Proceedings of the 41th International Symposium on Automation and Robotics in Construction

P

Lille, France, June 3-5, 2024

Extracting roof sub-components from orthophotos using deep-learning -based semantic segmentation

Jiajun Li¹, Boan Tao¹, Frédéric Bosché¹, Chris Xiaoxuan Lu² and Lyn Wilson³

¹School of Engineering, University of Edinburgh, UK ²School of Informatics, University of Edinburgh, UK ³Historic Environment Scotland, UK

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

Abstract -

Best practice for the detection and annotation of visible defects in slated roofs is by annotation of photos, ideally orthophotos. If such a process is to be effectively automated in support of emerging Digital Twinning solutions, it is necessary to first recognise the external sub-components of the roof in the orthophotos, in particular the slated and leadwork areas. Using a dataset composed of many photos from two historic buildings, this study develops and compares different deep-learning -based semantic segmentation models to segment roof orthophotos into slated areas, leadwork, and 'other' areas. Since orthophotos typically contain pixels which do not belong to the roof panel (black 'background' pixels), the method employs a subsequent 'background' label correction step. The best-performing model is found to be PointRend with Focal Loss: overall *aAcc* = 99, *mIoU* = 88.91, and mAcc = 92.77; for slate class, IoU and Acc is nearly 100; for leadwork class, IoU and Acc is around 90.

Keywords -

Semantic segmentation; Deep learning; Slated roof; Orthophoto

1 Introduction

For most traditional slated building roofs, they are mainly composed of different elements: slate roofing tiles, leadwork, stonework, and masonry [1]. Additional decorative or functional accessory components may also exist, e.g. ventilator, balustrade, satellite, etc. Slates occupy the largest area and are where deteriorations happen easily and frequently, especially due to weathering with time [2]. To better detect the visible defects in roof monitoring, the annotation of photos is a common practice. The best practice to date employs orthophotos due to their benefits for length and area quantification.

Within the emerging area of Digital Twinning in the built environment [3], there is great interest in automating data acquisition and processing for building fabric monitoring, in order to efficiently, robustly and safely detect and monitor defects, and support computer-aided maintenance decision making. In the context of slated roofs, we showed in [4] how orthophotos can be generated for individual roof panels, from UAV-acquired photogrammetric data and the building's 3D digital twin model. To ensure effective defect detection in each such orthophoto, it is next necessary to distinguish the sub-components inside the orthophoto, in particular the slated and leadwork areas.

Semantic segmentation was developed decades ago, and can be applied to different kinds of data, from 2D image to 3D point cloud, and even video data. When applied to 2D images, it enables assigning a class label to each pixel of an image, and such pixel-level semantic information can help make judgements or be leveraged by other tasks [5, 6]. With the development of deep learning, different deep architectures have been introduced, especially Conventional Neural Networks (CNNs). As a result, the performance of semantic segmentation models has been greatly improved, not only in accuracy but also in efficiency [7].

In construction informatics research dealing with 2D and 3D data, such as in Scan-to-BIM, semantic segmentation is increasingly regarded as an essential step after data collection, to provide further information useful to subsequent tasks such as object detection [8]. This technique has been applied to different types of data, including: (1) 2D image of indoor scene [9] and aerial images of different architecture [10]; (2) 3D point cloud of building interiors [8], plumbing and structural components [11], autonomous vehicles and robot navigation [12].

This paper reports on the development and comparison of well-established deep-learning -based semantic segmentation models for segmenting orthophotos of individual roof panels into 'background', 'slate', 'lead', and 'other' classes.

2 Related work

With its powerful pixel-level segmentation ability, semantic segmentation has been developed into a wellestablished tool in Computer Vision.

Most recently, different deep learning models have been proposed for 2D image data. *DeepLabV3*, a widely used

system developed by Google [13], can handle the problem of segmenting objects at multiple scales with cascaded module and Atrous Spatial Pyramid Pooling(ASPP). Compared to *DeepLabV3*, *PointRend* demonstrates better performance (higher mIoU value), by extracting the point features made up on fine-grained features and coarse prediction [14]. Specifically, *PointRend* handles better the problems of smooth region and blurry contours from bilinear sampling process. Thanks to the recognisable features, the object can be easily detected and classified for this task, but the contour detection should be more precise.

In recent years, Vision Transformer (ViT) has achieved superior performance to the commonly used CNNs, by splitting each image into patches. Due to their outstanding performance in image classification, these models have also been explored for semantic segmentation: *Segmenter* can capture global interaction between elements of a scene using transformer, rather than the traditional convolutions, which would easily cause information loss [15]. *Seg-Former* reaches better performance and efficiency by redesigning the Transformer encoder and uses a simple multilayer perceptron (MLP) decoder[16]. However, all authors also point out that ViT relies heavily on large-scale datasets to achieve good performances [17].

Within these main classes of deep learning models, different variants can be created by modifying different some components, such as the loss function. Focal Loss is a commonly used loss function for dealing with class imbalance and putting more focus on the hard and misclassified examples, by multiplying each class loss with a weighting factor [18]. Dice Loss can deal with the imbalance problem between foreground and background, by giving more importance on foreground than background, thus making the model region-related [19]. These two loss functions and the default Cross-Entropy Loss can be paired and used together in improving model training [20].

3 Method

We assume as input an orthophoto of a slated roof panel generated by the method presented in [4]. The generated orthophoto shows the roof panel, which is composed of different essential sub-components, slates and leadwork, as well as other less frequent elements (e.g. stone, grid, equipment, glass, ladder, etc.), as can be seen in Figure 5. Since the intention of this work is to find defects in slated areas as well as leadwork areas, we must first segment the orthophoto to isolate these particular subparts. For this, we explore different methods for (pixel-level) semantic segmentation.

3.1 Dataset

The data output of Duff House in Banff, Scotland by Li et al. [4] includes data for 36 panels. We split these into 27 panels for training and 9 panels for testing. For each panel, we retain maximum 10 unmerged orthophotos obtained from different photos of the same panel (with different angles and covering the panel in various ways). This dataset is used for developing and comparing different initial semantic segmentation models.

In addition, the data output of St Mary's Church, in Stirling, Scotland includes data for 18 panels. With this data, we created another dataset composed of one unmerged orthophoto per panel. This dataset is used solely to test the generalisability of the models created using the Duff House dataset.

Finally, a combined dataset using data from the two buildings is created. It is divided into a training dataset that is ~80% of the overall dataset (27 panel orthophoto data of Duff House, along with 13 panel orthophoto data of St Mary's Church) and a testing dataset that is the remaining ~20% (9 from Duff House, along with 5 from St Mary's Church). Here, only one unmerged orthophoto is kept for each panel. We select the orthophoto with the largest coverage of the panel.

All the orthophotos used in the datasets above are manually labelled in 4 classes: *background* (labelled as 0), *slate* (1), *leadwork* (2), and *other* (3). The *other* category contains: stone, grid, equipment, glass, ladder, etc. The class *other* normally occupies a very small proportion of pixels in orthophotos.

3.2 Deep learning model

As discussed earlier in 2, these are the state-of-art models that can be useful to the specific problem in this study: *DeepLabV3*, *PointRend*, *Segmenter*, *Segformer*. Basic hyper-parameter settings are tuned for best performance by comparing these deep learning models. Our settings are reported in Table 1. All models were pre-trained using their default weights.

Table 1. Experimental parameters								
Parameter	Value							
Batch size	4							
Max iteration	2400							
Validation interval	400							
Training vs. Validation	75% : 25%							

During training, the input images are not rescaled. This is because most of the roof orthophotos are rectangular with varying width-to-length ratios, and the resizing processing operations that are typically applied in semantic segmentation pipelines would result in information loss. But, to meet the requirements of input image size and computing capacity limits, all images are cropped (tiled) and 3.4 Semantic segmentation results read as $512 \times 1024 \times 3$ matrices.

As will be shown in the Experimental Validation (Section 4), PointRend with default backbone (ResNet) and loss function (Cross-Entropy Loss) achieves best performance all four initial models. As a result, further experiments are conducted by using different loss functions including Dice Loss (sensitive to region detection) and Focal Loss (sensitive to imbalance problem). With grid search as a tool for hyperparameter adjustment, specific weights between different losses are selected for best performance.

The models trained with the Duff House dataset were then tested with the St Mary's Church dataset to assess its generalisation ability. The results lead us to finally use the combined orthophoto dataset (Duff House + St Mary's Church) for training and testing a final model with best segmentation performance and generalisation ability.

3.3 **Background label correction**

In most deep learning semantic segmentation methods, precise delineation of segmentation boundary is a challenging problem, with many confusions arising at those boundaries. Figure 1 illustrates this issue with a typical output of the semantic segmentation models we explored. The figure highlights the discrepancies between the Prediction and Ground Truth (GT) for the different classes.

However, in this study, the set of background pixels is actually known a priori, because the panel boundary is defined by the Digital Twin 3D model projection (see [4] for details). In other words, the Ground Truth for the background class is known a priori. Therefore, an extra step is introduced to correct the False Positive (FP) and False Negative (FN) results for the *background* class:

- 1. False Negative (*Prediction* \neq 0 and *GT* = 0): In this case, the predicted label is simply changed back to '0' (i.e. background class).
- 2. False Positive ($GT \neq 0$ and Prediction = 0): In this case, the predicted label is changed from the background class to the most likely other class. As illustrated in Figure 2, for each FP pixel the non-zero label that appears most frequently in the 3×3 grid around it is selected as the new label. If the grid contains only background pixels (i.e. class '0'), then the grid is expanded by one pixel (i.e. 5×5 grid) and the process is repeated until a at least a non-zero label is found. As will be shown, this simple process works well in our context.

As will be shown in the Experimental Results (Section 4), while the baseline semantic segmentation results are already good, this process delivers some additional improvements.

The model trained in Section 3.2 is tested using individual orthophotos generated by the process described in [4], which may not necessarily cover entirely a given roof panel of interest or may overlap. To obtain one single orthophoto covering the overall panel with a unified semantic segmentation result, the labels of individual orthophotos must be merged. For this, for each pixel, the label that appears most frequently among the unmerged orthophotos is selected as the final label. In cases when two (or more) classes have the same frequencies for all unmerged orthophotos, the final label will be selected in the following order of priority: other, leadwork, slate. For example, if *slate* and *other* appear the same time for one pixel, then other would be chosen as the final label, because we observed that the other objects always lay above the *slate* surface.

Experimental results 4

In this section, evaluation metrics for semantic segmentation are first introduced. Then all the experimental results are reported. First, different models are trained using the Duff House training dataset and tested using the Duff House and St Mary's Church testing datasets. The best model is selected by comparing these results, and it is finally re-trained and tested using the combined dataset.

All the training and testing work is completed in the Google Colab Pro environment, with NVIDIA A100 GPU 40 GB.

4.1 Evaluation metrics

The segmentation results for each class are evaluated using two parameters: Intersection over Union (IoU): computed by contrasting the Prediction and Ground Truth segmentations; and Accuracy (Acc): calculated by dividing the sum of the True Positive pixels by the sum of the True Positive pixels and False Positive pixels. To compare the testing performance between different models, the following overall evaluation metrics are evaluated:

- *aAcc*: the Accuracy of all pixels, evaluating the classification accuracy.
- *mIoU*: the mean *IoU* of all classes; *mIoU* is an important indicator to measure the accuracy of overall semantic segmentation.
- *mAcc*: the mean *Acc* of all classes, evaluating the overall performance in pixel classification.

In the following, we report results for different models in the form of tables and confusion matrices. In the tables, we report for each model: the aAcc, mIoU and mAcc for both overall and overall (excl. background), and then IoU and Acc for each of the four classes. For the confusion



Figure 1. Colour coded confusion matrix (left), semantic segmentation model output (middle), illustrations of FP/FN from the perspective of *background* class(right).



Figure 2. Illustration of the correction of FP for the *background* class. Left: Example 1 where the label is corrected to '1' after one step; Right: Example 2 where the label is corrected to '2' after two steps.

matrices, we report both absolute (in pixel counts) and relative (in percentages) confusion matrices.

