decision-support framework incorporates three primary performance dimensions: worker safety, assembly efficiency, and construction cost. Each dimension was further split into 12 secondary evaluation criteria (four per dimension), and specific performance indicators were weighed from expert reviews. The framework integrates outputs from BIM-based constructability simulation, computer-vision-based ergonomic assessment, cost database analysis across geographical locations, and TOPSIS multi-criteria decision analysis for systematic technology comparison.

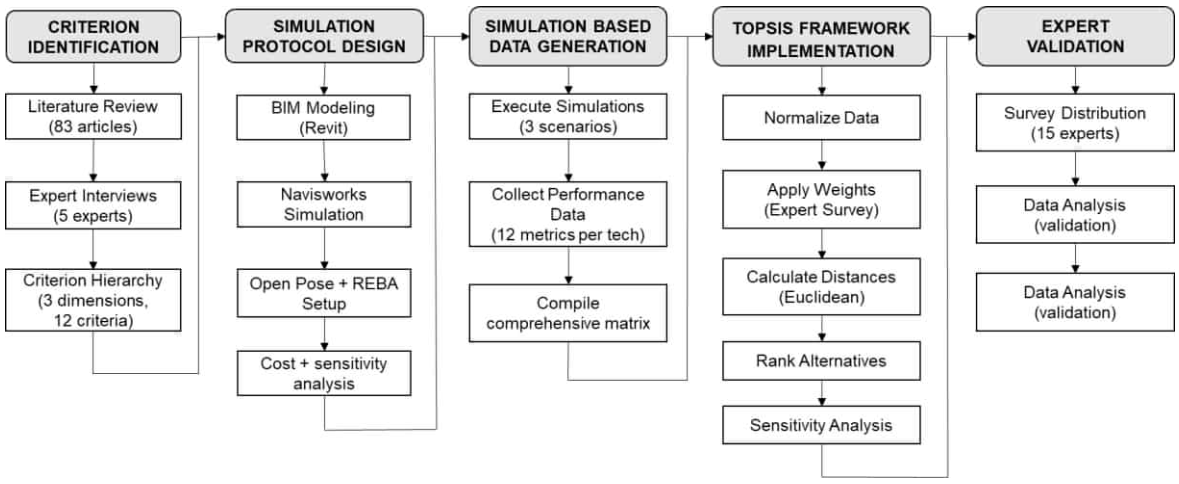


Figure 2. Conceptual Decision-Support Framework for Evaluating Steel Connection Systems

The proposed framework in Figure 2 addresses a critical gap in construction management by providing a structured, repeatable process for evaluating innovative steel connection systems, such as the intermeshed steel connection, against established traditional connection systems. The 3 dimensions of performance review offer practitioners a comprehensive overview of the technology when field performance data remains limited. The framework contributes to a systematic framework for integrating digital construction technologies with a thorough MCDM methodology, providing objective evaluation approaches. The proposed expert feedback weights from the survey and sensitivity analysis protocols enable the framework to accommodate several stakeholder priorities while maintaining robust decision recommendations. While this conceptual framework architecture establishes protocols and analytical procedures, further empirical data-based validation will be essential to confirm the framework's effectiveness and refine measurement approaches. Robust empirical data would also aid in verifying the criterion of relevance through expert review surveys and illustrating practical application through a pilot study.

Applying the survey-derived weights to the normalized performance matrix, the TOPSIS analysis yielded closeness coefficients of 0.664 for ISC and 0.344 for the bolted connection, ranking ISC as the preferred

technology across all 12 evaluation criteria. ISC outperformed the bolted connection on 11 of the 12 criteria, with the bolted system retaining an advantage only on the learning curve due to established industry familiarity. Sensitivity analysis under three alternative weight scenarios (equal weights, safety-focused, and cost-focused) confirmed ISC as the ranked-first technology in every scenario, demonstrating ranking robustness regardless of stakeholder priority orientation. Panelists evaluated the framework on comprehensiveness, criterion relevance, measurement feasibility, TOPSIS appropriateness, and practical applicability using a 5-point Likert scale, with an acceptance threshold of ≥ 4.0 on each dimension and ≥ 60% likelihood of adoption.

5 Conclusion

The study presents a conceptual decision-support framework for evaluating steel connection systems. This framework was tested through simulated performance data drawn from prior ISC research. Although the framework is conceptual in design, its development was informed by empirical and simulation-based evidence. A computer vision REBA assessment using OpenPose on steelworker image datasets quantified a 40% ergonomic risk deduction for ISC over bolted connections. 4D assembly simulations also documented a 29% reduction in installation time with a reduction in fall hazard exposure. RS means cost data analysis across five geographic locations confirmed a consistent 10% life-cycle cost advantage for ISC. These data-informed inputs strengthen the credibility of the TOPSIS ranking outputs and demonstrate how digital construction technologies can objectively support technology adoption decisions before full-scale empirical validation is available. The framework offers construction researchers and practitioners a replicable, evidence-based methodology for evaluating innovative steel connection systems. Future research should operationalize the framework through controlled laboratory or field-based pilot tests using live assembly data to validate the simulated performance metrics. A formal expert validation survey spanning professionals, academia, and industry practitioners will provide additional criterion-weighting feedback. The framework is also extensible to connection technologies beyond ISC and bolted systems, and automation through user-friendly TOPSIS-based decision tools would further support industry adoption.

Acknowledgment

Material discussed in this document is based upon work supported by the National Science Foundation (NSF) under award number 2222815. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

[1] Yap, J. B. H., Lam, C. G. Y., Skitmore, M., & Talebian, N. (2022). Barriers to the adoption of new safety technologies in construction: A developing country context. *Journal of Civil Engineering and Management*, *28*(2), 120-133.

[2] Almashhour, R., Al-Mhdawi, M. K. S., Daghfous, A., Qazi, A., & Ojiako, U. (2025). Traditional to sustainable risk management in the construction industry: a systematic literature review. *International Journal of Managing Projects in Business*.

[3] Naoum, S. G., & Egbu, C. (2016). Modern selection criteria for procurement methods in construction: A state-of-the-art literature review and a survey. *International Journal of Managing Projects in Business*, *9*(2), 309-336.

[4] AISC. (2017). Connection Designs - Design Requirement. <https://2cl405uvd.files.wordpress.com/2013/06/connection-design-e28093-design-requirements.pdf > (October 21, 2023).

[5] Almashhour, R., Al-Mhdawi, M. K. S., Daghfous, A., Qazi, A., & Ojiako, U. (2025). Traditional to sustainable risk management in the construction industry: a systematic literature review. *International Journal of Managing Projects in Business*.

[6] Karakhan, A. A., Rajendran, S., Gambatese, J., & Nnaji, C. (2018). Measuring and evaluating safety maturity of construction contractors: Multicriteria decision-making approach. *Journal of Construction Engineering and Mnagement*, *144*(7), 04018054.

[7] Ibrahim, A., Zayed, T., & Lafhaj, Z. (2024). Enhancing construction performance: A critical review of performance measurement practices at the project level. *Buildings*, *14*(7), 1988.

[8] Njemanze, B. C. (2024). *Implementation of an Evidence-Based Ergonomic Protocol to Decrease Patient Injuries in an Occupational Health Setting* (Doctoral dissertation, Aspen University).

[9] Adjei, K. O., Liphadzi, M., Bondinuba, F. K., Mewomo, C. M., & Abudu, H. D. (2025). Musculoskeletal Health Effects of Manual Sandcrete Block Handling Among Construction Artisans: Implications for Industry Practice and Training. *International Journal of Environmental Research and Public Health*, *22*(11), 1689.

[10] Wang, K., Ying, Z., Goswami, S. S., Yin, Y., & Zhao, Y. (2023). Investigating the role of artificial

intelligence technologies in the construction industry using a Delphi-ANP-TOPSIS hybrid MCDM concept under a fuzzy environment. *Sustainability*, *15*(15), 11848.

[11] Pidgeon, A., & Dawood, N. (2021). Bridging the gap between theory and practice for adopting meaningful collaborative BIM processes in infrastructure projects, utilising multi-criteria decision making (MCDM). *J. Inf. Technol. Constr.*, *26*, 783-811.

[12] Awolusi, I., & Sulbaran, T. (2021). Decision support framework for selecting wearable internet of things devices for safety management in construction. In *ASC* (pp. 320-309).

[13] Marzouk, M., & Sabbah, M. (2021). AHP-TOPSIS social sustainability approach for selecting supplier in construction supply chain. *Cleaner environmental systems*, *2*, 100034.

[14] Sacks, R., Eastman, C., Lee, G., & Teicholz, P. (2018). *BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers*. John Wiley & Sons.

[15] Tang, Y., Wang, G., Li, H., Cao, D., & Li, X. (2019). Comparing Project-Based Collaborative Networks for BIM Implementation in Public and Private Sectors: A Longitudinal Study in Hong Kong. *Advances in Civil Engineering*, *2019*(1), 6213694.

[16] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).

[17] Roberts, D., Torres Calderon, W., Tang, S., & Golparvar-Fard, M. (2020). Vision-based construction worker activity analysis informed by body posture. *Journal of Computing in Civil Engineering*, *34*(4), 04020017.

[18] Golparvar-Fard, M., Pena-Mora, F., & Savarese, S. (2015). Automated progress monitoring using unordered daily construction photographs and IFC-based building information models. *Journal of Computing in Civil Engineering*, *29*(1), 04014025.

[19] Al-Sabah, S., Laefer, D. F., Truong Hong, L., Phuoc Huynh, M., Le, J. L., Martin, T., ... & Dizon, R. (2020). Introduction of the intermeshed steel connection—a new universal steel connection. *Buildings*, *10*(3), 37.

[20] Ambekar, S., Dhaigude, A. S., Chatterjee, D., & Tapar, A. V. (2025). Identifying and ranking critical success factors for contractor value addition in the construction sector using the ISM-IRP technique. *Construction Innovation*.

[21] Taylor, J., & Levitt, R. (2004). Understanding and managing systemic innovation in project-based industries. *Innovations: Project management research*, *2004*, 83-99.

[22] Davies, A., Brady, T., Prencipe, A., & Hobday, M. (2011). Innovation in complex products and systems: implications for project-based organizing. In *Project-based organizing and strategic management* (pp. 3-26). Emerald Group Publishing Limited.

[23] Ingirige, B., & Sexton, M. (2006). Alliances in construction: investigating initiatives and barriers for long-term collaboration. *Engineering, Construction and Architectural Management*, *13*(5), 521-535.

[24] Rossoni, A. L., de Vasconcellos, E. P. G., & de Castilho Rossoni, R. L. (2024). Barriers and facilitators of university-industry collaboration for research, development and innovation: a systematic review. *Management Review Quarterly*, *74*(3), 1841-1877.

[25] Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of innovations. In *An integrated approach to communication theory and research* (pp. 432-448). Routledge.

[26] Mitropoulos, P., & Tatum, C. B. (2000). Forces driving adoption of new information technologies. *Journal of construction engineering and management*, *126*(5), 340-348.

[27] Gambatese, J. A., & Hallowell, M. (2011). Factors that influence the development and diffusion of technical innovations in the construction industry. *Construction Management and Economics*, *29*(5), 507-517.

[28] Naoum, S. G., & Egbu, C. (2016). Modern selection criteria for procurement methods in construction: A state-of-the-art literature review and a survey. *International Journal of Managing Projects in Business*, *9*(2), 309-336.

[29] Abidin, M. Z., Rusli, R., & Shariff, A. M. (2016). Technique for Order Performance by Similarity to Ideal Solution (TOPSIS)-entropy methodology for inherent safety design decision making tool. *Procedia Engineering*, *148*, 1043-1050.

[30] Balali, V., Zahraie, B., & Roozbahani, A. (2014). A comparison of AHP and PROMETHEE family decision making methods for selection of building structural system. *American Journal of Civil Engineering and Architecture*, *2*(5), 149-159.

[31] Gao, C., Elzarka, H., Yan, H., Chakraborty, D., & Zhou, C. (2023). A Hybrid TOPSIS-Structure Entropy Weight Group Subcontractor Selection Model for Large Construction Companies. *Buildings*, *13*(6), 1535. https://doi.org/10.3390/buildings13061535

[32] Wu, B., Lu, M., Huang, W., Lan, Y., Wu, Y., & Huang, Z. (2020). A case study on the construction optimization decision scheme of urban subway

tunnel based on the TOPSIS method. *KSCE Journal of Civil Engineering*, *24*(11), 3488-3500.
[33] Liu, S., Duffy, A. H., Whitfield, R. I., & Boyle, I. M. (2010). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*(3), 261-286.
[34] Wang, Y., & Luo, Y. (2023). Research on computer aided interaction design based on virtual reality technology. *Procedia Computer Science*, *228*, 619-628.
[35] Nguyen, A. T., Reiter, S., & Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. *Applied energy*, *113*, 1043-1058.
[36] Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). Concepts of model verification and validation.
[37] Oyeyipo, T., Avila, J., Awolusi, I., Schultz, A., Laefer, D., & Al-Sabah, S. (2025). BIM and Safety Integration for Steel Erection using Intermeshed Steel Connections. In CIB Conferences (Vol. 1, No. 1, p. 97).

Evaluation of Technical Readiness Level for Intelligent Construction Robots

Hongrui Yu[1] and Somin Park[2]

[1]Department of Civil & Environmental Engineering, Virginia Tech, USA
[2]Department of Civil Engineering, University of Texas Arlington, USA
hryu42@vt.edu, somin.park@uta.edu

Abstract –

This paper presents a systematic review of intelligent construction robots and their readiness for real-world deployment. Intelligent construction robots are defined as software-based or physically embodied agents that operate in a Perception–Reasoning–Action loop—sensing the environment, reasoning over observations and/or human commands, and executing actions through control or actuation. Using a structured search of Web of Science, the ASCE Library, and Scopus, 134 peer-reviewed journal and conference papers were identified and analyzed. A construction-oriented Technology Readiness Level (TRL) assessment framework was developed by adapting the classic TRL model and prior civil engineering TRL studies to evaluate technical maturity and deployability. Results reveal a step-up maturation trajectory: after an extended plateau near mid-level readiness, the literature shows accelerated growth in 2017–2025 with increasing dispersion in study quality. Only two studies achieved TRL 9 field validation, while most 2025 studies clustered at TRL 8. Findings highlight a persistent "valley of death" barrier between validation and deployment and motivate multi-condition field testing, safety assurance, and workflow integration to advance site-ready robotics.

Keywords –

Construction Robotics; Technical Readiness Level; Artificial Intelligence; Deep Learning; Reinforcement Learning

1 Introduction

The adoption of construction robots is increasingly necessary given the industry's severe labor shortage. Since 2023, the construction labor gap has reached 546,000 workers [1]. This sustained scarcity of skilled labor slows project delivery and substantially increases costs. A recent poll found that more than 90% of contractors report difficulty hiring qualified workers, which in turn raises operating expenses [2].

One prominent response is increased automation and robotics. Automated construction methods can reduce repetitive manual work by 25–90% and cut workers' exposure to hazardous activities by 72%, while improving accuracy and reducing rework [3]. Accordingly, construction robotics has been widely recognized as both necessary and productive [4,5]. Early diffusion, however, was constrained by high capital costs, the inherently unstructured nature of jobsites, and limited sensing and cognition in robotic platforms [6]. Today, simultaneous advances in artificial intelligence (AI), sensing, and building information modeling (BIM) are enabling a resurgence—from single-purpose tools toward more capable autonomous systems that can execute complex tasks. These developments are ushering in a new era of intelligent construction robots that can perceive, reason, and act autonomously to perform high-quality craft operations.

With rapid technological progress, construction robots increasingly leverage advanced perception and decision-making methods, including deep learning and reinforcement learning. Many successful prototypes have been demonstrated across a wide range of tasks, such as ceiling installation [7,8], window installation [9], drywall installation [10], inspection [11], welding [12], and human–robot interactive operations [13,14]. However, although these systems often show strong performance in high-fidelity simulation or controlled laboratory environments, the field still lacks a systematic evaluation framework that enables industry managers and decision-makers to assess the technical readiness of these technologies for deployment on real construction sites.

To address this gap, this paper adapts the widely used NASA Technology Readiness Level (TRL) framework [15] for construction robotics products and applies it to systematically evaluate the technical readiness of current intelligent construction robotics research. We analyze 134 papers published from 2017 to 2025 on intelligent construction robots and use the adapted framework to quantify their readiness levels. We further examine overall trends in readiness and product evolution to

characterize research progress, identify roadmaps, and highlight key challenges that must be addressed to accelerate real-world deployment.

2 Technical Readiness Level Framework

As mentioned in the Introduction section, this paper is theoretically supported by the classic TRL framework proposed by Nasa (15 Mankins 1995), which is shown in Table 1.

Table 1 Original TRL framework [15]

TRL Level	Definition (Operational Environment)
TRL 1	Basic principles observed and reported
TRL 2	Technology concept and/or application formulated
TRL 3	Analytical and experimental critical function and/or characteristic proof-of-concept
TRL 4	Component and/or breadboard validation in laboratory environment
TRL 5	Component and/or breadboard validation in relevant environment
TRL 6	System/subsystem model or prototype demonstration in a relevant environment (ground or space)
TRL 7	System prototype demonstration in a space environment
TRL 8	Actual system completed and "flight qualified" through test and demonstration (ground or space)
TRL 9	Actual system "flight proven" through successful mission operations

In this framework, the general hierarchy is from proposing the concept - component validation - system validation in controlled environments - system validation in true environment (space). It also fits the development of construction automation products. For construction robotics, it will also generally go through the similar phases: conceptualization (TRL)- prototyping in the lab- test in the field.

To understand how the traditional TRL scheme can be fit for construction technology development, the authors referred to two sources:

1) The core of development in NASA TRL: The authors conceptualized and abstracted the essence of TRL to three phase: conceptualization - prototyping in lab -test in the field. For construction robotics, it will also generally go through the similar phases: conceptualization (TRL)- prototyping in the lab- test in the field.

2) The previous work adapting TRL to civil engineering or construction related technologies. The following two representative TRL schemes were discovered, as shown in Table 2 and 3:

Table 2 TRL for construction and civil engineering [16]

TRL Level	Definition (Operational Environment)
TRL 1	Basic principles and research: Scientific principles observed
TRL 2	Applied Research: Application formulated and potential identified
TRL 3	Proof of Concept: Critical function validated in separate elements
TRL 4	Lab Validation (Low Fidelity): Component validation in a laboratory environment
TRL 5	Lab Validation (High Fidelity): Integrated components demonstrated in a laboratory environment
TRL 6	Prototype (Relevant Environment): Prototype demonstrated in a relevant environment
TRL 7	Prototype (Operational Environment): Prototype demonstrated in an operational environment
TRL 8	System Proven: Technology proven in an operational environment
TRL 9	Implementation: Technology refined and adopted

For a more abstract type, the following example on construction material development was also used in Table 3 [17].

Table 3 An abstract version of TRL for Civil Engineering [17]

TRL Phase	Construction-Specific

	Definition
TRL 1–3 (Concept)	Chemical formulation of the biocomposite (liquid wood) and basic mechanical testing (tension/compression) on small samples.
TRL 4–5 (Validation)	"Component Testing: Producing standard construction shapes (profiles, boards) and testing them for fire resistance, water absorption, and UV stability in a lab."
TRL 6–7 (Demonstration)	"Pilot Projects: Installing the material in a non-critical real-world structure (e.g., a park bench, facade cladding, or small shelter) to monitor weathering over 1–2 years."
TRL 8–9 (Deployment)	"Standardization: The material is included in national building codes (e.g., Eurocodes) and is available in commercial catalogs for structural use."

As summarized in Table 4, we aggregate the nine original TRL levels into four tiers to better reflect the typical development trajectory of construction robotics. Tier 1 (TRL 1–3) emphasizes conceptualization, including knowledge integration across disciplines (e.g., construction engineering, robotics, sensing, and AI) to clarify potential technological solutions to a construction-domain problem. Tier 2 (TRL 4–5) corresponds to component-level validation, where key subsystems are tested to verify that individual components can perform their intended functions within the proposed system architecture. Tier 3 (TRL 6–7) focuses on system-level prototyping and validation, in which integrated robotic systems are demonstrated and evaluated as complete solutions. Finally, Tier 4 (TRL 8–9) addresses deployment in relevant environments, where systems with established performance through prior validation are implemented and assessed under jobsite-representative conditions.

Building on these tier definitions, this paper proposes a TRL assessment framework specifically tailored to intelligent construction robotics (Table 4). This framework utilizes two primary criteria: Test Environment and Model Performance. The Test Environment (rated 1–6) evaluates the physical context of deployment, ranging from low-fidelity simulations to real-world construction sites. Simultaneously, Model Performance (rated 1–3) assesses the accuracy of perception and reasoning. To ensure reliability, if a robot's performance falls below 60% for any given rating, the overall TRL is downgraded by one level.

The specific criteria for each factor are as follows:

- Test Environment: 6 (field deployment on real-world construction sites); 5 (construction-similar lab environments); 4 (simplified lab environments); 3 (high-fidelity simulations); 2 (low-fidelity simulations); and 1 (non-construction environments with no direct task applicability).
- Model Performance: 3 (near-perfect; such as >95% object identification accuracy); 2 (moderate; such as 80% accuracy in human state inference); and 1 (low; such as ~60% accuracy in object identification).

Table 4 TRL for intelligent construction robots

TRL Level	Definition (Operational Environment)
TRL 1	Basic principles and research: Scientific principles observed
TRL 2	Applied Research: Application formulated and potential identified
TRL 3	Proof of Concept: Critical function validated in separate elements
TRL 4	Lab Validation (Low Fidelity): Component validation in a laboratory environment, such as perception results in general datasets or trajectory planning in simplified simulation environment
TRL 5	Lab Validation (High Fidelity): Integrated components demonstrated in a laboratory environment, such as perception results in task-relevant dataset or motion planning in high-fidelity simulation or simulations with task-relevant elements
TRL 6	Prototype (Relevant Environment): Prototype demonstrated in a relevant environment, such as integrated intelligent system in high-fidelity and task-relevant simulations
TRL 7	Prototype (Operational Environment): Prototype demonstrated in an operational environment, such as embodying all elements in a robot to perform construction tasks in simplified

	environment
TRL 8	System Proven: Technology proven in an operational environment, such as embodying all elements in a robot to perform construction tasks in lab environment that has all construction task elements but less cluttered
TRL 9	Implementation: Technology refined and adopted

However, it should be noted that the authors also considered the performance of the demonstrations. For example, if a lab prototyping of intelligent robot was demonstrated (TRL7) but the performance does not show its capabilities to perform high-quality construction tasks, the authors will rate this as TRL 6.

3 Data from Literature Review

This paper reviews intelligent construction robots that integrate deep learning (DL), reinforcement learning (RL), and other AI methods to perceive site conditions, interpret human intent or task specifications, and execute goal-directed actions. In this review, "intelligent robots" refers to agents: software-based or physically embodied, that operate in a Perception–Reasoning–Action loop: they sense the environment, reason over observations and/or commands, and generate actions through control or actuation [18].

Using this inclusive definition, the review covers two categories that are essential for understanding methodological progress in the field. First, it includes robots without physical embodiment (e.g., simulation-based or purely computational agents). Because construction robotics hardware deployments were relatively sparse prior to 2020, a substantial portion of foundational research validated learning-based perception and control using simulated platforms; these studies are therefore included for their algorithmic contributions and their role in establishing DL-based robotic control pipelines. Second, the review includes automated construction equipment under a broad robotics lens—ground robots, robotic manipulators, unmanned aerial vehicles, and sensorized construction machinery such as excavators or other articulated equipment. When such systems are equipped with environmental sensing, DL-driven reasoning modules, and actuation/control capabilities, they satisfy the criteria of intelligent construction robots.

The literature search was conducted across three major scholarly repositories: Web of Science (WoS), the American Society of Civil Engineers (ASCE) Library, and Scopus. The search was executed on August 10, 2025 and limited to publications from 2017 through 2025 to emphasize the most recent wave of advances in DL-enabled construction robotics. Although neural-network-based work in construction robotics predates this window (e.g., studies appearing as early as 2017), the selected timeframe reflects the rapid acceleration in DL architectures, data-driven control, and foundation-model capabilities that have materially changed the state of practice. The review scope was restricted to peer-reviewed journal articles and conference proceedings written in English. Searches were performed over titles, abstracts, and author keywords using three keyword groups combined with Boolean "AND": (i) robot-related terms ("robot" OR "robotic"), (ii) the domain term ("construction"), and (iii) learning-related terms, with "deep learning" as the primary keyword supplemented by "neural networks," "reinforcement learning" (given its growing role in robot control and long-horizon decision-making), and "large language model" (included because it is DL-based and increasingly leveraged for robotic reasoning and human–robot interaction).

Studies were excluded if the search terms appeared incidentally but the content was not substantively related to construction robotics or DL/RL methods, and review papers were also removed to avoid duplicating secondary syntheses. After screening and applying these criteria, 134 publications were retained for analysis (Appendix I). The year-by-year distribution shows a sustained rise in output over 2017–2025 (Figure 1), indicating expanding research attention at the intersection of robotics, construction, and learning-based methods. In terms of publication venues, Automation in Construction emerged as the dominant journal outlet, followed by leading conference proceedings including *the Construction Research Congress (CRC) and the International Conference on Computing in Civil Engineering (IC3E). Additional high-frequency sources included the Journal of Computing in Civil Engineering and the Journal of Field Robotics*, underscoring both the construction-engineering and robotics communities' increasing engagement with DL-enabled construction automation.

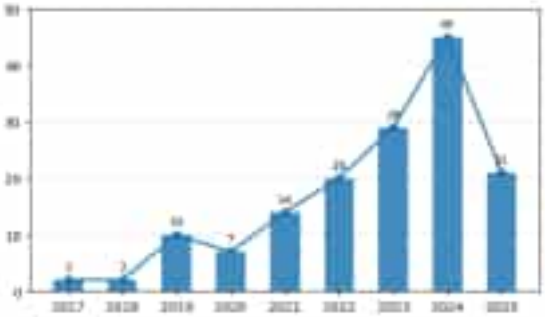


Figure 1. Publication trends and major publication venues.

4 Results

The TRL has largely improved from 2017 (averagely TRL3) to 2025 (averagely TRL 6.7). There are two studies testing the robotic products on construction sites and reached TRL 9. However, the majority does not show readiness to be directly deployed on construction sites. \

The authors also observed the large variation of the TRL in the prototypes or systems, all mean, median, and mode were used to show the patterns. The results are shown in Table 5.

To provide easier understanding of the trend, the visualization of the table is performed. The result is shown in Figure 2.

Table 5 Average TRL for intelligent construction robot studies 2017-2025

	Mean	Median	Mode
2017	3	3	3
2018	5	5	5
2019	5.625	5	5
2020	5.571	5	4.5
2021	6.25	5.5	5
2022	6	5	5
2023	6.444	6	6
2024	5.641	6	5
2025	6.667	7	8

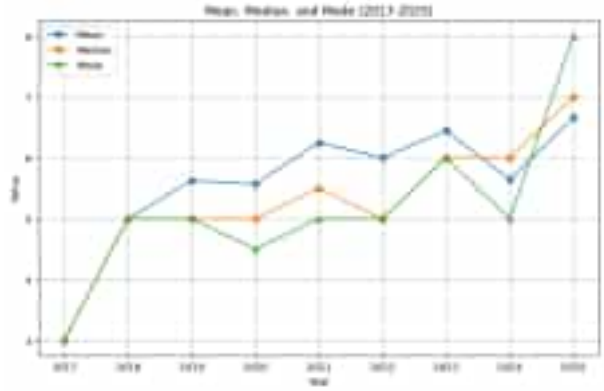


Figure 2. Average TRL for intelligent construction robot studies 2017-2025.

As can be seen in Figure 2, the TRL for intelligent construction robots illustrates a classic "step-up" maturation trajectory, characterized by a distinct phase of stabilization followed by a recent period of volatile growth. The jump from 2017 to 2018 established a solid baseline around a value of 5.0, where the technology or metric sat for several years with minimal deviation, suggesting a period of consistency or stagnation. However, beginning in 2021 and accelerating into 2023–2025, the data shows a breakout from this plateau. The divergence between the Mean, Median, and Mode in these later years indicates varying performance; specifically, the high Mode (8.0) in 2025 suggests that while the most common outcome is now high-performing, the lower Mean (6.667) indicates that significant low-performing outliers are still dragging the average down. This indicated the variation in the quality of research.

In addition, from the innovation perspective, the average TRL from 2020-2025 stabilized around 6. Innovation management to describe this as the difficulty of moving from TRL 4–6 (Research/Validation) to TRL 7–9 (Commercialization/Deployment) [19][20]. The "stagnation" seen in your data (flatline at 5.0 for years) is the textbook statistical signature of a project stuck in this valley. However, the majority of studies in 2025 has reached TRL 8, suggesting the completion of a full-scale robot embodiment in a construction-relevant lab environment.

In addition, the authors performed stratified analysis to further explore the reasons behind the TRL trend. The intelligent robots are categorized to: manipulators (robot arm and hand), soft robots, self-made robots (assembled from scratch), industrial equipment (modified excavator and cranes), combined robots (off-the-shelf solution modified), mobile robots, UAV, and 3D printing machines. The max and mode TRL were both used for each category, as shown in Figure 3.

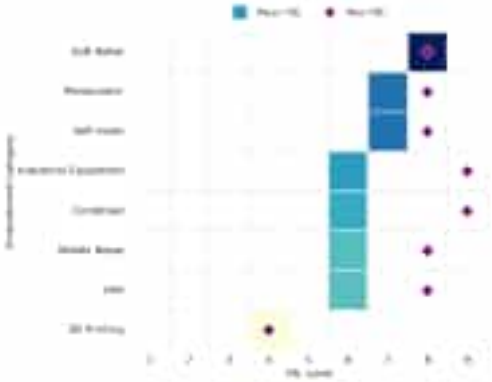


Figure 3. Stratified TRL analysis with eight robot types.

As illustrated in Figure 3, most construction robot categories now possess near-field-ready configurations. 3D printing remains the primary outlier with limited site deployability for intelligent systems. This is largely because 3D printing trajectories are typically pre-programmed to prevent structural defects and ensure material integrity, with relatively few studies successfully integrating adaptive, ML-based control. Conversely, both industrial equipment and combined robotic systems demonstrate viable prototypes for direct on-site deployment. It is important to note that while these specific categories reach high maturity, the high

volume of simulation-based research across the broader field pulls the aggregate mean TRL toward lower validation levels, highlighting how sample composition can mask the operational progress of specific sub-sectors.

5 Conclusions

This paper systematically reviewed 134 publications on intelligent construction robots and assessed their technical readiness for on-site deployment. Intelligent construction robots were defined broadly as software-based or physically embodied agents that implement a Perception–Reasoning–Action loop—sensing the environment, reasoning over observations and/or human commands, and executing actions through control or actuation [18]. Building on the classic Technology Readiness Level (TRL) framework [15], and informed by prior TRL adaptations in civil engineering research [16][17], this study developed a domain-appropriate, methodologically rigorous readiness assessment tailored to the realities of construction environments.

The readiness trends reveal a "step-up" maturation pattern rather than smooth linear progress. The TRL metrics show an extended period of stabilization around mid-level readiness (approximately TRL 5) through the late 2010s, followed by a clear breakout beginning around 2021 and accelerating through 2023–2025 (Figure 2). In the most recent years, the widening spread among central tendency measures (mean, median, and mode) indicates uneven progress across the literature: although the most common outcome has shifted to high maturity (mode near TRL 8 in 2025), the average remains noticeably lower, implying that a subset of studies still lag behind and pulls down overall performance. This dispersion suggests increasing variation in methodological rigor, validation quality, or the extent of real-world testing as the field scales.

Consistent with innovation management theory, the evidence also suggests that the field has faced a persistent transition barrier between research validation and deployment. Across 2020–2025, the average TRL hovering around ~6 aligns with the well-documented difficulty of moving from TRL 4–6 (research/validation) into TRL 7–9 (commercialization/deployment), often described as the "valley of death" in technology maturation [19][20]. In this context, the earlier multi-year plateau near TRL 5 can be interpreted as a statistical signature of technologies that demonstrate feasibility but struggle to secure the field access, integration pathways, safety assurance, and robustness needed for sustained deployment.

The readiness evaluation indicates that full site-ready maturity remains uncommon. Only two studies achieved TRL 9, meaning robotic systems were validated in real construction settings under varied conditions sufficient to represent real-world variability. Most of the remaining work clustered below that threshold, reflecting persistent barriers such as unstructured environments, safety constraints, limited access to field trials, and challenges in generalization across tasks, sites, and work practices. At the same time, the evidence points to rapid progress: by 2025, the dominant maturity level shifted to TRL 8, indicating that many systems are approaching deployability through late-stage validation and near-field testing, even if they have not yet demonstrated sustained performance across diverse live jobsite conditions.

Overall, the findings suggest the field has moved from early-stage feasibility demonstrations around 2017 (typically TRL ~3) to near-deployment prototypes within less than a decade, underscoring the accelerating impact of deep learning, reinforcement learning, and AI-enabled perception and planning. The primary gap is no longer proof-of-concept capability, but robust translation to the field—especially generalization, safety assurance, and integration into work-zone operations and workflows. Future research should therefore prioritize multi-site, multi-condition field evaluations, standardized benchmarking protocols, and human-centered deployment procedures that ensure reliability and safety while delivering measurable gains in productivity, constructability, and reduced disruption to construction operations.