4.2 Background label correction

Regardless of the model employed, the *Background model correction* step described in Section 3.3 can be applied to correct FP and FN for the *background* class. We thus demonstrate the benefits of this correction using one model (which we will see later performs well): PointRend (CEL+FL).

Table 2 and Figure 3 show the results obtained when training and testing this model on Duff House dataset, before and after applying the *background* label correction results.

While the baseline performance is already quite good (all metrics > 85 and most of them are > 95), the additional corrective step improves performance for all classes, in particular the *leadwork* and *other* classes, which had the lowest performance without this correction. The higher increases in mIoU (+1.81 to 94.64) and mAcc (+1.19 to 97.24) also indicate a reduced difference in performance among the different classes. Although anticipated, this improvement is welcome, because the *background* regions are often next to *leadwork* regions. And so, any correction of *background* class would most likely benefit the *leadwork* class. Nonetheless, the results demonstrate the good performance of the proposed *background* label correction method.

4.3 Initial models with Duff House dataset

All models are first trained using Duff House training dataset only, and tested with the Duff House testing dataset. Table 3 presents the performance of the different models. Generally, all models already show good performance: the evaluation metrics of overall performance are all > 90, most of them are > 95 and even nearly 100. Generally, the target class, *slate*, is segmented satisfactorily. Errors mostly come from the classes *leadwork* and *other*.

By comparing the first 4 rows, PointRend stands out with the highest values in all evaluation metrics. The variants in the last three rows are then developed based on PointRend, in an attempt to enhance performance with regard to specific challenges with our dataset, namely data class imbalance (the *leadwork* and *other* classes occupy much fewer pixels than *background* and *slate*) and region ambiguity.

Among all the variant models, PointRend(CEL+FL) achieves the best overall performance. Though it is close to the default PointRend model, it increases the values on *other* class both in IoU (+0.89 to 88.52) and Acc (+0.88 to 94.76). In comparison, PointRend(CEL+DL), by adding the Dice Loss function, also improves the performance on *leadwork* class, but with a sacrificial drop on *other* class and accordingly, a decreased overall performance. Therefore, from all the 3 variant models, we conclude that the Focal Loss more successfully improves the results and accounts for data imbalance better than Dice Loss.

4.4 Testing generalisation with St Mary's Church testing dataset

Table 4 reports the evaluation results of all models developed in Section 4.3 on the testing dataset of St Mary's Church. This enables an assessment of the model's generalisation, since no data from St Mary's Church was used to train those models. Table 4 shows a similar pattern as Table 3: all models can segment the *slate* area more accurately than *leadwork* and *other* classes, while the *other* class has the worst performance among all classes. However, compared to the performance on Duff House, there are general decreases in the overall performance of all models, such as a nearly 30 drop in mIoU and nearly 20



Table 2. Testing results before and after background label correction

Figure 3. Confusion matrices before (left) and after (right) applying the background label correction step.

	e													
Model	overall		overall	overall (excl. <i>background</i>)		background		slate		leadwork		other		
Wiodel	aAcc	mIoU	mAcc	aAcc	mIoU	mAcc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
DeepLabV3	99.24	93.54	96.73	98.24	91.39	95.64	100	100	98.37	99.25	89.49	92.48	86.31	95.18
PointRend	99.36	94.46	97.11	98.52	92.61	96.15	100	100	98.59	99.3	91.62	95.28	87.63	93.88
Segformer	99.17	93.43	96.63	98.08	91.23	95.51	100	100	97.99	98.97	89.6	94.05	86.11	93.51
Segmenter	99.03	92.06	94.73	97.75	89.41	92.97	100	100	97.81	99.31	86.9	92.22	83.52	87.39
PointRend(CEL+DL)	99.03	90.78	93.51	97.75	87.71	91.34	100	100	98.1	99.45	88.4	96.16	76.63	78.42
PointRend(CEL+FL)	99.36	94.64	97.24	98.52	92.86	96.32	100	100	98.57	99.3	91.49	94.9	88.52	94.76
PointRend(DL+FL)	99.33	94.46	96.91	98.45	92.61	95.88	100	100	98.45	99.35	91.48	94.37	87.9	93.93

Table 3. Testing results on Duff House

drop in mAcc. When comparing models in different rows, performance values for the *leadwork* and *other* classes are lower and more spread out than the *other* class and the same classes in Table 3. In general, there is no prominent model that stands out in all metrics specifically for St Mary's Church.

Therefore, even though aAcc remains close to 100% and the general performance is acceptable, the generalisation ability of the models is limited. Put another way, unsurprisingly the models, when trained using data from only one building (Duff House), work but not sufficiently well for other buildings. Therefore, more data, especially with diverse features of *leadwork* and *other* classes, should be used for training.

4.5 Models trained with combined dataset

Based on the results above, the best performing model, PointRend (CEL+FL), is retrained using the the combined training dataset, and tested the combined testing dataset.

The overall performance, reported in Table 5, is almost at the same level as that of Table 3, whose performance was already very high. Looking at individual classes, a slight improvement is achieved for the *slate* class with IoU (+0.04 to 98.61) and Acc (+0.25 to 99.55). However, some reductions in performance are observed for the *leadwork* and *other* classes. But, importantly, compared to the results in Table 4, all metrics show significant improvements. This implies that the new model has achieved a greater level of generalisability (it performs well on testing data from both Duff House and St Mary's Church) without significant drop in overall performance. Naturally, this does not mean the new model will work in all cases of slated roofs; much more diverse data would need to be collected for that. But, the selected model performs satisfactorily.

4.6 Visualise semantic segmentation results

After merging the result labels using the strategy in Section 3.4, the resulting confusion matrix is reported in Figure 4. It shows that the *slate* class segmentation accuracy is still high, but the confusion between *leadwork* and *other* is not insignificant.

This is further illustrated with three example roof panels in Figure 5 and Figure 6. The segmentation result of Panel A is nearly flawless. In Panel B, though the situation is more complex (containing *other* pixels), the slated area of this orthophoto is generally segmented satisfactorily. However, there are still some pixel misclassifications specifically caused by the *other* class, with confusions observed particularly between *slate* and *other* at the bottom of this panel. In the result of Panel C, there are noise pixels

) ~		-)			0			8		
Model	overall		overall	overall (excl. <i>background</i>)		background		slate		leadwork		other		
Widder	aAcc	mIoU	mAcc	aAcc	mIoU	mAcc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
DeepLabV3	96.48	62.74	69.51	91.84	50.32	59.35	100	100	96.42	98.56	31.3	37.51	23.23	41.97
PointRend	96.66	65.65	75.62	92.26	54.2	67.49	100	100	96.37	97.9	45.93	75.46	20.3	29.11
Segformer	96.89	66.72	76.74	92.79	55.63	68.99	100	100	97.25	98.14	43.48	65.41	26.15	43.41
Segmenter	97.21	66.58	74.05	93.53	55.44	65.4	100	100	96.91	99.1	53.94	78.69	15.48	18.42
PointRend(CEL+DL)	96.39	64.17	76.33	91.63	52.23	68.44	100	100	96.12	97.3	43.48	85.3	17.1	22.73
PointRend(CEL+FL)	94.79	64.31	80.45	87.92	52.41	73.93	100	100	92.94	93.89	41.93	78.25	22.35	49.66

Table 4. Testing results on St.Mary's Church by models trained using Duff House training dataset

Table 5. Testing results on combined dataset

Model	overall		overall (excl. <i>background</i>)		background		sla	slate		leadwork		ıer		
	aAcc	mIoU	mAcc	aAcc	mIoU	mAcc	IoU	Acc	IoU	Acc	IoU	Acc	loU	Acc
PointRend(CEL+FL)	99.0	88.91	92.77	97.68	85.22	90.35	100	100	98.61	99.55	85.93	91.99	71.11	79.52

at the top of the roof panel, the GT of which is *slate* but predicted as *other*. The reason for this mistake possibly lies in the biological growth on the slate surface, which may still confuse the deep learning model despite some of the training data containing it. This issue may nonetheless be addressed through a more extensive training dataset.



Figure 4. Confusion matrices of results after applying merging strategy

5 Conclusions

PointRend(DL+FL)

Different deep learning models for semantic segmentation are developed and compared using a dataset composed of data coming from two traditional buildings: Duff House and St Mary's Church. PointRend added with Focal Loss (PointRend(CEL+FL)), trained by the combined dataset is chosen as the most suitable when considering both datasets jointly. All of its evaluation matrices, except the ones in *other* class, are all > 85%.

The performance of all models is enhanced thanks to an extra *background* label correction steps: by eliminating the confusion between *background* and other classes (especially *leadwork* and *other*, which are usually the surrounding area of *slate*), the *background* accuracy is corrected to 100%, and the accuracies of other classes are shown to also increase. However, the confusion matrix in Figure 4 shows that the confusion between *leadwork* and *other* is still significant, affecting the accuracy of *other*. This is possibly because that the *other* class includes many kinds of objects. While enhancing the model robustness may be achieved with more data, it must be highlighted again that our main focus is the effective segmentation of the *slate* and *leadwork* classes.

16.

63

The proposed method focuses on the traditional building roofs, but the methodologies developed are equally applicable to more modern roofs with a slate or tile construction.

It should be noted that the 'best' model is just marginally better than the other ones, with all of them perform reasonably well. The difference of testing results on Duff House and St. Mary's Church indicates the risk of over-fitting, which can be addressed by getting more data involved. With more data collected and used for training, the strategy of selecting the best and most robust model can be improved. Future work can thus first look at collecting more building roof data in order to further validate and improve the orthophoto generation pipeline, and train semantic segmentation models with greater generalisability. These data shall include roofs with various forms and shapes, including slate laying methods and different components around the slated areas, and in various conditions, containing different levels of deteriorations.

Future work should also look at the next step of our proposed overall pipeline, starting with the detection of defects in the slated areas in particular, and the leadwork areas as well. Due to the fact that all the created orthophotos have the same orientation and uniform scale, we anticipate that this should ease the development of further machine learning models.

6 Acknowledgements

This paper was made possible thanks to research funding from Historic Environment Scotland (HES) and the Engineering and Physical Sciences Research Council (EP-SRC) [grant reference EP/W524384/1]. The views and



(c) Panel C.

Figure 5. Orthophotos of typical panels.





Figure 6. Semantic segmentation results of orthophotos of the typical panels shown in Figure 5

opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of HES and EPSRC. The authors would also like to acknowledge the HES Digital Documentation and Innovation team, Stirling City Heritage Trust for providing us with the data used in the experiments reported in this paper. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

- [1] Roger Curtis and Jessica Hunnisett Snow. guide-climate Short adaptation change for traditional buildings. 2016. URL https://www.historicenvironment.scot/ archives-and-research/publications/ publication/?publicationId= a0138f5b-c173-4e09-818f-a7ac00ad04fb.
- [2] J Walsh. Predicting the service life of natural roofing slates in a scottish environment. In 9th international conference on durability of building materi-

als and components. Brisbane: In House Publishing, 2002. URL https://inspectapedia.com/ roof/Scottish-Roof-Life-Walsh.pdf.

- [3] Min Deng, Carol C Menassa, and Vineet R Kamat. From bim to digital twins: A systematic review of the evolution of intelligent building representations in the aec-fm industry. *Journal of Information Technology in Construction*, 26, 2021. doi:10.36680/j.itcon.2021.005.
- [4] Jiajun Li, Frédéric Bosché, Chris Xiaoxuan Lu, and Lyn Wilson. Occlusion-free orthophoto generation for building roofs using uav photogrammetric reconstruction and digital twin data. In 40th International Symposium on Automation and Robotics in Construction, pages 371–378, 2023. doi:10.22260/ISARC2023/0051.
- [5] Yanming Guo, Yu Liu, Theodoros Georgiou, and Michael S Lew. A review of semantic segmentation using deep neural networks. *International journal* of multimedia information retrieval, 7:87–93, 2018. doi:10.1007/s13735-017-0141-z.