Overall, to address the lack of a standardized technical maturity baseline in current literature, this study employs the TRL framework as a vital "first-pass" screening tool. By establishing this technical baseline, we provide a curated starting point for the research community and industrial decision-makers to identify viable candidates for adoption. However, we acknowledge that TRL is a high-level metric with inherent limitations; it does not, on its own, guarantee safety, generalization, or robustness across diverse construction environments, which are all important operational factors to consider for a technology adoption decision. In this paper, the authors position TRL as a preliminary filter that allows stakeholders to judiciously apply resource-intensive evaluations: such as safety audits and cross-environment generalization tests, to a more targeted subset of technologies. This tiered framework bridges the gap between laboratory-based technical maturity and field-proven technical maturity.

References

[1] Associated Builders and Contractors. Title of the web content. On-line: https://www.abc.org/News-Media/News-Releases/entryid/20005/abc-construction-industry-to-need-more-than-500-000-additional-workers-to-meet-demand-in-2023, Accessed: 30/01/2026.

[2] Associated General Contractors of America. *2022 Construction Industry Workforce Survey*. On-line: https://www.agc.org/sites/default/files/users/user21902/2022_Workforce_Survey_National.pdf, Accessed: 30/01/2026.
[3] Brosque, C. and Fischer, M. Safety, quality, schedule, and cost impacts of ten construction robots. *Construction Robotics*, 6:163–186, 2022.
[4] Park, S., Yu, H., Menassa, C. C. and Kamat, V. R. A comprehensive evaluation of factors influencing acceptance of robotic assistants in field construction work. *Journal of Management in Engineering*, 39(3):04023010, 2023.
[5] Yu, H., Kamat, V. R. and Menassa, C. C. Generalizable skill learning for construction robots with crowdsourced natural language instructions, composable skills standardization, and a large language model. *ASCE OPEN: Multidisciplinary Journal of Civil Engineering*, 3(1):04025014, 2025.
[6] Linner, T. and Bock, T. Evolution of robotics in construction. In *Proceedings of the 29th International Symposium on Automation and Robotics in Construction (ISARC 2012)*, 2012.
[7] Liang, C. J., Kamat, V. R. and Menassa, C. C. Teaching robots to perform quasi-repetitive construction tasks through human demonstration. *Automation in Construction*, 120:103370, 2020.
[8] Yu, H., Kamat, V. R. and Menassa, C. C. Cloud-based hierarchical imitation learning for scalable transfer of construction skills from human workers to assisting robots. *Journal of Computing in Civil Engineering*, 38(4):04024019, 2024.
[9] Huang, L., Zhu, Z. and Zou, Z. To imitate or not to imitate: Boosting reinforcement learning-based construction robotic control for long-horizon tasks using virtual demonstrations. *Automation in Construction*, 146:104691, 2023.
[10] Park, S., Wang, X., Menassa, C. C., Kamat, V. R. and Chai, J. Y. Natural language instructions for intuitive human interaction with robotic assistants in field construction work. *Automation in Construction*, 161:105345, 2024.
[11] Yang, L., Li, B., Feng, J., Yang, G., Chang, Y., Jiang, B. and Xiao, J. Automated wall-climbing robot for concrete construction inspection. *Journal of Field Robotics*, 40(1):110–129, 2023.
[12] Lee, D. and Han, K. Vision-based construction robot for real-time automated welding with human-robot interaction. *Automation in Construction*, 168:105782, 2024.
[13] Yu, H., Kamat, V. R., Menassa, C. C., McGee, W., Guo, Y. and Lee, H. Mutual physical state-aware object handover in full-contact collaborative human-robot construction work. *Automation in Construction*, 150:104829, 2023.
[14] Amani, M. and Akhavian, R. Intelligent ergonomic optimization in bimanual worker-robot interaction: A reinforcement learning approach. *Automation in Construction*, 168:105741, 2024.
[15] Mankins, J. C. Technology readiness levels. 1995.
[16] Towery, N. D., Machek, E. and Thomas, A. *Technology Readiness Level Guidebook* (No. FHWA-HRT-17-047). 2017.
[17] Petrescu, T. C., Voordijk, J. T. and Mihai, P. Developing a TRL-oriented roadmap for the adoption of biocomposite materials in the construction industry. *Frontiers of Engineering Management*, 10(2):223–236, 2023.
[18] Beni, G. and Wang, J. Swarm intelligence in cellular robotic systems. In *Robots and Biological Systems: Towards a New Bionics?*, pages 703–712, Berlin, Heidelberg, 1993.
[19] Markham, S. K., et al. The Valley of Death as a context for role theory in product innovation. *Journal of Product Innovation Management*, 2010.
[20] Frank, C., et al. Surviving the 'Valley of Death': A comparative analysis. *Journal of Technology Transfer*, 1996.

Digital Games for Technology and Sustainability in the Human-centred Built Environment: a literature review

Ming Shan Ng[1], Jack Tam[1], Clara Cheung[2], Tsukasa Ishizawa[3], Yifan Xu[2] and Akilu Yunusa-Kaltungo[4]

[1]Center for the Possible Futures, Kyoto Institute of Technology, Japan
[2]Department of Civil Engineering and Management, The University of Manchester, United Kingdom
[3]Institute of Industrial Science, The University of Tokyo, Japan
[4]Department of Mechanical and Aerospace Engineering, The University of Manchester, United Kingdom

ng@kit.ac.jp, tam@kit.ac.jp, clara.cheung@manchester.ac.uk, ishizawa@iis.u-tokyo.ac.jp, yifan.xu@manchester.ac.uk, akilu.kaltungo@manchester.ac.uk

Abstract -

The AECO sector increasingly calls for digital transformation to support sustainable development and a more human-centred built environment. In this context, the Society 5.0 dimensions, which emphasise *human-centred values*, *cyber-physical intelligence*, *socio-technical embeddedness*, *societal challenge orientation* and *system-level impact potential*, provide a lens for examining how digital technologies are conceived and deployed in the sector. Despite rapid technological advancement, challenges remain in operationalising these principles within complex socio-technical systems. Gamification has been explored as a potential response, yet its contribution to Society 5.0-oriented transformation remains fragmented. This study conducts a systematic literature review of 98 publications from the past five years (2021-2025), each presenting a game or gamified application in the sector. The review analyses their correspondence to the Society 5.0 dimensions. The results indicate that many existing games are not primarily designed for engagement or enjoyment. Research-oriented games are predominantly targeted at students rather than industry practitioners or the general public. Most applications operate at the building scale and focus on the operation phase of the lifecycle. Furthermore, a heatmap analysis reveals systematic patterns: Amongst the minority that are engagement-oriented games, these tend to strengthen *human-centred values* and *socio-technical embeddedness*; decision-making games are more closely associated with *system-level impact*. Notably, no reviewed games address the end-of-life stage of buildings, while scenario-based interactive games often align with individual-level human values. The study concludes by identifying key gaps and outlining future gamification requirements to better align digital innovation in the AECO sector with the human-centred and system-oriented aspirations of Society 5.0.

Keywords -

Digital Games; Gamification; Society 5.0; Technology; Sustainability; Literature Review

1 Introduction

1.1 Society 5.0

Industry 5.0 has been emerging globally to rethink the role of humans and place humans at the centre of industrial development through sustainability, resilience and human-centredness. This aligns with a new overarching vision for society - Society 5.0 - that envisions a human-centred physical and cyber space with advanced technologies including artificial intelligence (AI), information and communication technologies (ICT) that enable sustainable development in environment, economy and society. Ageing societies, inclusiveness and equality are considered through technological development. In Europe, policy-makers and scholars are exploring a neighbouring concept to envision decentralised autonomous cities and human digital twins [1]. In Japan, the government has been pushing a human-centred system in Society 5.0 that includes the transdisciplinary concepts of digitalisation, integration and sustainability [2]

Nevertheless, there are clear challenges in Society 5.0 regarding humans, technologies and values in a sustainable society in the physical space. These multi-dimensional challenges include how to achieve *human-centred values*, such as health and wellbeing, safety and inclusiveness in society, as well as unknowns in *societal challenge orientations* to different extents. The interactions between humans, humans and technologies, physical and cyber space and amongst each other as complex systems within systems entail *socio-technical embeddedness* Also, the complexity within physical and cyber spaces with AI and intelligence-based technologies entails new forms of *cyber-physical intelligence*. In general, a system-level rethinking of the potential impacts in the physical world is necessary to adapt and optimise these challenges to foster sustainable development in Society 5.0 through digital transformation, which refers to an inter-connected set of discoveries and inventions [3].

1.2 Construction 5.0

The World Economic Forum reveals that the architecture, engineering, construction, and operations (AECO) sector is responsible for 13% of the global gross domestic product (GDP), employs more than 100 million people, and accounts for 79% of total greenhouse gas emissions and 88% of climate adaptation costs. However, the sector suffers from low digitalisation and innovation. It is one of the least digitised in the world and requires a systematic approach to digital transformation to foster a sustainable development. The industry embraces a new movement of Construction 5.0, in reference to Industry 5.0 and Society 5.0 in AECO, to drive digital transformation through digitalisation, integration and sustainability to different extents [4, 5].

In correspondence to the five dimensions of Society 5.0, the AECO sector has been trying to achieve *human-centred values* to improve human wellbeing such as worker health and safety, as well as continuous improvement in knowledge and skills [6, 7]. The industry has adopted automa-

tion and robotics in construction to address broader social challenges such as ageing society and labour shortages [8, 9]. In addition, recent research reveals the complex ontology networks and interactions between the different human roles, resources, conditions, artifacts, processes, attributes, values and risks in adopting automation technologies to achieve sustainability in AECO. This shows that *socio-technical embeddedness* is significant and requires more integrated management for Construction 5.0 [10]. Recent studies have also examined real-time digital twins for automation and robotics in construction management [11] *Cyber-physical intelligence* enables human-centred controls in, for example, human-robot collaborations [12].

1.3 Gamification

Given the complexity of the AECO sector, gamification has been one approach to encourage and assist the industry to comprehend and stimulate strategies through experiential learning to different extents. Suits [13] defines a game as 'the voluntary attempt to overcome unnecessary obstacles'. He suggested that a game possesses three components of prelusory goal, constitutive rules and lusory attitude. The term gamification was coined by Pelling in 2002 and hit the mainstream around 2010 [14]. The elements of a gamified system can be categorised into three groups [15]: (1) *dynamics* defines the big-picture aspect of the game and includes elements such as constraints, narratives, progression and relationships; (2) *mechanics* defines the processes that drive actions forward and includes elements such as challenges, chance, competition, cooperation, feedback, resource acquisition, rewards, transactions, turns and win states; and (3) *components* specify the corresponding instantiations of *dynamics* and *mechanics* such as quests and points [16]. Garneau [17] identifies fourteen forms of fun in games, ranging from beauty and immersion to competition, discovery and social interaction, while Heeter et al. [18] shows these more prevalent in commercial than educational games.

1.4 Gamification in AECO Research

Recently, researchers have been exploring gamification approaches for research and education on technology and sustainability to different extents. Yanai and Ishikawa [19] developed an immersive environment in VR to investigate the learning experience at Osaka Castle. Ng and Chen [7] developed a teaching game using the online platform *Slido* to investigate students and practitioners' perceptions of lean terminology. Tan et al. [12] reviewed the literature on augmented reality (AR) for human-robot interactions to investigate innovations in pedagogy. Ng and Hall [20] developed an online game to allow players to experience a round of traditional design-bid-build design process and a second round of lean-based Target Value Design for digital fabrication. Kastner et al. [21] designed an online game in 3D BIM to let players learn about lifecycle thinking for renovating a historic building in Switzerland. Harichandran et al. [22] presents a virtual reality (VR) game with digital twins to train workers about construction safety.

Some scholars had reviewed existing academic games or gamification in AECO. For example, Ilbeigi et al. [16] conducted a scoping review on Scopus of 103 game-based educational solutions for construction engineering in the past two decades. They differentiated the basic information of the article, such as titles, and analysed several factors, such as whether each study developed a novel gamified solution based on a learning theory and specific its specific area of domain. Bhatnagar et al. [23] reviewed 96 lean simulation games from a selection of 52 articles. Articles were analysed for basic information, such as authorship, using social-networks analysis. They found that the International Group for Lean Construction (IGLC) developed a large number of lean games in AECO. They show that a few authors collaborate extensively, while others in the periphery are beginning to publish and collaborate. Oke et al. [24] identified five challenges to the adoption of gamification for sustainable construction, including hesitation in adopting and lack of budgeting for innovation, with a questionnaire survey.

Nevertheless, no recent articles have reviewed academic games, in particular, about technology and sustainability in AECO.

1.5 Research Objectives

To help researchers and industry think about technology and sustainability in Society 5.0 and Construction 5.0, this work presents a systematic review of the literature of publications from the past five years to investigate the recent trends of digital games in AECO and their relevance to the promotion of technology and sustainability. This research possesses two overarching research questions of **(RQ1)** what are the recent trends in game development in AECO? and **(RQ2)** what can be the foci when researchers develop games for AECO in Society 5.0? In this regard, this work correlates gamification to the five dimensions in Society 5.0 of *human-centred value*, *socio-technical embeddedness*, *cyber-physical intelligence*, *societal challenge orientation* and *system-level impact potential*. The research objectives of this work include **(OB1)** identifying recent trends and features of recent game development in AECO through a systematic review of the literature; **(OB2)** to analyse the correlations between the indicated features and the five dimensions in Society 5.0; **(OB3)** to elaborate on potential topics in game development to foster Society 5.0 in AECO. This article summarises the preliminary of this systematic literature review and a further in-depth analysis will be presented in future publications.

The structure of this article is organised as follows: Section 2 describes the systematic literature review methodology; Section 3 presents the results of recent trends and features; Section 4 discusses the co-occurrence pattern analysis with Society 5.0 dimensions; and Section 5 concludes with limitations and future research directions.

2 Research Methodology

To conduct a systematic literature review, the first author of this work ran a first-round Scopus search with the keywords of keyword group 1, which comprises "gamification" or "digital games", and keyword group 2, which comprises "construction" or "building" or "architecture". The search includes journal articles and conference proceedings published in the past five years (2021-2025) in the fields of AECO only. This resulted in 1279 articles. The first author then screened the abstract of every article to eliminate those that do not include an original development of gamified solutions, those about research of game theory or game modelling, those that gamify a survey or other types of research but not a game development, those

Table 1. The basic information of the selected 98 articles in this work. "J" refers to a journal article; "C" refers to a conference proceeding.

Authors	J/C	Ref	Authors	J/C	Ref
2025 (25 literature)					
Boyacioglu et al.	J	[25]	Cao et al.	J	[26]
Emiroglu et al.	J	[27]	Hao et al.	C	[28]
Hardy et al.	C	[29]	Hu et al.	J	[30]
Hung et al.	J	[31]	Jacob et al.	C	[32]
Jang et al.	J	[33]	Jiang et al.	J	[34]
Kastner et al.	C	[21]	Kilis et al.	J	[35]
Ng & Chen	C	[7]	Ng et al.	C	[36]
Pappas et al.	C	[37]	de Sena et al.	J	[38]
Taha et al.	J	[39]	Turk & Bozkurt	J	[40]
Voinea	C	[41]	Wang et al.	J	[42]
Wang et al.	J	[43]	Yang et al.	J	[44]
Yanai et al.	C	[19]	Zhang et al.	J	[45]
Zielke et al.	C	[46]			
2024 (19 literature)					
Arbesser-Rastburg et al.	C	[47]	Arslan et al.	J	[48]
Caliskan et al.	C	[49]	Carrasco-Beltran et al.	J	[50]
Chen et al.	J	[51]	Elenany et al.	J	[52]
Fawad et al.	C	[53]	Fawad et al.	J	[54]
Ilbeigi et al.	J	[55]	Joy et al.	J	[56]
Kolokotsa et al.	J	[57]	Park et al.	J	[58]
Pickersgill et al.	C	[59]	Sa Alves et al.	J	[60]
Speiser et al.	C	[61]	Stevens et al.	C	[62]
Tay et al.	J	[63]	Vandenabeele et al.	J	[64]
Xiao et al.	J	[65]			
2023 (20 literature)					
Alexander et al.	C	[66]	Alp et al.	J	[67]
Angheluta et al.	J	[68]	Arefazar et al.	C	[69]
Bhat et al.	C	[70]	Borucka et al.	J	[71]
Cook et al.	J	[72]	Diaz-Aleman et al.	J	[73]
Ehab et al.	J	[74]	Ernstsen et al.	J	[75]
Jungmann et al.	C	[76]	Kalak et al.	C	[77]
Li et al.	J	[78]	Martinez et al.	J	[79]
Ospina-Bohorquez et al.	J	[80]	Sorensen et al.	J	[81]
Tabi et al.	J	[82]	Villamil et al.	J	[83]
Wang et al.	C	[84]	Speiser et al.	C	[85]
2022 (13 literature)					
Alves et al.	C	[86]	Anschutz et al.	C	[87]
Argiolas et al.	J	[88]	Avendano-Uribe et al.	J	[89]
Comai et al.	C	[90]	Getuli et al.	J	[91]
Goli et al.	J	[92]	Jiang et al.	J	[93]
Kassi et al.	C	[94]	Lee et al.	J	[95]
Potseluyko et al.	J	[96]	Shareef et al.	J	[97]
Vollmer et al.	J	[98]			
2021 (21 literature)					
Biotto et al.	C	[99]	Chen et al.	J	[100]
Ezzeddine et al.	J	[101]	Filippova et al.	C	[102]
Gangolells et al.	J	[103]	Goubran et al.	J	[104]
Gutierrez-Rodríguez et al.	C	[105]	Harichandran et al.	C	[22]
Heins et al.	C	[106]	Huang et al.	C	[107]
Jacobsen et al.	C	[108]	Kan et al.	C	[109]
Koutsabasis et al.	C	[110]	Martín-Hernández et al.	J	[111]
Meuser et al.	C	[112]	Ng & Hall.	C	[20]
Putz et al.	C	[113]	Riedlinger et al.	C	[114]
Schnieder et al.	C	[115]	Walls	J	[116]
Yan et al.	C	[117]			

that do not involve any digital tools, as well as those that review existing commercial games or those review literature about games. In case of those presenting the same game in multiple articles, only the most relevant article was selected. This resulted in 98 relevant articles as shown in Table 1.

To answer RQ1 and achieve the objective OB1, all authors reviewed each article in detail to identify the features of each of the 98 games from the selected literature. The features are identified under six categories of *target player*, *scale*, a *building lifecycle stage*, the *research goal*, any focused *digital features* significantly involved, as well as any *Society 5.0 concepts* involved [2, 4]. Moreover, this work includes an evaluation of the five dimensions presented in Section 1. The results of this analysis are explained in Section 3. Furthermore, this work analyses the correlations amongst the features through heatmap visualisation. The results answer RQ2 and achieve the objectives OB2 and OB3 are discussed in Section 4.

Each article was independently coded by at least two authors across six feature categories (target player, scale, lifecycle stage, research goal, digital features, Society 5.0 concepts) using binary present/absent labels. Multi-label coding was permitted, and disagreements were resolved through discussion or third-author adjudication. The heatmap in Figure 3 shows conditional co-occurrence percentages: each cell = (articles tagged with both the row and column features) /(articles with the row feature) × 100, rounded to the nearest integer. For example, 16 of the 23 'Cultural Preservation' games involve the 'Building' scale, yielding approximately 70%. These values represent co-occurrence patterns, not statistical correlations; normalisation is row-based, and multi-label articles contribute to all applicable categories.

3 Recent Trends and Features

This section presents the results of the identified features in the seven categories for the analysis.

Target Player: Based on systematic review of the literature, recent trends in digital games show that 45% are designed to target AECO students at universities, while 40% target professionals in the sector. For example, Arbesser-Rastburg et al. [47] developed an escape room game in VR to teach AECO students about building energy. In addition, 28% are targeted at visitors or end-users of the facilities involved in those games. For example, Yanai and Ishikawa [19]'s VR experiment aims to improve visitors' experience when visiting Osaka Castle in Japan. 6% target on general households. For example, Hao et al. [28] uses a game approach to help the decision-making of general households and governors, respectively, about the management of solid waste. 5% target kids and/or teenagers. For example, Emiroglu et al. [27] used minecraft to teach primary and secondary K-12 students about flood mitigation. 4% target on governors or policymakers; while only 1% target on elderly. For example, Hung et al. [31] gamified a cycling exercise for residents in long-term care homes.

Scale: The recent trends show eleven scales in game development. 41% of the games developed in the past five years are games with buildings. For example, Kastner et al. [21] developed an online 3D building game of the existing Hotel Schatzalp in Davos, Switzerland, for a case study of lifecycle thinking for architecture renovation. 28% present games with a project scenario. For example, Ng and Hall [20] developed a game to make a salad plate that mimics van Gogh's The Starry Night painting as a design project for students and professionals to learn about the lean-based Target Value Design approach for digital fabrication. 20% of the games developed comprise a campus, a site or a landscape in the gaming environment. For example, Turk and Bozkurt [40] presents an AI-based board game to teach AECO students about landscape architecture. 18% are about an individual person or a role. For example, Ng and Chen [7] developed a game for everyone to learn their own perception of lean terminology from Japan. 18% gamified user experience and 16% gamified urban and cityscape; while 14% games are about infrastructure assets such as bridges. 5% are community-based games. For example, Wang et al. [43] presents a community game in a smart green village in China. 2% of the gamifications are specifically about climate such as wind and furniture, respectively, such as the gamified bike in [31]. Only 1% gamified softscape such as audio.

Lifecycle Stage: This work also presents which of the commonly known six lifecycle stages are more frequently presented in a game. 38% and 27% of the games reviewed are developed for the operational stage and the construction stage, respectively. For example, Boyacioglu et al. [25] presents a game to address the delays of construction projects. 19% developed games for the design stage, while 12% developed games for the pre-design planning and initial design stage. 9% games are designed for urban

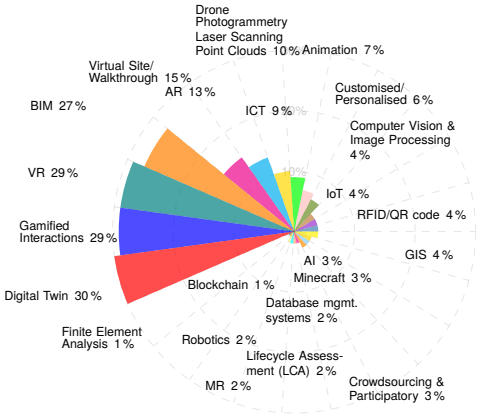


Figure 1. Recent 23 trends of digital features in digital games in AECO.

planning. Nevertheless, no game has been developed in the last five years for the end-of-life stage (afterlife) of a building or infrastructure.

Research Goal: To assist researchers and game developers understand how games could be developed in the future, this work also reveals that there are 18 research goals that have been addressed in the past five years. 62% of the games were developed for educational and communication purposes. Surprisingly, only 37% of the 'games' were gamified for people to play. Many of the developments in the reviewed literature do not aim for fun. In this regard, a further discussion is provided in Section 4. 26% and 20% of the games aim to assist decision-making and improve efficiency, respectively, such as reducing project delays in Boyacioglu et al. [25]. 19% aim for co-creation and collaboration amongst players. For example, de Sena et al. [38] developed a participatory planning system to collect spatial data from minecraft and improve collective governance in urban planning for co-creation. 18% of the games were developed to explore cyber-physical interaction and resource management, respectively. 17% were developed to provide information to players for knowledge sharing. 10% aim to improve site management and safety training, such as Harichandran et al. [22], respectively. 9% gamified, for example, on the urban scale, to mitigate safety and resilience. 7% aim for collective governance; while 7% promote modular construction and 6% explore new processes or innovation in the built environment. For example, Jang et al. [33] presents a gamification of crane lifting tasks in digital twins to improve a prefabricated bridge deck assembly process. The salad game in Ng and Hall [20] facilitates integrated design processes to stimulate systemic innovation. 5% aim to explore acceptability amongst players. For example, Hung et al. [31] and Ng and Chen [7] used gamification to explore people's perceptions. 5% were designed to perform seismic damage assessment, while 3% were aimed at boosting business-or-finance-related activities. Surprisingly, only 3% promote a clean and green built environment.

Digital Features: Furthermore, in response to emerging needs in digital transformation in AECO, this work reveals the trends of 23 digital and technological features in recent digital game development. Based on the literature review, 30% of the gamification research developed a digital twin such as Jang et al. [33]; while 29% comprise gamified interactions such as Hardy et al. [29]. 29% use VR, 13% use AR and 2% use Mixed Reality (MR) features, respectively. 27% include BIM such as Kastner et al. [21]. 15% include virtual site or virtual walkthrough such as Zhang et al. [45], Only 6% of the games include customised or personalised and only 2% are games about robotics. For example, Hu et al. [30] developed a customised experience in the archaeological site museum through gamification.

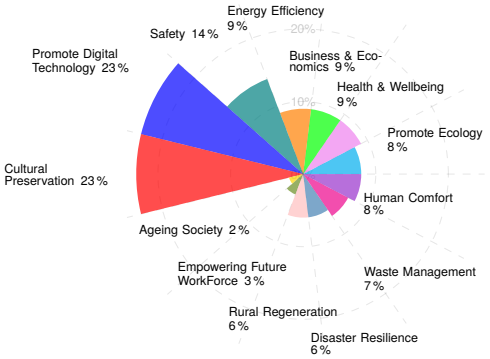


Figure 2. Recent trends in games relevant to the concepts in Society 5.0 .

Society 5.0 Concepts: To answer RQ2, this work identifies 13 Society 5.0 concepts that can be found in recent game developments based on the reviewed literature. Amongst all, 23% of the games have been developed for cultural preservation and promoting digital technologies, respectively. Games about safety in AECO are also relatively popular, such as Harichandran et al. [22]. Energy efficiency, health and wellbeing, and human comfort can also be found in around 8-9% of the publications reviewed. Games that simulate business related activities can be found in 9% of the games. For example Zielke et al. [46] developed a serious digital game about lean-based logistics to improve efficiency and therefore cost effectiveness. Very few games explore future workforce and ageing society respectively.

Five Dimensions in Society 5.0: To achieve OB2, this work further investigates a more detailed overview of the five dimensions in Society 5.0 as described in Section 1. The differentiation of the recent games are relatively well distributed across these dimensions. 33% of the games explore *human-centred values*; 28% explore *cyber-physical intelligence* through, for example, digital twins in the game development such as Voinea [41]. 27% delve into *socio-technical embeddedness*, while 23% look at *societal challenge orientation*, respectively. 20% explore *system-level impact potential*.

4 Discussions about Games for Society 5.0

The analysis of heatmap-based co-occurrence patterns in Figure 3 provides a systematic overview of how current

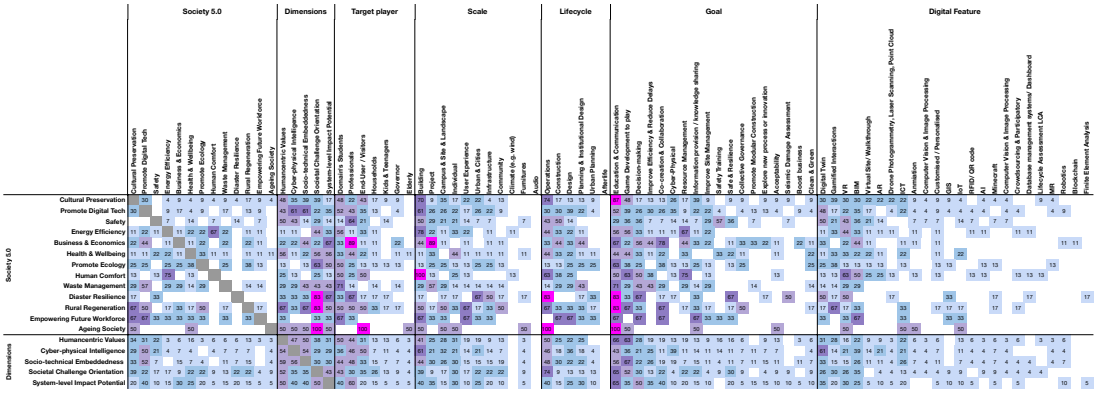


Figure 3. Heatmap of row-normalised conditional co-occurrence percentages across game features, Society 5.0 concepts and the five dimensions. Each cell shows the proportion of articles with the row feature that also exhibit the column feature, rounded to the nearest integer. The number in each cell refers to the percentage of the correlations. For example, amongst all 23 games about cultural preservation, 16 of them, i.e. rounded-up to 70%, involves gamification of buildings.

digital game-based developments in the AECO sector align with the broader aspirations of Society 5.0. While certain themes such as safety training, cultural preservation, and digital technology promotion are well represented, other critical dimensions remain underexplored. This reveals both the strengths and the limitations of current gamification practices in supporting a more human-centred and system-oriented built environment. This work achieves OB3 to discuss potential game developments to foster Society 5.0 in AECO.

4.1 Dominant Clusters and Underexplored Needs

Amongst the most prominent trends, safety-related games exhibit strong integration across multiple lifecycle stages, particularly construction and operations. These games frequently use digital features such as gamified interactions and immersive environments, with the primary goal of supporting safety training and hazard awareness. This is reflected in the heatmap, where the 'Safety' row shows relatively high co-occurrence values for the 'Construction' lifecycle stage (50%), as well as 'Gamified Interactions' (21%) and 'VR' (43%) amongst digital features. Cultural preservation is another area of significant focus, with many games using VR, digital twins and walkthroughs to enhance user experience in heritage contexts. Similarly, promoting digital technology is a recurring theme, often linked to BIM, gamified learning and educational communication. These clusters suggest a tendency to deploy games in domains where the benefits of technology can be clearly demonstrated, risks are manageable and regulatory constraints are lower.

Nevertheless, despite the emphasis on human-centredness in Society 5.0, certain societal needs are noticeably absent. The concept of ageing society, for instance, is scarcely addressed in the reviewed games. Very few applications are designed with elderly users, children or general households in mind. Most games target students and professionals, thereby limiting their societal reach. Similarly, lifecycle stages such as the end-of-life stage of buildings, covering topics such as *system-level impact* and *socio-technical embeddedness* in deconstruction, reuse and circular design, receive almost no attention, despite their importance for sustainability and long-term planning. These gaps point to a missed opportunity for games to contribute to inclusive design, intergenerational equity and circular economy transitions. Several factors may explain these gaps: recruiting vulnerable populations (elderly, children, non-expert households) poses ethical and methodological challenges unfamiliar to technically oriented AECO teams; the end-of-life stage lacks commercial incentive relative to design or operations; and disciplinary silos between AECO and social science or gerontology limit cross-pollination of methods. To address this, cross-disciplinary co-design workshops involving diverse end-users, partnerships with care facilities and community organisations, as well as policy-driven agendas linking circular economy mandates to gamification could broaden both target audiences and thematic scope.

4.2 Mechanics and Embeddedness

The analysis also highlights the importance of game goals and mechanics in shaping their alignment with Society 5.0 dimensions. While a majority of the games are developed for educational and communicative purposes, only a minority are designed to be engaging or enjoyable in a conventional sense. This suggests that many academic games prioritise knowledge transmission over player experience, which may hinder their impact. Interestingly, games developed for decision-making and co-creation tend to correlate more strongly with dimensions like *system-level impact*. These findings imply that the transformative potential of games lies not only in their content or digital features, but also in how they frame interaction, collaboration and agency. The heatmap supports this: 'System-level Impact Potential' co-occurs more strongly with 'Professionals' (60%) than 'Domain's Students' (40%) as target players, and shows relatively strong co-occurrence with the Society 5.0 concept of 'Business & Economics' (30%) compared to other concepts.

Socio-technical embeddedness, in particular, emerges as

a critical but underutilised dimension. It is most effectively realised in games that involve collective governance, participatory planning and stakeholder negotiation, especially at urban, community or site scales. However, many games still treat social contexts, institutional structures and power dynamics as external conditions rather than integral parts of gameplay. Even when advanced technologies such as VR or BIM are used, their contribution to embeddedness is limited unless supported by mechanics that model real-world complexity, such as role-based decision-making or scenario-based negotiation.

To support the broader transformation envisioned in Society 5.0 and Construction 5.0, future game development must go beyond technology demonstrations or visual simulations. There is a clear need for games that support participatory and experiential learning, that model trade-offs and systemic consequences, and that engage a broader spectrum of users beyond students and professionals. Addressing the end-of-life stage of buildings, designing for inclusivity and accessibility, and embedding socio-technical complexity into gameplay will be essential steps. Only by expanding the scope, depth, and diversity of game-based applications can the AECO sector fully leverage gamification as a tool for human-centred, sustainable, and system-level innovation.

4.3 Limitations and Future Research Directions

This review is restricted to Scopus-indexed peer-reviewed publications from 2021-2025. Industry-developed games, commercial applications and grey literature are excluded. The findings may under-represent applied solutions that influence practice but fall outside the academic literature. This study has several limitations that should be considered when interpreting the findings. First, the analysis is based on thematic co-occurrence and correlation patterns derived from a systematic review. While this approach enables a structured comparison across a large body of literature, it does not assess the depth, quality or effectiveness of individual game implementations. Moreover, recent AECO game developments rarely engage explicitly with the fourteen forms of fun. As such, the results indicate patterns of emphasis rather than evidence of impact or causal relationships between game design features and societal outcomes. Second, the classification of game features, Society 5.0 concepts, and the five dimensions necessarily involves a degree of interpretive judgment. Although a systematic and consistent coding framework was applied, some dimensions, particularly *socio-technical embeddedness* and *system-level impact potential*, are conceptually broad and may be operationalised differently across studies. The resulting heatmap therefore reflects relative emphases rather than definitive or exhaustive categorisations. Finally, each game was treated as a single unit of analysis, without accounting for differences in maturity, scale of deployment, duration of use or institutional embedding. Early-stage prototypes and long-running platforms are weighted equally, which may obscure important variations in adoption, learning outcomes or long-term influence.

Although this study provides a structured overview of how digital games relate to Society 5.0 concepts in the built environment, the analysis is primarily based on thematic correlations and does not evaluate the depth or effectiveness of individual implementations.

Future studies should move beyond the current concentration on technology demonstration, safety training and cultural preservation. They should explore how games can support more complex social challenges. In particular, there is a strong need to address under-represented themes such as ageing society and inclusive participation, by designing game-based applications for elderly users, households and non-expert stakeholders. Sustainability oriented games should also evolve from technically focused simulations toward participatory and experiential formats that allow players to engage with environmental trade-offs, long-term impacts, and collective responsibility. Moreover, the near absence of afterlife stage applications highlights an important opportunity for future research to develop games that address deconstruction, reuse, and circular economy strategies. Finally, expanding game-based approaches towards decision-making, governance and system-level exploration may enhance their potential to support the broader socio-technical transformation envisioned in Construction 5.0 and Society 5.0.

5 Conclusions

This paper presented a systematic review of 98 digital games in the AECO sector, analysing their alignment with the five dimensions of Society 5.0. The results reveal that while current games often focus on safety, cultural heritage and digital technology promotion using tools such as VR, BIM and digital twins, critical areas like ageing society, inclusive participation and the afterlife stage of buildings remain largely overlooked. Most games are developed for students and professionals, with limited engagement of diverse or vulnerable user groups. Moreover, while many games employ advanced digital features, few fully support *socio-technical embeddedness* or *system-level impact*, often lacking participatory or collaborative mechanics. To better support the goals of Society 5.0 and Construction 5.0, future gamification efforts should go beyond education and demonstration. There is a clear opportunity to design more inclusive, participatory, and systems-oriented games that engage broader audiences and address complex societal challenges in the built environment.

References

[1] C. Fontes, D. Carpentras, and S. Mahajan. Human digital twins unlocking society 5.0? approaches, emerging risks and disruptions. *Ethics and Info. Tech.*, 2024.

[2] JST (Japan Science and Technology). Science, technology and innovation for sdgs japanese actions in sti for sdgs. *Technical Report*, 2021.

[3] The World in 2050. Transformations to achieve the sustainable development goals. *Technical report*, IIASA, Laxenburg, Austria, 2018., 2018.

[4] M.S. Ng, T. Bock, F. Kastner, and S. Langenberg. Digital transformation, sustainability and construction 5.0. In *Proc. ISARC 2024*, pages 1222–1229, 2024.

[5] M. Skibniewski. The present and future of smart construction technologies. *Engineering*, 44:21–23, 2025.

[6] A. Babalola, P. Manu, C. Cheung, A. Yunusa-Kaltungo, and P. Bartolo. A systematic review of the application of immersive technologies for safety and health management in the construction sector. *J. Safety Research*, 85, 2023.

[7] M.S. Ng and Q. Chen. Perceptions of lean concepts from japan: using an interactive game to teach lean terminology. In *Proc. IGLC33*, 2025.
[8] T. Bock and T. Linner. *Robot-Oriented Design*, vol. 2(3). Cambridge University Press, New York, 2015.
[9] M.S. Ng, . . . , and B.T. Adey. Designing for digital fab-rication: an empirical study of industry needs, perceived benefits and strategies for adoption. *J.Mgmt.Eng.*, 2022.
[10] M.S. Ng, . . . , and T. Bock. Identifying enablers and rela-tional ontology networks in design for digital fabrication: a systematic literature review. *AutoCon*, 2022.
[11] K.S.D. Ravi, M.S. Ng, J. Medina Iba´n˜ez, and D.M. Hall. Real-time digital twin of on-site robotic construction pro-cesses in mixed realit. In *Proc. ISARC*, 2021.
[12] T. Tan, . . . , and D.M. Hall. Designing human-robot in-teraction interfaces with augmented reality: A systematic review of frameworks, applications, and future directions in construction. *Frontiers of Eng. Mgmt.*, 2026.
[13] B. Suits. *The grasshopper: Games life and utopia*. Broad-view Press, Peterborough, ON, Canada, 2014.
[14] N. Pelling. The (short) prehistory of gamification, 2011.
[15] K. Werbach and D. Hunter. *For the win: How game thinking can revolutionize your business*. 2012.
[16] M. Ilbeigi, D. Bairaktarova, and A. Morteza. Gamification in construction eng. education: A scoping review. *J.of Civil Eng. Education*, 149(2):04022012, 2022.
[17] P.A. Garneau. Fourteen forms of fun, 2001.
[18] C. Heeter et al. Comparing 14 plus 2 forms of fun in commercial versus educational space exploration digital games. In *Proc. DiGRA*, 2003.
[19] Y. Yanai and T. Ishikawa. Immersive learning of cultural heritage through vr: Evaluating interest development and educational outcomes in the osaka castle case study. In *Proc. ICVARS*.
[20] M.S. Ng and D.M. Hall. Teaching target value design for digital fabrication in an online game: overview and case study. In *Proc. IGLC29*, page 249–258, 2021.
[21] F. Kastner et al., . . . , and O. Kasap. Transdisciplinary development of a prototype game on longterm building perspectives. In *Proc. EECA*, 2025.
[22] A. Harichandran, . . . , and J. Teizer. A conceptual frame-work for construction safety training using dynamic vr games and digital twins. In *Proc. ISARC*, 2021.
[23] S. Bhatnagar, . . . , and R. Obulam. A systematic review of lean simulation games in the construction industry. *J.of Civil Eng. Education*, 19:701–719, 2023.
[24] E. Oke, . . . , and T. Abayomi. Application of gamifi-cation for sustainable construction: an evaluation of the challenges. *Construction Innovation*, 2024.
[25] S.E. Boyacioglu, . . . , and A. Parry. Enhancing foren-sic analysis of construction project delays through digital interventions. *Buildings*, 15:2391, 2025.
[26] Y. Cao et al. Research on the regenerative design and practice of the residential architecture atlases in jiangnan during the ming and qing dynasties: a case study of the mingyue bay village in suzhou. *J.AABE*, 2025.
[27] E. Emiroglu, . . . , and I. Demir. Floodcraft: Game-based interactive learning environment using minecraft for flood mitigation for k-12 education. *Intl. J. DRR.*, 2025.
[28] X. Hao et al. Unpacking the decision-making evolution mechanism of governance actors towards smart municipal solid waste (msw) management: an insight from game approach. *Tech. Forecasting & Social Change*, 2025.
[29] J. Hardy et al. Playing with rural design: Design probe failure and community disconnections. In *ICCT*, 2025.
[30] Q. Hu et al. A spatio-temporal overlap narrative expe-rience model for archaeological site museums: A case study of the panlongcheng archaeological site museum. *Buildings*, 2025.
[31] L. Hung, ..., and Y. Zhao. Feasibility and acceptability of gamified cycling exercise for residents in a long-term care home: A qualitative study. *Plos One*, 20:0335679, 2025.
[32] J. Jacob and P. Vaidya. Phantom load: An immersive 3d energy detective game. In *Proc. BuildSys*, 2025.
[33] D. Jang et al. Simulation-based optimization of crane lifting position and capacity using a construction digital twin for prefabricated bridge deck assembly. *Bldgs*, 2025.
[34] J. Jiang et al. Game engine-empowered digital twin for seismic damage simulation of buildings. *Eng. Str.*, 2025.
[36] W.L. Ng et al. Expansion of architectural heritage conservation through digital interactions and gamification: A case study of sg thian hock keng. In *ISPRS Archives*, 2025.
[37] R.A. Pappas et al. Co-minecrafting hermoupolis with children: a minecraft-based digital twin for architectural heritage education. In *Proc. Greece ACM*, 2025.
[38] ´I. S. de Sena et al. A playful participatory planning sys-tem (p-pps): A framework for collecting and analyzing player-generated spatial data from minecraft worlds. *In-ternational J.of Geo Information*, 2025.
[39] T. Taha et al. Enhancing safety training and hazard identification in construction projects using 360-degree panoramic images. *J.ITC*, 2025.
[40] A. Turk and M. Bozkurt. Let's play the ai-based board game: New era for landscape architecture education. *J.of Digital Landscape Architecture*, 10:678–685, 2025.
[41] A.E. Voinea. Gamification in architecture heritage education. virtual interaction with architectural heritage from cluj county, romania. In *Proc. ISPRS*, 2025.
[42] D. Wang et al. Spatial narrative optimization in digitally gamified architectural scenarios. *Buildings*, 2025.
[43] W. Wang et al. Ecological enhancement through smart green village development: Strategic options, key influencing factors, and simulation evidence from hunan province, china. *Sustainability*, 17:6041, 2025.
[44] T. Yang et al. Intelligent interaction system development for civil eng. education. In *Proc. ICETT*, 2025.
[45] R.P. Zhang et al. Supportive communication between apprentices and supervisors: development of a digital role play game. *ECAM*, 2025.
[46] P. Zielke et al. Development of a serious game in construction logistics considering digital and lean tools. In *Proc. IGLC33*, 2025.
[47] G. Arbesser-Rastburg, . . . , and J. Pirker. Project beyond: An escape room game in virtual reality to teach building energy simulations. In *Proc. IEEE CoG*, 2024.
[48] K. Arslan and N. Karakus. Environmental teaching sup-ported by web 2.0-based digital games for a sustainable life. *Sustainability*, 2024.
[49] A. Caliskan. Construction education through virtual re-ality: A game-based approach for interior architecture students. In *Proc. ECGBL*, 2024.
[50] D. Carrasco-Beltran et al. Digital transformation in uni-versity architecture: Optimizing construction processes and user experience through campus 2.0 at pontificia uni-versidad javeriana. *Buildings*, 2024.
[51] Q Chen et al. Intelligent systems integrating bim and vr for urban subway microenvironmental health risks man-agement. *Buildings*, 14:1912, 2024.
[52] Y. Elenany and V. Ahmed. A digital simulation game for resource management in construction projects. *Electronic J. E-Learning*, 2024.
[53] M. Fawad et al. Development of an immersive digital twin framework to support infrastructure management: a case study of bridge asset health monitoring. In *Proc. ISARC*, 2024.
[54] M. Fawad et al. Integration of bridge health monitor-ing system with augmented reality application developed using 3d game engine-case study. *IEEE Access*, 2024.
[55] M. Ilbeigi et al. A gamified method for construction eng. education: Learning through guided active exploration. *J.of Civil Eng. Edu.*, 2024.
[56] E. Joy et al. Digital 3d modeling for preconstruction real-time visualization of home interior design through virtual reality. *Construction Innovation*, 2024.
[57] D. Kolokotsa et al. The intersection of the green and the smart city: A data platform for health and well-being through nature-based solutions. *Smart Cities*, 2024.
[58] S. Park et al. Participatory framework for urban pluvial flood modeling in the digital twin era. *SCS*, 2024.
[59] S. Pickersgill et al. Virtual site visits through gamifica-tion for aec students: Perspectives on practice. In *Proc. SIGGRAPH*, 2024.
[60] G.B. Sa Alves et al. An fpga integrated 2d graphic processor for enhanced digital design and computer architecture education. *IEEE RITA*, 2024.
[61] K. Speiser and J. Teizer. Automatic creation of personalised virtual construction safety training in digital twins.

[63] Tay et al. Design for safety training for construction professionals: A digital game-based learning approach. *Safety Science*, 2024.

[64] L. Vandenabeele et al. Deep learning for the segmentation of large-scale surveys of historic masonry: A new tool for building archaeology applied at the basilica of st anthony in padua. *Intl. J.AH*, 2024.

[65] J. Xiao et al. Fine-grained digital twin sharing framework for smart construction through an incentive mechanism. *Intl. J.PE*, 2024.

[66] Alexander et al. Applying video game design to building digital twin creation. In *Proc. EC3*, 2023.

[67] N.C. Alp et al. Augmented reality experience in an architectural design studio. *MTA*, 2023.

[68] L.M. Angheluta et al. A web-based platform for 3d visualization of multimodal imaging data in cultural heritage asset documentation. *Heritage*, 2023.

[69] Y. Arefazar et al. Development and testing of a digi-tal lean tool to sharpen motion and transportation waste recognition. In *Proc. IGLC31*, 2023.

[70] V.K. Bhat et al. Understanding bim through a simulation game - case study of indian students subjected to this course. In *Proc. ISARC*, 2023.

[71] J. Borucka and S. Parrinello. Vrea project-a digital curator for architecture and digital perspectives for heritage management and enhancement. *Proc. ISPRS Archives*, 2023.

[72] J. Cook et al. The cranky uncle game—combining humor and gamification to build student resilience against climate misinformation. *Envr. Ed. Research*, 2023.

[73] M.D. Diaz-Aleman et al. Nanite as a disruptive technology for the interactive visualisation of cultural heritage 3d models: A case study. *Heritage*, 2023.

[74] A. Ehab and T. Heath. Exploring immersive co-design: Comparing human interaction in real and virtual elevated urban spaces in london. *Sustainability*, 2023.

[75] S.K.N. Ernstsen et al. Technology cards: A design game for navigating in a future of digital technologies. *Intl. J.of Design*, 2023.

[76] M. Jungmann et al. A combined digital twin and location-based management system. In *Proc. IGLC31*, 2023.

[77] D. Kalak et al. Experiencing cultural heritage through gamification: Mardin orphanage. In *Proc. eCAADe*, 2023.

[78] W. Li et al. Development of a mixed reality method for underground pipelines in digital mechanics experiments. *Tunnelling and Underground Space Technology*, 2023.

[79] J.G. Martinez et al. Situational awareness in construction using a serious game. *J.Constr.Eng.Mgmt.*, 2023.

[80] A. Ospina-Bohorquez et al. A digital twin for monitoring the construction of a wind farm. *Infrastructures*, 2023.

[81] J.V. Sorensen et al. Potentials of game engines for wind power digital twin development: an investigation of the unreal engine. *Energy Informatics*, 2023.

[82] S. Tabi and Y. Ikeda. Location hunting game: Developing an application to promote gameful hybrid machi-aruki town exploration. *Urban Science*, 2023.

[83] C. Villamil et al. Implementing sustainability in product portfolio development through digitalization and a game-based approach. *Sus. Prod. Consu.*, 2023.

[84] Z. Wang and J. Li. Designing interactive games based on intangible cultural heritage: A case study on chinese nuo culture. In *Proc. ACM ICPS*, 2023.

[85] K. Speiser and J. Teizer. An ontology-based data model to create virtual training environments for construction safety using bim and digital twins. In *EG-ICE*, 2023.

[86] S. Alves et al. Lean simulation game with bim-based progress monitoring for takt control. In *IGLC30*, 2022.

[87] C. Anschutz et al. Free ride in rush-hour traffic - designing gamified smart mobility systems for sustainable use. *Proc. Icis*.

[88] R. Argiolas et al. Virtual environments to communicate built cultural heritage- a hbim based virtual tour. *ISPRS*, 2022.

[89] B. Avendano-Uribe et al. Playing with uncertainty: Facilitating community-based resilience building. *Urban Planning*, 2022.

[90] S. Comai et al. Simulation modelling in a bim environ-

[91] V. Getuli et al. A smart objects library for bim-based con-struction site and emergency management to support mo-bile vr safety training experiences. *Constr. Innov.*, 2022.

[92] A. Goli et al. Architectural design game: A serious game approach to promote teaching and learning using multi-modal interfaces. *Edu.Info.Tech.*, 2022.

[93] Y. Jiang et al. Digital twin-enabled smart modular inte-grated construction system for on-site assembly. *Comput-ers in Industry*, 2022.

[94] D. Kassi et al. An educational cultural game for an indus-trial urban landscape. In *Proc. ECEE*, 2022.

[95] A. Lee et al. A geospatial platform to manage large-scale individual mobility for an urban digital twin platform. *Remote Sensing*, 2022.

[96] L. Potseluyko et al. Game-like interactive environment using bim-based virtual reality for the timber frame self-build housing sector. *Automation in Construction*, 2022.

[97] S.S. Shareef and H.L. Rauf. Learning technical courses in architectural education through gamification. *CAEE*, 2022.

[98] M. Vollmer et al. Modeling urban complexity in point clouds and sound. *J .Digital Landscape Arch.*, 2022.

[99] C.N. Biotto et al. Virtual parade game for lean teaching and learning in students from brazil and chile. In *Proc. IGLC*, Lima, Peru, 2021.

[100] H. Chen et al. Development of bim, iot and ar/vr tech-nologies for fire safety and upskilling. *AutoCon*, 2021.

[101] A. Ezzeddine and B. Garcia de Soto. Connecting teams in modular construction projects using game engine tech-nology. *Automation in Construction*, 2021.

[102] L.B. Filippova et al. Using 3d graphics and interfaces in training games. In *Proc. IOP CSEES*, 2021.

[103] M. Gangolells et al. Exploring the potential of a gamified approach to reduce energy use and carbon emissions in the household sector. *Sustainability*, 2021.

[104] S. Goubran et al. Eyes on the goal! exploring interac-tive artistic real-time energy interfaces for target-specific actions in the built environment. *Sustainability*, 2021.

[105] A. Gutierrez-Rodr´ıguez et al. Industry 4.0 and digital transformation in higher education through the perspec-tive of smart cites. In *Proc. ICAIW*, 2021.

[106] C. Heins et al. Gamification and bim - the didactic guid-ance of decentralised interactions of a real-life bim busi-ness game for higher education. In *Proc. ISARC*, 2021.

[107] X. Huang et al. Visualization of ancient buildings: Virtual simulation for online historical architecture learning. In *Proc. IEEE ICET*, 2021.

[108] E.L. Jacobsen et al. Lean construction in a serious game using a multiplayer virtual reality environment. In *Proc. IGLC29*, 2021.

[109] P. Kan et al. Building information monitoring via gamifi-cation. In *Proc. International Joint Conference on Com-puter Vision Imaging and Computer Graphics Theory and Applications*, 2021.

[110] P. Koutsabasis et al. Location-based games for cultural heritage: Applying the design thinking process. In *Proc. ACM ICPS*, 2021.

[111] P. Martin-Hernandez et al. Fostering university stu-dents' engagement in teamwork and innovation behaviors through game-based learning. *Sustainability*, 2021.

[112] T. Meuser et al. Netskylines: Digital twins for evaluating disaster communication. In *Proc. IEEE GHTC*, 2021.

[113] C. Putz et al. Potential of gamification for lean construc-tion training: An exploratory study. In *Proc. IGLC29*, 2021.

[114] U. Riedlinger et al. Digital support for bridge inspectors through mixed reality bim data visualizations. In *Proc. ASCE ICCCE*, 2021.

[115] M. Schnieder et al. Using gamification and flipped class-room for remote/virtual labs for eng. students. In *Proc. ECELE*, 2021.

[116] W. Walls. Teaching urban landscape microclimate design using digital site visits: A mosaic method of embedding data, dynamics, and experience. *J. Digital Landscape Arch.*, 2021.

[117] J. Yan et al. Developing a web-based bim asset and facility management system of building digital twins. In *Proc. ASCE ICCCE*, 2021.

Artificial Intelligence and Blockchain in Construction Supply Chains: A Bibliometric and Thematic Analysis

Bryce Meekes[1] , Melissa Chan[1] , Wei Yang[2], Vaughan Coffey[3] and Bambang Trigunarsyah[4]

[1]College of Sport, Health and Engineering, Victoria University, Melbourne, Australia
[2] College of Architecture & Civil Engineering, Beijing University of Technology, Beijing, China
[3] School of Architecture & Built Environment, Queensland University of Technology, Brisbane, Australia
[4] School of Property, Construction and Project Management, RMIT University, Melbourne Australia
bryce.meekes@live.vu.edu.au, melissa.chan@vu.edu.au, wei.yang@bjut.edu.cn,
v.coffey@qut.edu.au, bambang.trigunarsyah@rmit.edu.au

Abstract

This study examines the emerging role of Artificial Intelligence (AI) and blockchain in construction supply chain research through bibliometric and thematic analysis of the existing literature. Using VOSviewer, the study identifies key research trends, thematic clusters, and knowledge structures related to supply chain transparency, automation, optimisation, and adoption barriers. The analysis highlights interconnections between blockchain-enabled traceability, supply chain management, and AI-driven optimisation in construction projects.

The study also synthesises reported applications, including blockchain for secure transaction tracking and smart contract automation, and AI for predictive analytics in resource planning and risk management. Findings indicate growing scholarly attention to the complementary role of AI and blockchain in improving transparency, coordination, and decision-making across construction supply chains, while highlighting persistent challenges such as implementation costs, regulatory uncertainties, technological integration, and organisational resistance.

By mapping the current research landscape, this study provides a structured overview of key themes and emerging research directions, offering insights for researchers and practitioners advancing digital transformation in the construction industry.

Keywords –

Artificial intelligence; Blockchain technology; Construction supply chain; Thematic review; Bibliometric analysis

1 Introduction

The construction industry, crucial for global economic growth, consistently grapples with inefficiencies, transparency issues, and delays in supply chain processes, leading to significant cost overruns and reduced productivity. These technologies offer promising solutions to long-standing issues in the construction supply chain, including inefficiencies, delays, and a need for more transparency. This study aims to provide a comprehensive analysis of how AI and blockchain can be applied to construction supply chains, focusing on research gaps and opportunities for practical implementation.

Traditionally reliant on labour-intensive methods prone to human error, the industry has been slow to adopt new technologies. AI technologies, including machine learning and predictive analytics, enhance project planning and resource allocation by analysing historical and real-time data to anticipate delays, optimise logistics, and reduce material wastage. For instance, AI-driven predictive maintenance can forecast equipment failures, preventing costly downtime (Zheng et al., 2023; Wang et al., 2020).

Blockchain technology provides a decentralised ledger system that enhances transparency and traceability across supply chains by recording transactions and material movements, fostering stakeholder trust. Smart contracts automate agreements, streamline procurement, and support timely payments (Cheng et al., 2021). While integrating AI and blockchain can improve decision-making, reduce operational costs, and enhance project reliability, challenges such as high implementation costs, skills shortages, and regulatory barriers hinder adoption. This study reviews recent literature to identify key trends, benefits, and challenges in construction supply chains.

2 Literature Review

Integrating AI and blockchain technology in construction supply chains is a relatively new study area. However, it holds immense potential for transforming the construction industry's operations. Both technologies

have significantly improved efficiency, transparency, and decision-making when applied individually. However, their combined potential still needs to be explored in current research. This study will explore the role of AI and blockchain in construction supply chains, examining critical research gaps, previous studies and recent technological advancements.

2.1 Previous Work and Study Gaps

Research into using AI and blockchain technology within construction supply chains has evolved considerably in recent years. This field of study addresses critical challenges in construction, such as inefficiencies in resource management, transparency issues and the need for better project coordination. While construction has historically been slower to adopt advanced technologies than manufacturing, the sector's increasing demand for productivity, cost reduction, and adherence to regulatory standards has led to growing interest in leveraging AI and blockchain to enhance supply chain processes. Studies have explored AI and blockchain separately, focusing on their respects and challenges, with only a few addressing their combined potential.

2.2 AI in Construction Supply Chains

The application of AI in construction primarily centres on improving project planning, resource optimisation, and risk management. Researchers have explored how machine learning algorithms can analyse vast datasets to forecast project timelines, identify resource bottlenecks, and predict delays. Elghaish et al. (2023) reviewed predictive analytics in construction supply chains, emphasising how AI-driven forecasting tools enable managers to plan proactively and allocate resources more effectively, reducing the likelihood of costly overruns.

Other studies have highlighted AI's potential to streamline project monitoring and control. Technologies such as drones, real-time sensors, and AI algorithms allow continuous tracking of construction progress against predefined benchmarks. Hamledari and Fischer (2020) showed that AI-powered reality capture technologies provide real-time project updates by comparing site data with project plans, enabling prompt responses to deviations and improving site efficiency.

AI's role in resource optimisation has also received significant attention in construction research. Predictive algorithms can analyse past projects to forecast material requirements, allowing firms to plan procurement schedules more precisely. Research shows AI can optimise inventory and supply chain flows, reducing wastage, streamlining procurement, and lowering costs in dynamic construction environments.

2.3 Need for Integration of AI Applications

The integration of Artificial Intelligence (AI) and blockchain technologies has increasingly been discussed in the digital transformation of construction supply chains. AI techniques such as machine learning and predictive analytics can analyse historical and real-time project data to forecast material demand, optimise inventory, and anticipate supply chain disruptions. However, the effectiveness of AI-based systems depends on reliable and trusted data exchange among stakeholders, which remains challenging due to fragmented project structures and limited transparency.

Blockchain has been proposed as a complementary solution due to its ability to provide secure, transparent, and tamper-resistant data records across decentralised networks (Turk and Klinc, 2017). Studies show that blockchain can enhance supply chain traceability, strengthen stakeholder trust, and automate contractual transactions through smart contracts (Li et al., 2019). When integrated with AI analytics, blockchain can provide a secure infrastructure for sharing predictive insights and supporting collaborative decision-making across project participants.

Another promising area of integration is quality assurance and compliance monitoring. AI-based computer vision and sensor analytics can detect material defects and monitor construction processes in real time, while blockchain can maintain immutable records of inspection results and quality certifications for audit and dispute resolution. Despite this potential, existing research largely examines AI and blockchain separately, leaving limited exploration of integrated AI–blockchain frameworks for construction supply chains.

3 Methodology

This study adopts a bibliometric and thematic review approach to analyse the development of research on Artificial Intelligence (AI) and blockchain in construction supply chains. Relevant literature was identified through systematic searches in major academic databases, including Web of Science, Scopus, and Google Scholar, focusing on peer-reviewed publications published between 2010 and 2025. Keywords related to artificial intelligence, blockchain, construction and supply chain were used to retrieve relevant studies.

The inclusion criteria were articles published in peer-reviewed journals; studies focusing on Artificial Intelligence (AI) and blockchain applications within construction supply chains; and papers discussing technological applications or implications related to supply chain efficiency, transparency, automation, or management. The exclusion criteria were articles not published in English; studies not directly related to

construction supply chains; and papers without substantive discussion of AI or blockchain in the construction context. Following the initial search, duplicate and irrelevant records were removed through a screening process based on title, abstract, and relevance to the study scope A total of 412 records were initially identified through database searches in Google Scholar, PubMed, and Web of Science. After removing duplicates, 368 records remained and were screened based on their titles and abstracts.

Following this screening process, 283 records were excluded due to lack of relevance, leaving 85 full-text articles assessed for eligibility. Of these, 46 articles were excluded based on the predefined exclusion criteria, resulting in 39 studies being included in the final bibliometric analysis. The final dataset was analysed using bibliometric mapping techniques with VOS viewer to identify publication trends, keyword co-occurrence networks, and thematic clusters within the literature. These bibliometric insights were complemented by a narrative synthesis of selected studies to interpret emerging research themes, technological applications, and reported challenges associated with AI and blockchain adoption in construction supply chains. This approach enables the study to systematically map the knowledge structure, research trends, and thematic evolution of AI and blockchain applications in construction supply chain research.

3.1 Quantitative analysis

The quantitative analysis examines case studies of construction projects implementing AI and blockchain technologies, focusing on metrics such as project delivery times, cost reductions, and supply chain efficiency to identify the tangible benefits of these digital tools in construction supply chain management. Evidence from the literature suggests that AI-enabled predictive analytics and blockchain-based platforms can enhance decision-making, coordination, and resource allocation across project supply chains. For instance, the case study by Elghaish et al. (2023), which analysed projects using AI-driven predictive analytics for resource planning, reported a 20% reduction in project delivery times and a 15% decrease in material waste compared with projects that did not implement AI. These improvements were attributed to AI's ability to accurately forecast resource requirements and enable just-in-time delivery of materials and equipment, thereby reducing inventory costs and minimising idle time on-site.

Similar benefits of AI-enabled predictive analytics in construction have been highlighted in broader studies on AI adoption in construction management, which report improvements in productivity, scheduling accuracy, and resource planning efficiency (Abioye et al., 2021). In addition, blockchain-based supply chain systems have been shown to improve transparency and coordination among project stakeholders, contributing to more efficient supply chain operations (Perera et al., 2021). However, empirical studies providing detailed quantitative performance metrics for integrated AI and blockchain applications in construction supply chains remain limited. Therefore, the reported figures should be interpreted as illustrative evidence rather than universally generalisable outcomes, highlighting the need for further empirical case studies and industry validation.

3.2 Qualitative analysis

The qualitative analysis is conducted using a desktop-based approach. This approach utilises insights from existing literature, case studies and industry reports rather than primary data collection through interviews and surveys. This approach allows for a comprehensive examination of the documented experiences, perspectives, and challenges associated with adopting AI and blockchain technologies in construction supply chain.

The qualitative analysis focuses on synthesising findings from multiple sources to understand the real-world applications and perceived barriers to implementing AI and blockchain in construction. Studies indicate that traditional practices and a lack of digital expertise among construction professionals contribute to a cautious approach toward technological transformation. This study thus contextualises AI and blockchain adoption within the construction industry, synthesising existing research to conclude how these technologies can address operational inefficiencies while acknowledging barriers to their widespread adoption.

4 Analysis of Network Visualisations

The two VOSviewer network visualisations provided are bibliometric analyses based on research related to AI, blockchain and construction supply chains. Both maps, created at different resolution levels, offer insights into the dominant themes, clusters, and connections within the literature "Influences of AI and Blockchain Technology on the Construction Supply Chain". The first map, with a resolution of 1.00, presents a broad view of relationships across various keywords. In contrast, at a reduced resolution of 0.80, the second map focuses on fewer but more robust connections, highlighting the core themes and indicating critical gaps in the literature.

4.1 VOSviewer First Mapping

Figure 1. offers visualization of five primary clusters, each represented by a different colour, indicating major thematic areas and subfields within literature.

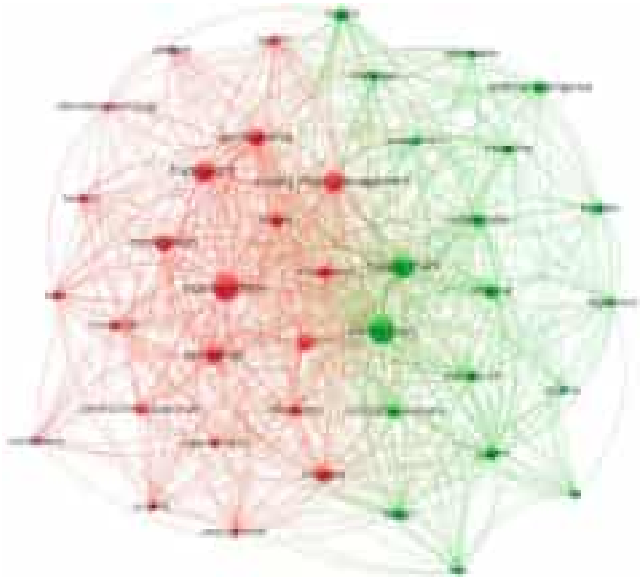


Figure 1. VOSviewer Network Visualisations at Resolution 1.00

4.1.1 Clusters and Key Themes

Red Cluster (Blockchain and Supply Chain): The red cluster centres on "blockchain" and "supply chain," suggesting that blockchain's applications within supply chain management are a focal point in the literature. This cluster includes terms like "big data," "IoT" (Internet of Things), "digital twin," and "Industry 4.0," highlighting that blockchain is often discussed in the context of digital transformation technologies that enhance supply chain transparency, traceability, and data security. The term "sustainability" in this cluster emphasises blockchain's potential to support sustainable supply chains, which is particularly relevant in construction, where resource management and waste reduction are critical concerns.

Green Cluster (Management and Technology Implementation): This cluster is anchored by terms such as "management," "innovation," "trust," and "coordination," with a focus on the operational and practical aspects of adopting AI and blockchain in construction. The frequent connection to the "construction supply chain" reflects an interest in how these technologies affect management processes, particularly in coordinating complex projects with multiple stakeholders. This cluster underscores the literature covering technical and human factors, like stakeholder trust and collaboration, necessary for successful implementation.

Blue Cluster (Construction and Integration): The blue cluster contains terms such as "integration," "construction," "information," and "supply chain management," suggesting a focus on incorporating digital technologies, like AI and blockchain, into existing construction processes. The term "integration" is particularly prominent, implying that researchers are examining the difficulties and strategies of embedding new technologies in traditional, often fragmented, construction supply chains.

Yellow Cluster (AI and Optimisation): This cluster is built around "artificial intelligence," "optimisation," "performance," and "model." It highlights the literature's interest in AI's ability to optimise performance and efficiency within construction supply chains. AI is commonly discussed in terms of predictive analytics and resource allocation, which are vital for minimising project delays and enhancing productivity in construction.