- survey on semantic segmentation with deep learning. Neurocomputing, 406:302-321, 2020. ISSN 0925-2312. doi:10.1016/j.neucom.2019.11.118.
- [7] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. arXiv, 2017. doi:10.48550/arXiv.1704.06857.
- [8] Jong Won Ma, Thomas Czerniawski, and Fernanda Leite. Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic bim-based point clouds. Automation in Construction, 113:103144, 2020. ISSN 0926-5805. doi:10.1016/j.autcon.2020.103144.
- [9] Liu Yang and Hubo Cai. Cost-efficient image semantic segmentation for indoor scene understanding using weakly supervised learning and bim. Journal of Computing in Civil Engineering, 37(2):04022062, 2023. doi:10.1061/JCCEE5.CPENG-5065.
- [10] Biswajeet Pradhan Abolfazl Abdollahi and Abdullah M. Alamri. An ensemble architecture of deep convolutional segnet and unet networks for building semantic segmentation from high-resolution aerial images. Geocarto International, 37(12):3355-3370, 2022. doi:10.1080/10106049.2020.1856199.
- [11] Chao Yin, Boyu Wang, Vincent J.L. Gan, Mingzhu Wang, and Jack C.P. Cheng. Automated semantic segmentation of industrial point clouds using responntnet++. Automation in Construction, 130:103874, 2021. ISSN 0926-5805. doi:10.1016/j.autcon.2021.103874.
- [12] Hanyu Shi, Guosheng Lin, Hao Wang, Tzu-Yi Hung, and Zhenhua Wang. Spsequencenet: Semantic segmentation network on 4d point clouds. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4573-4582, 2020. doi:10.1109/CVPR42600.2020.00463.
- [13] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. arXiv, 2017. doi:10.48550/arXiv.1706.05587.
- [14] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. Pointrend: Image segmentation as rendering. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 9796-9805, 2020. doi:10.1109/CVPR42600.2020.00982.

- [6] Shijie Hao, Yuan Zhou, and Yanrong Guo. A brief [15] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pages 7242-7252, 2021. doi:10.1109/ICCV48922.2021.00717.
 - [16] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems, volume 34, pages 12077-12090. Curran Associates, Inc., 2021. URL https://proceedings. neurips.cc/paper_files/paper/2021/file/ 64f1f27bf1b4ec22924fd0acb550c235-Paper. pdf.
 - [17] Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zihang Jiang, Francis E. H. Tay, Jiashi Feng, and Shuicheng Yan. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pages 538-547, 2021. doi:10.1109/ICCV48922.2021.00060.
 - [18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 2999-3007, 2017. doi:10.1109/ICCV.2017.324.
 - [19] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Fully convolutional neural Ahmadi. V-net: networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV), pages 565-571, 2016. doi:10.1109/3DV.2016.79.
 - [20] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics, 95:102026, 2022. ISSN 0895-6111. doi:10.1016/j.compmedimag.2021.102026.

A Step from Virtual to Reality: Investigating the Potential of a Diffusion-Based Pipeline for Enhancing the Realism in Fully-Annotated Synthetic Construction Imagery

Sina Davari, Ali Tohidifar, and Daeho Kim

Department of Civil and Mineral Engineering, University of Toronto, Toronto, ON, Canada

sina.davari@mail.utoronto.ca, ali.tohidifar@mail.utoronto.ca, and civdaeho.kim@utoronto.ca

Abstract

In the rapidly evolving field of artificial intelligence (AI), synthetic data generation has become increasingly crucial, particularly in domains where real-world data is scarce, expensive, or sensitive. In this study, we introduce BCGen, a novel image realism enhancement pipeline that integrates proprietary synthetic construction our data generation and autonomous labeling engine, BlendCon, integrated with Generative AI. Leveraging the graphical capabilities of Blender and the deep learning prowess of the ControlNet model, BCGen represents a novel approach to synthesizing and enhancing construction site imagery. Our methodology narrows the reality gap, delivering images with increased realism and diversity while preserving the full annotations. The paper delineates our approach, methodology, and the broader implications of our findings. Through meticulous hyperparameter tuning and an innovative postprocessing technique, we demonstrate the enhanced realism and diversity of the generated images, pointing towards the vast potential of synthetic data in visual AI applications within construction.

Keywords

Artificial Intelligence in Construction, Synthetic Data, ControlNet, Reality Gap

1 Introduction and Background

1.1 Overcoming Data Acquisition Challenges in Construction with Synthetic Solutions

Since 2017, the construction industry, known for its complex operations and dynamic environments, has increasingly harnessed Deep Learning (DL) to overcome unique challenges in data acquisition and processing across various domains, such as safety, road surveys, bridge inspection, and site operation monitoring [1]. Despite the significant role of DL in revolutionizing traditional practices, such as object detection, instance segmentation, and pose/activity recognition—often surpassing human capabilities— its integration into the construction sector faces hurdles, notably, the scarcity of fully-annotated data [2]. This scarcity stems from the high costs and time requirements for manual collection and labeling, logistical challenges in deploying sensors, and confidentiality issues, highlighting a critical barrier to leveraging data-hungry DL tools within the sector [3].

To address the limitation mentioned, synthetic data generation, bolstered by advances in rendering engines and GPUs, offers a promising solution, especially for tasks difficult to label in real-world settings. The adoption of Blender [4] and video game engines for data generation demonstrates this potential, particularly in facilitating the creation of human-centric images [5-7]. This strategy addresses the disparity in the volume of training images available in computer science versus construction studies, where the former often uses multimillion-image datasets while the latter struggles with far fewer images [8]. These rendering engines enable the generation of synthetic data, addressing data scarcity and streamlining the deployment of DL techniques in construction contexts more effectively.

1.2 An Introduction to our Synthetic Data Generation Engine, BlendCon

In a forthcoming publication, we introduce BlendCon, a computational framework that addresses the need for high-quality, diverse data in the construction industry, particularly for the advancement of Deep Neural Networks (DNNs)-based visual AI. The framework utilizes the graphics engine, Blender, to generate synthetic, fully-labeled data, which is a step forward in overcoming the limitations associated with manual data collection and privacy concerns. BlendCon creates a virtual construction site, producing diverse synthetic images by randomizing key elements of the scene. This approach has been validated in two key areas: trainability and scalability. For instance, Yolov7 models trained with data generated by BlendCon showed comparable or superior performance to models trained with real data.

BlendCon distinguishes itself by integrating dynamic elements, such as mobile equipment and human workers, into synthetic image generation and by offering simulations from multiple perspectives, breaking away from the limitations of single-viewpoint models and enhancing diversity in synthetic data. Furthermore, it is equipped with multimodal label generation capability, producing various types of labels for each image, namely depth maps, semantic masks, and 2D and 3D bounding boxes (BBs) and key points for construction workers.

While BlendCon has proven effective in synthetic data generation and the trainability of DNNs, it still faces a pivotal challenge: the "reality gap." In the subsequent section, we delve deeper into this challenge and elaborate on how this gap, i.e., the divergence in the perceptual and contextual quality between synthetic and real-world imagery, is a crucial factor influencing the effectiveness of DNNs trained on synthetic data in real-world scenarios.

As we delve into the realm of synthetic data generation and the pursuit of enhanced image realism, we arrive at an inquiry that guides our research trajectory: How can we further improve BlendCon? By enhancing image realism through generative AI, we propose the next generation of BlendCon, aimed at offering a scalable solution to the scarcity of labeled construction datasets and facilitating the adoption of deep learning within the construction industry.

1.3 Reality Gap in Synthetic Data Generation

The concept of the reality gap emerges from the inherent differences between the distributions of real and synthetic images. Synthetic images, while beneficial in overcoming data scarcity, often lack nuanced details and contextual variability present in real-world scenarios [9]. This disparity can lead to a decrease in the effectiveness of DNN models when applied to actual construction environments. Effectively bridging this gap is thus a critical objective in enhancing the utility of synthetic data for DNN training, a process referred to as domain adaptation in machine learning [10,11]. Two primary methods have emerged to address the reality gap: enhancing realism and domain randomization.

A) Enhancing Realism: This approach focuses on making synthetic images closely mimic real-world scenarios. Studies suggest that the heightened realism in synthetic data significantly improves the performance of DNNs, allowing them to better adapt to real-world applications [12]. By refining textures and lighting conditions, and incorporating real-world irregularities, this method aims to reduce the perceptual differences between synthetic and real imagery while preserving full annotations [13,14]. This approach is not without its limitations; significantly, the process of enhancing realism in synthetic images can be both time-consuming and costly, and the subjective nature of 'realism'—what is deemed realistic—remains ambiguously defined.

B) Domain Randomization: Contrasting with the pursuit of realism, domain randomization adopts an unconventional tactic -intentionally making synthetic images more abstract or less realistic. This method involves introducing a high degree of variability in the synthetic images, which paradoxically can lead to the development of more robust DNN models. The premise is that by exposing the DNN to a wide range of variations, the model learns to focus on the most critical features, becoming more adaptable to real-world variability. A notable study in this area demonstrated the effectiveness of this approach, where severely randomized images contributed to the training of robust models capable of bridging the reality gap [15]. This approach, however, requires significant computational resources and carries the risk of overgeneralizing DNN models, potentially leading to models that, while robust in handling diverse scenarios, may not perform optimally in specialized tasks, such as construction applications.

The choice between these approaches depends on the specific requirements of the application and the nature of the tasks the DNN models are expected to perform. This paper marks the beginning of our exploration into the first approach, laying the foundation for a future study that is planned to investigate both strategies with the objective of evaluating their impact and effectiveness in optimizing the utility of synthetic data for training DNN models in construction applications.

1.4 ControlNet for Enhanced Image Realism

In this study, we introduce BCGen, a pipeline for integrating the power of generative AI, specifically the Stable Diffusion (SD) model, ControlNet [16], into our automated construction image synthesis and labeling framework, BlendCon. This marks the first application of ControlNet in the construction domain for image-toimage translation, generating more realistic images from synthetic ones while maintaining full annotations.

ControlNet is designed to integrate spatial conditioning controls into large, pre-trained text-toimage diffusion models. By leveraging robust and deep encoding layers of these models and applying zero convolutions, it finetunes the imagery while avoiding the introduction of deleterious noise. Its ability to handle various inputs, such as edges and human poses, and its robust performance across datasets of differing scales, make ControlNet an essential tool for generating realistic images from synthetic data, particularly in applications where maintaining accurate annotations is as crucial as image quality itself [16].

ControlNet distinguishes itself by its ability to fine-

tune the realism of generated images while preserving annotations [16]. This capability is crucial in applications like construction site imaging, where maintaining the accuracy of annotations is as important as the visual realism of the images. By integrating ControlNet, we hypothesize that synthetic images could become more adaptable for real-world applications, thereby potentially bridging the reality gap. This assertion, however, remains to be validated in future phases of our research.

1.5 Related Work on Enhancing the Realism of Synthetic Images

Deep learning, particularly Generative Adversarial Networks (GANs) [17] and Diffusion Models [18] has been pivotal in enhancing image realism. GANs, by their adversarial nature, refine images to closely resemble real photographs. Diffusion Models, demonstrate remarkable capabilities in text-to-image generation and synthesizing photorealistic images, offering a significant leap in image quality and diversity [19]. GANs, while effective, may struggle with ensuring stability during training, producing artifacts [20]. Diffusion Models, however, characterized by their gradual process of image formation, offer higher stability and image quality, albeit at the cost of increased computational complexity.

A recent study leveraged diffusion models, including ControlNet, to enhance the FFHQ-Aging dataset [21], producing synthetic images that exhibit a diverse array of facial expressions, ethnicities, and lighting conditions, thereby advancing the realism and quality of synthetic imagery for facial image augmentation [22]. Furthermore, in the medical domain, diffusion models have been extensively utilized for various applications, including realistic endoscopic image generation [23] and synthesizing MRI sequences and thoracic X-ray images [24].

Several of the mentioned studies have implemented established photorealism metrics to gauge the quality of

the enhanced images. Metrics such as the Inception Score [25], Fréchet Inception Distance [26], Kernel Inception instance [27], Structural Similarity Index Measure [28], Learned Perceptual Image Patch Similarity [29], and Contrastive Language-Image Pre-training (CLIP) [30]-based metrics [31] have been pivotal in assessing the realism of synthetic imagery. However, despite such evaluations, there remains an underexplored area in the existing literature: assessing the cost-effectiveness of employing these advanced generative models. To the best of the authors' knowledge, a systematic examination of the cost-benefit analysis of utilizing such sophisticated techniques for enhancing synthetic image realism has not yet been documented.