Purple Cluster (Adoption and Information Technology): The purple cluster includes terms such as "adoption," "information technology," and "barriers," reflecting an emphasis on understanding the obstacles to implementing AI and blockchain. The presence of terms like "information technology" and "barriers" indicates that technological and infrastructural limitations are primary concerns in the literature.

4.1.2 Central Nodes and Connections

Key terms such as "blockchain," "supply chain," "management," and "artificial intelligence" appear as central nodes with dense interconnections. "Blockchain" and "supply chain" are heavily connected, suggesting blockchain's potential to enhance supply chain management is a central theme. "Management" also shows high connectivity, indicating that operational and managerial aspects are widely explored within the context of technology adoption in construction. Additionally, "artificial intelligence" is well-connected, especially with "optimisation" and "performance," showing its application in improving efficiency within construction processes.

4.1.3 Overall insight

Dominant Research Themes: The VOSviewer network visualisation highlights distinct clusters representing dominant research themes in AI, blockchain, and construction supply chains. In the first mapping, the red cluster centred on blockchain and supply chain forms one of the most densely connected areas, with strong linkages to IoT, big data, digital twin, and Industry 4.0. This indicates that blockchain research in construction is largely framed within broader digital transformation efforts to improve transparency, traceability, and data security across supply chains. In contrast, the yellow cluster centred on artificial intelligence connects strongly with optimisation, performance, and model, reflecting AI's focus on predictive analytics and operational efficiency in construction projects.

The presence of these clearly defined thematic clusters suggests that the literature has developed around specific technological capabilities rather than integrated digital supply chain systems, indicating that research on AI and blockchain remains largely compartmentalised.

Fragmented Integration Between Technologies: Although both technologies appear as central nodes in the network, the visualisation shows relatively limited direct linkages between AI-related keywords and blockchain-related keywords. Instead, AI terms are primarily connected to optimisation and performance-oriented research, while blockchain terms are strongly associated with supply chain transparency and traceability. This structural separation suggests that the two technologies are frequently studied as independent technological solutions rather than as integrated systems within construction supply chains. The limited co-occurrence of keywords such as integration, framework, or combined systems further supports this interpretation. Consequently, the bibliometric structure indicates that while both technologies are recognised as important digital innovations in construction, research exploring their combined application remains relatively underdeveloped, highlighting a gap in integrated AI–blockchain supply chain studies.

Practical Implementation and Adoption Challenges: The network visualisation also reveals clusters related to management, coordination, trust, and barriers, particularly within the green and purple clusters. While these terms indicate that the literature acknowledges organisational and technological challenges associated with adopting AI and blockchain, their relatively weaker connections compared to core technological terms suggest that practical implementation issues receive less attention than conceptual technological benefits. In addition, keywords associated with empirical application such appear less central within the network. This pattern indicates that although the theoretical potential of AI and blockchain in construction supply chains is widely discussed, empirical studies demonstrating real-world implementation remain limited. The visualisation therefore supports the conclusion that future research should place greater emphasis on implementation frameworks, case-based evidence, and practical strategies for integrating AI and blockchain technologies within construction supply chains.

4.2 VOSviewer Second Mapping

Figure 2. offers the second VOSviewer map, with a reduced resolution removes weaker links and secondary themes, providing a distilled view of the literature's core focus areas. The clustering has been consolidated, highlighting fewer but stronger relationships between keywords and reinforcing the most central themes in AI and blockchain research related to construction.

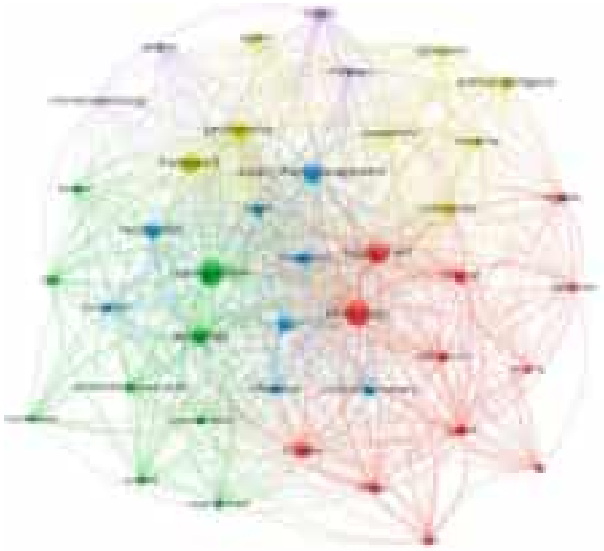


Figure 2. VOSviewer Network Visualisations at Resolution 0.8

4.2.1 Clusters and Key Themes

Red Cluster (Management, Technology, and Performance): This cluster consolidates terms around "management," "technology," "framework," and "performance." The intense focus on management and technological frameworks indicates that much research concentrates on understanding AI and blockchain's organisational impacts. This consolidation reveals that while managerial aspects are well-studied, the interaction between blockchain, AI and specific construction management practices is less pronounced.

Green Cluster (Blockchain, Supply Chain, and AI): The green cluster centres around "blockchain," "supply chain," and "artificial intelligence." AI and blockchain are closely connected in this reduced resolution map, suggesting that the literature identifies these technologies as complementary tools for optimising supply chain management. However, terms like "integration" and "implementation" are less prominent in this simplified view, pointing to a possible gap in research focused on applying these technologies in real-world construction.

Consolidation of Key Nodes The reduced resolution places even greater emphasis on "blockchain," "supply chain," and "management" as central nodes, suggesting that these are the predominant themes driving research. Blockchain’s link to the "supply chain" remains strong, reinforcing its role in enhancing transparency and trust in construction projects. The increased emphasis on these connections also suggests that, at a fundamental level, the literature is more focused on broad themes of blockchain's role in supply chain improvement and AI’s optimisation potential than on specific case applications in construction.

4.2.2 Overall insight

Research Themes and Structural Patterns: The VOSviewer keyword network reveals several prominent thematic clusters centred on blockchain, artificial intelligence, and construction supply chain management. In the higher-resolution map (1.00), strong linkages are observed between blockchain and terms such as supply chain, IoT, and digital twin, indicating that blockchain research in construction is primarily associated with improving transparency, traceability, and data sharing across project stakeholders. In contrast, AI-related terms cluster around optimisation, performance, and resource allocation, suggesting that AI research is mainly directed toward predictive analytics and operational efficiency. The separation of these clusters indicates that the literature tends to examine AI and blockchain as distinct technological approaches rather than as components of an integrated digital supply chain system. This structural pattern suggests that although both technologies are widely recognised as key enablers of digital transformation in construction, their combined application within construction supply chains remains relatively underexplored in existing research.

Barriers and Adoption Challenges: The comparison between the two visualisations also highlights differences in the prominence of adoption-related terms. In the higher-resolution map, keywords such as trust, coordination, and barriers appear more clearly connected to the broader network, indicating that the literature acknowledges organisational and technological challenges associated with adopting AI and blockchain in construction supply chains. However, in the lower-resolution map (0.80), these terms become less visible and exhibit weaker connections within the network. This pattern suggests that while barriers to adoption are recognised within the literature, they may receive less sustained analytical attention compared with themes related to technological capabilities or optimisation benefits. As a result, the bibliometric structure indicates that research has focused more heavily on conceptual technological potential than on systematically addressing the organisational, regulatory, and operational challenges associated with implementation.

Gap in Practical Integration and Implementation: Another notable observation from the network visualisations is the limited prominence of terms directly associated with integration and real-world implementation, such as implementation, framework, or deployment. These keywords appear with relatively lower occurrence frequencies and weaker link strengths compared with dominant technological terms such as blockchain, AI, and supply chain. This structural pattern indicates that much of the existing literature focuses on conceptual discussions of technological capabilities rather than empirical demonstrations of how these technologies can be integrated into operational construction supply chains. Consequently, the visualisation suggests that while the theoretical benefits of AI-driven optimisation and blockchain-enabled transparency are well established, practical frameworks for implementing these technologies together within construction supply chains remain limited. This gap highlights an important opportunity for future research to explore integrated implementation strategies and case-based evidence supporting AI–blockchain adoption in construction practice.

5 Emerging Themes

The following section analyses key trends and themes emerging from the bibliometric analysis.

5.1 Transparency and Traceability

Blockchain technology's primary advantage in construction supply chains is its unparalleled transparency and traceability. It records every transaction in an immutable, tamper-proof ledger, ensuring that all project activities, from procurement to delivery and payment, are documented precisely. This transparency can significantly reduce risks in construction, where delays, miscommunication, and disputes are common.

As mentioned previously, blockchain can be used to track the origin of materials, ensuring compliance with environmental regulations and certifications such as LEED. Wang et al. (2020) demonstrated how blockchain could track precast components, from manufacturing to installation, providing real-time visibility into their status and ensuring that all quality standards were met. This level of traceability is critical for large-scale construction projects where multiple suppliers are involved and materials are sourced from different regions.

Additionally, blockchain's transparency helps build trust among stakeholders. When all parties in a construction project have access to the same verified information, disputes over deliveries, payments, and project milestones are less likely to occur. Smart contracts further enhance this by automating payments when specific conditions, material delivery or project completion, are met Yoon et al. (2022). This drastically reduce delays caused by payment disputes, improve cash flow, and ensure projects remain on schedule.

5.2 Automation and Efficiency

AI has the potential to automate many repetitive and time-consuming tasks in construction supply chains, including scheduling, resource allocation, and material procurement. By leveraging machine learning algorithms, AI can analyse historical and real-time project data to predict future trends and optimise supply chain decisions.

One area where AI has demonstrated significant potential is project scheduling, where AI-powered project management tools can analyse project scope, timelines, and resource availability to generate optimised schedules that minimise delays and improve overall efficiency. Similarly, AI can support procurement by predicting material needs based on project progress, enabling timely ordering while reducing excess inventory and costs.

When integrated with blockchain, these AI-driven capabilities can be further enhanced through secure data sharing and automated transaction execution. For example, AI-generated predictions such as forecasts of material demand or schedule adjustments can be recorded on blockchain-based distributed ledgers, ensuring transparency, traceability, and data integrity across project stakeholders. In addition, smart contracts can interface with machine learning outputs to automatically trigger procurement orders, verify supplier deliveries, or release payments once predefined project milestones are achieved. This integration enables more reliable information exchange, reduces manual intervention in supply chain processes, and improves overall operational efficiency in project delivery.

5.3 Stakeholder Engagement

The construction industry is known for its fragmentation, with multiple stakeholders, including contractors, subcontractors, suppliers and clients, often working in silos. This lack of collaboration can lead to miscommunication, delays, and cost overruns. Blockchain and AI technologies can help address these challenges by fostering greater cooperation and ensuring all stakeholders access the same information.

Blockchain's decentralised ledger provides a single source of truth for all project stakeholders, ensuring that everyone is working from the same set of verified data. This transparency reduces the likelihood of disputes over project milestones, material deliveries, and payments, as all parties can access a consistent and trustworthy record of transactions. By enabling real-time data sharing, blockchain enhances collaboration among stakeholders, helping to minimise delays and miscommunication that can arise from discrepancies or incomplete information Wang et al. (2020), Elghaish et al. (2023).

AI can also improve collaboration by providing real-time insights into project progress and potential risks. AI can also analyse data from previous projects to identify patterns and suggest best practices for future projects, helping stakeholders make more informed decisions Hamledari et al. (2021).

5.4 Smart Contracts

Another significant theme in the literature is the use of smart contracts to automate processes in the construction supply chain. Smart contracts are self-executing contracts where the terms of the agreement are written directly into code. Once certain conditions are met, such as the completion of a project milestone, the smart contract automatically executes, triggering payments Yoon et al. (2022). In construction, smart contracts can use intelligence to streamline procurement processes, ensuring that payments are released as soon as materials are delivered and verified. This reduces the need for manual interventions and helps ensure that suppliers are paid on time, improving cash flow and reducing delays caused bypayment disputes Wang et al. (2020). Smart contracts can also automate compliance checks, ensuring all project requirements are met before payments are released Elghaish et al. (2023).

6 Discussion

While the potential benefits of AI and blockchain in construction supply chains are well-documented, several barriers exist to their widespread adoption such as high implementation costs, regulatory challenges, and resistance to change.

6.1 High Implementation Costs

One of the main barriers to adopting AI and blockchain in construction is high cost of implementation. Both technologies require significant upfront investments in hardware, software and training costs can be prohibitive if the return on investment (ROI) is not immediately apparent Yoon et al. (2022). However, long-term benefits of AI and blockchain such as improved efficiency, reduced delays and enhanced transparency can offset initial costs. For example, blockchain can help construction firms save time and money by automating procurement processes and reducing likelihood of payment disputes. Similarly, AI's ability to predict risks and optimise resource allocation can lead to significant cost savings by preventing delays and material waste Elghaish et al. (2023).

6.2 Regulatory Challenges

Another significant barrier to the adoption of blockchain in construction is the need for clear regulations surrounding the use of smart contracts. While intelligent contracts offer numerous advantages—such as automating payments and compliance checks—the legal framework for their use in construction is still underdeveloped. For example, there are questions about the enforceability of smart contracts in different jurisdictions, particularly when resolving disputes Yoon et al. (2022). Blockchain’s decentralised nature raises concerns about data privacy and security. Construction firms must comply with regulations such as the General

Data Protection Regulation (GDPR) when using blockchain to store and share data. Failure to comply can result in significant penalties, making it essential for firms to work closely with legal experts to ensure their use of blockchain is compliant Elghaish et al. (2023).

6.3 Resistance to Change

The construction industry traditionally needs to adopt new technologies faster, but management and workers often resist change. Many construction firms are risk-averse, preferring to stick with traditional methods rather than invest in new technologies that they perceive as unproven. There is also a fear that adopting AI and blockchain could lead to job losses, particularly among administrative and clerical staff whose roles could be automated Hamledari et al. (2021). To overcome this resistance, construction firms must demonstrate the tangible benefits of AI and blockchain. By providing training and education programs, firms can help workers understand how these technologies can improve rather than replace their jobs. By automating administrative tasks, AI and blockchain can free workers to focus on more value-added activities Elghaish et al. (2023).

7 Conclusion

Integrating artificial intelligence and blockchain technology into construction supply chains holds enormous potential to revolutionise the industry. These technologies offer solutions to many challenges that have plagued the construction sector. AI's predictive capabilities enable construction firms to optimise project planning, resource allocation, and risk management, while blockchain ensures transparency, traceability, and security across the supply chain. However, despite the apparent benefits, significant barriers exist to the widespread adoption of AI and blockchain in construction. These include high implementation costs, regulatory challenges, and resistance to change. Overcoming these barriers will require a concerted effort from industry stakeholders and a deeper understanding of integrating these technologies into existing supply chain management practices. Future research should focus on real-world case studies that demonstrate practical implementation of AI and blockchain in construction. These studies would help build confidence in these technologies and encourage broader adoption across the industry. Additionally, research should explore how AI and blockchain can be integrated with other emerging technologies, such as the Internet of Things (IoT) and Building Information Modelling (BIM) to create fully digitalised and interconnected construction supply chains. Finally, construction firms must work closely with regulators to ensure blockchain and smart contracts are legally enforceable and compliant with data protection laws. By addressing these challenges, construction industry can fully leverage the benefits of AI and blockchain to create more efficient, transparent and collaborative supply chains.

References

[1] Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. Journal of Building Engineering, 44, 103299.

[2] Cheng, M., Liu, G., Xu, Y., & Chi, M. (2021). When blockchain meets the AEC industry: Present status, benefits, challenges, and future research opportunities. Buildings, 11(8), 340.

[3] Elghaish, F., Abrishami, S., & Hosseini, M. R. (2023). Digitalised circular construction supply.

[4] chain: An integrated BIM-blockchain solution. Automation in Construction, 132, 103988. https://doi.org/10.1016/j.autcon.2021.103988

[5] Hamledari, H., & Fischer, M. (2021). Construction payment automation using blockchain-enabled smart contracts and robotic reality capture technologies. Automation in Construction, 132, 103926.

[6] Li, J., Greenwood, D., & Kassem, M. (2019). Blockchain in the built environment: analysing current applications and future opportunities. Automation in Construction.

[7] Perera, S., Nanayakkara, S., Rodrigo, M., Senaratne, S., & Weinand, R. (2021). Blockchain technology: Is it hype or real in the construction industry? Journal of Industrial Information Integration.

[8] Turk, Ž., & Klinc, R. (2017). Potentials of blockchain technology for construction management. Automation in Construction

[9] Wang, J., Wu, P., Wang, X., & Shou, W. (2020). Blockchain technology in the construction industry: A review of applications, challenges, and future directions. Automation in Construction, 115, 103181. https://doi.org/10.1016/j.autcon.2020.103181

[10] Yoon, Jong Han & Pishdad-Bozorgi, Pardis. (2022). State-of-the-Art Review of Blockchain-Enabled Construction Supply Chain. Journal of Construction Engineering and Management. 148. 10.1061/(ASCE)CO.1943-7862.0002235.

[11] Zheng, Z., Wang, F., Gong, G., Yang, H., & Han, D. (2023). Intelligent technologies for construction machinery using data-driven methods. Automation in Construction, 147, 104711. https://doi.org/10.1016/j.autcon.2021.103926

Development and Implementation of an OpenBIM–LCA Workflow for Factory Planning

Leonie Große-Wilde[1], Luca Philipp[2], Tina Esmaeilzadeh[1], Hannah Leisten[1], Sven Mackenbach[1], Matthias Schmidt[2], Katharina Klemt-Albert[1]

[1] Institute of Construction Management, Digital Engineering and Robotics in Construction, 52070 Aachen, Germany
[2] Institute of Production Systems and Logistics, 30823 Garbsen, Germany

grosse-wilde@icom.rwth-aachen.de, philipp@ifa.uni-hannover.de, tina.esmaeilzadeh@rwth-aachen.de; hannah.leisten@rwth-aachen.de, mackenbach@icom.rwth-aachen.de, klemt-albert@icom.rwth-aachen.de

Abstract –

Factories are a key part of industrial value creation and have a significant impact on the environment throughout the entire supply and production chain. Therefore, a sustainability assessment of both the building and the production system requires a holistic, interdisciplinary perspective on the factory system, which cannot be achieved using conventional methods employed for other building types. This study therefore addresses a research gap concerning the automated integration of Building Information Modeling (BIM) and Life Cycle Assessment (LCA) in the context of factory planning. The objective is to address the key challenge of linking multiple data sources for the very different interdisciplinary requirements. Hence, based on literature current approaches for both building LCA and factory planning processes are analyzed, and a factory-specific openBIM workflow for ecological assessment is proposed and implemented using Desite md. To meet the requirements of the various disciplines involved in the factory planning project, the workflow considers processes at both the disciplinary and coordination levels. To integrate the databases containing the energy demand and LCA profiles of the machinery and building materials used, the implemented approach connects individual production system databases, as well as building-specific databases such as the German ÖKOBAUDAT. The findings demonstrate the current potential and challenges of applying the workflow in Desite md, revealing practical insights for assessing the ecological sustainability of factory buildings. Ultimately, this supports factory operators in making more sustainable decisions during the design phase, leading to significantly more sustainable building operations.

Keywords –

Building Information Modeling (BIM), Life Cycle Assessment (LCA), sustainability, openBIM

1 Introduction

The environmental sustainability of industrial value chains is gaining relevance in light of advancing climate change, stricter legal frameworks and growing social expectations [1–3]. In the context of factory planning in particular, there is considerable potential to reduce environmental impacts across all life cycle phases [4]. In practice, Life Cycle Assessments (LCA) of factories, including the building, technical building systems (TBS) and production equipment, are challenging due to the extensive effort required to capture all environmentally relevant factors and their interactions over the entire life cycle [5]. In contrast to the planning of conventional buildings, factory planning requires a multidimensional approach that takes into account not only the building and TBS, but also production processes [6]. Another significant difficulty stems from the substantial amount of data generated by the project's interdisciplinary nature and the requirement for consistent documentation and traceability of results within such a complex system as a factory [7]. Building Information Modelling (BIM) offers the possibility of capturing, linking and systematically evaluating information across domains [8]. This raises the central research question of how a BIM-based approach to the ecological sustainability assessment of factories is designed so that it serves as a basis for sustainable decisions throughout the entire life cycle of the factory. Although various ecological assessment approaches exist [9, 10], they primarily address conventional buildings. Existing methods for production facilities consider only the production system and are not yet fully digitally integrated [11, 12]. Consequently, this study aims to develop a holistic, BIM-supported methodology for life cycle–based ecological sustainability assessment of factory planning projects.

2 Background and related works

2.1 BIM-based LCA

Several approaches to integrating BIM and LCA already exist for building construction and infrastructure projects. The predominant classification of these approaches was defined by WASTIELS AND DECUYPERE, who identified five overarching strategies for integrating BIM and LCA. The first strategy involves the manual export of the bill of quantities to a spreadsheet, wherein the LCA is calculated either directly or in a LCA software. Materials and quantities are manually linked to corresponding LCA profiles, representing the current state of practice [13]. The second strategy adopts an openBIM approach by exporting an Industry Foundation Classes (IFC) file to LCA software, transferring geometry, material names, and Global Unique Identifiers, after which materials are manually associated with LCA profiles. The third strategy, also based on openBIM, introduces an intermediate BIM viewer to link materials to LCA profiles, prior to performing the assessment in a LCA software. The fourth strategy employs a closed BIM approach, in which a plug-in within the native BIM authoring environment links BIM objects to LCA databases and visualizes results directly in the 3D model. Finally, the fifth strategy entails the direct allocation of LCA data to BIM objects, facilitated by either database identifiers or explicit indicator values. This approach finds application in both openBIM and closedBIM workflows, employing either plug-ins or IFC-based LCA software [14]. There is an absence of research approaches integrating BIM-LCA in factory planning, which is necessary because current approaches do not allow for an integration of the production system and are not addressing key requirements of factory planning. The key requirements of this use case are outlined in the subsequent subchapter.

2.2 Requirements for a BIM-based LCA workflow for factories

Factory planning is highly interdisciplinary and presents specific requirements for integrating BIM and LCA. In factory planning projects, up to 100 different disciplines may be involved. These disciplines can be divided primarily into the domain of factory planning (e.g. layout or logistics) as well as the domain of object planning (e.g. architecture, TBS). Their collaboration requires an error-free exchange of large and heterogeneous data sets [15–17]. This heterogeneity is mainly caused by discipline-specific data formats, software applications and modeling logics [16, 17]. In addition, several of these disciplines are required to collaborate within a shared or coordinated model environment rather than exchanging data sequentially, in order to enable a holistic sustainability assessment across building-related and process-related elements. Consequently, technical interoperability constitutes a central requirement for a BIM–LCA workflow. Standardized data exchange via open formats, such as IFC, is required for enabling an interdisciplinary flow of information.

In addition to technical interoperability, a suitable workflow must consider the existing IT systems of the parties involved in the planning process. Especially in small and medium-sized enterprises (SMEs), IT infrastructures are often centralised, inflexible, and modernised only incrementally [18, 19]. This is largely due to high software and implementation costs [20, 21]. This challenge intensifies in factory planning because individual planning tasks are often outsourced to external offices while the manufacturing company retains overall coordination [22]. Therefore, a BIM–LCA workflow must integrate with existing software portfolios and operate without fundamentally restructuring the IT landscape to be practical, particularly for SMEs.

Another key requirement is the integration of LCA datasets for factory-specific objects. While standardized and high-quality environmental product declaration (EPD) databases such as ÖKOBAUDAT are already available for the building perspective [23], data availability for the process perspective is highly individualized. Factories differ from other types of buildings, especially due to the integration of production equipment and associated media supply systems. These systems significantly contribute to the overall environmental impact [24, 25]. Accordingly, a suitable BIM–LCA workflow must support the simultaneous integration of multiple and heterogeneous LCA data sources to ensure that all relevant factory elements can be assigned appropriate environmental profiles. This includes the combination of standardized databases for building-related components with non-standardized, manufacturer- or project-specific datasets for production-related elements. Hence, the integration of factory-specific LCA data cannot rely on a single database but requires a flexible data integration concept that accommodates diverse data structures and levels of detail.

Finally, the project-based nature of factory planning requires flexibility in influencing the underlying calculation principles of the LCA. Each factory represents an individual planning project [26], meaning system boundaries, input-output flows, service lives of individual factory elements, and the temporal scope of the assessment must be defined on a project-specific basis [25, 27, 28]. Therefore, a BIM–LCA workflow must allow for the adaptation of the calculation logic and must not rely exclusively on proprietary default assumptions of LCA software, as this could otherwise lead to simplified or non-representative results [29].

In conclusion, the existing strategies are only suitable for implementation in factory planning to a limited extent. Consequently, to address the factory-specific requirements, a new workflow for BIM-based LCA for factory buildings is proposed in the following section.

3 OpenBIM Workflow for the LCA of a Factory

This workflow builds on the fundamental elements of BIM-LCA strategies that have been established in the construction sector but extends them to meet the requirements of factory planning, so that a holistic LCA of a factory can be performed during the various stages of the planning phase. This ultimately results in enhanced decision-making processes with regard to sustainability, thereby optimizing the entire life cycle of a factory.

3.1 Sustainability assessment

Prior to the development of the workflow, it is necessary to outline the general conditions for the environmental sustainability assessment of a factory, which include the considered life cycle phases and ecological indicators.

This workflow considers the whole life cycle of the factory and is based on the life cycle modules used in the construction industry, that are defined in EN 15978 [30]. Based on this standard the following life cycle modules are included: the production stage (A1–A3) covering raw material extraction, transport, and manufacturing; the construction process stage (A4–A5) addressing transport to site and construction activities; the use stage (B1–B7) comprising operation, maintenance, replacement, and operational energy and water use; and the end-of-life stage (C1–C4) including deconstruction, waste processing, and disposal [30].

The ecological impact of a factory system can be measured using various midpoint indicators, as in a typical LCA. This workflow emphasizes the most critical indicators for factories but can be expanded to include more. An analysis of German certification systems like German Sustainable Building Council [31] and Assessment System for Sustainable Building [32], as well as research approaches from Gebler [33] reveals that the environmental indicators global warming potential (GWP), ozone depletion potential (ODP), acidification potential (AP), eutrophication potential (EP), ozone creation potential (POCP), renewable primary energy demand (PERT), and non-renewable primary energy demand (PENRT) are classified as relevant or necessary across all approaches considered. Therefore, it can be concluded that these indicators can act as a basis for ecological assessment in the context of factory planning.

Furthermore, the selected indicators are suitable for incorporation into a BIM-LCA because they can be accessed in standardized life cycle assessment databases, such as ÖKOBAUDAT [34, 35].

3.2 Interoperability and openBIM

In order to meet the requirement of open data exchange and thus facilitate a comprehensive evaluation of the coordination model, the openBIM approach is employed. This approach ensures interdisciplinary information exchange via standardized data formats, particularly IFC [36]. While WASTIELS AND DECUYPERE identify three openBIM strategies, this workflow integrates the fifth strategy, which involves the semantic enrichment of BIM objects and their export to IFC, with the third strategy, which incorporates a BIM viewer. While many LCA software applications support importing IFC files, several studies demonstrate that differences in data structures between BIM and LCA systems often result in inefficient data exchange or lead to functional limitations [18–20]. Hence, utilizing a BIM viewer as a substitute and implementing the LCA calculation there can result in enhanced interoperability and minimal disruption to the BIM-based planning workflow. Additionally, implementing a BIM viewer fulfils the requirement for seamless integration with existing software portfolios, since BIM viewers are widely used for BIM-based planning.

In this workflow the semantic enrichment entails the addition of unique identifiers (ID) of LCA datasets to BIM objects, as opposed to the assignment of indicator values. Specifically, this implies that the materials contained in the respective discipline models are assigned corresponding IDs from environmental databases, for example the Universally Unique Identifier (UUID) in the ÖKOBAUDAT. This method reduces the impact on the quantity of data contained within the BIM model since fewer attributes need to be assigned.

Due to the large number of disciplines involved in factory planning projects, this workflow suggests a mapping process in which planners add the IDs to their discipline models in their authoring tools. Once this assignment has been done manually, the corresponding LCA indicator values can be automatically linked to the individual components in the model. A key advantage of this process is that the challenge of assigning the right data set is already solved at the level of the individual discipline models. The early and automated assignment of LCA IDs within the respective BIM objects ensures the consistency of the environmental data basis before the discipline models are merged into an overall coordination model. This is especially true in the context where the disciplines of factory and object planning interact, and a lack of expertise exists among sustainability experts to map the LCA data sets for both domains. Furthermore, as the transition to standardized, semantically enriched

object libraries in BIM authoring tools continues, this effort can be minimized in the future [37].

3.3 Integration of multiple databases

Due to the different planning domains described in section 2.2, the integration of various databases is required. The following section explains which databases are considered in this workflow.

This workflow focusses on the database ÖKOBAUDAT as an environmental database for the domain of building and technical building systems. The ÖKOBAUDAT is the official database for German sustainability certification systems and comprises over 2,100 data sets that have been specially prepared for construction products [35, 38]. As a material-based database, the ÖKOBAUDAT provides specific environmental factors for individual materials, allowing for precise life cycle assessments based on material-specific data [39].

In contrast to the construction sector, there are no established environmental databases for production equipment in the factory planning domain. Therefore, manufacturer-specific EPDs are critical for ensuring consistent and robust LCAs in factory contexts. Instead of a standardized database, company-specific databases or Excel spreadsheets are used to assess the necessary data for evaluating the production system. Enterprise resource planning (ERP) systems, such as SAP, can also provide important data on machines and their media usage [11]. An ERP system is a business management software solution that aims to automate and integrate a company's central business processes and link them via a company-wide database. It enables access to real-time information, avoids data redundancy and ensures data integrity by providing a single source of truth. [40] Manufacturing companies use ERP systems as central data repositories for managing project-related information. This system stores and provides structured access to general project information and environmental data, technical parameters, cost overviews, maintenance intervals and measures, as well as process management information [11]. These data are necessary for sustainability assessments. For this reason, integrating different databases and ERP systems is recommended for this workflow to meet the aforementioned requirement of integrating databases from both planning domains. Although the implementation described below primarily uses CSV files, a direct connection via Application Programming Interfaces (API) is planned for future work to ensure accuracy.

3.4 Integration in planning process

Having described the theoretical basis of the key elements of the workflow in previous sections, this section summarizes the complete workflow and its integration into the planning process. The foundation of this workflow is based on existing standards and guidelines.

In accordance with the guideline of the Association of German Engineers [41], planners assign the ID parameter to match the LCA data sets of their BIM objects, as explained in chapter 3.2. After enriching the model, planners are responsible for conducting plausibility and validation checks to ensure that all BIM objects are assigned the correct LCA datasets. This includes verifying the completeness and correctness of the required datasets. Subsequently, the BIM coordinators of each discipline upload the approved IFC files to the Common Data Environment (CDE) [41]. The overall BIM coordinator is responsible for merging the IFC files in a coordination software (e.g., Desite md) and performing the second validation step. This involves a rule-based model check within the coordination software that detects missing attributes, such as IDs, and documents them in the form of a report. These reports are serve as the foundation for the corrective measures by the involved planners. Subsequent to validation and quality testing, a holistic LCA calculation is performed based on the enriched coordination model by the LCA expert. Consequently, the overall coordination model is automatically linked to the databases via an API within the coordination software. Based on this, the rule-based LCA calculation is performed. Figure 1 provides a graphical summary of the workflow.

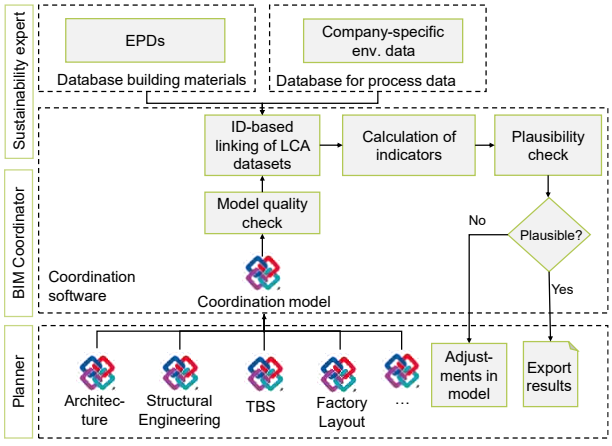


Figure 1 BIM-based factory sustainability assessment

3.5 Modelling requirements

For this workflow to be executed, it is essential that certain modeling requirements are met. First, the aforementioned ID is written as a shared parameter and, in the context of IFC data exchange, is assigned to a user-defined Pset, such as 'Pset_LCA'. It should also be noted that the UUIDs are version-dependent and that both the

UUID and the version number (e.g. as LCA_Version) must be stored in the model during the mapping process. In addition, the type and material information should be assigned for the purpose of simplifying the plausibility check later on.

While this procedure applies to all elements of the object planning domain, additional aspects must be considered for the technical building equipment. According to the findings of GENOVA [42], the technical elements can be divided into two main categories: technical routing and technical devices. The technical routing system covers all components such as pipes, air ducts, electrical cables, lighting systems and cable trays and their fittings. These are typically characterised by the fact that they are generated by extruding a cross-section along a line path. These elements are usually not represented in detail geometrically, while the information content is nevertheless retained in full. However, a technical deficit arises from the fact that, for example, authoring tools like Revit do not allow direct access to volumetric geometry data for these components. This makes it considerably more difficult to calculate material masses. A practical solution is to determine the area resulting from the thickness of the pipe and the circumference and multiply it by the length of the corresponding reference line. Based on this, mass values of the components can be calculated [42]. The second category, technical devices, includes all industrially manufactured products installed in a building whose dimensions and material composition are usually specified by the manufacturer per piece. The correct reference flow of the matched LCA data set should therefore be automatically read out so that no errors occur in the calculation and the correct unit is assigned.

In the scenario where SAP is used as the company-specific data repository and therefore needs to be integrated with BIM, the BIM objects, along with their properties such as dimensions and material definitions, are assigned to the corresponding SAP identifiers and attributes. This data mapping process between BIM and SAP is carried out using a classification system. The use of classification systems supports the assignment of BIM data to corresponding structures in the SAP system [43]. The classification system must be adapted for factory planning, as established classification systems are designed for building construction in accordance with the ISO 12006-2 standard [44]. For this purpose, it is recommended that the ECLASS classification system is used. ECLASS is an internationally recognized standard for the uniform description and structuring of products, machines, components and services [45, 46].

For a consistent BIM-SAP link, it is necessary to store the ECLASS codes as type parameters in the BIM model. This makes it possible to uniquely identify, for example all CNC milling machines in the model using an eight-digit code. When exporting to IFC, these parameters must be assigned in IfcClassificationReference [47]. The ECLASS codes and versions stored in the SAP system can be exported in the form of structured data records (e.g. as CSV or Excel files). These usually contain the material number, the equipment ID, a textual product description, the associated ECLASS code and the ECLASS version used. Integration with the BIM model is achieved by mapping the data exported from SAP to the model objects available in the authoring software.

4 Implementation

For the exemplary implementation of the calculation part of the workflow, Desite md and therein WebForms are used to demonstrate the automatic linking of LCA data sets and BIM objects based on their common ID for elements of the building as well as the production system.

For this purpose, two types of databases are integrated: the ÖKOBAUDAT and a spreadsheet-based database containing LCA data for production equipment, both provided in CSV format. The ÖKOBAUDAT is available as a CSV file on their website. [35] The second CSV acts as examples of ERP system exports or other company-specific data that could be queried via an API in the future. The CSV requirements and IFC export information can be found here [48].

The internal functioning of the WebForm is based on a multi-stage processing sequence. This sequence includes reading the CSV files, recognizing and aggregating the relevant data records, mapping them to the model objects, and assigning the sustainability indicators via the Desite md API. The logic is controlled entirely by JavaScript, while the user interface is implemented in HTML. Both components interact closely with each other via the Desite md API, so that all data is processed directly in the model. The script links the environmental indicators from the ÖKOBAUDAT with the corresponding component instances in the model in Desite md. This is based on reading the UUIDs from the ÖKOBAUDAT and the associated identification attributes in the model. The relevant field names for identification and assignment parameters can be defined by the user directly in the user interface. This allows the script to be flexibly adapted to different model structures, data exports and attribute names.

First, the UUID column is searched for the ID in the model and the indicator columns for example for GWP and PERT are identified. In the next step, the BIM objects are read into Desite md. Here, either all objects or only the currently selected elements can be processed, so only a selection of relevant objects can be evaluated as well. Objects without valid identification are given the status ‘no match’. If there is a valid match, the corresponding sustainability values are assigned to the BIM objects. The

process is the same for the production system, except a company-specific equipment ID is used to link the database.

During processing, a tabular overview of the processed objects is created. This overview contains not only object names and IDs, but also the mapping status and all transferred sustainability indicator values, their reference units, and the indicator totals. Since the reference unit of the LCA data set is identified automatically, the indicator total is calculated by automatically multiplying the indicator value by the correct unit of measurement. In addition, several key figures are calculated automatically. These include the total number of objects in the model, the number of objects with a valid UUID, the number of objects without a match, and the sum of the calculated GWP values, separated by buildings and production equipment, as well as the total. While there are 4,202 objects in the tested model, 14.8% of them are unmatched. This is because some of the objects, such as rooms or small parts of the façade, do not have a UUID. The result can be saved as a CSV file using the export function. The key figures are then written back to each BIM object as attributes and are therefore suitable as a basis for variant comparisons, optimization and heatmap visualizations.

To test and evaluate this WebForm, an example model containing objects of building construction and production equipment is used.

The result is an evaluation of the entire factory, including the building and the production equipment. For this evaluation, multiple databases are linked to the BIM model, allowing the results to be written back to the BIM objects.

4.1 Discussion

The results of this study are to be interpreted with consideration of the following limitations, which also represent potential for further research. First, there is currently insufficient availability of environmental datasets to consistently establish one-to-one links between BIM objects and LCA data using unique identifiers, such as UUIDs. Consequently, future work will require a mixed linking approach combining identifier-based assignment with direct assignment of LCA indicator values to BIM objects. Second, component service lives are not yet integrated into the assessment framework. As a result, the replacement of building elements and technical components over the life cycle is not yet accounted for. Addressing this limitation will require either an enrichment of the underlying databases with service life information or the introduction of classification-based matching using external databases for typical life spans of different construction and equipment types. Third, energy demand is considered exclusively at the level of production equipment. Energy consumption associated with technical building systems beyond machinery is not included, leading to an unrealistic representation of operational impacts. Nevertheless, this exemplary implementation indicates the potential and feasibility of the proposed workflow.

Despite existing limitations, the workflow fulfills the initially outlined requirements by enabling a holistic assessment of a factory, including the building and the production system. Additionally, the logic can be transferred and expanded to various factory elements. Overall, the proposed LCA data set matching workflow based on unique identifiers in the IFC model can significantly reduce the effort required for a factory sustainability assessment.

5 Conclusion

This work introduces a semi-automated openBIM workflow for the LCA of factories, integrating building construction and production systems within a digital process. By embedding LCA-relevant information directly into BIM objects through standardized identifiers, the workflow enables a consistent linkage between planning models and environmental datasets. The exemplary implementation using Desite md and WebForms illustrates the technical feasibility of identifier-based data linking for both building components and production equipment using a BIM viewer. However, the workflow must be tested on several different, more complex BIM models as a next step. Future work also requires incorporating component service life data into the employed databases and integrating technical building systems. Implementing these changes will improve the accuracy and comprehensiveness of the proposed workflow and build on the demonstrated potential and feasibility of this study.

Acknowledgements

The project “Evaluation of the ecological sustainability of factories with Building Information Modeling (BIM) (ecoFIM)” [01IF23418N] is funded by the Federal Ministry of Economic Affairs and Climate Action as part of the “Industrielle Gemeinschaftsforschung” program based on a resolution of the German Bundestag.

6 References

[1] S. Kadner, J. Kobus, M. R. Stuchtey, and T. Weber, "Circular Economy Roadmap für Deutschland.," acatech Studie. R. acatech/SYSTEMIQ, München/London, 2021.

[2] M. Landherr, M. Neumann, J. Volkmann, and C. Constantinescu, "Digitale Fabrik," in *Digitale*

Produktion, E. Westkämper, Ed., Berlin, Heidelberg: Springer Vieweg, 2013. [Online]. Available: https://doi.org/10.1007/978-3-642-20259-9_12

[3] M. Roesch, J. Köberlein, M. Goldmanns, and A. Hohmann, "Green-Lean-Digital als Leitbild für die nachhaltige Fabrik der Zukunft," *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 116, 1-2, pp. 29–33, 2021, doi: 10.1515/zwf-2021-0006.

[4] L. Philipp, L. Große-Wilde, L. Rieke, S. Hartmann, M. Schmidt, and K. Klemt-Albert, "Environmental Sustainability Assessment Of Factories With BIM: Relevance And Challenges For A Holistic Approach," in *Conference on Production Systems and Logistics*, Lima, Peru, 2025.

[5] J. H. Franz, *Nachhaltige Entwicklung technischer Produkte und Systeme: Der Ingenieurberuf im Wandel,* 1st ed. Wiesbaden: Springer Fachmedien Wiesbaden; Imprint Springer Vieweg, 2021.

[6] L. Rieke, J. Kerstgens, M. J. Kaufmann, C. Blanke, J. Reichardt, and M. Schmidt, "Bewertung der Zirkularität und Minimierung von CO_2-Emissionen im Gesamtsystem Industriebau – Case Study einer Gebäudebewertung mit madaster," (in ger), 2024, doi: 10.15488/17979.

[7] H.-P. Wiendahl, J. Reichardt, and P. Nyhuis, *Handbuch Fabrikplanung: Konzept, Gestaltung und Umsetzung wandlungsfähiger Produktionsstätten,* 3rd ed. München: Hanser Verlag, 2024.

[8] U. Bracht, D. Geckler, and S. Wenzel, "Anwendungen der Digitalen Fabrik," in *VDI-Buch, Digitale Fabrik: Methoden und Praxisbeispiele*, U. Bracht, D. Geckler, and S. Wenzel, Eds., 2nd ed., Berlin, Heidelberg: Springer Vieweg, 2018, pp. 259–373.

[9] Deutsche Gesellschaft für Nachhaltiges Bauen, *Das DGNB-Zertifizierungssystem.* [Online]. Available: https://www.dgnb-system.de/de/system/index.php (accessed: May 20 2022).

[10] Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen, Ed., "Bewertungssystem Nachhaltiges Bauen (BNB)," Accessed: Feb. 18 2025. [Online]. Available: https://www.bnb-nachhaltigesbauen.de/bewertungssystem/buerogebaeude/

[11] D. Schneider, P. Jordan, J. Dietz, M. F. Zaeh, and G. Reinhart, "Concept for Automated LCA of Manufacturing Processes," *Procedia CIRP*, vol. 116, pp. 59–64, 2023, doi: 10.1016/j.procir.2023.02.011.

[12] T. Heinemann, S. Thiede, K. Müller, B. Berning, J. Linzbach, and C. Hermann, "Life cycle evaluation of factories: Approach, tool and case study," in *11th Global Conference on Sustainable Manufacturing*, 2013, pp. 479–484.

[13] R. Schumacher *et al.,* "Analysis of current practice and future potentials of LCA in a BIM-based design process in German," *E3S Web Conf.*, no. 349, 2022. [Online]. Available: https://doi.org/10.1051/e3sconf/202234910004

[14] L. Wastiels and R. Decuypere, "Identification and comparison of LCA-BIM integration strategies," *IOP Conference Series: Earth and Environmental Science*, no. 323, 2019. [Online]. Available: doi://10.1088/1755-1315/323/1/012101

[15] A. Meckelnborg, *Integrative Fabrikplanung durch effiziente Koordinationsmodelle*. Zugl.: Aachen, Techn. Hochsch., Diss., 2015, 1st ed. Aachen: Apprimus Verlag, 2015.

[16] C.-G. Grundig, *Fabrikplanung: Planungssystematik – Methoden – Anwendungen,* 7th ed. München: Hanser, 2021. [Online]. Available: https://www.hanser-elibrary.com/doi/book/10.3139/9783446470064

[17] P. Burggräf and G. Schuh, *Fabrikplanung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021.

[18] D. Lindner and C. Leyh, "Digitalisierung von KMU – Fragestellungen, Handlungsempfehlungen sowie Implikationen für IT-Organisation und IT-Servicemanagement," *HMD*, vol. 56, no. 2, pp. 402–418, 2019, doi: 10.1365/s40702-019-00502-z.

[19] D. Lindner and C. Leyh, "Organizations in Transformation: Agility as Consequence or Prerequisite of Digitization?," in *Lecture Notes in Business Information Processing, Business Information Systems*, W. Abramowicz and A. Paschke, Eds., Cham: Springer International Publishing, 2018, pp. 86–101.

[20] F. H. Abanda, C. Vidalakis, A. H. Oti, and J. Tah, "A critical analysis of Building Information Modelling systems used in construction projects," *Advances in Engineering Software*, vol. 90, pp. 183–201, 2015, doi: 10.1016/j.advengsoft.2015.08.009.

[21] A. Criminale and S. Langar, "Challenges with BIM Implementation: A Review of Literature," *Conference: 53 rd ASC Annual International Conference Proceedings*, 2017.

[22] J. Mohr, N. Schmidtke, and F. Behrendt, "Design of Matrix Production Systems: New Demands on Factory Planning Methods," *Procedia Computer Science*, vol. 232, pp. 1972–1981, 2024, doi: 10.1016/j.procs.2024.02.019.

[23] S. Theißen *et al.,* "BIM integrated automation of whole building life cycle assessment using German LCA data base ÖKOBAUDAT and Industry Foundation Classes," in *IOP Conference Series: Earth and Environmental Sciences*, 2020.

[24] M. Despeisse, P. D. Ball, S. Evans, and A. Levers, "Industrial ecology at factory level – a conceptual model," *Journal of Cleaner Production*, vol. 31, pp. 30–39, 2012, doi: 10.1016/j.jclepro.2012.02.027.

[25] M. Gebler, J. F. Cerdas, S. Thiede, and C. Herrmann, "Life cycle assessment of an automotive factory: Identifying challenges for the decarbonization of automotive production – A case study," *Journal of Cleaner Production*, vol. 270, 2020, doi: 10.1016/j.jclepro.2020.122330.

[26] M. Schenk, S. Wirth, and E. Müller, *Fabrikplanung und Fabrikbetrieb: Methoden für die wandlungsfähige, vernetzte und ressourceneffiziente Fabrik,* 2nd ed. Berlin, Heidelberg: Springer Vieweg, 2014.

[27] M. Perau, N. Laubach, T. Schröer, W. Boos, and G. Schuh, "Information-based Integration of Life Cycle Assessment into IT Landscapes of Manufacturing Companies," *IFAC-PapersOnLine*, vol. 58, no. 19, pp. 1204–1209, 2024, doi: 10.1016/j.ifacol.2024.09.089.

[28] A. Dér, L. Hingst, P. Nyhuis, and C. Herrmann, "A review of frameworks, methods and models for the evaluation and engineering of factory life cycles," *Advances in Industrial and Manufacturing Engineering*, vol. 4, p. 100083, 2022, doi: 10.1016/j.aime.2022.100083.

[29] C. Bueno and M. M. Fabricio, "Comparative analysis between a complete LCA study and results from a BIM-LCA plug-in," *Automation in Construction*, vol. 90, pp. 188–200, 2018, doi: 10.1016/j.autcon.2018.02.028.

[30] *DIN EN 15978 - Nachhaltigkeit von Bauwerken – Bewertung der Umweltleistung von Gebäuden – Methodik,* DIN-Normenausschuss Bauwesen (NABau), 2024.

[31] Deutsche Gesellschaft für Nachhaltiges Bauen - DGNB e.V., *Sustainable building with the DGNB: Europe's largest network for sustainable building, the globally recognised DGNB certification and much more.* [Online]. Available: https://www.dgnb.de/en (accessed: Dec. 19 2025).

[32] BBSR, *Evaluation Criteria for Laboratory Buildings.* [Online]. Available: https://www.bnb-nachhaltigesbauen.de/en/assessment-system/laboratory-buildings/ (accessed: Dec. 19 2025).

[33] M. Gebler, *Planung und Gestaltung einer Positivfabrik als Beitrag für eine absolute Nachhaltigkeit*. Wiesbaden: Springer Vieweg, 2022.

[34] M. Lambertz, S. Theißen, J. Höper, R. Wimmer, A. Meins-Becker, and M. Zibell, "Ökobilanzierung und BIM im Nachhaltigen Bauen," Forschungsendbericht im Auftrag des Bundesinstituts für Bau-, Stadt- und Raumforschung im Bundesamt für Bauwesen und Raumordnung, 2019.

[35] Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen, "ÖKOBAUDAT," Accessed: Feb. 28 2025. [Online]. Available: https://www.oekobaudat.de/

[36] buildingSMART international, *OpenBIM.* [Online]. Available: https://www.buildingsmart.org/about/openbim/ (accessed: Dec. 19 2025).

[37] A. Hollberg, G. Genova, and G. Habert, "Evaluation of BIM-based LCA results for building design," *Automation in Construction*, vol. 109, 2020. [Online]. Available: https://doi.org/10.1016/j.autcon.2019.102972

[38] K. Forth, J. Abualdenien, and A. Borrmann, "Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing," *Energy and Buildings*, vol. 284, p. 112837, 2023, doi: 10.1016/j.enbuild.2023.112837.

[39] R. Ayman Mohamed, Z. Alwan, M. Salem, and L. McIntyre, "Automation of embodied carbon calculation in digital built environment- tool utilizing UK LCI database," *Energy and Buildings*, vol. 298, p. 113528, 2023, doi: 10.1016/j.enbuild.2023.113528.

[40] H. Madathala, G. Yeturi, V. Mane, and P. D. Muneshwar, "Navigating SAP ERP Implementation: Identifying Success Drivers and Pitfalls," in *International Conference on Intelligent Data Communication Technologies and Internet of Things*, India, 2025, pp. 75–83.

[41] *VDI 2552 Sheet 7: Building information modeling — Processes*, 2552, VDI – The Association of German Engineers.

[42] G. Genova, "BIM-based LCA throughout the design process: A dynamic approach," *WIT Transactions on The Built Environment*, Vol 192, pp. 45–56, 2019, doi: 10.2495/BIM190051.

[43] V. Sagarkar, *Bridging the Gap: How BIM Integration with SAP Transforms Construction Project Management into 5D BIM process.* [Online]. Available: https://www.desapex.com/blog-posts/bridging-the-gap-how-bim-integration-with-sap-transforms-construction-project-management (accessed: Dec. 16 2025).

[44] *Building construction - Organization of information about construction works - Part 2: Framework for classification*, 12006-2, DIN German Institute for Standardization, Jul. 2020.

[45] ECLASS e.V., *ECLASS.* [Online]. Available: https://eclass.eu/ (accessed: Dec. 20 2025).

[46] Mittelstand Digital Zentrum, *ECLASS product data standard: Standardized classification and description of products and services.* [Online]. Available: https://www.mittelstand-digital-wertnetzwerke.de/interoperabilitaet/standards-im-e-business/klassifikation/eclass/ (accessed: Dec. 20 2025).

[47] *Building Information Modeling Klassifikationssysteme: Blatt 9*, 2552, Verein Deutscher Ingenieure e. V., 2020.

[48] *External Link for CSV and IFC export settings.* [Online]. Available: https://www.dropbox.com/scl/fo/u1kom999hajeah6ejxjo0/AITDWfCj9pc7ZNPj_Rd2w0M?rlkey=0wo30l1kwhxwv7qm7oss1iy7x&st=3i9rxv2q&dl=0

Methods for finding photos of parts of buildings in virtual space

Ishida Kosei[1]

[1]Associate Professor, Department of Architecture, Faculty of Creative Science and Engineering, Waseda University, Doctor of Engineering
k_ishida@waseda.jp

Abstract

We created a support system for inspecting the exterior appearance by loading a three-dimensional photogrammetric model into a virtual space. This system shows a photograph of what the observer sees in virtual space.

To realize this system, we devised a method to represent buildings as voxels, check whether the voxels are reflected from shooting points in virtual space, and register shooting points as voxels. The system allows the viewer to scan the voxel in front of their line of sight in virtual space and then displays a photo of the shooting point registered to the voxel.

We created software that implements this mechanism and confirmed that the necessary photos can be obtained promptly.

Keywords –

Photogrammetry; 3D model; Voxel; Inspection; VR

1 Introduction

The technology for creating three-dimensional (3D) models based on photographs is used in various situations in Japan. Many people use 3D maps, such as Google Maps daily. Three-dimensional models created using photogrammetry are essential for building inspections and patrols. Research is also being conducted on the technology to perform inspections in a virtual space.

Inspection using photographs confirms the disaster situation. Research has been conducted to predict earthquake damage using images and photometry [1]. In addition, to reduce the environmental impact, there are methods for checking the status of heat loss using photographs and thermography [2][3]. We considered the digitalization of buildings by combining UAV, photogrammetry, and BIM [4]. Photogrammetry and deep learning have been used to detect cracks and perform exterior inspections of buildings [5]. Similarly, this method has been applied to the external inspection of bridge piers [6]. Photogrammetry has been used to survey the occupancy status of buildings [7]. Furthermore, some studies have been conducted on visual inspection of tunnels [8] and remote inspection of wind power generators [9]. Photogrammetry has also been used for inspecting dam bodies [10].

However, as the surface texture of 3D models created using photogrammetry is rough, referring to an original photograph is often desirable. Because photogrammetry uses various photographs, it is often difficult to identify a desired photograph. Therefore, we aimed to develop a technology that would allow us to smoothly obtain detailed photographs of parts of 3D models.

2 Methods for finding photos of parts of buildings in virtual space

2.1 Procedure to search for photos that depict parts of buildings in virtual space

We plan to create a system to search for photographs of 3D models of buildings prepared in virtual spaces. Because this 3D model was created using photogrammetry, an original photograph exists. We devised a method to search for photos of a specific part of a building's surface while observing the exterior of the building in a virtual space. A conceptual diagram of this concept is shown in Figure 1.

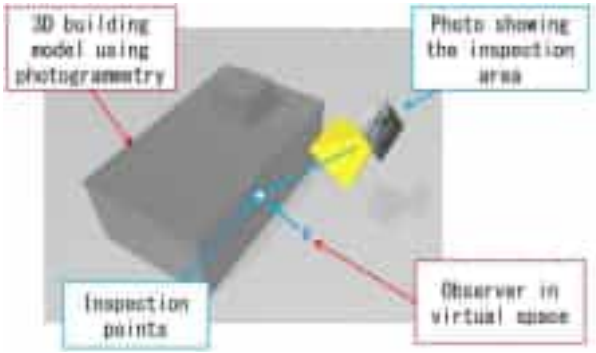


Fig 1 Conceptual diagram for finding a photo of a specific part of a building's surface in virtual space

The following procedure shows how to search for photographs depicting parts of buildings in a virtual space.

(1) Load the 3D model created using photogrammetry and photographic points.
(2) Replace the 3D model with voxels using photogrammetry
(3) Set up a camera at a shooting point in the virtual space and find all voxels within the range of the camera view.
(4) Record all shooting points on the voxel side.

2.2 Registration of shooting points to voxels indicating buildings

As described in Section 2.1, the buildings were made into voxels, and photographic points were registered for these voxels. When observing a structure in a virtual space and looking for a photograph of a part of a building, one can search based on the shooting points registered in the voxels.

Subsequently, we explain how to register shooting points for each voxel of this building.

This voxel registration method is straightforward: it reproduces the camera's shooting point and direction in the virtual space, captures a photo from the camera in the virtual space, and searches for voxels that can be observed from the camera. This approach is currently being implemented.

Figure 2 shows a photograph of the registered camera shooting points on the voxels of the building.

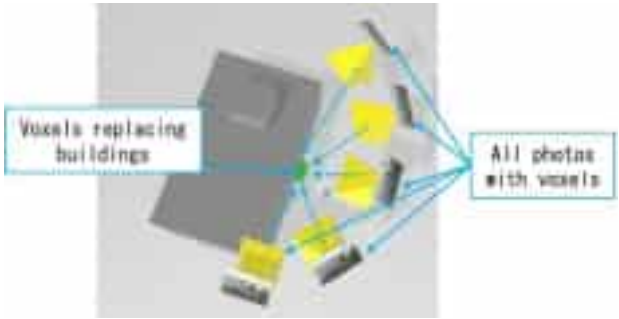


Fig 2 Image of registering camera shooting points to building voxels.

To register shooting points to voxels, the system reproduces the actual angle of view of each camera in the virtual space. We implemented a ray-casting approach where rays are emitted from the camera position with a sampling interval of 50 pixels in both horizontal and vertical directions across the image plane. This process identifies all visible voxels for each photograph and records the relationship in a pre-computed lookup table. By shifting the computational burden to this pre-processing stage, the system ensures high performance during real-time interaction.

2.3 Method for an observer in virtual space to search for photos of parts of a building

We explain how an observer can search for a photo that shows the observation point of a building in a virtual space using a 3D model in which photo points are registered in the voxels of the building.

When an observer in a virtual space observes a part of a building, the ray is irradiated in the direction of the center of their line of sight. Next, a list of photograph shooting points associated with the voxel hit by the ray is obtained. This list of photo points shows the parts of the building being observed so that one can efficiently search for photos used during photogrammetry.

Figure 3 shows how to search for the original photo when observing a building in a virtual space.

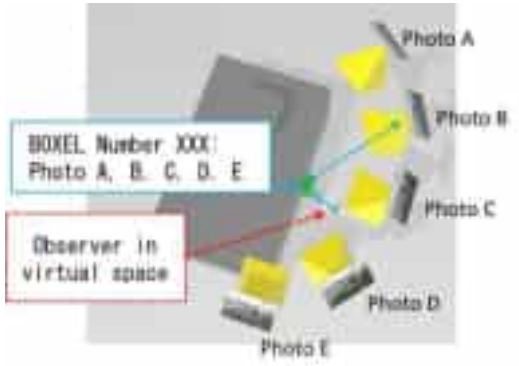


Fig 3 Searching for the original photo when observing a building in a virtual space.

3 Development of the finding system for photos of parts of buildings in virtual space

3.1 Creating a 3D model using photogrammetry

We create a system to search for photos in a virtual space, as described in Section 2. To achieve this, we created a 3D model using photogrammetry. The photograph used is shown in Figure 4. We cut the aerial videos captured using a drone into 960 still photographs.

Figure 5 shows the 3D model created using this photograph. Bentley context-capture photogrammetry software was used.



Fig 4 Examples of photos used to create 3D models



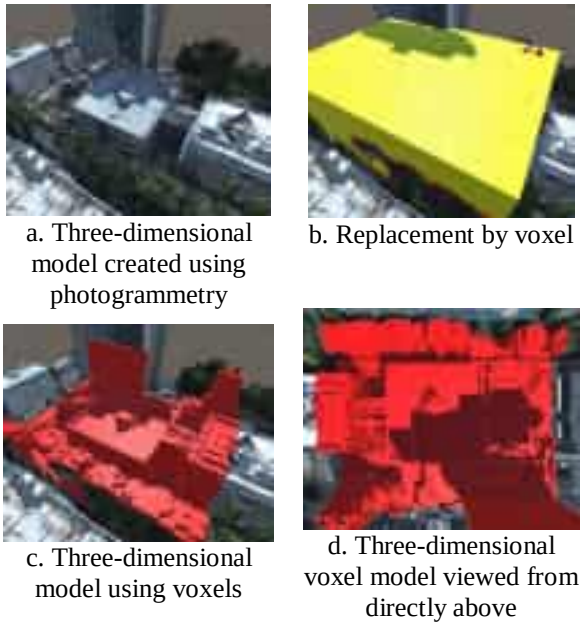
Fig 5 Three-dimensional model created using photogrammetry.

3.2 Replace the 3D model by photogrammetry with voxels

Subsequently, the 3D model created using photogrammetry was replaced with voxels. Using voxels with a side length of 1 m, the shape of the building and its surroundings was reproduced. We created a 3D model using voxels by the following steps:

(1) Place a 1 m-sized cube (voxel) at the location where the 3D model created using photogrammetry exists.
(2) If the 3D model and 1 m size voxel overlap, place the voxel.

Figure 6 illustrates the procedure used to replace the voxels. The system placed 1 m-sized cubes without gaps, as shown in Figure 6b, in a space containing the 3D model shown in Figure 6a. The yellow cubes represent space, and the red cubes represent voxels placed in the 3D model. Figures 6c and d show the results of replacing the 3D model with voxels.



a. Three-dimensional model created using photogrammetry

b. Replacement by voxel

c. Three-dimensional model using voxels

d. Three-dimensional voxel model viewed from directly above

Fig 6 Results of replacing a 3D model of a building created using photogrammetry with voxels

3.3 Set up a camera at a shooting point in virtual space and find all voxels within the range that the camera can see

Next, our system searches for voxels of buildings that are visible from the point where the photo was captured. The procedure is as follows:

(1) Reproduce the photographing point and direction of a photograph in a virtual space
(2) A camera that reproduces the actual angle of view in the virtual space is created and installed at the shooting point.
(3) The shooting range is irradiated with rays to clarify the range seen from the camera in the virtual space.
(4) The photo number in the voxel hit by the ray is recorded.

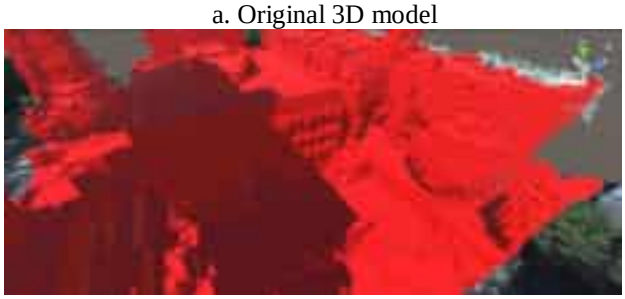
We used the data estimated using photogrammetry software for the camera shooting point and direction.

Figure 7 shows a situation in which the range

captured by the camera in the virtual space is confirmed. Figure 7b shows the range observed by the camera in the virtual space. Figure 7c shows a situation in which a ray is irradiated to investigate the range that the camera is viewing.



a. Original 3D model



b. Result after replacing the building with voxels



c. Situation where Ray is emitted from the camera

Fig 7: Range visible from a camera in virtual space is being investigated.

Figure 8 shows a situation in which the camera in the virtual space is moved to other shooting points, and the shooting range is investigated from each shooting point.

3.4 Record all the shooting points that you can see on the voxel

The number of photographs captured for each voxel was recorded by repeatedly searching for voxels within the visible range of the camera. In this section, we present the results of recording the shooting points in the voxels of buildings using the proposed system. Figure 9 shows the number of photographs captured for each voxel. The blue voxels indicate images captured at 50 or more locations. Figure 10 shows the number of photographs recorded for each voxel.

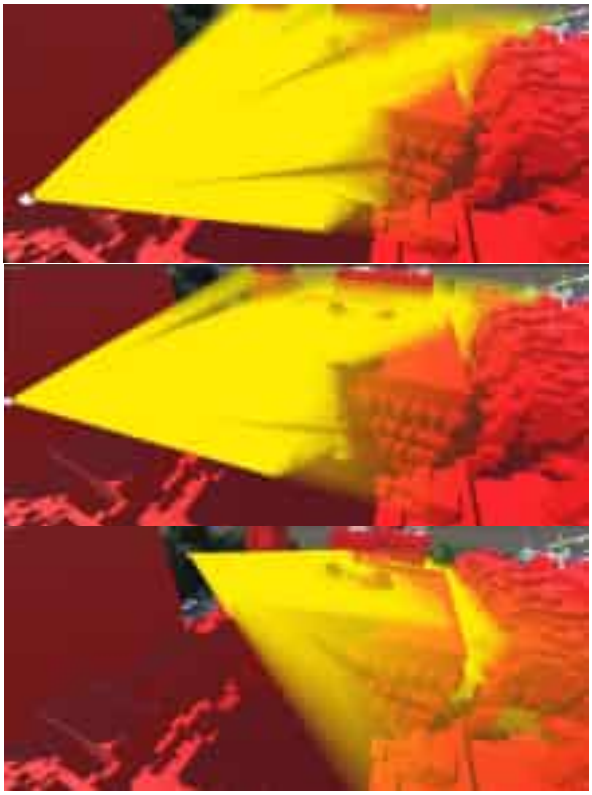


Fig 8 Situation where the shooting range is being investigated from each shooting point

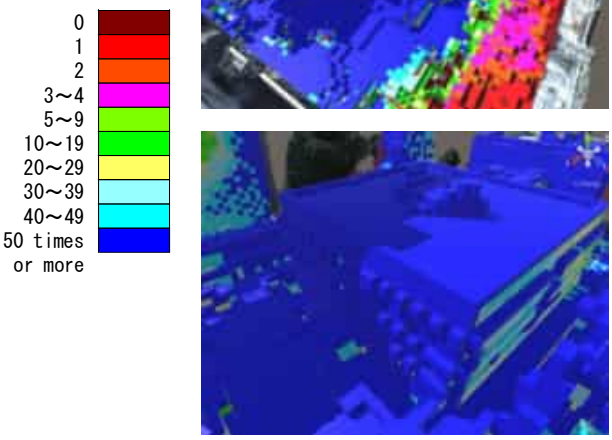


Fig 9 Number of times a voxel appeared in a photo.



Fig 10 Results of recording the number of photo shoots for each voxel

The proposed system was implemented as a Unity-based WebGL application. During the inspection in virtual space, when an observer's gaze hits a voxel, the system retrieves the associated photo list near-instantaneously. Since the spatial relationship is already indexed in a pre-built data structure, the retrieval complexity is $O(1)$, allowing for a seamless transition from the 3D model to original high-resolution photographs without any noticeable latency.

4 System for an observer in virtual space to search for photos of parts of a building

4.1 Building a system to search for photos of building surfaces in virtual space

Using the method described in Section 3, we registered the shooting points within the voxels. Next, we created a system that smoothly searched for photos captured at the observation points while observing a 3D model in a virtual space.

First, a 3D model created using photogrammetry was loaded, and then the point at which the photo was captured was reproduced in the virtual space. Finally, the voxels described in Section 3 were placed. This voxel was placed in a transparent state—place voxel as shown in Figs. 11 and 12.