2 Method

2.1 Architecture of BCGen: BlendCon with Integrated Generative AI

Our proposed framework, BCGen, leverages a threepart pipeline to enhance image realism. Figure 1 demonstrates the BCGen pipeline, encapsulating the endto-end process from image synthesis with BlendCon, through realism enhancement via the ControlNet pipeline, to the final avatar cut and paste for anatomical accuracy, ensuring the retention of high-quality annotations.

To further elucidate, initially, BlendCon synthesizes RGB construction site images, along with their corresponding depth maps, semantic masks, and precise annotations of 2D and 3D bounding boxes and key points for construction workers, using inputs such as horizon, processed scenes, animated avatars, and lighting and camera configuration. The outputs are fed into the ControlNet pipeline, where the RGB images, depth maps, and semantic masks—alongside a text prompt—are processed by ControlNet to create more realistic images.

Our text prompt is "a high-quality, high-resolution



Figure 1. The proposed image realism enhancement pipeline

image of a construction site." We also pass "blurry, blurred, bad anatomy, low quality" as the negative prompt.

This integration of ControlNet paves the way for enhanced realism while meticulously preserving all annotations, including worker key points and 2D and 3D bounding boxes. Hyperparameter tuning is employed to refine this process and optimize the outputs, a topic we will explore in detail in the subsequent section. In the final stage, to ensure anatomical accuracy and maintain the integrity of our key point annotations, we initiated a process termed 'Avatar cut and paste,' which involves extracting the worker avatars from the initial BlendCongenerated images and superimposing them onto the images enhanced by ControlNet.

2.2 ControlNet Hyperparameter Tuning

Given ControlNet's extensive range of adjustable parameters, such as the degree of reliance on the input images, conditions, and text prompts, we embarked on a rigorous hyperparameter tuning exercise employing a grid search methodology. This allowed us to identify the most ideal settings for our particular use case. Moreover, it was during this tuning process that we encountered scenarios where, despite explicitly excluding poor anatomy and low quality in our negative text prompt, the output sometimes exhibited compromised structural integrity, especially in the anatomy of the construction workers, and as outlined previously, to counteract this, we resorted to avatar cut and paste.

We investigated six ControlNet hyperparameters, namely output image size, conditioning scales—which determine the weight of our conditions, i.e., depth maps and semantic masks, classifier-free guidance scale (CFG) —which dictates the influence of the text prompt on image generation, number of inference or denoising steps for the diffusion model, input image strength—which determines the input image weight, and the choice of diffusion model noise scheduler, across three different random seed initializations. The results were scrutinized, and the most effective hyperparameter combinations were selected through visual comparisons.

The initial phase of our study involved a qualitative assessment of the synthetic images generated by our pipeline, relying on visual observation to evaluate the quality. We considered any image unrealistic, blurry, distorted, or exhibiting anatomical inaccuracies and abstract backgrounds as unsatisfactory. Figure 2 examples where the showcases interplay of hyperparameters resulted in suboptimal results, such as compromised human anatomies, abstract backgrounds, and blurred images, underscoring the inherent challenges and complexities of synthetic data generation.

3 Results, Discussions, and Limitations

3.1 Hyperparameter Tuning Results

Our analysis underscored the significant impact of image size on output realism, diversity, and quality, investigating two sizes of 512 by 512 and 1280 by 1280, which revealed that larger images notably enhanced all aspects. Our investigation into the number of inference steps, specifically examining 40, 80, and 150 steps, revealed its critical significance: fewer than 50 steps often resulted in blurry and structurally unsound images, while exceeding 100 steps did not notably improve quality but extended runtime unnecessarily.

The investigated eight schedulers are linear multistep (LMSDiscrete), denoising diffusion implicit (DDIM), denoising diffusion probabilistic (DDPM), multistep diffusion probabilistic (DPMSolverMultistep), Euler (EulerDiscrete), pseudo numerical (PNDM), Euler with ancestral sampling (EulerAncestralDiscrete), and unified predictor-corrector scheduler (UniPCMultistep) [32]. The unified predictor-corrector noise scheduler was identified as the most effective for our task.

The interplay of CFG, Strength, and Conditioning Scalehyperparameters, and their impact on image realism was further explored. We charted the instances where these parameters harmonized to produce satisfactory outputs, yielding realistic images with minimum blurs, and anatomical inaccuracies, as demonstrated in Figure 3. The graph illustrates the frequency of the satisfactory outcomes across various configurations of CFG, strength, and conditioning scales, with marker size indicating the occurrence count.

We experimented with conditioning scales of [0.3, 0.8], [0.5, 0.5], [0.8, 0.3], and [0.8, 0.8], input image strengths of 0.5, 0.7, and 0.9, and CFGs of 5, 7.5, 10, and 12.5. Ultimately, we selected a conditioning scale of 0.8 for both depth maps and semantic masks, a strength of 90 percent, and a CFG of 12.5.

3.2 BCGen Results and Discussions

The application of ControlNet to the original synthetic images from BlendCon has resulted in enhancements in realism and diversity, as evidenced in Figure 4. The before-and-after comparisons illustrate the ControlNet-induced changes, with noticeable improvements in texture detail, lighting fidelity, and the incorporation of realistic environmental effects. These images not only demonstrate an enriched visual diversity but also indicate a substantial narrowing of the reality gap, affirming the potential of our approach in creating realistic images for use in AI training and other construction industry applications.

While generative DL models today can produce hyperrealistic images, our contribution lies in the unique

capability of our pipeline to maintain original annotations.

The importance of annotated images cannot be overstated, as manual labeling of 2D, 3D, key points, semantic masks, and depth maps involves considerable costs, time, and potential for errors. Our effort ensures that the synthetic images generated are not only visually compelling but also maintain full annotations, making them immediately useful for DNN training and other applications within the construction industry.

3.3 Limitations of The Study

This research is a stepping stone, highlighting the necessity for verification and validation methods tailored to the unique requirements of enhanced synthetic image evaluation. The study presents several limitations that inform its theoretical implications:

A) Generalization of Results: The findings, although promising, are not yet generalizable across all potential input scenarios, indicating that further research is required to broaden the applicability of the results. For instance, the suboptimal results in indoor environments with a high degree of clutter, as seen in the last row of Figure 4, suggest that the model may struggle with overly complex indoor construction scenes. This issue may arise due to the lack of specific information in the textual prompt and a heavy reliance on it. This aspect will be further investigated in future studies.

B) Evaluation of Results: Established photorealism metrics could play a pivotal role in evaluating the quality of enhanced images, enabling the creation of businessoriented key performance indicators that measure the efficacy and cost-efficiency of synthetic data generation and enhancement. However, the investigation of these established metrics for evaluating the efficacy and costefficiency of the synthetic data enhancement process was not carried out in the current stage of our study.

As previously discussed in our methodology, the initial phase of our study was dedicated to a qualitative assessment of the synthetic images generated by our pipeline, where we relied on visual observation to determine the quality of the output. This subjective method highlights a limitation in our evaluation process, underscoring the need for developing objective criteria and metrics to assess realism and AI training applicability. C) Computational Resources: The computational demand varies with the conditions set and image size, with the current setup requiring about 40 seconds per 1280 by 1280 image on two NVIDIA RTX 3090 GPUs, which could be a limiting factor for scalability.

D) Variability in Text Prompts: Our exploration of variability in text prompts was limited to a few variations of main and negative prompts, restricting our understanding of their precise impact on the results. Further detailed prompt engineering is designated for future research.



Figure 3. Visualization of ControlNet hyperparameter tuning through grid search



Figure 2. Examples of suboptimal image generations: (a): BlenCon's synthetic images, (b): a satisfactory result, (c) compromised human anatomy, (d): a generated image suffering from both blurriness and compromised human anatomy, (e): occurrence of both anatomical inaccuracies and an abstract background.



Figure 4. Contrast between the original synthetic construction site image from BlendCon (leftmost) and its enhanced iterations by ControlNet, showcasing diversity improvement and realism enhancements.

4 Looking Ahead: Future Directions

Building upon the work presented in the previous section, the development of business-oriented performance indicators metrics for measuring synthetic data quality and degree of enhancement is essential. These metrics should not only assess the visual fidelity of the images but also quantify the cost-benefit of enhancing realism within a corporate context. Upcoming studies will focus on validating DNN trainability, performance, and affordability with enhanced images, crucial for practical AI applications.

Further exploration into additional input modalities for ControlNet, such as human poses or edge maps as conditions or main inputs, is planned. Leveraging human key point detection models, such as OpenPose [33], could allow for a more nuanced representation of worker anatomies. This aligns with methods like Control-GPT [19], which combines programmatic sketches with textto-image generation, a technique that could be adapted to enrich our dataset diversity and control [19]. Additionally, we plan to explore the implementation of alternative generative models besides ControlNet to assess their performance and facilitate comparative analysis.

By experimenting with multiple prompts and conditions, we aim to refine our generative model's output further, ensuring that the synthetic images not only serve the construction industry's current needs but also pave the way for emergent AI-driven solutions.

In future research, we aim to employ photorealism metrics for developing business-oriented key performance indicators, thereby measuring the efficacy and cost-efficiency of synthetic data generation and enhancement. This initiative seeks to standardize the validation of synthetic image quality within the AI field.

Additionally, we will evaluate the performance of DNNs trained on limitedly available real-life construction datasets, synthetic data generated by stateof-the-art models, such as Midjourney [35], and domainrandomized synthetic imagery. This evaluation is crucial to verify our method's cost-effectiveness and practicality in real-world applications, shedding light on the economic viability of leveraging such advanced techniques in the construction industry.

5 Conclusion

In this study, focusing on the investigation of the reality gap in synthetic data generation, we introduced a pipeline incorporating the stable diffusion-based model, ControlNet, within our synthetic construction data generation and labeling engine, BlendCon. This pipeline paves the way for generating a diverse range of enhanced synthetic images, while preserving their full annotations, i.e., depth maps, semantic masks, and 2D and 3D bounding boxes and key points for construction workers. Our investigation reveals that ControlNet's hyperparameters critically influence the enhancement of realism, prompting us to conduct a thorough search across over 2,300 hyperparameter combinations, evaluating them through visual observation. This exhaustive process underscored the significance of quantifying realism via photorealism metrics and highlighted the need to balance the costs associated with synthetic image generation and enhancement. In conclusion, our study demonstrates that cutting-edge, controllable diffusion-based generative models hold significant potential for the construction industry, enabling the creation of realistic, fully annotated synthetic imagery by narrowing the reality gap.

References

- [1] Xu Y., Zhou Y., Sekula P., and Ding L. Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, 6:100045, 2021.
- [2] Gupta S., Agrawal A., Gopalakrishnan K., and Narayanan P. Deep Learning with Limited Numerical Precision. In *Proceedings of the International Conference on Machine Learning*, pages 446–454, 2015.
- [3] Liu J., Luo H., and Liu H. Deep learning-based data analytics for safety in construction. *Automation in Construction*, 140:104302, 2022.
- [4] Blender Foundation. Blender. On-line: https://www.blender.org, Accessed: 22/02/2024.
- [5] Tang H. and Jia K. A New Benchmark: On the Utility of Synthetic Data with Blender for Bare Supervised Learning and Downstream Domain Adaptation. In *Proceedings of the Computer Vision* and Pattern Recognition, pages 15954–15964, 2023.
- [6] Neuhausen M., Herbers P., and König M. Using Synthetic Data to Improve and Evaluate the Tracking Performance of Construction Workers on Site. *Applied Sciences*, 10(14):4948, 2020.
- [7] Fabbri M., Brasó G., Maugeri G., Cetintas O., Gasparini R., Ošep A., Calderara S., Leal-Taixé L., and Cucchiara R. MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking?. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), pages 10849–10859, 2021.
- [8] Krizhevsky A., Sutskever I., and Hinton G. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [9] Gamal O., Rameshbabu K., Imran M., and Roth H. Bridging the Reality Gap: Investigation of Deep Convolution Neural Networks Ability to Learn

from a Combination of Real and Synthetic Data. In Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2020), pages 446–454, 2020.