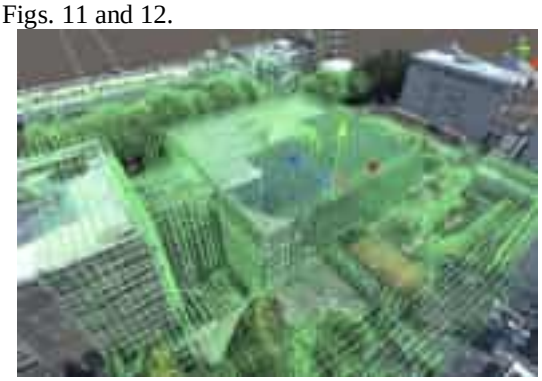


Fig 11 Result of placing transparent voxels

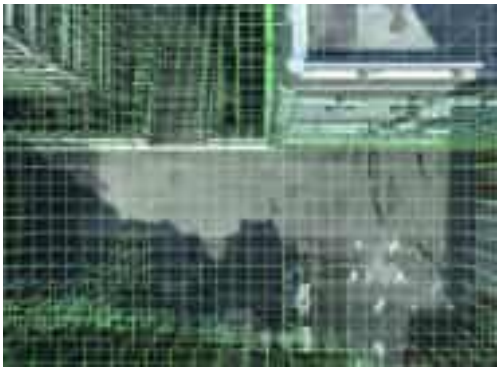


Fig 12 Close-up view of the result of placing transparent voxels

4.2 Search results for all photos depicting observation points

Based on the results obtained thus far, we created a system that shows all the photographs captured at the observation point when a building is observed in a virtual space.

Figure 13 shows the results of running the created system. In Figure 13, the voxel where the observer is looking in the virtual space is displayed, all shooting points registered to this voxel are shown, and a line is drawn between the voxel and the shooting point.



Fig 13 Result of the observation system execution

Figure 14 shows the viewpoint of the observer in the virtual space. In particular, voxels viewed by the observer are shown. The purple line extends to the shooting point that captures the voxel ahead of the line of sight.



Fig 14 Observer's perspective

Figure 15 shows an example of acquiring voxel numbers on a building surface as observed by an observer. By accepting this voxel number, a list of photographs showing the voxels can be obtained, as shown in Figure 16. A photo number can be obtained by selecting any photo from the photo list. Next, as shown in Figure 17, the directory in which the photo was stored was obtained based on the photo number, and the photograph was displayed. The desired photograph is shown in Figure 18.

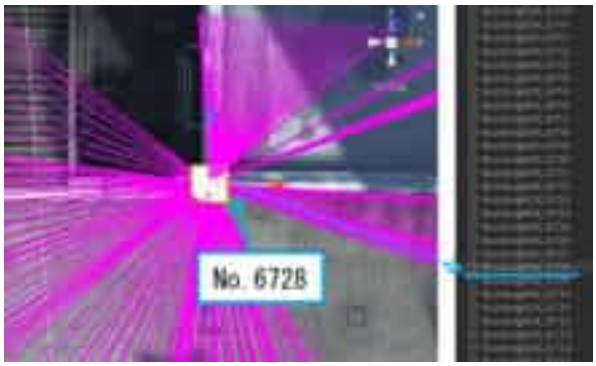


Fig 15 Example of obtaining the number of the voxel ahead of the line of sight



Fig 16 List of photo numbers that capture the voxel ahead of your line of sight



Fig 17 Situation to obtain photo storage directory from photo number

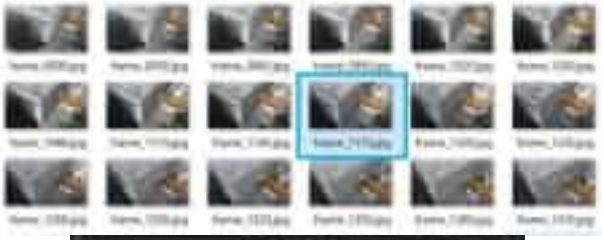




Fig 18 Example of displaying the desired photo

We selected a voxel size of 1 m for the building representation. This resolution is strategically determined based on the practical requirements of building exterior inspections, such as identifying cracks or concrete spalling. A 1 m grid provides a sufficient level of detail to localize defects on a building scale while ensuring that each voxel is covered by an adequate number of overlapping photographs from various drone shooting positions, which is essential for a reliable photogrammetric inspection.

5 Conclusion

We created a system that searches for all photographs captured from observation points when a 3D model created using photogrammetry is observed in the virtual space. Using the developed system, we obtained an original photograph of the location observed in virtual space. This results in more efficient inspections using building models in a virtual space.

References

[1] Ewelina Rupnik, Francesco Nex, Isabella Toschi and Fabio Remondino: Contextual classification

using photometry and elevation data for damage detection after an earthquake event, European Journal of Remote Sensing, 51(1), pp.543–557, 2018.5, https://doi.org/10.1080/22797254.2018.1458584

[2] M. Hasan Shariq and Ben Richard Hughes: Revolutionising building inspection techniques to meet large-scale energy demands: A review of the state-of-the-art, Renewable and Sustainable Energy Reviews, Volume 130, 2020, https://doi.org/10.1016/j.rser.2020.109979

[3] Mariluz Gil-Docampo, Juan Ortiz Sanz, Ignacio Cañas Guerrero and Manés Fernández Cabanas :UAS IR-Thermograms Processing and Photogrammetry of Thermal Images for the Inspection of Building Envelopes, Appl. Sci. 13(6), 3948, 2023, https://doi.org/10.3390/app13063948

[4] Didit Gunawan Prasetyo Jati: UAV-BASED PHOTOGRAMMETRY DATA TRANSFORMATION AS A BUILDING INSPECTION TOOL: APPLICABILITY IN MID-HIGH-RISE BUILDING, Jurnal TEKNIK SIPIL, 16(2), pp.113–121, 2021, https://doi.org/10.24002/jts.v16i2.4777

[5] P. Ko a, S. A. Prieto a, B. García de Soto :ABECIS: an Automated Building Exterior Crack Inspection System using UAVs, Open-Source Deep Learning and Photogrammetry, 38th International Symposium on Automation and Robotics in Construction, pp.637-644, 2021, https://doi.org/10.22260/ISARC2021/0086

[6] Sara Zollini, Maria Alicandro, Donatella Dominici, Raimondo Quaresima and Marco Giallonardo: UAV Photogrammetry for Concrete Bridge Inspection Using Object-Based Image Analysis (OBIA), Remote Sens., 12(19), 3180, 2020, https://doi.org/10.3390/rs12193180

[7] Seung Hyeon Lee, Seung Hak Woo, Jung Rim Ryu, and Seung Yeon Choo: Automated building occupancy authorization using BIM and UAV-based spatial information: photogrammetric reverse engineering, Journal of Asian Architecture and Building Engineering, Volume 18, Issue 2, Pages 151–158, 2019, https://doi.org/10.1080/13467581.2019.1631172

[8] F. Panella, N. Roecklinger, L. Vojnovic, Y. Loo, and J. Boehm: COST-BENEFIT ANALYSIS OF RAIL TUNNEL INSPECTION FOR PHOTOGRAMMETRY AND LASER SCANNING, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020 XXIV ISPRS Congress (2020 edition), https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1137-2020

[9] Dayi Zhang, Robert Watson, Gordon Dobie, Charles MacLeod, Aamir Khan, Gareth Pierce: Quantifying impacts on remote photogrammetric inspection using unmanned aerial vehicles, Engineering Structures, Volume 209, 15 April 2020, 109940, https://doi.org/10.1016/j.engstruct.2019.109940

[10] Sizeng Zhao, Fei Kang, Junjie Li, Chuanbo Ma:Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction, Automation in Construction, Volume 130, 2021, 103832, https://doi.org/10.1016/j.autcon.2021.103832

A Spatial Multi-Criteria Framework for Assessing Last-Mile Delivery Resilience in Modular Construction Supply Chains: A Network-Based Analysis of Lille

Saad EL BABIDI[1] **and Zoubeir LAFHAJ**[1]

[1]École Centrale de Lille, University of Lille, France

saad.el-babidi@centralelille.fr, zoubeir.lafhaj@centralelille.fr

Abstract -

The resilience of construction supply chains has become a critical issue in dense urban environments, particularly for modular construction systems that rely on large and time-sensitive deliveries. While existing studies have extensively addressed construction logistics and supply-chain resilience at strategic and regional levels, the last-mile delivery phase within urban areas remains insufficiently analyzed. This paper proposes a spatially explicit methodological framework to assess last-mile delivery resilience for modular construction in urban contexts. The framework is based on a grid-based, multi-criteria analysis integrating road access suitability, truck-feasible network redundancy, and urban morphology operability. The city of Lille was selected as a representative dense European urban last-mile context and was discretized into a regular grid to ensure spatial comparability and reproducibility. Open geospatial data were used to model road networks and urban form, and normalized indicators were computed for each grid cell. The results showed that, although large portions of the study area remained physically accessible for modular deliveries, the truck-feasible road network exhibited very low redundancy, indicating a high sensitivity to local disruptions. Urban morphology was found to be a secondary limiting factor compared to network structure. The proposed framework enabled the identification of structurally fragile zones and critical corridors, providing a basis for targeted operational, territorial, and regulatory recommendations to improve last-mile delivery resilience for modular construction.

Keywords -

Modular construction; Last-mile delivery; Supply-chain resilience; Urban road networks; GIS-based multi-criteria analysis

1 Introduction

Urban construction systems increasingly face constrained spatial conditions, environmental requirements, and regulatory frameworks. In response, modular construction has gained growing attention as an alternative to conventional methods due to its potential to improve productivity, quality, and sustainability by shifting a significant share of activities from on-site execution to off-site manufacturing [1, 2]. This shift, however, increases project dependency on reliable logistics and precise delivery operations, particularly in dense urban environments.

In this context, the last mile of the construction supply chain—defined as the final segment connecting the urban transport network to the construction site—plays a critical role in project performance. Modular construction relies on transporting large and often oversized components under strict sequencing and time constraints, making last-mile accessibility and reliability decisive factors. Unlike other time-critical construction deliveries such as ready-mixed concrete or crane-assembled components, modular systems impose a compounded constraint: components are dimensionally fixed and non-substitutable, assembly sequences are predetermined and cannot tolerate partial or delayed deliveries, and off-site fabrication schedules are decoupled from the site, eliminating any on-site buffering capacity. A single delivery failure in a modular sequence thus propagates directly to installation downtime, with no on-site workaround available. Urban freight studies consistently identify last-mile operations as the most complex and disruption-prone segment due to congestion, access restrictions, and competition for limited urban space [3], with similar constraints observed in urban construction logistics, where delivery failures directly affect on-site productivity and installation schedules [4].

Supply-chain resilience has been widely studied in manufacturing and logistics and, more recently, in construction, emphasizing the ability of systems to absorb, adapt to, and recover from disruptions. Research on modular and prefabricated supply chains highlights transportation and logistics as critical determinants of resilience [2, 5]. However, most contributions remain project-centric or conceptual and rarely translate resilience principles into spatially explicit indicators applicable at the urban scale.

In parallel, geographic information systems (GIS) and multi-criteria decision analysis (MCDA) have been widely used in urban planning and infrastructure assessment to integrate heterogeneous spatial data within transparent and reproducible frameworks [6, 7]. Three GIS capabilities

are central to the present framework: (i) network analysis, enabling graph-based computation of route redundancy and alternative-path availability; (ii) spatial overlay operations, allowing heterogeneous data layers (road hierarchy, building density) to be aggregated over a common grid; and (iii) grid-based spatial discretization, ensuring that indicators are comparable and transferable across different urban contexts. Despite their potential, these methods have seldom been applied to characterize last-mile delivery resilience for modular construction, particularly through indicators explicitly accounting for road-network structure and redundancy.

As a result, a methodological gap persists at the intersection of modular construction, last-mile logistics, and urban system resilience: spatially explicit and network-informed frameworks capable of assessing last-mile delivery resilience using interpretable and transferable indicators remain limited. To address this gap, this paper proposes a GIS-based, network-informed multi-criteria framework integrating three complementary dimensions: road access suitability, truck-feasible network redundancy, and urban morphology operability. Indicators are computed over a regular spatial grid to enable comparability and transferability across urban contexts. The city of Lille (France) is used as a pilot case representative of dense European urban environments with pronounced last-mile constraints for construction logistics.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature. Section 3 presents the methodology and data sources. Section 4 reports the results for the Lille case study, Section 5 discusses the findings and recommendations, and Section 6 concludes and outlines future research perspectives.

2 State of the Art

This section reviews prior research at three nested levels: (i) supply-chain resilience in construction and modular/prefabricated delivery systems, (ii) last-mile logistics and construction logistics constraints in dense urban environments, and (iii) spatially explicit assessment methods (GIS-based indicators and multi-criteria approaches) enabled by open geospatial data. The review follows a funnel logic from general concepts to the specific gap addressed by this paper.

2.1 Supply-chain resilience in construction and modular systems

Supply-chain resilience is commonly defined as the capability of a supply network to anticipate, absorb, adapt to, and recover from disruptions while maintaining essential functions. In construction, resilience discussions have intensified due to increasingly complex procurement structures, fragmented stakeholders, and exposure to shocks (supply shortages, transport disturbances, regulatory changes). Recent synthesis works highlight that disruption research remains fragmented and that resilience operationalization often lacks consistent and measurable indicators comparable across contexts [8, 9].

Modular and prefabricated construction systems strengthen the coupling between manufacturing, transportation, and on-site assembly, thereby increasing sensitivity to delivery reliability and urban access constraints. Quantitative and conceptual contributions have addressed modular logistics through optimization models, buffering and sourcing strategies, and conceptual frameworks. For instance, two-stage stochastic programming has been used to optimize modular logistics decisions across manufacturing–inventory–transportation under demand variability, illustrating the value of mathematical decision-support primarily at the project planning level rather than at an urban last-mile territorial scale [10]. Systematic works further emphasize disruption drivers (including transportation and procurement as critical levers) and resilience mechanisms (e.g., redundancy, flexibility, collaboration, digitalization), yet rarely translate these concepts into spatially explicit indicators supporting local last-mile deployment [1, 2, 5].

2.2 Last-mile logistics and construction logistics in dense cities

Last-mile logistics is widely characterized as the most complex and least efficient segment of urban supply chains due to high uncertainty, limited space, conflicts with other urban functions, and dynamic regulatory constraints. Literature reviews highlight recurring barriers such as congestion, limited curb space, access restrictions, and the need to balance economic, environmental, and social performance [3]. Cities nevertheless often under-integrate last-mile logistics into planning despite its growing impact, calling for locally adapted operational and regulatory instruments [11].

Construction logistics intersects with city logistics but exhibits distinct characteristics: heavy and oversized deliveries, temporal criticality aligned with installation windows, safety constraints, and site-interface limitations. Empirical studies on construction logistics centers (CLCs) show that logistics strategies (consolidation vs. just-in-time deliveries) significantly influence delivery patterns and disturbances, while implementation barriers remain substantial [4, 12]. Despite these advances, most last-mile research still focuses on retail and e-commerce flows, whereas construction logistics studies often remain project-centric. Modular construction, whose installation is delivery-dependent and schedule-critical, therefore requires last-mile analysis that is both urban-system oriented

and operationally grounded.

2.3 Spatially explicit assessment and GIS-based multi-criteria approaches

Supply-chain resilience and accessibility challenges in dense urban contexts are increasingly addressed through GIS-based indicators and multi-criteria decision analysis (MCDA), enabling heterogeneous data layers (transport, morphology, constraints) to be integrated into comparable territorial scores [6, 7]. In parallel, network science provides measures (e.g., redundancy, alternative routing potential) to describe transport network robustness under disruption. However, in construction logistics, and particularly in modular last-mile delivery, these approaches are still rarely instantiated as reproducible indicator sets supporting urban zoning and actionable recommendations. Data availability constitutes a practical challenge: OSM, widely used for road-network extraction, reaches sufficient quality for urban analytical uses when limitations are explicitly acknowledged, though quality varies by geography and feature type [13, 14].

2.4 Synthesis of gaps and positioning of this paper

Three gaps emerge: modular supply-chain studies rarely operationalize last-mile constraints spatially [2, 5]; urban last-mile research targets retail flows rather than construction [3, 11]; and GIS-MCDA frameworks lack a bridge to modular last-mile resilience metrics.

3 Methodology and Data

A spatially explicit, network-informed multi-criteria framework was developed to assess urban last-mile delivery resilience for modular construction. The framework integrated (i) road access suitability, (ii) truck-feasible network redundancy, and (iii) urban morphology operability into normalized indicators computed over a regular grid. The approach followed established principles from GIS-based multi-criteria analysis and urban freight/logistics assessment, while explicitly operationalizing network redundancy as a resilience property [3, 7, 4].

3.1 Study area and spatial discretization

The city of Lille (France) was selected as a pilot case representative of dense European urban conditions where modular construction deliveries face significant last-mile constraints. The study area was defined as a polygonal extent within which all computations were performed.

The area was discretized into a regular grid of square cells of side length $\Delta = 250$ m. Let $\mathcal{G} = \{g_1, \ldots, g_N\}$ denote the set of grid cells, with $A_g = \Delta^2$ the area of a cell. A grid-based formulation was adopted to ensure spatial comparability and transferability across cities, as commonly required for spatial decision-support and multi-criteria mapping [7].

3.2 Open data sources and preprocessing

OpenStreetMap (OSM) data were used for road and building layers to ensure transparency and reproducibility, while explicitly acknowledging quality and fitness-for-use limitations [14, 13].

Road geometries and attributes (including the `highway` classification) were extracted and clipped to the study boundary. Building footprints were extracted using OSM `building` tags and filtered to retain polygonal geometries. All layers were re-projected to a metric coordinate reference system (Lambert-93 EPSG:2154) to enable consistent length and area computations.

3.3 Road-network representation and truck-feasible subnetwork

The road network was represented as a graph $\mathcal{N} = (V, E)$, where V denotes intersections and E road segments. Each edge $e \in E$ had a geometric length L_e and an OSM road class $h(e)$.

Truck feasibility was operationally defined based on OSM road classes by excluding residential, service, and pedestrian roads, consistent with typical heavy and oversized vehicle constraints in urban construction logistics. This filtering approximates the access constraints applicable to the standard semi-trailer configuration used for modular deliveries in France: total length 16–18 m, width 2.5 m, maximum authorized gross weight 44 tonnes. Such vehicles are structurally incompatible with residential and service roads due to geometric (turning radius, lane width) and load-bearing constraints. Vertical clearance restrictions (e.g., overpasses, underpasses) were not modeled at this stage, as OSM does not systematically provide `maxheight` attributes across the study area; this constitutes an explicitly acknowledged limitation (Section 6).

A truck-feasible subnetwork $\mathcal{N}_T = (V_T, E_T)$ was derived by retaining edges satisfying a feasibility rule $\phi(e) = 1$ based on road class and intended modular delivery constraints. This subnetwork enabled redundancy analysis under truck-feasible conditions, which is central to resilience assessment [2, 5].

3.4 Indicator definition and mathematical formulation

Three criteria were computed per grid cell $g \in \mathcal{G}$ and normalized to $[0, 1]$ for comparability. Their selection was grounded in the specific constraints of modular last-mile delivery. C1 (road access suitability) captures the physical compatibility of the road environment with oversized

vehicles—the necessary entry condition for any modular delivery. C2 (truck-feasible network redundancy) directly operationalizes resilience as the availability of alternative routing options under disruption, which is the defining structural vulnerability of modular logistics in dense urban networks. C3 (urban morphology operability) captures site-interface constraints related to building density, conditioning on-site maneuvering and staging capacity. Other potential factors—including real-time traffic, time-window regulations, and low-emission zones—were excluded because they require dynamic or proprietary data unavailable across cities; their influence is partially captured through road hierarchy classification and acknowledged in Section 6.

3.4.1 C1: road access suitability

C1 quantified the suitability of the local road environment for modular deliveries using road hierarchy as a proxy for capacity and maneuverability. Each road segment e intersecting grid cell g was assigned a suitability score $s(h(e)) \in [0, 1]$ based on its road class $h(e)$. Let E_g denote the set of edges intersecting g and $L_{e,g}$ the length of edge e within g.

The raw suitability score was defined as the length-weighted mean:

$$C1^{raw}(g) = \frac{\sum_{e \in E_g} s(h(e))\, L_{e,g}}{\sum_{e \in E_g} L_{e,g} + \varepsilon}, \tag{1}$$

where $\varepsilon > 0$ avoided division by zero. This formulation ensured proportional contribution of longer segments and boundedness in $[0, 1]$ prior to normalization.

3.4.2 C2: truck-feasible network redundancy

C2 operationalized local redundancy of the truck-feasible network as a resilience property using two complementary proxies: (i) truck-feasible edge-length density and (ii) branching opportunity density [8].

Let $E_{T,g}$ denote the set of truck-feasible edges intersecting cell g and define the edge-length density:

$$\rho_L(g) = \frac{\sum_{e \in E_{T,g}} L_{e,g}}{A_g}. \tag{2}$$

Let $V_{T,g}$ be the set of truck-feasible nodes within g, with degree $\deg_T(v)$ in $\mathcal{N}_T$. Define the branching node set $B_g = \{v \in V_{T,g} : \deg_T(v) \geq 3\}$ and branching density:

$$\rho_B(g) = \frac{|B_g|}{A_g}. \tag{3}$$

The raw redundancy score was computed as:

$$C2^{raw}(g) = \left(\frac{\rho_L(g)}{\max_{g' \in \mathcal{G}} \rho_L(g') + \varepsilon} \right) \left(\frac{\rho_B(g)}{\max_{g' \in \mathcal{G}} \rho_B(g') + \varepsilon} \right), \tag{4}$$

yielding high values only where sufficient truck-feasible infrastructure and multiple branching opportunities coexist.

3.4.3 C3: urban morphology operability

C3 characterized urban operability using building footprint density as a proxy for spatial constraints. Let $\mathcal{P}_g$ denote the set of building polygons intersecting grid cell g, and $A_{p,g}$ the area of polygon p within g. The built-up ratio was defined as:

$$\beta(g) = \frac{\sum_{p \in \mathcal{P}_g} A_{p,g}}{A_g}. \tag{5}$$

Higher built-up ratios implied tighter spatial conditions; operability was therefore defined via inversion and normalization (see Section 3.5).

3.5 Normalization and direction harmonization

All criteria were mapped to a common scale $[0, 1]$ using min–max normalization:

$$\mathcal{N}(x(g)) = \frac{x(g) - \min_{g' \in \mathcal{G}} x(g')}{\max_{g' \in \mathcal{G}} x(g') - \min_{g' \in \mathcal{G}} x(g') + \varepsilon}. \tag{6}$$

Final normalized indicators were defined as:

$$C1(g) = \mathcal{N}(C1^{raw}(g)), \tag{7}$$

$$C2(g) = \mathcal{N}(C2^{raw}(g)), \tag{8}$$

$$C3(g) = 1 - \mathcal{N}(\beta(g)). \tag{9}$$

Higher values therefore always corresponded to more favorable conditions for modular last-mile delivery resilience.

3.6 Linking the three criteria for zoning and decision support

To support interpretable zoning while avoiding black-box indices, a transparent composite score was optionally defined:

$$R(g) = w_1\, C1(g) + w_2\, C2(g) + w_3\, C3(g), \quad \sum_{i=1}^{3} w_i = 1,\ w_i \geq 0. \tag{10}$$

A baseline equal-weight configuration $w_1 = w_2 = w_3 = \frac{1}{3}$ was adopted as the epistemically neutral prior: in the absence of empirical weighting evidence (e.g., expert survey

data or revealed logistics preferences for the Lille context), equal weights avoid introducing arbitrary priorities that cannot be validated against field data at this pilot stage. A sensitivity check varying each w_i within [0.10, 0.60] confirmed that the spatial ranking of the most critical zones remained stable under all tested configurations [7].

3.7 Implementation and replicability

All steps (data extraction, preprocessing, spatial overlays, indicator computation, normalization, and visualization) were implemented in Python using open-source geospatial and network libraries. The workflow was fully scripted to ensure reproducibility given the same boundary definition, grid resolution, and OSM data snapshot.

4 Results

This section presents the results of the spatial multi-criteria analysis applied to the city of Lille. Results are reported separately for each criterion in order to distinguish the respective effects of road access suitability, truck-feasible network redundancy, and urban morphology operability on last-mile delivery resilience for modular construction. Both spatial patterns and statistical distributions are presented.

4.1 Road access suitability (C1)

Criterion C1 quantified the physical suitability of the local road network for modular construction deliveries based on road hierarchy and length distribution. Figure 1 shows the spatial distribution of C1 values across the 250 m grid covering the study area.

The results revealed a strong spatial heterogeneity in road access suitability. Grid cells located along major urban corridors and near primary road infrastructures exhibited high C1 values, indicating favorable access conditions for heavy and oversized vehicles. In contrast, dense residential areas characterized by a predominance of local and service roads were associated with low C1 values.

From a statistical perspective, the distribution of C1 values was highly asymmetric, with a large proportion of grid cells exhibiting very low access suitability. The median C1 value was equal to zero, reflecting the structural predominance of constrained road segments in the urban fabric. High C1 values were limited to a small number of grid cells, mainly aligned with major transport axes.

4.2 Truck-feasible network redundancy (C2)

Criterion C2 assessed the redundancy of the truck-feasible road network by evaluating the availability of alternative routes and branching opportunities within each grid cell. The spatial distribution of C2 values is presented in Figure 2.



Figure 1. Spatial distribution of road access suitability (C1) for modular construction last-mile delivery in Lille (250 m grid). Normalized scale 0–1: $C1 < 0.2$ (red) indicates low suitability (predominantly residential and service roads); $C1 > 0.8$ (green) indicates high suitability (primary arterial corridors); intermediate values reflect mixed road-class environments.

The results showed that network redundancy was extremely limited across most of the study area. A very large majority of grid cells exhibited C2 values close to zero, indicating the absence of locally available alternative truck-compatible routes. Higher C2 values were observed only in a small number of locations, generally near major intersections and high-capacity road segments.

The statistical distribution of C2 values was strongly skewed toward zero, with a mean value close to zero and very limited dispersion. This pattern is also structurally expected from the multiplicative formulation of C2 (Eq. 4): high values require both truck-feasible edge-length density and branching node density to be simultaneously non-negligible within the same cell, a condition that is intrinsically rare in a dense urban network where the truck-feasible subnetwork is sparse by construction. This distribution highlighted a structurally fragile last-mile network configuration, in which physical access did not necessarily imply resilience to disruptions such as road works, congestion, or temporary access restrictions.

4.3 Urban morphology operability (C3)

Criterion C3 evaluated the operability of the urban morphology by analyzing building density and footprint concentration as proxies for on-site spatial constraints. Figure 3 presents the spatial distribution of C3 values over the study area.

The results indicated generally high C3 values across the majority of grid cells, reflecting favorable urban morphology conditions for last-mile construction operations.



Figure 2. Spatial distribution of truck-feasible network redundancy (C2) in the Lille last-mile context. Normalized scale 0–1: $C2 < 0.2$ (dark red) indicates no or negligible locally available alternative truck-compatible routes; $C2 > 0.8$ indicates high redundancy, observed only near major high-capacity intersections. The near-zero distribution reflects both the structural scarcity of truck-feasible routes from road-hierarchy filtering (Section 3.3) and the multiplicative formulation of C2, which requires simultaneous non-negligible edge density and branching density (Section 3.4.2).

Lower C3 values were mainly concentrated in areas characterized by high building density and compact urban forms.

The statistical distribution of C3 values showed a concentration toward high operability levels, with a high mean value and limited dispersion. This pattern suggested that, at the city scale, urban morphology imposed fewer constraints on modular construction last-mile operations than road access and network structure.



Figure 3. Spatial distribution of urban morphology operability (C3) in the Lille study area. Normalized scale 0–1: $C3 < 0.2$ indicates low operability (dense, compact built-up areas with high building footprint coverage); $C3 > 0.8$ indicates high operability (sparse or open urban fabric with low building coverage).

4.4 Statistical distribution of resilience indicators

Figure 4 presents the statistical distributions of the three normalized criteria (C1, C2, and C3) computed over the 250 m grid covering the city of Lille. The distributions highlight strong contrasts between access suitability, network redundancy, and urban morphology operability in the last-mile delivery context.

C1 showed a highly asymmetric distribution with a median of zero, confirming the structural predominance of constrained roads, while approximately 25% of cells exceeded moderate suitability along major corridors. C2 was strongly skewed toward zero, more than 90% of cells exhibited near-zero redundancy, revealing a structurally fragile network where most locations depend on a single truck-compatible route. C3 concentrated toward high values (median > 0.80), indicating that urban morphology imposed fewer constraints than network structure. Together, these distributions confirm that last-mile resilience in Lille is primarily limited by network-related factors.

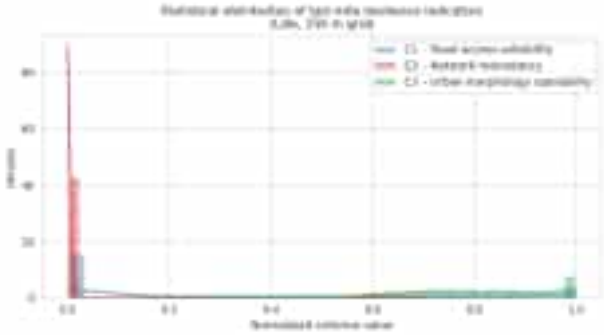


Figure 4. Statistical distributions of road access suitability (C1), truck-feasible network redundancy (C2), and urban morphology operability (C3) across the Lille study area (250 m grid, kernel density estimates).

5 Discussion and Recommendations

5.1 Network structure as the primary driver of last-mile resilience

The results demonstrated that last-mile delivery resilience for modular construction in Lille was primarily governed by road network structure rather than by local accessibility alone. Although road access suitability (C1) exhibited marked spatial heterogeneity, truck-feasible network redundancy (C2) remained extremely low across the majority of the study area. This interpretation is directly supported by the near-zero median values of C2 and its very limited dispersion (Figure 4), in contrast with the extended upper tail observed for C1 along major corridors.

The near-zero values observed for C2 in most grid cells reflected a structural dependence on a limited number of truck-compatible routes. Such configurations increased sensitivity to localized disruptions, including road works, congestion, and temporary access restrictions. This vulnerability is consistent with the spatial pattern of C2 (Figure 2), where higher redundancy is confined to a small number of cells near major intersections and high-capacity segments. These findings are consistent with resilience theory in urban systems, which emphasizes redundancy and alternative paths as key properties for disturbance absorption and recovery. In dense European urban contexts, historical road hierarchies and planning priorities have often concentrated freight traffic on a small number of corridors, thereby limiting redundancy at the local scale.

5.2 Limited influence of urban morphology in dense last-mile contexts

Urban morphology operability (C3) exhibited generally high values across the study area, with limited dispersion compared to access- and network-related criteria. This result is directly supported by the concentration of C3 toward high values (median exceeding 0.80 in Figure 4) and by the spatial distribution in Figure 3, where low operability is limited to compact and highly built-up areas. These patterns suggested that spatial constraints related to building density and footprint concentration were not the dominant limiting factors for last-mile modular construction operations in Lille.

This finding can be explained by the fact that conventional construction practices already operate under constrained urban conditions, relying on regulated temporary occupation of public space, staged deliveries, and adaptive on-site logistics. As a result, morphology-related constraints can often be operationally mitigated at the site interface, whereas limitations imposed by road network structure are more difficult to compensate. The results therefore indicated that, in dense urban environments, improving last-mile resilience requires a stronger focus on network-level characteristics rather than solely on site-scale spatial availability.

5.3 Distinguishing accessibility from resilience in modular construction logistics

A key outcome is the distinction between accessibility (C1, nominal conditions) and resilience (C2, functionality under disruption). The divergence between these criteria—heterogeneous C1 versus near-zero C2 in over 90% of cells—shows that accessibility alone is insufficient as a resilience indicator (Figure 4). For modular construction, where alternative routes are non-negotiable given dimensional constraints, the framework formalizes this distinction spatially, addressing a gap in existing supply-chain resilience studies.

5.4 Operational, territorial, and regulatory recommendations

Based on the identified spatial patterns, several recommendations can be derived. At the operational level, delivery planning for modular construction should explicitly account for local network redundancy, rather than relying solely on shortest-path or nominal access criteria. In practice, low-C2 areas (Figures 2 and 4) should be treated as disruption-sensitive zones where contingency routing, time-buffering, and coordination with on-site installation windows become critical. Identifying critical corridors and intersections can support more robust routing strategies and contingency planning.

At the territorial level, targeted interventions on the few high-C2 links and junctions may yield significant resilience improvements without extensive network modifications. At the regulatory level, urban access policies and temporary road occupation permits could be adapted to reflect network resilience considerations. Coordinated scheduling of road works and construction activities along critical corridors may reduce cumulative disruption effects, particularly in contexts where redundancy is structurally low (C2 near zero across most cells). More broadly, integrating last-mile resilience diagnostics into urban planning and construction permitting processes could support the deployment of modular construction systems in dense cities.

6 Conclusion

This paper proposed a GIS-based, network-informed multi-criteria framework to assess last-mile delivery resilience for modular construction in dense urban environments, integrating three complementary indicators: road access suitability (C1), truck-feasible network redundancy (C2), and urban morphology operability (C3), demonstrated on the city of Lille.

The results revealed that favorable access for modular deliveries was concentrated along a limited number of major corridors (C1), while truck-feasible network redundancy (C2) emerged as the most critical limiting factor, with near-zero values across most of the study area. Urban morphology operability (C3) remained comparatively favorable. These findings emphasize that accessibility alone does not guarantee last-mile resilience, and that road network redundancy plays a decisive role for modular construction logistics.

Several limitations must be acknowledged: reliance on open geospatial data with possible local inaccuracies; road-class filtering without explicit modeling of vertical clearance or axle load restrictions (reference configuration: semi-trailer, 16–18 m, 44 t); exclusion of regulatory

constraints such as time windows and low-emission zones; and a static spatial scope that does not capture temporal variability.

Future research should incorporate regulatory and temporal dimensions, empirical validation using real logistics data, vehicle-specific network filtering, and application to other urban contexts to support transferable guidelines for modular construction last-mile resilience.

Acknowledgment

This paper was originally presented at ISARC 2026 and published in the Proceedings of the 43rd ISARC, 2026, Singapore.

References

[1] Yang Liu, Jianjun Dong, and Ling Shen. A conceptual development framework for prefabricated construction supply chain management: An integrated overview. *Sustainability*, 12(5):1878, 2020. doi:10.3390/su12051878.

[2] Minghao Lin, Wei Zhang, and Shuang Liu. Analyzing resilience influencing factors in the prefabricated building supply chain based on SEM–SD methodology. *Scientific Reports*, 14:65271, 2024. doi:10.1038/s41598-024-65271-2.

[3] Techane Bosona. Urban freight last mile logistics—Challenges and opportunities to improve sustainability: A literature review. *Sustainability*, 12(21):8769, 2020. doi:10.3390/su12218769.

[4] Mats Janné and Anna Fredriksson. Construction logistics in urban development projects—Learning from, or repeating, past mistakes of city logistics? *The International Journal of Logistics Management*, 33(1):49–68, 2022. doi:10.1108/IJLM-03-2020-0128.

[5] Xiaojuan Li and Jiajing Lu. Resilience of prefabricated building industry chains based on genetic algorithm and PROMETHEE. *Engineering, Construction and Architectural Management*, pages 1–24, 2025. doi:10.1108/ECAM-07-2024-0892.

[6] Jacek Malczewski. GIS-based multicriteria decision analysis: A survey of the literature. *International Journal of Geographical Information Science*, 20(7): 703–726, 2006. doi:10.1080/13658810600661508.

[7] Selçuk Şişman. Using GIS-based multi-criteria decision analysis techniques in urban planning: An implementation workflow. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XLIV-4/W3-2020, pages 383–390, 2020. doi:10.5194/isprs-archives-XLIV-4-W3-2020-383-2020.

[8] Jinjing Li, Haizhe Yu, and Xiaopeng Deng. A systematic review of the evolution of the concept of resilience in the construction industry. *Buildings*, 14 (9):2643, 2024. doi:10.3390/buildings14092643.

[9] M. Ghufran, M. Khan, et al. Key enablers of resilient and sustainable construction supply chains: A systems thinking approach. *Sustainability*, 14(19): 11815, 2022. doi:10.3390/su141911815.

[10] Pei-Yuan Hsu, Panagiotis Angeloudis, and Marco Aurisicchio. Optimal logistics planning for modular construction using two-stage stochastic programming. *Automation in Construction*, 94:47–61, 2018. doi:10.1016/j.autcon.2018.05.029.

[11] Andrii Galkin, Libor Švadlenka, Radek Vrba, and Kinga Kijewska. Navigating the future of urban logistics: Conceptual framework for dynamic freight management. *Transportation Research Part D: Transport and Environment*, 147:104956, 2025. doi:10.1016/j.trd.2025.104956.

[12] Cindy Guerlain, Samuel Renault, and Francesco Ferrero. Understanding construction logistics in urban areas and lowering its environmental impact: A focus on construction consolidation centres. *Sustainability*, 11(21):6118, 2019. doi:10.3390/su11216118.

[13] Milad Moradi, Stéphane Roche, and Mir Abolfazl Mostafavi. Exploring five indicators for the quality of OpenStreetMap road networks: A case study of Québec, canada. *Geomatica*, 76(1):1–16, 2022. doi:10.1139/geomat-2021-0012.

[14] Ming Wang, Qingquan Li, Qingwu Hu, and Meng Zhou. Quality analysis of OpenStreetMap data. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XL-2/W1, pages 155–158, 2013. doi:10.5194/isprsarchives-XL-2-W1-155-2013.

Event-based Subsurface Utility Data Collection Framework

Vignesh Vijayalakshmi Palanisamy[1] **and Senthilkumar Venkatachalam**[1]

[1]Department of Civil Engineering, Indian Institute of Technology Palakkad, Kerala, India

102214002@smail.iitpkd.ac.in, senthil@iitpkd.ac.in

Abstract -

Underground utility information is frequently incomplete, inconsistent, or of uncertain confidence, leading to safety risks, rework, and inefficiencies during construction and maintenance activities. Although existing standards, sensing technologies, and digital platforms define data quality concepts and investigation methods, their application within construction workflows remains fragmented and largely project-specific. In particular, limited guidance exists on when utility data should be collected across the construction lifecycle and how confidence in existing information should inform construction decisions. This paper proposes an event-based utility data collection approach that aligns subsurface information acquisition with recurring construction and maintenance events. It integrates lifecycle events, utility data categories, technology classes, and confidence considerations to support transparent, confidence-aware decisions on investigation, escalation, and opportunistic data capture. A visual representation and supporting tables illustrate how construction methods and access conditions influence achievable confidence and data availability, while highlighting construction and maintenance activities as high-value opportunities for reliable data capture. By providing an operational structure for applying existing standards in real construction contexts, the proposed approach supports coordinated decision-making, reduces reliance on ad-hoc investigations, and enables incremental improvement of utility information over time, contributing to safer construction practices and more reliable long-term subsurface asset management.

Keywords -

Underground utilities; Subsurface utility engineering; Construction lifecycle; Data collection framework; Utility data confidence; Event-based data capture; Infrastructure asset management; Utility mapping

1 Introduction

Underground utilities constitute a critical yet largely invisible component of the built environment. Utilities such as water, sewerage, gas, electricity, and telecommunication networks are routinely affected during construction and maintenance activities, resulting in service disruptions, safety hazards, schedule delays, and cost overruns [1]. Despite advances in sensing technologies and digital information systems, utility strikes and subsurface conflicts remain prevalent, particularly in dense urban corridors where multiple assets coexist within constrained space [2].

A key reason for these recurring issues is the misalignment between construction workflows and subsurface information practices [3]. Construction and asset management progress through interconnected stages, planning, design, installation, operation, maintenance, and decommissioning, yet underground utility information is commonly treated as static, captured once and reused irrespective of subsequent interventions [4]. As a result, critical construction decisions are often made using outdated or low-confidence information, while opportunities to improve data quality during direct interaction with the subsurface are frequently missed.

In practice, subsurface utility information is fragmented across organizations and project phases. Planning teams rely on legacy records and GIS databases; designers commission selective investigations; contractors expose utilities during construction; and maintenance crews uncover assets during routine and emergency maintenance works. These repeated encounters are rarely coordinated through a unified, lifecycle-oriented data logic, and records are infrequently updated after sites are reinstated, perpetuating uncertainty and rework across successive projects [1]. These challenges manifest differently across greenfield and brownfield contexts: while greenfield projects face uncertainty due to limited baseline records, brownfield environments are dominated by legacy utilities of unknown location, condition, and confidence, making coordinated construction particularly risk-prone.

At the same time, the construction sector has widely adopted detection, surveying, and documentation technologies such as ground-penetrating radar, electromagnetic locating, GNSS surveying, laser scanning, etc., [2, 5]. While these technologies have improved detection and documentation capabilities, their use remains largely project-driven and reactive rather than integrated across the asset lifecycle. Decisions regarding when to investigate, how far to escalate verification, and what information to capture are typically left to individual project teams, leading to inconsistent outcomes across projects and jurisdictions [4].

These observations indicate that subsurface utility challenges are not only technological but also procedural, highlighting the need to align utility data collection with construction and maintenance events while explicitly accounting for information confidence. This paper addresses this need by proposing an event-based utility data collection framework that operationalizes existing standards and technologies within real construction workflows. It focuses on defining when, how, and under what construction conditions subsurface utility data should be captured to maximize long-term data value, safety, and reliability.

2 Background and related works

2.1 Persistent challenges in subsurface utility data

A consistent finding across subsurface utility research is that existing utility records are frequently inaccurate, incomplete, outdated, or of unknown confidence. Comparative studies demonstrate that traditional utility records, including as-built drawings and One-call systems (Call before you dig), often significantly under-represent underground infrastructure and exhibit substantial positional inaccuracies when compared with Subsurface Utility Engineering (SUE) investigations [1]. Empirical evidence indicates that conventional records may represent up to 160% less utility infrastructure than identified through systematic investigation, and that only a minority of mapped utilities fall within acceptable positional tolerances [3].

Similar challenges are reported across transportation projects, airports, and dense urban developments, where the depth information is missing, reference systems are inconsistent, and records reflect as-designed rather than as-built conditions [6, 2]. These shortcomings result in increased project costs, repeated site investigations, and heightened safety risks during excavation [7, 8]. The literature consistently identifies fragmented data capture practices and the absence of systematic updating mechanisms as key contributors to poor data quality [4].

2.2 Quality-based approaches and SUE

Subsurface Utility Engineering (SUE), formalized through standards such as ASCE 38 and complemented by ASCE 75 and PAS 128 with PAS 256, is widely recognized as a structured engineering approach for improving subsurface utility data quality using defined quality or confidence levels. Higher SUE quality levels are associated with increased positional accuracy, improved completeness, and reduced risk of utility conflicts during construction [9].

A key strength of these standards lies in its explicit linkage between investigation methods and achievable confidence, enabling informed engineering decisions during planning and design. However, prior studies note that such standards are predominantly applied during pre-design and design stages as a project-specific investigation and provides limited guidance on systematically capturing and updating utility data during installation, repair, diversion, or replacement activities [1].

2.3 Utility investigation technologies

A wide range of technologies and methods has been investigated for subsurface utility detection and documentation. Non-destructive geophysical methods, including ground-penetrating radar and electromagnetic locating, are widely used to detect buried utilities under varying site and soil conditions [5]. Geospatial surveying technologies such as GNSS, total stations, and terrestrial laser scanning enable accurate spatial referencing of exposed utilities and support integration with digital records [2].

Intrusive verification methods, including trial trenches and vacuum excavation, remain the most reliable means of achieving high-confidence positional and depth information, particularly where non-destructive methods are constrained. While the technical performance of investigation technologies is well documented, studies consistently note that achievable accuracy is limited by soil conditions, congestion, and utility material, and that investigation methods are rarely sequenced or escalated within construction workflows [5, 2]. Consequently, technology selection remains reactive and project-specific rather than embedded within a lifecycle-oriented strategy.

2.4 Data uncertainty and digital integration

Prior studies emphasize that uncertainty is an inherent characteristic of subsurface utility data and must be explicitly represented to support risk-informed decision-making [10]. Probabilistic mapping and uncertainty-aware visualization approaches acknowledge that complete information is rarely achievable and instead aim to represent confidence envelopes around detected utilities [11].

Parallel work on BIM–GIS, CAD-GIS integrations, digital twins, and augmented reality seeks to improve coordination and situational awareness for utility management [12, 13]. While these platforms offer substantial potential, their effectiveness is highly dependent on the quality and confidence of input data. Many such approaches implicitly assume the availability of reliable baseline utility data or the feasibility of systematic data collection independent of construction activities. In practice, this assumption is rarely valid for legacy infrastructure, limiting their effectiveness in real construction contexts.

2.5 Lifecycle and event-based data capture

A smaller but critical body of literature explicitly argues that subsurface utility data quality should be improved incrementally across the asset lifecycle, rather than through

large-scale resurveying efforts [14, 1]. These studies highlight the impracticality of repeated full exposure and advocate leveraging recurring construction and maintenance events, such as installation, diversion, repair, and replacement, as high-value data capture opportunities. While this work establishes the importance of lifecycle-based improvement, it does not provide a detailed operational structure defining event priorities, data requirements, technology applicability, or confidence management over time.

3 Research gap and objectives

Despite advances in underground utility data standards, sensing technologies, and digital platforms, construction practice still lacks a clear operational structure for aligning utility data collection with recurring construction and maintenance events. Existing approaches define data quality concepts, investigation methods, and uncertainty representation; however, their application across the construction lifecycle remains fragmented and largely project-specific. In practice, decisions regarding when to undertake additional investigation, which methods are appropriate under given access conditions, and how confidence in existing information should influence construction activities are often left to individual judgement. As a result, high-confidence data capture opportunities during routine construction and maintenance activities are frequently missed, perpetuating uncertainty and repeated investigations.

To address this gap, this study proposes an event-based utility data collection framework that situates utility information acquisition within construction workflows. The framework integrates lifecycle events, utility data categories, technology classes, and confidence considerations into a single construction-oriented structure. Specifically, it aims to (i) align utility data categories and applicable technology classes with key lifecycle events, and (ii) support transparent, confidence-aware decisions related to verification, escalation, and opportunistic data capture during construction and maintenance activities.

4 Methodology and framework development

This section summarizes the methodological approach used to develop the event-based subsurface utility data collection framework. The framework is conceived as a structuring and decision-support artefact that synthesizes concepts from existing utility data standards (e.g., ASCE 38-22, ASCE 75-22, PAS 128, PAS 256), investigation methods, and construction practice into a coherent, construction-oriented logic suitable for application within real work zones.

The framework was developed based on the premise that physical access to underground utilities is governed primarily by construction and maintenance events rather than by continuous monitoring or isolated survey campaigns. Accordingly, methodology adopted an event-based perspective: recurring lifecycle events are treated as the primary opportunities for interaction with subsurface assets and for capturing durable data.

Framework design was guided by five principles: (i) construction-centricity, aligning with real workflows; (ii) lifecycle orientation, treating utility information as an evolving asset; (iii) spatial anchoring, applying the framework to defined work zones containing multiple utilities; (iv) confidence awareness, explicitly accounting for information reliability; and (v) technology neutrality, grouping methods by functional capability rather than specific tools. These principles ensure compatibility with existing standards while maintaining adaptability across project and regulatory contexts.

Eight representative lifecycle events (E1–E8) were identified through synthesis of typical planning, construction, and operational workflows in utility-dense environments. These events span records review, reconnaissance, non-intrusive investigation, intrusive verification, open-cut and trenchless installation, maintenance or repair, and decommissioning. Events are non-linear and non-mandatory, reflecting construction practice in which activities may recur, overlap, or be bypassed depending on project intent and risk.

To support standardized capture, utility information requirements were synthesized into six high-level categories: P (positional & geometric), M (material), F (functional), C (condition & performance), S (survey metadata), and L (lifecycle metadata). Investigation methods were abstracted into five technology classes (T1–T5) reflecting record sources, optical/visual methods, geospatial surveying, geophysical/NDT, and intrusive verification.

A key methodological distinction is made between *current confidence*, representing the reliability of existing information for a given utility within the spatial extent, and *achievable confidence*, representing the maximum reliability attainable during a given event using appropriate methods. This distinction underpins the framework's escalation logic: construction or maintenance activities should proceed only when current confidence is appropriate for the intended action; otherwise escalation to earlier or more invasive events is recommended.

Finally, events involving physical exposure of utilities, such as intrusive verification, open-cut installation, maintenance repairs, and decommissioning, are explicitly identified as high-value data capture opportunities. During these events, comprehensive utility information can be obtained with minimal additional effort relative to the con-

struction activity itself. The framework therefore adopts an opportunity-based perspective that promotes systematic capture and recording whenever exposure occurs, supporting incremental improvement of utility records over time and reducing reliance on repeated standalone investigations.

5 Event-based utility data collection framework

This section presents the proposed event-based utility data collection framework and articulates its conceptual structure and decision logic. The framework integrates lifecycle events, utility data categories, technology classes, and confidence considerations into a construction-oriented model for subsurface utility data acquisition. The visual representation formalizes relationships between construction events, data availability, and achievable confidence, providing a structured basis for transparent and auditable decision-making across construction and maintenance contexts.

5.1 Framework overview and representation

Figure 1 illustrates the framework for a defined spatial extent (e.g., road segment, project corridor, or work zone), which constitutes the unit of application. Within this spatial extent, multiple underground utilities may coexist, each characterized by different levels of data availability and confidence. The framework is therefore spatially anchored rather than asset-isolated, reflecting real construction conditions in which decisions affect all subsurface assets within the work area.

The outer ring denotes eight recurring lifecycle events (E1–E8), derived from synthesis of construction workflows reported in literature, subsurface utility standards, and typical installation, maintenance, and diversion practices, and representing repeatable opportunities to consult, verify, or capture utility information. Events are nonlinear and may be entered, skipped or revisited depending on project intent, access and risk. The inner region groups utility information into six high-level categories (P, M, F, C, S, L); implementers should rely on Table 1 for attribute-level definitions and capture templates. The figure marks exposure-based events as high-value data capture opportunities (starred) where the highest achievable confidence is typically attainable.

When used, the framework encodes a simple decision logic: practitioners populate a side-panel register with existing information and assign current confidence per utility; they define the planned activity and its required confidence; they consult Table 2 to identify technology classes and events that can raise achievable confidence; and they escalate to those events when current confidence is insufficient. This logic is intentionally generic and auditable and supports selective escalation rather than prescriptive sequencing.

5.2 Utility data categories

Table 1 formalizes the six data categories used throughout the framework. The attributes listed are representative and not intended to be exhaustive or restrictive. This consolidated category structure is intended for attribute dictionaries, data capture templates, contractual clauses and permit conditions, and provides a consistent lens for linking data requirements to construction events. These categories define what information may be captured; when and how it is captured depends on lifecycle events and access conditions.

5.3 Lifecycle events and data capture capability

Each lifecycle event (E1-E8) provides a different degree of physical or informational access to underground utilities and therefore supports different subsets of abovementioned data categories. Early events such as record review and reconnaissance primarily support baseline understanding and risk screening, with limited achievable confidence. Non-intrusive investigation events enable indicative detection and alignment estimation but remain constrained by site conditions and utility characteristics.

Events involving direct exposure, such as intrusive verification, open-cut installation, maintenance repairs, and decommissioning, offer the highest achievable confidence across multiple data categories. These events allow direct measurement and observation of geometry, material, condition, and installation context. The framework explicitly identifies such exposure-based events as high-value data capture opportunities, encouraging systematic recording rather than incidental documentation.

Open-cut and trenchless construction methods are treated distinctly. Open-cut installation inherently enables comprehensive data capture prior to backfilling, whereas trenchless installation significantly restricts post-installation verification. Consequently, the framework emphasizes achieving sufficient confidence through earlier events before undertaking trenchless works.

5.4 Confidence Assessment and Escalation

A core feature of the framework is the distinction between achievable confidence and current confidence. Achievable confidence represents the maximum reliability that can reasonably be attained during a given event using appropriate methods. Current confidence represents the reliability of existing information for a specific utility within the spatial extent, based on available records, prior investigations, and documented exposures. Current

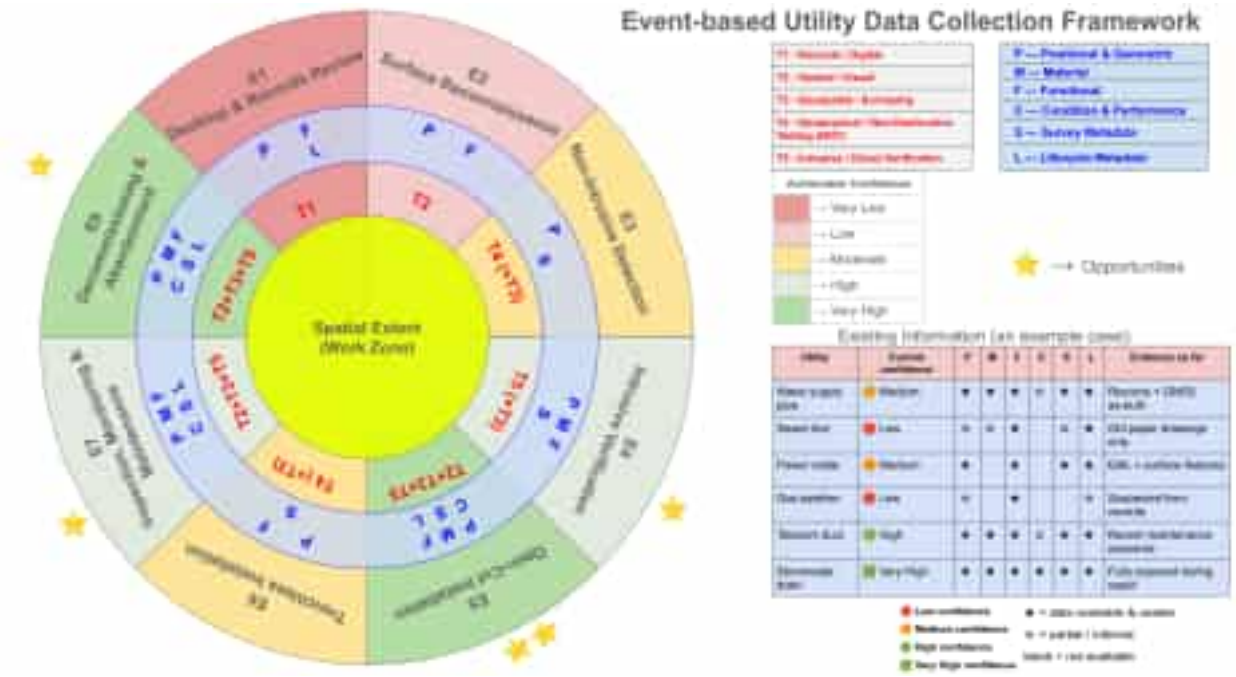


Figure 1. Event-based utility data collection framework

Table 1. Utility data categories and representative (non-exhaustive) attributes

Data Category	Example Attributes / Parameters
P - Positional & Geometric	Horizontal coordinates (X, Y), elevation/depth, spatial reference system, alignment, length, diameter, horizontal and vertical profile,burial depth, network connectivity, proximity to other utilities
M - Material	Material type (PVC, steel, concrete), wall thickness, grade/specification, lining/coating type, corrosion protection, backfill/bedding type, manufacturer, applicable material standards, installation age
F - Functional	Asset type (water pipe, gas pipe, duct, etc.,), service type, design capacity, pressure rating, flow direction, load class, operating limits, functional role in network, redundancy or criticality, control function (e.g., valve)
C - Condition & Performance	Condition rating, structural integrity score, corrosion level, deformation, leakage/blockage status, hydraulic/performance metrics, failure history, remaining useful life, inspection findings, inspection frequency/last inspection date
S - Survey Metadata	Survey method (e.g., GPR, GNSS, etc), instrument, survey date, surveyor/organization, accuracy, resolution, QA/QC status, detection limitations, standards compliance, data format/encoding
L - Lifecycle Metadata	Installation date, construction method (open-cut or trenchless), project/contract ID, ownership, operation start date, maintenance/repair/upgrade history, decommissioning date, regulatory compliance, replacement cost, asset status, risk/criticality score, warranty

confidence may be assessed using established quality classification systems such as SUE quality levels defined in ASCE 38-22, or equivalent local practices where available. Where formal classifications are not available, confidence may be assigned based on the data source, investigation method, and recency of verification.

Construction and maintenance decisions are informed by comparing these two measures. Where current confidence is sufficient for the intended activity, work may proceed. Where it is insufficient, the framework supports escalation to earlier investigation events until acceptable confidence is achieved. This logic is particularly critical for trenchless works and high-risk excavations, where post-construction verification is limited or impossible. Importantly, the framework does not mandate escalation through all events; it provides transparent guidance on when escalation is justified.

5.5 Technology classes and applicability constraints

Investigation and documentation methods are grouped into five functional classes (T1–T5). Table 2 lists representative methods as sub-rows, summarizes where each method typically performs well, and identifies known limitations (for example, material dependency of GPR and tracer-wire reliance for EML). This abstraction allows technology constraints to be considered at the decision level without embedding site-specific optimization into the framework.

Where different technology classes provide conflicting information, a simple precedence logic is applied. Geo-

Table 2. Technology tiers for subsurface utility data collection and applicability

Tier	Typical Methods / Data Sources	Works Well For	Known Limitations
T1 – Records	GIS databases & utility atlases	Baseline topology, planned layouts, design intent, ownership context	Often incomplete or outdated; may represent design not as-built
	As-built drawings & owner records	Approximate alignment, Material specifications, installation dates, and historical context	Fragmented across agencies; inconsistent georeferencing; unknown coordinate datums
	Historical maps & archived plans	Identifying legacy or abandoned utilities	Poor accuracy; unknown reference systems; unverified current existence
T2 – Optical / Visual	Photos, video logs, & field documentation	Chronological record of exposed assets, materials, and construction activities	Unstructured; lacks spatial reference; no subsurface detection
	UAV & aerial imagery	Surface context, obstructions, site topography, and progress monitoring	Cannot detect buried utilities; weather/lighting dependent; requires airspace clearance
	CCTV & pipeline inspection	Internal condition (corrosion, cracks, blockages) of accessible pipelines	Limited to pipe interiors; lacks external positional data
T3 – Geospatial / Surveying	GNSS (RTK / PPK)	Centimeter-level positioning of exposed utilities and survey control	GNSS-denied environments; multipath errors; requires clear sky view
	Total station survey	Millimeter-level precision for exposed utility geometry	Line-of-sight constraints; limited range
	Laser scanning (TLS/MLS)	High-density 3D capture of exposed infrastructure (TLS) or rapid corridor mapping (MLS)	Requires exposure; setup effort (TLS); reduced accuracy in cluttered environments (MLS)
T4 – Geophysical / NDT	Ground penetrating radar (GPR)	Detection of shallow utilities; non-metallic pipes	Performance dependent on soil conditions; interpretation uncertainty
	Electromagnetic locating (EML)	Tracing metallic utilities and plastic pipes with tracer wire	Ineffective without intact tracer
	Acoustic leak detection	Identifying pressurized leaks in utility networks	Ineffective for gravity or low-pressure systems
	Fiber-optic & strain sensing	Distributed monitoring of instrumented pipelines	Requires pre-installed sensors; does not locate unmapped utilities
T5 – Intrusive / Verification	Vacuum excavation (hydrovac)	Non-destructive exposure for verification and safe clearance	Cost and access constraints
	Test pits & trenches	Direct confirmation of position, depth & material	High disruption; reinstatement required
	Boreholes	Stratigraphy and groundwater assessment for design planning	Risk of utility strikes; limited lateral information
	Manhole & chamber surveys	Direct measurement of invert elevations, pipe sizes, and connectivity	Limited to existing access points; confined space hazards

physical survey results (T4) are treated as the primary source for positional attributes (P) within the surveyed area, as they reflect physical detection rather than historical records (T1). Record-based sources (T1) remain relevant for material (M) and functional (F) attributes not directly inferred from geophysical methods. Such discrepancies are recorded in survey metadata (S) and act as a trigger for escalation to intrusive verification (E4/T5) where conflicts occur within the planned work zone.

5.6 Framework use and interpretation

The framework is intended to be applied iteratively. For a given spatial extent, users first assess existing information and current confidence, define the intended construction or maintenance activity, and identify the required confidence level. The framework then guides whether existing information is sufficient, whether escalation is warranted, and which events represent high-value opportunities for durable data capture.

By structuring utility data collection around construction and maintenance events, the framework supports incremental improvement of subsurface information over time. It enables authorities, designers, and contractors to move beyond static records and ad-hoc investigations toward a coordinated, confidence-aware approach aligned with real construction workflows.

5.7 Illustrative application scenarios

To illustrate how the framework supports decision-making, Table 3 summarizes representative, illustrative scenarios. These scenarios demonstrate how the framework guides escalation, technology classes, and opportunistic data capture under different construction intents and site conditions. These examples provide operational clarity without prescribing rigid procedures.

6 Discussion and Implications

This study proposes an event-based utility data collection framework that reframes subsurface information management as a construction- and maintenance-driven process, rather than a one-time or pre-design activity. The primary contribution lies not in introducing new sensing technologies or data standards, but in operationalising when, under what construction conditions, and with what confidence utility data should be acquired within real construction workflows.

Existing standards and guidance have established data quality concepts, investigation methods, and documentation requirements, significantly improving awareness of subsurface uncertainty. The proposed framework complements these efforts by providing an explicit operational layer that links such guidance to recurring lifecycle events. By distinguishing between current confidence and

Table 3. Illustrative application scenarios

Scenario	Context	Framework Application	Expected outcome
Baseline understanding and risk screening for planned works	Planned surface works (e.g., resurfacing) in a utility-dense corridor using primarily GIS and legacy as-builts	Framework applied at corridor level. Assessment begins at E1 with current confidence assigned per utility. Low-confidence utilities are flagged, and escalation to E2 or E3 is triggered only where construction risk or congestion justifies it.	Selective escalation improves confidence where needed, avoiding unnecessary investigations and supporting construction safety.
New utility installation - open-cut	Open excavation for new line installation	Baseline from E1; open-cut treated as opportunistic high-value capture event (E5). Comprehensive as-built data captured using T2, T3, T5 during exposure before backfilling	High-confidence as-built update eliminates need for future E2–E4 investigations in this zone
New utility installation - trenchless (HDD)	New HDD installation beneath a road where post-installation verification is limited	Trenchless installation (E6) require pre-construction escalation to E3/E4 to verify existing utility positions, as post-installation correction is not feasible	Prevents utility strikes and alignment errors that cannot be rectified after drilling
Maintenance and repair as opportunistic capture	Routine or emergency maintenance requiring temporary exposure of utilities	Maintenance (E7) treated as opportunistic capture event. Exposed utility attributes systematically recorded using T2, T3, T5 and integrated into authoritative records to update confidence	Maintenance events incrementally improve authoritative records at marginal cost.
Multi-utility coordination in brownfield corridors	Fragmented ownership, inconsistent records, conflicting stakeholder data	Framework applied at spatial-extent level. Current confidence is assessed per utility; missing or conflicting records are treated as low confidence, triggering selective escalation to E3/E4. Confidence is updated per utility to support coordinated decision-making across stakeholders.	Improved coordination, reduced disputes, and increased corridor-level confidence.
Incremental information improvement over time	Authority lacks resources for comprehensive resurveying	Framework is institutionalized to require data capture at high-value events such as E4, E5, and E7. Each intervention incrementally updates records and confidence without dedicated survey campaigns	Progressive update of utility information supports long-term asset management & reduces repeated investigations.

achievable confidence, the framework supports transparent and defensible escalation decisions without assuming that higher investigation levels are always necessary.

Importantly, the framework does not prescribe rigid sequences or mandatory investigations. Instead, it formalizes a decision logic that allows events to be skipped, revisited, or escalated based on construction intent, access conditions, and risk. This flexibility reflects real construction practice and avoids the inefficiencies associated with uniform or overly conservative investigation requirements.

From a construction perspective, the framework supports risk-informed decision-making by clarifying when existing information is sufficient and when escalation is justified, particularly for high-risk activities such as trenchless installations. For asset owners and authorities, it provides a consistent basis for defining permit conditions, inspection requirements, and data capture obligations without prescribing specific technologies or vendors.

A key practical implication is the explicit recognition of construction, maintenance, and decommissioning activities as high-value data capture opportunities. Systematic recording during these events enables incremental improvement of utility records over time, reducing reliance on repeated standalone investigations and supporting long-term asset information quality.

The use of technology classes rather than specific instruments is a deliberate design choice. Known applicability constraints, such as material dependency of geophysical methods or access limitations for surveying, are addressed through technology applicability tables and embedded escalation logic. The framework therefore guides how technology limitations should influence decisions, rather than attempting to optimize sensor selection.

This study does not claim to eliminate uncertainty or replace engineering judgement. Site-specific conditions, contractual constraints, and safety considerations remain critical and must be addressed by practitioners. The framework has been developed through synthesis of standards, prior research, and construction practice and illustrated through representative scenarios. Large-scale field validation and quantitative assessment are identified as important areas for future work.

7 Conclusion

This paper presented an event-based utility data collection framework that aligns subsurface utility information acquisition with construction and maintenance workflows. By structuring utility data collection around lifecycle events, data categories, technology classes, and confidence considerations, the framework provides practical guidance on when data should be collected, when escalation is warranted, and when construction activities should be leveraged as durable data capture opportunities.

The proposed approach shifts utility data management from a predominantly record- or survey-driven paradigm toward a workflow-integrated, confidence-aware process. Rather than requiring comprehensive resurveying, the framework enables incremental improvement of utility information through routine construction and maintenance activities, supporting safer construction, improved coordi-

nation among stakeholders, and more reliable long-term asset records.

Future work may focus on implementing the framework within digital platforms, evaluating its application across diverse project types, and exploring automated or semi-automated support for confidence assessment and escalation decisions. Nevertheless, the framework provides a structured foundation for operationalising existing standards and technologies within real construction contexts.

8 Acknowledgment

This research was funded by the Department of Science and Technology (DST), India through the Global Innovation & Technology Alliance (GITA) under Grant No. GITA/DST/TWN/P-104/2022.

References

[1] National Academies of Sciences, Engineering, and Medicine. *Strategies to Address Utility Issues During Highway Construction*. The National Academies Press, Washington, DC, 2024. ISBN 0-309-72245-4. doi:10.17226/27859.