- [10] Kouw W. and Loog M. An introduction to domain adaptation and transfer learning. *Machine Learning*, 2019.
- [11] Lemberger P. and Panico I. A primer on domain adaptation. *Machine Learning*, 2020.
- [12] Wood E., Baltrušaitis T., Hewitt C., Dziadzio S., Cashman T. J., and Shotton J. Fake It till You Make It: Face Analysis in the Wild Using Synthetic Data Alone. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
- [13] Shen T., Zhao G., and You S. A Study on Improving Realism of Synthetic Data for Machine Learning. In *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [14] Atapattu C. and Rekabdar B. Improving the realism of synthetic images through a combination of adversarial and perceptual losses. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, pages 14– 19, 2019.
- [15] Tremblay J., Prakash A., Acuna D., Brophy M., Jampani V., Anil C., To T., Cameracci E., Boochoon S., and Birchfield S. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. In *Proceedings of the Computer Vision and Pattern Recognition* (CVPR), 2018.
- [16] Zhang L., Rao A., and Agrawala M. Adding Conditional Control to Text-to-Image Diffusion Models. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2023.
- [17] Ozair S., Courville A. C., Bengio Y. Generative Adversarial Networks. Machine Learning, 2014.
- [18] Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. *Machine Learning*, 2020.
- [19] Zhang T., Zhang Y., Vineet V., Joshi N., Wang X. Controllable Text-to-Image Generation with GPT-4. In *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [20] Saxena D. and Cao J. Generative Adversarial Networks (GANs Survey): Challenges, Solutions, and Future Directions. *Machine Learning*, 2023.
- [21] Or-El R., Sengupta S., Fried O., Shechtman E., Kemelmacher-Shlizerman I. Lifespan Age Transformation Synthesis. In *Proceedings of the Computer Vision and Pattern Recognition*, 2020.
- [22] Zhang J., Hsiung L., and Hsu G. Unleashing Stable Diffusion for Multi-Modal Facial Image Augmentation. In Proceedings of the International Conference on Advanced Robotics and Intelligent Systems (ARIS), pages 1–5, Taipei, Taiwan, 2023.

- [23] Kaleta J., Dall'Alba D., Płotka S., and Korzeniowski P. Minimal data requirement for realistic endoscopic image generation with Stable Diffusion. *International Journal of Computer Assisted Radiology and Surgery*, 19:531–539, 2023.
- [24] Kazerouni A., Khodapanah Aghdam E., Heidari M., Azad R., Fayyaz M., Hacihaliloglu I., and Merhof D. Diffusion Models for Medical Image Analysis: A Comprehensive Survey. *Image and Video Processing*, 88:102846, 2023.
- [25] Salimans T., Goodfellow I. J., Zaremba W., Cheung V., Radford A., and Chen X. Improved techniques for training GANs. *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.
- [26] Heusel M., Ramsauer H., Unterthiner T., Nessler B., and Hochreiter S. GANs trained by a two timescale update rule converge to a local Nash equilibrium. In *Proceedings of the 31st Conference* on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2018.
- [27] Binkowski M., Sutherland D., Arbel M., and Gretton A. Demystifying MMD GANs. In Proceedings of the International Conference on Learning Representations (ICLR), 2018.
- [28] Wang Z., Bovik A., Sheikh H., and Simoncelli E. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [29] Zhang R., Isola P., Efros A., Shechtman E., and Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [30] Radford A., Kim J., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., and Sutskever I. Learning Transferable Visual Models from Natural Language Supervision. In *Proceedings of the Computer Vision* and Pattern Recognition (CVPR), 2021.
- [31] Hessel J., Holtzman A., Forbes M., Le Bras R., and Choi Y. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. In *Proceedings of the Computer Vision and Pattern Recognition*, 2022.
- [32] Hugging Face. Schedulers. On-line: https://huggingface.co/docs/diffusers/en/api/sched ulers/overview, Accessed: 22/02/2024.
- [33] Cao Z., Hidalgo Martinez G., Simon T., Wei S., and Sheikh Y. A. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2019.
- [34] Midjourney, Inc. Midjourney. On-line: https://docs.midjourney.com/, Accessed: 24/02/2024.

Integration of Construction Progress Monitoring Results using AI Image Recognition from Multiple Cameras onto A **BIM**

Chang-Cheng Hsieh¹, Hung-Ming Chen², Wan-Yu Chen³ and Ting-Yu Wu³

¹Ph.D. Candidate, Department of Civil & Construction Engineering, National Taiwan University of Science and Technology

²Professor, Department of Civil & Construction Engineering, National Taiwan University of Science and Technology

³Graduate Research Assistant, Department of Civil & Construction Engineering, National Taiwan University of Science and Technology

E-mail: D10605009@mail.ntust.edu.tw, hungming@mail.ntust.edu.tw, M10905508@mail.ntust.edu.tw, M11005501@mail.ntust.edu.tw

Abstract

This study merges artificial intelligence (AI) image recognition technology with Building Information Modeling (BIM), to develop a prototype system for the automation and visualization of construction site progress control. Our focus is the strategic deployment of multiple construction site surveillance cameras using a BIM model to encompass the entirety of the construction site. Following the capture of camera images, the application of object detection techniques within AI image recognition locates all constructed objects in the actively images, subsequently identifying the construction phases to which these objects belong. By integrating multiple camera perspectives from the site into a BIM model, the results of AI detection are automatically inputted into the corresponding components of the model. Finally, real-time on-site progress information obtained from the BIM model is compared with the progress schedule, and the comparative results are visually presented on the BIM model components in distinct colors. Through this visual approach, managerial personnel can intuitively and instantly control the construction progress.

Keywords -

Deep Learning, Image Recognition, Construction **Progress Management, Build Information Modeling,** Automation, Visualization

1 Introduction

In the project management of the traditional construction industry, monitoring the progress of the project has always been an important task [1]. In addition to having a profound awareness of the construction environment, the on-site engineer must also understand drawings. Progress data must be collected on-site, and real-time progress information must be presented in the form of text and data for project managers to refer to [2]. For less experienced site engineers, there may be cognitive standards for different construction phases. Moreover, it is not easy for people who are not familiar with engineering to convert two-dimensional drawings into three-dimensional scenes, which will cause differences in information transmission.

In order to fully support the life cycle of construction projects and interpret engineering information models through computer programs, BIM came into being. The application of BIM covers all stages of the building life cycle, including planning, design, procurement, construction, operation and maintenance, etc. [3]. In the construction stage, BIM is commonly integrated with project timelines to create a 4D model for construction simulation. Although the dynamic model exists, on-site engineers still need to update the schedule data of the model components by comparing animations with actual on-site construction conditions to achieve progress control.

Using a huge management manpower to collect and organize complex data, this traditional management method no longer seems efficient enough. Many studies have surveyed how to improve complex data processing procedures that rely on manpower [4]. In order to allow managers to perform progress management tasks more quickly and ensure that progress evaluation standards are unified, thereby making the overall management process smoother.

As technology advances, more research in artificial intelligence (AI) has made significant strides in recent years. Machine learning is a method of learning from past data and experiences to identify operational rules. Deep learning, a subset of machine learning that applies multilayered neural networks to simulate human neuron functions, has achieved notable breakthroughs in the fields of images, videos, and speech [5-6].

The rapid development of AI image recognition technology has led to its expanding applications in the field of engineering, particularly in construction industries for the management of construction machinery, personnel, and materials. This project aims to apply AI image recognition to construction progress control, focusing on the use of object detection technology in AI image recognition to achieve automated recognition of various work progress in construction sites.

However, constrained by factors such as the site's scope, layout, component obstructions, and camera wide angle, AI image detection can only address detection within a single image, making it challenging to cover the overall area. Even with multiple cameras set up to encompass the overall area, integrating the detected results from these cameras and automating the comparison with the construction planning schedule still requires the development of effective solutions.

Therefore, this project further integrates AI image detection with BIM technology and develops a prototype system. This system, utilizing multiple cameras, applies an AI image recognition model to recognize the construction status of work items within the images. Subsequently, through the BIM model, it achieves image alignment and identifies the corresponding components, inputting the construction progress of the respective components into the BIM model. Finally, in the application program, a visual representation using a color concept is employed to present different progress states, enabling project managers to control the construction progress in real-time.

This study will establish an image-based construction progress detection model applicable to the construction phase, providing project managers with a method to assess construction progress states through image object detection. Adopt the BIM model for the integration of component construction states, automatically inputs the corresponding component progress states into the BIM model. This not only avoids variations in construction progress judgment among different personnel but also enhances the utilization of human resources, thereby improving the efficiency and convenience of on-site progress management.

Based on the background, this study plans to propose a progress object detection module used in the construction phase. By collecting images of the progress of each stage of construction, it can be used as training data. By combining common model architecture, an object detection model suitable for identifying construction progress is selected. Apply the transfer learning method and adjust model parameters to improve the recognition rate, and then find the most suitable model for construction progress detection. Finally, integrating with the BIM model at the application end will achieve component positioning, integrate multiple images of the same component for detection input, and address differences in progress judgment due to manual input and optimize human resource utilization.

The study will be divided into four phases. First is the collection of construction progress image data, followed by the training and testing of the object detection model, optimization and validation of the object detection model, and finally, the integration of BIM for automated and visualized construction progress monitoring.

In terms of data collection, cameras will be installed at the construction site to collect image data, supplemented by collecting relevant construction progress photos from online sources.

The establishment of the object detection model adopted transfer learning. This pre-trained model will undergo training and testing to seek an optimal object detection model. The parameters of the model will be systematically optimized and adjusted to enhance the accuracy of the detection model.

During the model validation and testing phase, images or videos directly obtained from construction sites will be utilized. Image recognition will be applied to identify the construction activities in the data source.

Adopting the BIM model as the data integration hub, simulating camera deployment conditions, utilizing image recognition technology to detect the construction status of components within the coverage area, and inputting it into the model components. The integrated construction progress of components can be automatically input into the scheduling system.

The prototype system will compare with the original schedule, and using different progress visualization methods, categorize the results by color. Through visualization, management personnel can grasp the progress concretely and intuitively in real time.



Figure 1.The process diagram of image-based construction progress detection model



Figure 2. Demo diagram of integrates AI image detection with BIM technology

2 **Related Research**

The research topics related to this study include realtime progress monitoring and AI image recognition, especially for construction progress monitoring. The literature review for these topics are as follows.

2.1 **Real-time Progress Monitoring**

Past research has utilized various Field Data Capturing Technologies (FDCT) [7] to collect progress data from construction sites, integrating them with Building Information Modeling (BIM) to enhance construction progress monitoring. These include image recognition of construction elements, alignment with 4D models for comparison, RFID for tracking personnel and materials, UWB positioning systems for progress tracking, and laser scanners for construction environment scanning to create 4D As-Built BIM models, compared with 4D Designed BIM models [8-11].

These technologies still have limitations for improvement in progress management applications, such as the need for confirmation of work status for UWB and RFID, like verifying completion and installation. Point cloud analysis requires more time for processing to the object level for project progress.

2.2 AI Image Recognition for Construction **Progress Monitoring**

In recent years, the application of image recognition in construction through deep learning has been steadily increasing, encompassing the identification of elements like construction workers, materials, and machinery [12-13]. The studies of applying AI image recognition, for construction progress monitoring are relatively few.

Utilizing image recognition for automated progress monitoring in construction projects involves extracting features through deep learning to detect and identify construction status. Zheng et al. [14] utilized R-CNN model to automatically detect modules and identify their status, such as Hooking, Lifting, and Final Positioning.

With the affordability of cameras, gathering rich information from construction sites to achieve automated visual monitoring of construction sites becomes feasible [15]. Martinez et al. [16] utilized low-resolution CCTV images and combined deep learning methods (R-CNN) with Finite State Machines (VFSM) to identify labor and key equipment in floor manufacturing. The study presents the calculation of task duration and working hours, providing managers with clear and real-time insights into workstation progress.

The study adopts deep learning-based recognition technology to propose a framework suitable for on-site construction progress management. Data collection is conducted using fixed on-site cameras to improve the costly investment in equipment and manpower. The use of YOLO and transfer learning enhances efficient detection, particularly in scenarios with limited construction photos. Furthermore, this study aims to define major operational steps in structural engineering, such as rebar binding, formwork assembly, and concrete pouring, for effective control and management of construction progress. Additionally, suggestions for integrating progress management systems will be provided, improving existing research.

3 Methodology

To achieve these objectives, we designed a prototype system, which includes the required model architecture, on-site install flow, and functional displays. Building the progress object detection module in the construction phase, configuring site cameras, data integration and presentation will be detailed in the following sections.

3.1 Building Progress **Object** Detection **Module in the Construction Phase**

The study proposes the utilization of a progress object detection module for the construction phase. By collecting images of construction progress at various phases and applying transfer learning, the model parameters will be adjusted to enhance recognition. A comparison of various models will be conducted to identify the most suitable one for construction progress detection.

The implementation is divided into three stages: image classification and data collection, selection of the object detection model, and training/testing of the object detection model.

Image Classification and Data Collection 3.1.1

To collect datasets for training the model, this study focuses on collecting relevant images from architectural projects, specifically targeting the structural construction phase. To mitigate redundancy in the training data that may result in high feature similarity and potentially impact the model's training outcomes, photos are collected through three distinct approaches: daily

progress photos, time-lapse cameras on site, and online sources.

In this study, focusing on architectural structures, the collected progress images are mainly categorized into rebar tying, formwork assembly, and concrete pouring. Recognizing variations in the construction sequence between columns and walls, which proved to be confusing in initial tests, the construction phases are dissected into distinct phases: rebar tying of columns, pre-rebar tying of walls, completion of rebar tying of walls, formwork assembly of walls, formwork assembly of columns and concrete pouring.

3.1.2 Selection of Object Detection Model

The object detection model is composed of three parts: input, convolutional neural network layers, and detection layers. The convolutional neural network layers serve as the main network backbone, responsible for extracting image features and producing a feature map by merging the extracted features through pooling layers. The detection layers are responsible for the final prediction of object categories and generating candidate boxes.

To explore and seek a suitable combination of detection models, this study chose one-stage and twostage object detection models, and selected models that have performed well in most related studies, YOLOv5 represents the one-stage model, while Faster R-CNN represents the two-stage model. Then, five groups of convolutional neural network layers and detection layers are constructed.

Table 1 Comparison of object detection model combinations

	Convolutional Neural Network Layers	Detection Layers	mAP(%)	FPS
1	ResNet50	YOLOv5	27.6	11.7
2	ResNet50	Faster R-CNN	31.8	4.3
3	DenseNet121	YOLOv5	31.1	10.3
4	DenseNet121	Faster R-CNN	38.3	3.9
5	CSPDarknet53	YOLOv5	51.1	17.6

The training conditions are set to 300 epochs, with 16 samples per batch, and image pixels of 640x640. The pixel size of the images utilizes the maximum value allowed by the computer hardware to avoid suboptimal training results. Regarding optimization parameters, the original default values of the model will be used, and to examine the fundamental performance of the model, no transfer learning pre-trained weights will be adopted. The model will be trained from scratch to obtain the detection model's training results.

Among the selected detection model combinations, the model with the highest individual accuracy will be compared. Considering conditions such as detection box overlap and detection speed, YOLOv5 - CSPDarknet53 will be adopted as the detection model for this study.



Figure 3. Compare prediction results and candidate boxes of detection models

3.1.3 Training and Testing of the Object Detection Model

The quantity of the dataset can impact the accuracy. During the initial stages, obtaining a clear and adequate number of construction progress photos posed challenges. This is mainly due to construction activities developing in different phases, and the collection of construction photos progresses gradually with the advancement of the project, making it difficult to rapidly and substantially increase the data volume.

To address this, the study utilizes the mosaic feature proposed in YOLOv4 [17] as a form of data augmentation to generate additional photo data, aiming to increase the number of the dataset. Additionally, the study adjusts hyperparameters during the training process to enhance the detection accuracy of the model. In this phase, transfer learning is applied, utilizing the dataset named MS COCO (Microsoft Common Objects in Context) [18-19] to train the YOLOv5 model. This source was designed to detect and segment common objects like humans, cars, and buses in daily life. The dataset consists of 328K images and 80 object categories.

Through transfer learning and hyperparameters optimization, the optimized model improves 20.4% accuracy over the original model, with no significant degradation in FPS (frames per second, FPS) performance. On average, the optimized model shows better loss values and performance compared to the initial architecture.

Table 2 Model optimization performance

YOLOv5	Accuracy (%)	Recall rate (%)	mAP (%)	FPS
Initial model	63.1	74.67	51.1	17.6
Optimization model	83.5	71.74	57.9	17.4

3.2 Site Cameras Deployment and Data Integration and Presentation

Upon the completion of the detection model, many surveillance cameras are deployed on the construction

site to record real-time video during the ongoing construction. This process is aimed at further recognizing the construction progress of each component. The positioning and alignment of cameras are constrained by some factors such as the site location and installation conditions. Therefore, systematic positioning and alignment are adopted to facilitate recognition and utilization of image data. In addition, image capture is performed from multiple angles, many components will be recorded repeatedly. This study provides a procedure to integrate images and presents a visual approach that enables management personnel to intuitively and promptly comprehend the construction progress.

3.2.1 Site Cameras Deployment

The input source for this study is captured from the surveillance camera at the site. To ensure alignment between the input and the system's camera, two sets of camera initialization, positioning, and alignment configurations must be executed. The methods vary depending on whether the on-site surveillance cameras are already installed or are to be set up based on parameters. In this study, the coordinate system in Unity is utilized as a reference for calculating the relative spatial position after importing the model. Different processes and methods for installation are planned for two scenarios, as illustrated in the figures below.



Figure 4. Case of on-site surveillance cameras are already installed



Figure 5. Case of on-site surveillance cameras are set up based on system parameters

3.2.2 Mechanism of Construction Image Detection and Input

The return of images captured by the surveillance cameras at the site to the system's main server, the image detection model is utilized for recognition. The recognition results are outputted as a txt format, including information such as the coordinates of the detection box's center point, length, height, recognized progress phase, and mAP value.

The consistency perspective has been established between Unity and site cameras, the photos captured by the cameras can be considered aligned with the perspective of the model. Therefore, use the Unity Physic. Raycast function for component selection to find the component corresponding to the recognition result.



Figure 6. Demo of Physic. Raycast function for component selection

Because the system integrates multiple cameras, it may recognize different phases for the same component. This study proposes three solutions, allowing users to choose the most suitable method for their engineering project. These are prioritized based on accuracy (mAP), construction phase, and detection frequency. This approach enables the provision of distinct update principles based on the condition of the engineering project.

3.2.3 Visual Presentation and the User Interface

Through the automated progress update function, it can reduce the human resources required for project management. On the other hand, visual presentation allows operators to understand the construction progress through screen presentation, thereby effectively improving work efficiency. This study proposes two visual presentation methods, explained as follows:

Users can select components through the Physic. Raycast function in the operation interface. This will query and display different colors corresponding to the construction phases detected by image recognition, simulating the on-site construction situation.



Figure 7. Query and display different colors corresponding to the construction phase

Another visual presentation method focuses on the overall progress control. Users can update component progress through the progress update interface. When this function is enabled, the system will compare and analyze the planned and actual completion dates for each component, presenting the results through the 3D model display area.



Figure 8 Process of schedule update and model presentation

4 Application Scenarios

Based on the prototype system developed, this section will conduct tests through various scenarios to showcase the functionalities developed to achieve the research objectives. The following sections will categorize the system's operations and provide detailed demonstrations and explanations for each application mode.

4.1 The Camera Deployment of the Prototype System

In the initial phase, this study utilized a simulated construction site in an interior parking floor of the building to validate whether the camera deploy functionality of the prototype system aligns with the requirements of typical construction scenarios. The redhighlighted area in the layout represents the scope of the research tests. Four surveillance cameras with corresponding field-of-view lenses were strategically installed both on-site and within the system, facilitating subsequent progress detection through the detection function and verifying its capability to cover the entire testing area.



Figure 9. The layout of the research tests scope and camera deploys

The test has already pre-installed multiple perspective cameras in the system, and the desired camera position data will be obtained within the prototype system. In the system interface's 3D model display area, the selected perspective camera position will be shown. Users can choose a reference component closest to this camera by clicking with the mouse and calculating the X and Y-axis displacement. Finally, at the site, using this reference component as the origin, surveillance cameras will be set up using relative displacement.



Figure 10. Demonstration of Camera set up on site and in model

4.2 Recognize Various Construction Phases

This study utilized an existing building to simulate the site environment and whether the detection model can recognize various construction phases. The study modified the final images to depict scenarios such as rebar tying of columns, formwork assembly of walls, and concrete pouring.

Through the execution of image recognition functions, the prototype system of this study successfully recognized the construction phases of components, except for components that have been obstructed.



Figure 11. Simulate the completion phase of rebar tying and formwork assembly



Figure 12. Simulate the completion phase of concrete pouring

4.3 Multi-angle View Detection and Results Integration

After the AI conducts object detection and outputs the results, the prototype system automatically reads the information. Subsequently, from four different perspectives, the BIM model is interactively selected using the Unity Physic. Raycast function. Based on these selected components, progress data is updated. The BIM model, representing the defined testing area, adjusts its color presentation according to the different construction phases. The extraction of the updated results from the selected components confirms the effective coverage of the entire construction area using multiple cameras.



Figure 13. The completion phase visualization of rebar tying and formwork assembly



Figure 14. The completion phase visualization of concrete pouring

4.4 Use Surveillance Cameras for Recognize Testing during the Construction Stage

After validating the relevant functions indoors, the system proceeded to on-site testing in a construction setting. In this case, focusing on a construction project, CCTV images from the construction site were used as the image source, and the BIM model for the project was constructed for system testing. Initially, the prototype system deployed surveillance cameras. After measuring and calculating the data for the coordinates of the system's perspective camera at the site, the system completed the installation of the perspective camera. The comparison between the system's perspective view and the construction site image after deployment is shown in the following figure.



Figure 15. Comparison of system perspective and construction site images

Subsequently, the construction image recognition function was executed. The prototype system, after detection, retrieved the results and utilized the functionality to select the components to be updated. The comparison between the recognition results and the system interface is illustrated in the following figure. This presentation of results demonstrates the feasibility of various functions of the prototype system in practical cases.



Figure 16. Comparison of recognized results and visualization component screen

5 Conclusions

This study proposes a prototype system that integrates AI image recognition, BIM, visualization technology, and on-site construction image monitoring. It not only provides real-time monitoring of construction site progress but also establishes an automated and visual management system. The AI detection and recognition integration mode, based on BIM, realizes an automated and visual construction progress management platform.

The system automates the integration of detection results from multiple surveillance cameras, ensuring comprehensive progress control over the entire construction area. It also utilizes a BIM model to integrate construction schedules, achieving automated updates to the schedule, thus reducing the operational loading on project managers.

In terms of functionality, the system integrates a BIM model and 4D construction management, using visualization technology to present different construction phases according to the schedule. Additionally, the system uses different colors on the component to show detection results, simultaneously comparing planned schedules with actual schedules. This color-coded representation indicates whether the construction progress of components is ahead or behind, providing project managers with a more concrete and intuitive understanding of construction progress.

With the continuous development of mixed reality technology, this study plans to integrate wearable mixed reality devices in the future. This involves presenting the model on MR devices to assist less-experienced on-site engineers in quickly familiarizing themselves with the site conditions. Additionally, by replacing camera installations with wearable mixed-reality devices, the aim is to achieve real-time image detection and updates.