[2] R. van Son, S. W. Jaw, J. Yan, V. Khoo, R. Loo, S. Teo, and G. Schrotter. A framework for reliable three-dimensional underground utility mapping for urban planning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:209–214, 2018. doi:10.5194/isprs-archives-XLII-4-W10-209-2018.

[3] J. Adebiyi, J. Anspach, and R. Sturgill. Evaluating completeness and positional accuracy of traditional utility records compared to subsurface utility engineering: A comparative case study analysis. *Transportation Research Interdisciplinary Perspectives*, 32:101539, 2025. doi:10.1016/j.trip.2025.101539.

[4] L. H. Hansen, R. van Son, A. Weiser, and E. Kjems. Addressing the elephant in the underground: An argument for the integration of heterogeneous data sources for reconciliation of subsurface utility data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 43–48, 2021. doi:10.5194/isprs-archives-XLVI-4-W4-2021-43-2021.

[5] C. Peng, C. Wang, and Z. Li. Review of geophysical data acquisition methods for underground feature detection and future trends. *Tunnelling and Underground Space Technology*, 163:106731, 2025. doi:10.1016/j.tust.2025.106731.

[6] C. Quiroga, J. Anspach, P. Scott, and E. Kraus. Feasibility of mapping and marking underground utilities by state transportation departments. Technical report, United States. Federal Highway Administration, 2018.

[7] N. Metje, B. Ahmad, and S. M. Crossland. Causes, impacts and costs of strikes on buried utility assets. *Proceedings of the Institution of Civil Engineers – Municipal Engineer*, 168(3):165–174, 2015. doi:10.1680/muen.14.00035.

[8] L. O. Makana, N. Metje, I. Jefferson, M. Sackey, and C. D. F. Rogers. Cost estimation of utility strikes: towards proactive management of street works. *Infrastructure Asset Management*, 7(2):64–76, 2020. doi:10.1680/jinam.17.00033.

[9] H. S. Jeong, D. M. Abraham, and J. J. Lew. Evaluation of an emerging market in subsurface utility engineering. *Journal of Construction Engineering and Management*, 130(2):225–234, 2004. doi:10.1061/(ASCE)0733-9364(2004)130:2(225).

[10] L. olde Scholtenhuis, X. den Duijn, and S. Zlatanova. Representing geographical uncertainties of utility location data in 3d. *Automation in Construction*, 96: 483–493, 2018. doi:10.1016/j.autcon.2018.09.012.

[11] Q. Dou, Z. Lin, D. R. Magee, and A. G. Cohn. 3d mapping from partial observations: An application to utility mapping. *Automation in Construction*, 117: 103229, 2020. doi:10.1016/j.autcon.2020.103229.

[12] M. Wang, Y. Deng, J. Won, and J. C. P. Cheng. An integrated underground utility management and decision support based on bim and gis. *Automation in Construction*, 107:102931, 2019. doi:10.1016/j.autcon.2019.102931.

[13] V. Vijayalakshmi Palanisamy and S. Venkatachalam. Leveraging multi-source data for enhanced utility mapping and conflict prevention in construction. In *Proceedings of the International Conference on Smart and Sustainable Built Environment*, volume 844 of *Lecture Notes in Civil Engineering*. Springer, 2026. doi:10.1007/978-981-95-8489-5_43.

[14] R. van Son, S. W. Jaw, and A. W. A data capture framework for improving the quality of subsurface utility information. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W15:97–104, 2019. doi:10.5194/isprs-archives-XLII-4-W15-97-2019.

MR-based Immersive Co-Design of Housing for Climate Displacement

Viktoriia Bezverkhnia[1], Bo Su[2], Muhammad Fawad[2], and Qian Chen[2]

[1]Faculty of Architecture, Construction, Decorative and Applied Arts, Yuriy Fedkovych Chernivtsi National University, Ukraine.
[2]School of Engineering, Faculty of Applied Science, University of British Columbia, Okanagan Campus, Canada.
bezverkhnia.viktoriia.work@gmail.com, bo.su@ubc.ca, muhammad.fawad@ubc.ca, qian.chen@ubc.ca

Abstract –

The rapid change in global climatic conditions and the increasing number of severe climate events have urged the demand for rapid, installable, customized, and sustainable housing solutions for climate-affected and displaced populations. To address this concern, this study develops a modular housing system tailored for climate displacements that integrates prefabricated bio-panel construction with an immersive digital design and evaluation workflow. The proposed design approach is based on the Building Information Modelling (BIM) framework to enable the systematic parametric integration of architectural quality, structural coherence, environmental performance, and human-centered design criteria.

An example of a housing design deliverable is studied based on dry-assembled prefabricated bio-panel components to support rapid on-site installation, reduced construction time, and limited dependence on skilled labor. Prefabrication design logic allows individual units to operate independently or be assembled as a large housing compound, which will facilitate expandable displacement community developments across diverse climatic and geographical regions. The BIM design is integrated with the Low-Latency Mixed Reality Design Platform (LMRDP) through a customized Revit API–based workflow, enabling high-fidelity design visualization to engage users. This integration enables seamless data transfer to an almost real-time MR application deployed to an MR headset. Based on the example demonstration, this study showcases a pathway to the combined potential of modular construction, BIM, and MR to enhance early-stage design evaluation, spatial usability assessment, and user engagement to support sustainable housing solutions for climate-affected displaced populations.

Keywords –

Climate displacement; Modular housing design; Building Information Modeling (BIM); Mixed Reality (MR); WebSocket

1 Introduction

Climate change has increased the number of people to lose their homes due to floods, fires, droughts, and other natural disasters. Every year, millions of people are forced to move and become climate refugees. Many of them need safe and comfortable housing that can be quickly built and used for different lengths of time. Traditional emergency shelters are often constructed in a hurry, using materials that are not environmentally friendly and do not provide adequate living comfort.

Modern construction methods allow for the design of modular prefabricated houses that can be manufactured in a factory, easily transported, and quickly assembled on site. These systems allow buildings to be reused or relocated as needed, reducing waste and promoting sustainable construction. Previous studies show that modular and off-site construction supply chains can save time, lower costs, and improve building quality [1][2] in post-disaster situations.

This study is designed based on the climate refugees' community needs for temporary homes that are easy to build, move, energy efficient, and comfortable to live in. The design uses EcoCocon straw panels as the main material for walls and roofs. These panels provide good thermal insulation, allow the building to “breathe,” and store carbon throughout their lifetime [3]. The use of natural materials helps to create healthy indoor conditions while reducing the environmental impact of construction. Several types of modular units, including small and medium-sized residential buildings, multi-room family modules, office and technical units, as well as shared spaces such as playgrounds, are proposed. Each module can be combined with others or used separately, depending on the number of residents and the specific location.

Thus, the study presents proof of concept design and system development efforts that integrate climate-

responsive design, modular construction, and a mixed reality (MR)-based digital toolkit. Overall, the study indicates how modular construction and biomaterials can help create climate-resilient homes that are fast to build, easy to transport, and pleasant to live in. For this purpose, a digital toolkit using MR and WebSocket is developed to enhance the visualization experience. The BIM model of the housing design, with local climate data (temperature, wind, etc.) adapted to it and available cost-effective materials, is integrated into an MR environment through a self-developed low-latency MR design platform using Revit API, WebSocket protocol, and the Microsoft MRTK-3. By bridging the gap between static architectural data and immersive environments, the toolkit provides a platform for stakeholders to visualize, iterate, and optimize climate refugee housing solutions in real-time.

2 Literature review

The literature and market survey indicate that cost-effective material selection for climate refugee housing is closely linked to prefabrication, dry assembly, and lifecycle performance. Bio-based prefabricated panels, engineered timber systems, and Structural Insulated Panels (SIP) technologies are repeatedly highlighted as materials that reduce construction time, labor intensity, and waste through factory-controlled production [4][5][6]. Studies emphasize that lower embodied carbon materials, such as straw-based panels and timber products, contribute to long-term environmental savings while supporting rapid deployment [4][7]. Industrial panel systems benefit from standardization, predictable performance, and reduced on-site errors, further improving cost efficiency [6][8]. The literature emphasizes that cost effectiveness must be evaluated across the full life cycle, including transport, assembly, reuse, and potential disassembly [9][10].

Comparative analysis identifies three main material groups relevant to accelerated housing construction: bio-based prefabricated panels, engineered wood and CLT systems, and SIP or sandwich panel technologies [4][5][6]. Bio-based panels, such as straw- and hemp-based systems, prioritize low embodied carbon, carbon sequestration, and healthy indoor microclimates while remaining suitable for rapid dry assembly [4][11]. Engineered timber and CLT panels provide higher structural capacity, scalable panel dimensions, and predictable mechanical performance, enabling applications from single units to multi-story prefabrication [5][7]. SIP and sandwich panel systems focus on speed, airtightness, and high thermal resistance, offering reduced assembly time compared to conventional masonry construction [6][8].

The reviewed literature highlights that panelized and modular construction systems are well-suited to BIM-based housing design due to standardized geometries, defined tolerances, and repeatable connection logic [5][6]. Manufacturer documentation for CLT, SIP, and proprietary panel systems provides technical data that can be directly integrated into BIM workflows, supporting coordination between structure, enclosure, and building services [5][10]. Prefabrication supported by BIM improves logistics planning, panel sequencing, and quality control, reducing construction errors and material waste [7]. Standardized fastening systems, membranes, and connection components further support digital coordination and reassembly potential [7][12] [13].

The literature consistently emphasizes that temporary housing must focus on human comfort, health, and adaptability alongside technical performance. Bio-based panels are known for their positive impact on indoor air quality and hygrothermal behavior, supporting healthy indoor environments for displaced populations [4][11]. High thermal resistance, airtightness, and moisture control are identified as critical factors for maintaining comfort across different climates [6][8]. Modular panel systems allow flexible layouts and scalable unit combinations, supporting various household sizes and shared community spaces [7]. Material selection is therefore shown to influence not only environmental performance and reusability but also occupants' sense of safety, comfort, and acceptance of temporary housing solutions [9][10].

The integration of BIM and MR represents a shift from traditional, asynchronous file-based exchanges toward synchronous, model-based collaborative environments [14]. BIM parametric designs are useful for managing the prefabrication supply chains and mass customization of houses [15]. MR distinguishes itself from other immersive technologies by overlaying virtual models onto physical contexts to facilitate real-time decision-making and on-site coordination [16][17]. Current literature suggests that integrating immersive technologies with parametric BIM models allows for the simulation of complex variables, such as energy performance and material impacts, within a shared digital-physical space [18][19].

To overcome the interoperability limitations of traditional formats like IFC and gbXML, recent research emphasizes the necessity of real-time, bidirectional protocols [19]. By utilizing low-latency frameworks such as WebSocket alongside specialized toolkits like Microsoft MRTK-3, designers can establish an open-source toolkit. This integration ensures data continuity and enables multiple users to visualize, edit, and validate housing designs simultaneously, ultimately reducing material wastage and ensuring that climate-responsive shelters can be optimized based on their specific geographic and economic constraints [20][21][22].

Conventional web communication protocols, such as HTTP and TCP/IP, are reliable but limited for real-time collaboration between BIM and MR [23]. They rely on request-response models that require established and closed connections, which causes non-negligible latency [24]. In HTTP, every client's request starts a new connection. It receives the server response after transmitting header data, creating significant delays for real-time connections required by multi-BIM collaboration [25][26]. As for TCP/IP, it requires frequent handshakes and acknowledgments, making it inefficient in terms of low-latency synchronization scenarios [24].

On the other hand, the WebSocket protocol solves these problems by providing a persistent, bi-directional channel over a single TCP connection, allowing clients and servers to exchange messages and data simultaneously without repeatedly building connections [27]. Once the initial handshake occurs, both client and server can push updates instantly, which reduces latency and resource burden. The research conducted by Łasocha and Badurowicz shows that when transmitting more than 100 copies of data using the WebSocket protocol (compared to HTTP), performance can be increased by several hundred percent [24]. While latency tests run by Pandey and Gill present that under high concurrency (5,000 users), the WebSocket application maintained an average latency of 30ms, compared to an average of 120ms for the HTTP long-polling application [28].

3 BIM-based modular housing design

3.1 Conceptual design of the house in BIM

The proposed system of temporary modular housing for climate refugees is grounded in the principles of mobility, modularity, and sustainability. Its primary aim is to deliver dignified, energy-efficient housing that can be deployed quickly and adaptable in different climatic conditions and social settings.

Each unit, Residential Single, Residential Double, Residential Multi-room, Office, and technical modules, is designed as part of a flexible and adaptable assembly system. Depending on the emergency situation, modules may function as standalone units or be combined to form compact settlements. Residential units are intended for two- and four-person households, while larger multi-room and dormitory configurations allow for shared facilities such as communal kitchens and sanitary cores. The detailed floor layout of the Residential Double module is shown in Figure 1, while Figure 2 presents a three-dimensional schematic view of the module, illustrating its spatial organization and structural logic. This modular logic supports the development of scalable settlements that encourage social interaction and community cohesion.

The housing modules are conceived as self-contained units that can also be interconnected through shared energy and water infrastructure where appropriate. This approach allows communities to function semi-autonomously while maintaining resilience in situations where existing infrastructure may be disrupted by natural disasters.

The modular system enables flexible reconfiguration across different terrains, climates, and settlement scales, with units operating independently or forming larger housing clusters. To accommodate varied site conditions, the housing units are designed to be installed on pile foundations, which allow construction on uneven or challenging terrain with minimal ground intervention. Lightweight construction and dry-assembly methods allow efficient installation, dismantling, and relocation. This approach corresponds with research on modular and off-site construction, which highlights improvements in response speed, cost efficiency, and sustainability in post-disaster housing [1].

3.2 Design philosophy and architectural criteria for climate refugee value

The design philosophy is based on a human-centered and regenerative approach that addresses both immediate emergency needs and long-term recovery, with a clear focus on architectural value for climate refugees. The primary objective is to create architecture that supports dignity, safety, and a sense of identity under temporary living conditions.

Dignity and psychological comfort are addressed through architectural criteria that prioritize privacy, daylight, and ventilation. Each housing unit provides a well-ventilated and daylit interior while avoiding the repetitive uniformity typical of refugee camps. Instead, the layouts emphasize personal space and establish a

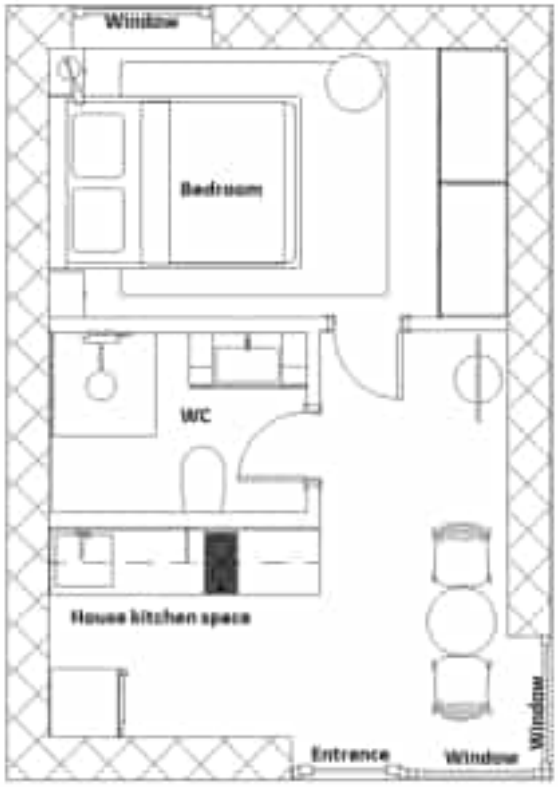


Figure 1. Detailed layout of the residential module



Figure 2. 3D view of the residential double module

hierarchy between private, semi-private, and shared zones, reinforcing residents' well-being and autonomy. Generous window openings, glazed entrance doors, and a roof-level window in the sanitary unit enhance the quality of natural light and improve overall spatial comfort without relying on artificial systems.

Furthermore, research in the fields of post-disaster architecture and humanitarian architecture emphasizes that community-oriented spatial planning, such as communal kitchens, open gathering spaces, and shared sanitation modules, can strengthen social cohesion and community-led recovery processes, transforming temporary shelters into vibrant social infrastructures. In line with this approach, the proposed modular system not only acts as a response to emergency housing needs but also as a platform for community resilience, allowing residents to regain a sense of normality, activity, and belonging in transitional conditions.

Aesthetic and spatial identity contribute to the architectural value of temporary housing by reinforcing a sense of home. The architectural expression is intentionally restrained, prioritizing functional clarity and emotional warmth. An asymmetrical roof form introduces visual variation while reinforcing a sense of shelter, while the combination of natural wood surfaces and anthracite window frames creates a familiar and calming domestic atmosphere that supports dignity and acceptance in temporary living environments.

3.3 Material selection to incorporate human aspects

Material selection is essential for addressing human needs in temporary housing for climate refugees. In addition to structural performance, materials influence indoor comfort, health, and psychological well-being. Studies in humanitarian architecture show that healthy indoor environments and natural materiality support emotional stability and adaptation of displaced populations [24][25].

Bio-based materials such as wood and straw contribute to improved indoor air quality and moisture regulation due to their breathable properties [7][9]. These characteristics help maintain stable indoor conditions and reduce dependence on mechanical systems, which is especially important in temporary and emergency housing. Research also indicates that natural materials create a more familiar and non-institutional atmosphere, positively affecting residents' mental well-being [26].

The use of EcoCocon straw-based prefabricated panels supports a human-centered design approach by combining high thermal insulation with vapor permeability and low environmental impact [3]. Thermal stability reduces stress caused by climate extremes, while renewable, low-carbon materials align with ethical and sustainability principles relevant to long-term displacement scenarios.

The structural and envelope system is based on EcoCocon straw-prefabricated panels used for both walls and roofing [4]. These panels are carbon-negative, offer high thermal insulation performance (U-value $\approx$ 0.12 W/m²K), and remain vapor-permeable while achieving a high level of airtightness. Interior partition walls are designed as lightweight gypsum board systems, allowing

flexible space planning and ease of modification over time. Construction is based on a dry-assembly process, eliminating the need for wet construction methods or heavy machinery. This enables fast on-site installation as well as straightforward disassembly when relocation is required, aligning with the project's sustainability and circularity objectives.

3.4 User experience: Contractors, owners, and Humanitarian organizations

User experience in temporary housing systems for climate refugees involves several stakeholder groups, including contractors, owners, and humanitarian organizations. For contractors, modular construction systems must ensure clear assembly logic, standardized components, and minimal reliance on specialized equipment. Prefabricated panel systems, such as straw-based EcoCocon panels, support dry construction processes, which simplify on-site work and improve construction reliability in emergencies and remote contexts [4][7].

From the owners and project coordinators' perspective, adaptability and resource efficiency are key factors. Modular housing solutions that allow reconfiguration, reuse, and relocation improve long-term value and operational flexibility. This approach also results in cost savings for owners, as the housing units are designed for repeated assembly and multiple life cycles rather than single-use deployment. Research in modular and off-site construction highlights that standardized systems reduce maintenance complexity and support scalable deployment in humanitarian housing projects [6].

Humanitarian organizations act as coordinators between residents, local authorities, and technical partners. Organizations such as Okanagan 4 Ukraine focus on community support, integration, and well-being of displaced persons through coordinated services, partnerships, and local engagement [8]. Their activities demonstrate that housing solutions must not only provide shelter but also enable social interaction, access to support services, and community-building. Therefore, ease of deployment, flexibility, and compatibility with humanitarian operations are essential design criteria for temporary housing systems.

4 Development of a digital design toolkit

4.1 Integration of BIM and MR

The integration of BIM and MR is executed through a self-developed Low-latency Mixed Reality Design Platform (LMRDP), which is discussed in detail in 4.2. The BIM model of the climate refugee housing, together with local climate data (temperature, wind, etc.), adapted to it and available cost-effective materials, is transferred to Unity using a customized Revit add-in and finally deployed to Microsoft HoloLens-2 for visualization and interaction in the MR environment. This add-in utilizes the Revit API to extract high-fidelity geometry and material data, exporting them as OBJ and MTL files to a central cloud WebSocket server.

The integration is completed within a 3D game engine (Unity), where the digital assets are reconstructed into a digital dashboard. By leveraging the Microsoft MRTK-3, the platform enables spatial awareness and hand-tracking interactions, allowing users to manipulate the model (scale, rotate, and translate) in a real-world physical context. This workflow ensures that complex architectural and environmental data from the BIM environment is not only visualized but is also interactive and actionable for on-site decision-making.

Furthermore, the toolkit integrates the digital dashboard directly into the MR application. This dashboard can link the immersive model to any web-based platform, offering the automation features of the dashboard. The interface of this digital dashboard is shown in Figure 3.



Figure 3. Digital design dashboard of the integrated BIM system in a 3D game engine

4.2 Open-source design toolkit

The LMRDP is developed as an open-source toolkit to facilitate collaborative design among architects, engineers, and humanitarian organizations. The architectural backbone of the toolkit is a Node.js WebSocket server hosted on a remote Ubuntu instance on the Amazon Web Services platform, which enables persistent, real-time communication between distributed clients. This configuration replaces traditional, asynchronous file transfers with a low-latency pipeline capable of transmitting binary data in sequential 8KB chunks, ensuring that model updates are synchronized across the design team instantly.

To lower the barrier of entry for the stakeholders, the

toolkit includes a customized "WebSocket Connector" plugin for Unity. This plugin provides a dockable interface that allows users to list, preview, and import BIM models from the server with a single click, automatically parsing the OBJ/MTL data into prefabricated Unity assets, as shown in **Error! Reference source not found.**. By providing a unified interface for real-time data exchange and environmental analysis, the LMRDP democratizes access to sophisticated design tools, allowing for the rapid development of affordable and sustainable climate refugee housing solutions in humanitarian contexts.



Figure 4. Customized Unity plugin dockable panel interface to enable WebSocket connection

4.3 Immersive user experience

The BIM model of the climate refugee housing design was imported into the Unity engine using a customized Revit add-in and a self-developed Unity plugin. This workflow allows both geometry and basic information to be transferred from Revit to a real-time 3D environment. The Unity interface was then integrated with the MRTK-3 and deployed to the MR headset (HoloLens). This setup enables MR visualization and interaction with the housing model in the real world.

The model can be visualized in Figure 5 in reality at a 1:1 scale. The virtual model of the building is aligned with the physical environment, so its size and position match real-world conditions. Users can move and rotate the model using hand gestures supported by the MR headset without any physical controllers. This makes the interaction more natural and easier to learn. Users can quickly test different positions and orientations of the building on site.

After placing the model, users can either walk around it to view the exterior from different angles or walk directly inside the building. The experience is similar to entering a real house. The interior includes key elements such as chairs, tables, WC fixtures, and a complete kitchen layout. All objects are shown at real scale. This helps users understand room size, ceiling height, and the distance between furniture items.

The immersive view allows users to better understand space and physical features. Circulation paths and room connections become easier to understand. Users can check whether the living space feels comfortable and practical. This is difficult to achieve with 2D paper drawings or screen-based 3D models. For climate refugee housing, this is especially important because such projects require efficient use of space and must provide safe and dignified living conditions.

Overall, the MR system turns the BIM model into an interactive and reality-based experience. The housing design is no longer only a digital file. It becomes a life-size virtual model that can be explored with realistic features. This enhances communication between designers, contractors, and owners. It also helps users to better understand the design layout. By combining BIM, Unity, and an MR headset, the workflow provides a practical way to evaluate spatial quality and usability at an early design stage.

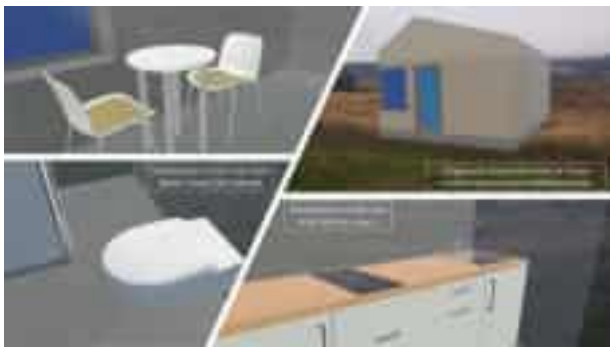


Figure 5. Visualization of the climate-displacement housing design model in MR at 1:1 scale

5 Results and discussion

The study evaluates the performance of a climate-resistant housing design in terms of adaptability to varying climatic and topographical conditions, life-cycle flexibility, and logistical efficiency. The potential outcomes of the proposed approach illustrate that the use of bio-based prefabricated panels facilitates multi-cycle reuse, simplifies transportation and assembly processes, and enhances environmental sustainability. Furthermore, the integration of modular planning and off-site fabrication improves construction efficiency while contributing to improved living conditions, including thermal comfort, spatial flexibility, and social adaptability for occupants.

The study also emphasizes that the proposed modular housing system effectively combines architectural, environmental, and human-centered criteria to meet the needs of climate refugees. The analysis indicates that the use of prefabricated bio-panel systems allows for rapid

deployment of housing, adaptation to different climatic and topographical conditions, and use over multiple life cycles. Modular logic allows housing units to function independently or as part of larger settlements, supporting scalability and flexible community formation depending on emergencies.

In addition, the research also focuses on the human aspect of design and how to make people feel safe, comfortable, and part of a community even after losing their homes. To enhance user experience in the design approach, a digital toolkit using MR and WebSocket has been developed. A high-fidelity BIM model of housing design, with local climate data (temperature, wind, etc.), is integrated into an MR environment through a self-developed low-latency MR design platform using Revit API, WebSocket protocol, and the Microsoft MRTK-3. Thus, the integration of this digital design toolkit emphasizes that the system offers clear advantages for contractors, owners, and humanitarian organizations by reducing on-site labor requirements, simplifying logistics, and enabling cost-effective reuse. These findings are consistent with existing research on modular and off-site construction, suggesting the relevance of the proposed approach for scenarios of population displacement following disasters and due to climate change.

Moreover, a detailed comparative analysis of MR-based design review against conventional methods, including user studies and quantitative performance metrics, has been conducted as part of a parallel study and will be reported separately to maintain the focus of the current work.

6 Conclusions

The increasing displacement of populations in climate-affected areas requires rapid, adaptable, and sustainable housing solutions. To address this challenge, this research developed an immersive (Mixed Reality-enabled) digital toolkit that integrates modular design principles, BIM parametric designs, and immersive visualization technologies.

The study utilized modular housing design based on prefabricated bio-panel systems as an effective response to the challenges associated with the temporary accommodation of climate refugees. The proposed approach increases construction efficiency through dry assembly methods, reduces on-site labor requirements, and allows for rapid deployment and relocation of buildings in different climatic and topographical conditions. Moreover, modular design supports the principles of reuse, scalability, and circular construction, while the choice of architecture and materials enhances indoor comfort and social well-being.

In addition, a digital toolkit using MR and WebSocket is developed to enhance the visual experience of stakeholders. The parametric housing design, including the local climate data (temperature, wind, etc.) and available cost-effective materials, is integrated into an MR environment through the low-latency MR design platform using Revit API, WebSocket protocol, and the Microsoft MRTK-3. By bridging the gap between static architectural data and immersive environments, the toolkit provides a platform for stakeholders to visualize, iterate, and optimize climate refugee housing solutions in real-time.

The proposed design workflow allows stakeholders to interact with housing designs in real time, enabling interactive exploration and collaborative decision-making, especially related to cost-effective material selection. By using MR-based solutions, the toolkit provides an intuitive and engaging platform for designing climate-resilient housing that is both cost-effective and context-sensitive.

Although the study has presented some promising outcomes, it still has some limitations. The proposed system has been evaluated at a conceptual and prototype level, and it still lacks some field implementation or stakeholder-based validation. To address these limitations, future work will focus on conducting field testing, engaging stakeholders (e.g., Okanagan 4 Ukraine, Habitat for Humanity), and performing quantitative evaluations to further evaluate the field implementation and validation of the proposed approach.

References

[1] P. Dallasega, E. Rauch, and M. Frosolini. A lean approach for real-time planning and monitoring in engineer-to-order construction projects. *Buildings*, 8(3):38, 2018.

[2] F. Kedir, Q. Chen, D. M. Hall, B. T. Adey, and R. Boyd. Formative scenario analysis of the factors influencing the adoption of industrialised construction in countries with high housing demand – the cases of Ethiopia, Kenya, and South Africa. *Construction Management and Economics*, 40(9):690–710, 2022.

[3] EcoCocon. FAQ. On-line: https://ecococon.eu/faq, Accessed: 27/03/2026.

[4] EcoCocon. EcoCocon – Sustainable prefabricated straw panels for high-performance buildings. On-line: https://www.ecococon.eu, Accessed: 27/03/2026.

[5] CLT Rezult. Technical characteristics. On-line: https://clt-rezult.com/tekhnichni-kharakterystyky/, Accessed: 27/03/2026.

[6] Super SIP. SIP panels: What is a SIP panel? Technical information. On-line: https://www.supersip.house/sip-panels/, Accessed: 27/03/2026.

[7] L. Tupenaite, L. Kanapeckiene, J. Naimaviciene, A. Kaklauskas, and T. Gecys. Timber construction as a solution to climate change: A systematic literature review. *Buildings*, 13(4):976, 2023.

[8] SIP Building Systems. *Technical papers*. On-line: https://www.sipbuildingsystems.co.uk/technical-papers.php, Accessed: 27/03/2026.

[9] C. Pezzica, V. Cutini, C. Bleil de Souza, and D. Aloini. The making of cities after disasters: Strategic planning and the Central Italy temporary housing process. *Cities*, 131:104053, 2022.

[10] Nexii. Nexii green construction technology. On-line: https://www.nexii.com/, Accessed: 27/03/2026.

[11] Ecohome / Hempbud UA. Kithemp wooden frame construction set — Product listing. On-line: https://www.ecohome.net/en/products/1992/hempbud-ua-kithemp-wooden-frame-construction/, Accessed: 27/03/2026.

[12] Element5. CLIPs product information. On-line: https://elementfive.co/products/clips/, Accessed: 27/03/2026.

[13] Hanit. Hanit - Our material. On-line: https://www.hanit.ca/Our-material/hanit, Accessed: 27/03/2026.

[14] M. Johansson and M. Roupé. Real-world applications of BIM and immersive VR in construction. *Automation in Construction*, 158:105233, 2024.

[15] S. Metvaei, K. Aghajamali, Q. Chen, and Z. Lei. Developing a BIM-enabled robotic manufacturing framework to facilitate mass customization of prefabricated buildings. *Computers in Industry*, 164:104201, 2025.

[16] M. Fawad, M. Salamak, Q. Chen, M. Uscilowski, K. Koris, M. Jasinski, P. Lazinski, and D. Piotrowski. Development of immersive bridge digital twin platform to facilitate bridge damage assessment and asset model updates. *Computers in Industry*, 164:104189, 2025.

[17] B. Su, M. Fawad, Q. Chen, and M. Salamak. Mixed reality-based digital twinning of building circularity: A co-design approach for sustainable buildings. In *Proceedings of the 42nd International Symposium on Automation and Robotics in Construction*, pages 1590–1596, Montreal, Canada, 2025.

[18] Y. Huang. Evaluating mixed reality technology for architectural design and construction layout. *Journal of Civil Engineering and Construction Technology*, 11(1):1–12, 2020.

[19] K. Zhou. Innovative applications of XR in architectural sustainability: Current trends and future prospects. *Academic Journal of Science and Technology*, 10(3):69–73, 2024.

[20] A. Ciccozzi, T. De Rubeis, D. Paoletti, and D. Ambrosini. BIM to BEM for building energy analysis: A review of interoperability strategies. *Energies*, 16(23):7845, 2023.

[21] H.-E. Cheng, T.-H. Le, and J. J. Lin. Multi-user mixed-reality collaboration framework for onsite MEP coordination. In *Proceedings of the Construction Research Congress*, pages 329–338, Des Moines, IA, USA, 2024.

[22] C. Chen, L. Cheng, X. Zhao, and W. Zhang. Websocket-based light quantitative BIM models for instant linkage building. In *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence*, pages 434–438, Dongguan, China, 2024.

[23] N. Ivaki and F. Araujo. Fault-tolerant bi-directional communications in web-based applications. In *Proceedings of the 20th IEEE International Conference on Parallel and Distributed Systems*, pages 833–836, Hsinchu, Taiwan, 2014.

[24] W. P. Łasocha and M. Badurowicz. Comparison of WebSocket and HTTP protocol performance. *Journal of Computer Sciences Institute*, 19:67–74, 2021.

[25] M. Marzouk, S. Azab, and M. S. Metawie. BIM-based approach for optimizing life cycle costs of sustainable buildings. *Journal of Cleaner Production*, 188:217–226, 2018.

[26] K. Lu, X. Deng, X. Jiang, B. Cheng, and V. W. Y. Tam. A review on life cycle cost analysis of buildings based on building information modeling. *Journal of Civil Engineering and Management*, 29(3):268–288, 2023.

[27] V. Sharma, M. K. Pandey, J. Gupta, and H. Kishan. Peer to peer communication using Web Sockets and Web RTC. In *Proceedings of the 2024 International Conference on Emerging Innovations and Advanced Computing*, pages 627–634, Sonipat, India, 2024.

[28] B. Pandey and D. P. Gill. Enhancing real-time web applications with WebSockets: Performance and responsiveness. International *Journal of Novel Research and Development*, 9(6):a810–a812, 2024.