6 References

- Son H. and Kim C., 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring. *Automation in Construction*, 19(7):844-854, 2010.
- [2] Golparvar-Fard M., Peña-Mora F., Savarese S., Application of D4AR - A 4-Dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of Information Technology in Construction*, 14:129-153, 2009.
- [3] Volk R., Stengel J., Schultmann F., Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, 38:109-127, 2014.
- [4] Elbeltagi E. and Dawood M., Integrated visualized time control system for repetitive construction projects. *Automation in Construction*, 20(7):940-953, 2011.
- [5] LeCun Y., Bengio Y., Hinton G., Deep learning. *Nature*, 521:436–444, 2015.
- [6] Schmidhuber J., Deep learning in neural networks: An overview. *Neural Networks*, 61:85-117, 2015.
- [7] Alizadehsalehi S., Yitmen I., The Impact of Field Data Capturing Technologies on Automated Construction Project Progress Monitoring. *Procedia Engineering* 161:97-103, 2016.
- [8] Golparvar-Fard M., Peña-Mora F., Savarese S., Application of D4AR - A 4-Dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of*

Information Technology in Construction 14:129-153, 2009.

- [9] Costin A., Pradhananga N., Teizer J., Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project. *Automation in Construction* 24:1-15, 2012.
- [10] Shahi A., West J. S., Haas C. T., Onsite 3D marking for construction activity tracking. *Automation in Construction* 30:136-143, 2013.
- [11] Pučko Z., Šuman N., Rebolj D., Automated continuous construction progress monitoring using multiple workplace real time 3D scans. *Advanced Engineering Informatics* 38:27-40, 2018.
- [12] Baduge S. K., Thilakarathna S., Perera J. S., Arashpour M., Sharafi P., Teodosio B., Shringi A., Mendis P., Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction* 141: 104440, 2022.
- [13] Reja V. K., Varghese K., Ha Q. P., Computer vision-based construction progress monitoring. *Automation in Construction* 138:104245, 2022.
- [14] Zheng Z., Zhang Z., Pan W., Virtual prototypingand transfer learning-enabled module detection for modular integrated construction. *Automation in Construction* 120:103387, 2020.
- [15] Cheng J. C. P., Wong P. K. Y., Luo H., Wang M., Leung P. H., Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Automation in Construction* 139:104312, 2022.
- [16] Martinez P., Barkokebas B., Hamzeh F., Al-Hussein M., Ahmad R., A vision-based approach for automatic progress tracking of floor paneling in offsite construction facilities. *Automation in Construction* 125:103620, 2021.
- [17] Bochkovskiy A., Wang C. Y., Liao H. Y. M., Yolov4: Optimal speed and accuracy of object detection. arXiv 2004.10934, 2020.
- [18] Lin T. Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick L., Microsoft COCO: Common Objects in Context. *Proceedings* of the 2014 European Conference on Computer Vision, 740-755, 2014.
- [19] cocodataset.org. COCO dataset. On-line:<u>https://</u> cocodataset.org/, Accessed:16/02/2024

Transformer-based Pavement Crack Tracking with Neural-PID Controller on Vision-guided Robot

Jianqi Zhang^{1,2}, Xu Yang^{2,3,*}, Wei Wang¹, Ioannis Brilakis⁴, Hainian Wang³ and Ling Ding⁵

¹School of Information Engineering, Chang'an University, Xi'an, 710064, China

²School of Future Transportation, Chang'an University, Xi'an, 710064, China

³School of Highway, Chang'an University, Xi'an, 710064, China

⁴Laing O'Rourke Centre, Engineering Department, Cambridge University, Cambridge, CB2 1PZ, United Kingdom ⁵School of Transportation Engineering, Chang'an University, Xi'an, 710064, China

jqzhang@chd.edu.cn, yang.xu@chd.edu.cn, wei.wang@chd.edu.cn, ib340@cam.ac.uk, wanghn@chd.edu.cn

Abstract -

Pavement crack tracking in unstructured road environments has been and continues to be a crucial and challenging task, playing a vital role in achieving accurate crack sealing for automated pavement crack repair. However, slender cracks suffer from insufficient feature extraction and low tracking efficiency. In this article, a hybrid adaptive control scheme combined with a self-tuning neural network and proportional-integral-derivative (PID) is proposed for dynamic visual tracking of pavement cracks. Specifically, the scheme extracts crack features on the road image plane based on a S2TNet system and determines an optimal control input to guide the robot. S2TNet cross-integrates the global features through the multi-head attention module. It also adaptively recalibrates the channel responses of partial feature maps for fusion operations with the transformer module. Moreover, the Neural-PID controller is designed for adaptive adjustment of control parameters, and the scheme was validated on a physical robot platform. Extensive experimental results showed that the effectiveness of the proposed method in achieving real-time tracking for pavement cracks.

Keywords -

Crack Tracking; Crack Segmentation; Transformer; Neural-PID Control; Mobile Robot

1 Introduction

Pavement cracks are prevalent and hazardous defects that significantly impact driving safety in highway transportation. They primarily arise from a range of factors, such as heavy traffic loads, subpar construction practices, the influence of climate, and inadequate drainage[1, 2]. Failure to promptly repair pavement cracks can lead to accelerated deterioration of the pavement structure through the ingress of rainwater. Even a small crack can rapidly degrade into a pothole overnight, posing a significant hazard to high-speed driving[3, 4]. Hence, regular maintenance and repair of pavement cracks are imperative to prevent crack deterioration and ensure traffic safety[5, 6]. Manual sealing is the conventional approach for repairing pavement cracks. However, manual pavement repair proves to be time-consuming, expensive, and subjective. Therefore, there is a growing demand for automated and efficient repair methods in pavement crack tracking.

Recent studies have primarily focused on the development of crack segmentation with convolutional neural network (CNN)-based methods in road environments. For instance, [7] constructed a novel crack segmentation network called CrackW-Net, and designed the skip-level round-trip sampling block, which can be easily used in various network structures. [8] developed mobile robot system can effectively segment pavement cracks in real scenarios at a speed of 25 frames per second. [9] used a 3D printer as a crack-filling machine. In recent years, path tracking research based on mobile device motion control has become popular. A crack sealing system was designed to control the experimental three-dimensional (3D) printer to repair cracks[2]. [10] proposed the cross-entropy-based adaptive fuzzy control for crack tracking with VT-UMbot.

The insufficient feature extraction is significantly triggered by the limited receptive field in the CNN segmentation model and it often leads to a coarse segmentation of the cracks. Over the years, researchers have proposed various techniques to improve object detectability. These approaches include encoder-decoder[11], multiscale attention[12], and multi-scale feature extraction[13]. Additionally, efforts have been made to enhance object feature representation[14] and fusion[15]. However, despite these advancements, challenges still persist in the field, such as inadequate detection of detailed features and susceptibility to background lighting conditions. On the other hand, low tracking efficiency is also caused by Slender pavement cracks have extreme length-width ratio and complex topology, which lead to irregular paths. Path tracking research mainly focuses on distribution rules and trajectory obeying certain rules. Recent tracking control methods range from traditional PID to various optimized and improved PID such as fuzzy control[10], genetic algorithms[16] and ant colony algorithms[17]. However, challenges related to tuning of control parameters in specialized environments significantly impact the performance of path tracking.

This article presents a pavement crack tracking framework that enhances tracking efficiency in unstructured road scenarios by fusing real-time crack video context features with transformer-based segmentation and proposing Neural–PID control strategies in the crack tracking. To address the insufficient feature extraction and low tracking efficiency, extensive experiments are conducted and verified. The contributions of this work are fourfold:

- Aiming at the problem of pavement crack tracking, a joint transformer-based fusion model and Neural–PID tracking control scheme is proposed. This algorithm successfully achieves stable real-time tracking for pavement crack.
- Enhancing the performance and effectiveness of crack segmentation in challenging road conditions with insufficient feature extraction. This article Introduces a transformer-based fusion model, which leverages multi-fusion strategies to address the challenges posed by coarse crack feature extraction.
- Considering pavement cracks with slender shape and irregular path, a Neural–PID tracking control method is proposed to improve the performance of tracking. Specifically, adaptive adjustment of control parameters is achieved by neural network.
- Conducting extensive experiments on self-created S2T-Crack dataset, the proposed algorithm is successfully deployed in self-developed vision-guided robot. The results show that our method achieved State-of-The-Art.

The structure of this article is organized as follows. Section 2 provides the existing related work. Section 3 outlines the detailed design of our methodology. Section 4 presents the experimental validation of our approach. Finally, Section 5 summarizes the article and discusses future directions.

2 Related works

This section reviews the literature relevant to our proposed pavement crack tracking.

Crack Segmentation. Crack segmentation is a crucial distress inspection technique for different infrastructures, including roads, bridges, tunnels, airports and buildings. There are numerous crack segmentation methods developed based on deep learning. YOLOv5[18] is a single-stage object detection model known for its architectural

features such as the incorporation of Cross-Stage Partial (CSP) and Spatial Pyramid Pooling-Fast (SPPF) methods in the backbone network, as well as the utilization of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) in the Neck network. A lightweight pavement crack detection model is proposed to realize the dual tasks of object detection and semantic segmentation[19].

However, CNN models primarily focus on local feature extraction, which may result in information ambiguity and coarse segmentation when dealing with long-range dependency relationships. Therefore, this research aims to fuse YOLOv5 with Transformer to achieve effective crack segmentation.

Vision Transformer. Thanks to strong representation capabilities, researchers are looking at ways to apply transformer to computer vision tasks. In various visual benchmarks, the performance of the transformer-based model is similar to or better than other CNN types of networks. [20] classified these visual transformer models according to different tasks, and analyzes their advantages and disadvantages, so as to review them. A new video instance segmentation framework based on Transformer is proposed, called VisTR, which regards the VIS task as a direct end-to-end parallel sequence decoding / prediction problem[21]. [22] designed a segmentation model called SEgmentation TRansformer (SETR). A large number of experiments show that SETR has achieved competitive results on Cityscapes.

Compared to CNN, transformer incurs higher computational costs and longer training times. Given the subtle nature of crack features, achieving fine-grained segmentation of cracks is crucial. Therefore, this research introduces self-attention and cross-attention mechanisms to enhance feature extraction.

PID Control. PID control is widely used in path tracking control of mobile robots. In the absence of robot knowledge, the PID controller may be the best controller because it is model-free and its parameters can be easily adjusted separately. However, the parameters depend on artificial empirical values, and parameter optimization is an existing challenge. [23] used the adaptive PID controller to adjust the error to adjust the front wheel angle. A robust PID controller for flight control of four-rotor aircraft is proposed[24]. An adaptive fuzzy control (CEAFC) method based on cross entropy is proposed for PID parameter tuning[10].

Traditional PID controllers are susceptible to external disturbances when it comes to parameter adjustments, leading to convergence issues and system uncertainty. To address these challenges, this study proposes the Neural-PID approach to ensure effective tracking performance.



Figure 1. General framework of our proposed scheme for pavement crack tracking on vision-guided robot. It mainly includes two separate modules: transformer-based crack segmentation (including two branches and three fusion modules), Neural-PID crack tracking (containing three layers networks). All modules are implemented based on the unified YOLOv5 framework, and the details of each module are shown in Figure 2. It is worth noting that both the input video images and tested results were conducted on the S2TCrack dataset.

3 Methodology

This work first describes related issues of pavement crack tracking systems. Additionally, it is deployed on vision-guided robot to achieve crack tracking. This section presents the details of our proposed method.

3.1 Framework

This article focuses on two key aspects of crack tracking in road environment. Firstly, it addresses the challenge of achieving accurate crack segmentation in pavement scenarios characterized by slender crack and complex background. Secondly, it examines the low tracking efficiency of crack tracking control methods in limited parameters tuning conditions. To address these challenges, a crack tracking framework is proposed that ensembles transformer-based fusion network and Neural-PID tracking control algorithm. This framework, illustrated in Figure 1, comprises two main modules: transformer-based crack segmentation and Neural-PID tracking control. The feature fusion module employs the volov5 under the popular transformer to encode and decode crack video images, enabling the fusion of image pixels at the feature level. In order to adaptively tune the tracking controller parameters more quickly, a three-layer structured neural network is used. A detailed overview of the framework is presented in the subsequent subsections.

3.2 Crack Segmentation with Transformer

The proposed module employs the yolov5 under the popular transformer to encode and decode crack video images, enabling the fusion of image pixels at the feature level. In contrast to the initial iteration of YOLOv5, this study presents a novel approach that incorporates a two-branch convolutional neural network backbone. This backbone is illustrated by the light-green modules in Figure 2, and it is designed to extract crack features between video frames from a vision-guided robot. In the context of fusion utilizing FT modules, the fusion process occurs at three distinct stages, facilitating the integration of fused characteristics that comprise both coarse-grained and finegrained semantic information.

A common layer in the encoder and decoder structure is multi-head attention, which consists of multiple parallel self-attention mechanisms. In Self-Attention, Q, K, and V are three vectors calculated on the same input (such as a word in a sequence). Specifically, Q, K, and V can be obtained by applying a linear transformation (e.g., using a fully connected layer) to the original input word's embedding. The dimensions of these three vectors are usually the same and depend on the decisions made during the model design. During the computation of Self-Attention, Q, K, and V are used to calculate attention scores, representing the relationship between the current position and other positions. Attention scores are obtained by taking the dot product of Q and K, dividing the scores by 8, and applying softmax normalization. This process yields weights for each position. Next, these weights are used to compute the weighted sum of V, resulting in the output for the current position. In order to illustrate the effectiveness of our proposed FT fusion module, the feature extraction network of YOLOv5 is extended and redesigned as a backbone composed of two streams to achieve modal fusion and interaction.

3.3 Neural-PID Control for Crack Tracking

In the process of actual pavement crack path tracking motion control, due to the complex control environment and the nonlinear and time-varying characteristics of the controlled object, the conventional PID control can not adjust the adaptive parameters and achieve good adaptability. Using the error back propagation technology, the multi-layer feedforward neural network is called to become a back propagation neural network. Because of its properties, it has excellent performance in nonlinear mapping, such as function approximation and pattern recognition. There are three layers in the back propagation neural network model: input layer, hidden layer and output layer. The input layer processes the type and quantity of input. By controlling the number of layers and activation functions, the hidden layer introduces the possibility of nonlinear mapping. The output layer is responsible for generating some information. The output of the neuron model structure is usually expressed as a nonlinear combination of input and weight.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
(1)

The three non-negative gain parameters of the PID control scheme are output by the BP neural network, so the sigmoid function and other functions without negative output values are applied.

$$g(x) = u \cdot \frac{1}{1 + e^{-x}}$$
 (2)

$$h(x) = \min(\max(0, x), u) \tag{3}$$

$$t(x) = u \cdot \frac{e^x}{e^x + e^{-x}} \tag{4}$$

where u is upper bound of the output. It is used to regulate the output range.

Back propagation neural network nonlinearly maps the input, output and error to the three parameters kp, ki and kd of the PID controller. In addition, the BP neural network has three neuron points for the input layer, five neuron points for the buried layer, and three neuron points for the output layer. The commonly used Tanh function is used in the hidden layer. Combined with BPNN and PID control algorithm, the online self-tuning of PID control parameters can be realized, and the optimal pavement crack tracking motion control effect can be achieved. The structure of the Neural–PID scheme is shown in Fig.1.

4 Experiments

This section focuses on evaluating the proposed method through representative benchmarks and validation. The first aspect covers the experimental settings. Then, the crack segmentation results are analyzed and discussed. Subsequently, our Neural–PID method is deployed on a vision-guided robot to achieve real-time tracking of pavement cracks.

4.1 Experimental Setting

The model training experiments were conducted on an Intel(R) i9-13900K(F) CPU running at 5.8 GHz, along with an NVIDIA GeForce RTX4090 GPU (24 GB) and the following software versions: CUDA v10.2, cuDNN v8.0.1, Pytorch v2.0, and Python v3.8. The unmanned wheeled robot is equipped with an embedded Nvidia Jetson AGX Xavier computer, serving as the main processor with the following specifications: 512 CUDA cores and 64 tensor cores within an Nvidia Volta GPU, v8.2 ARM CPU with 8 cores, and 32 GB DDR4 memory. To acquire pavement crack video images in the front view scene of the unmanned wheeled robot, a front-mounted Realsense D435i camera with a 135-degree field of view (FOV) and an RGB-D perception unit is utilized. The embedded environment includes Jetpack 4.4, PyTorch 1.8, Linux Ubuntu 18.04, and ROS Melodic, as shown in Figure 3.

The evaluation metrics utilized to assess the performance of our proposed method are Precision (B), Precision (M), Recall (B), Recall (M), and AP (Average Precision). Furthermore, the AP incorporates mAP0.5 (B), and mAP0.5 (M), which represent the AP with an IoU threshold greater than 0.5, and mAP0.5:0.95 (B), mAP0.5:0.95 (M), which pertain to the average AP with an IoU threshold ranging from 0.5 to 0.95 in increments of 0.05. The



Figure 2. The architecture of YOLOv5 uses a fusion transformer method that encompasses four separate components: backbone, neck, head, result.



Figure 3. Working conditions of our vision-guided robot under different perspectives are displayed.

notation (B) represents the metric of the predicted boundary frame, corresponding to crack detection. Similarly, the notation (M) represents the metric of the binary mask, corresponding to crack segmentation.

4.2 Results of Crack Segmentation

This section presents an approach to significantly enhance the performance of crack segmentation using the proposed method. The experimental results are analyzed on the open data set CFD and the self-built data set S2TCrack.

4.2.1 CFD Dataset

CFD is utilized for evaluation. The CFD dataset comprises 118 pavement crack images, each with dimensions of 480 pixels by 320 pixels. These images were captured by individuals standing on the road using an iPhone. The ground truths were meticulously annotated at the pixel level, a task that demands significant labor. The images exhibit high quality with a smooth and clean background. Table 1 compares the performance of YOLOv5, our method (Ours), on the pre-trained models n, s, m, l, x. Our method, using the different pretrained model, demonstrated improved performance on the CFD dataset. The following best performance metrics are: [Precision(M) = 0.6818, Recall(M) = 0.5178, APval0.5(M)=0.5304, APval0.5:0.95(M)=0.2453]. Moreover, based on the comprehensive results obtained from the CFD dataset, our proposed method exhibits significantly better performance and versatility, showcasing its exceptional ability in pixel-level crack segmentation tasks.

4.2.2 S2T-Crack Dataset

This section also includes a comparative experiment on the self-built S2TCrack dataset, as presented in Figure4. Our method demonstrates superior segmentation performance in the pretrained model 's', which boasts a mere 6.7M parameters and 15.2M GFLOPs. Meanwhile, the segmentation accuracy is moderately acceptable. Along with the segmentation results of three scenes from the self-built S2TCrack dataset, YOLOv5 roughly splits the cracks, ignoring certain subtle features, which may result in incomplete masks, leading to fractures or local losses. Our method effectively generates masks that appropriately cover the target cracks, thanks to the utilization of SA and CA. To further enhance the performance, FT modules are integrated to fuse crack features. Our method is capable of generating highly accurate binary masks, making it suitable for various complex scenes.

4.3 Online Tuning of PID Parameters

This section presents an approach to significantly enhance the performance of crack tracking using the proposed method. The experimental results are analyzed on different control algorithms.

4.3.1 Comparison of Tracking Control

As shown in Figure 5, compared with CEAFC, the Neural-PID control scheme approaches the ideal solution with a faster convergence rate at iteration 200, indicating that the Neural-PID has stronger deterministic global search ability and faster high-dimensional optimal solution discovery speed. The results show that the Neural-PID control algorithm is superior to the other three methods.

GFLOPs Pretrained Params Method Batch Size Precision(B) Precision(M) Recall(M) $mAP^{val}0.5(B) \quad mAP^{val}0.5(M) \quad mAP^{val}0.5:0.95(B) \quad mAP^{val}0.5:0.95(M)$ Recall(B) Model /M /M 32 0.6688 0.4424 0.4523 0.3158 0.4339 0.2301 0.1644 0.0393 1.9 6.7 0.7254 0.4621 0.4474 0.4474 0.4944 0.3875 0.2456 0.0854 7.4 25.7 16 YOLOv5 0.7326 0.4562 0.4645 0.3947 0.4750 0.3631 0.2586 0.0545 69.8 21.7 0.7289 0.4637 0.4737 0.4211 0.5032 0.3849 0.2561 0.0911 47.3 146.4 0.4726 0.486 0.3756 0.7288 0.5256 0.421 0.2871 0.082 88.2 264 0.5958 0.5033 0.5723 0.3264 0.7153 0.4167 0.4548 0.196 2.0 6.9 0.4943 16 0.7982 0.5653 0.4817 0.5921 0.4906 0.3485 0.1873 7.5 25.7 0.7705 0.5831 0.5736 0.4524 0.5257 0.5187 0.3356 0.2294 21.8 69.9 Ours 0.7657 0.6024 0.5975 0.4688 0.5354 0.5018 0.3721 0.2453 47.4 146.7 0.7724 0.6818 0.5487 0.5178 0.5677 0.5304 0.3953 0.2102 88.4 265

Table 1. Real-time segmentation results in the CFD dataset.



Figure 4. Visualization of segmentation results using YOLOv5 and our proposed method of our created S2T-Crack dataset.

According to the convergence curve, the Neural-PID algorithm needs 60 iterations to find the local optimal solution and 90 iterations to get rid of the local optimal solution. Compared with the 150 iterations required by the CEAFC method, this is a huge reduction. Therefore, Neural-PID can eliminate the local optimal solution and improve the robustness of crack tracking control.



Figure 5. The comparison results of algorithm optimization.

4.3.2 Analysis of Tracking Error

Table 2. The comparison results of crack tracking error.

Creak ID	Sogmontation Model	Control Method							
Crack ID	Segmentation Model	PID	Fuzzy PID	CEAFC	Neural-PID				
	n	9.71	5.81	4.68	4.47				
	S	9.57	5.73	4.54	4.12				
#1	m	9.93	6.07	4.73	4.59				
	1	10.24	6.44	4.91	4.75				
	х	11.86	6.76	5.16	5.01				
	n	13.12	6.21	5.03	4.94				
	S	12.86	6.19	4.85	4.63				
#2	m	13.38	6.58	5.17	4.86				
	1	13.89	6.91	5.43	5.21				
	х	14.31	7.35	5.79	5.57				
	n	15.08	7.73	6.25	5.94				
	S	14.59	7.51	5.86	5.67				
#3	m	15.36	8.09	6.57	6.29				
	1	16.18	8.67	6.93	6.76				
	х	16.85	7.28	7.06	6.81				

Experiments are performed on real roads to verify the

performance of road crack tracking, as shown in Table 2. This average absolute error is used as a performance evaluation index. The unmanned wheeled robot uses the proposed method to compare the results of road crack tracking error with other control methods during the tracking process. Crack #1 is a straight pavement crack. In the case of crack #1, our algorithm achieves the smallest average crack tracking absolute error in the pre-trained model 's', with a measured value of 4.12 mm. Crack #2 is a curved pavement crack. For the case of crack #2, our algorithm achieves the smallest average absolute error in the pre-trained model 's', with a measured value of 4.63 mm. Crack #3 is a continuous turning pavement crack. Our algorithm achieves the minimum mean absolute error in the pre-trained model 's', and the measured value is 5.67 mm.

5 Conclusions

This article addresses two critical issues in road crack tracking: insufficient feature extraction and low tracking efficiency. To overcome these challenges, the research primarily focuses on enhancing the pavement crack feature extraction from crack video images using our transformerbased crack segmentation method. By combining SA and CA, and leveraging FT model, the performance of binary masks in segmentation instances is significantly improved, enabling fine-grained segmentation of pavement cracks. Through the proposed Neural-PID, our method is deployed on NVIDIA AGX Xavier to enable real-time tracking of actual pavement cracks on a vision-guided robot. In future research, the utilization of road crack depth images will be considered, along with the exploration of alternative control methods to enhance the accuracy and robustness of the tracking control algorithm. The developed visionguided robot can be integrated with repair mechanisms to accomplish road crack repairs.

Acknowledgements

The study presented in the article was partially supported by the National Key Research and Development Program of China (No.2021YFB2601000), National Natural Science Foundation of China (No.52078049, No.52378431), Natural Science Foundation of Shaanxi Province (2022JM-193), Fundamental Research Funds for the Central Universities, CHD (No.300102210302, No.300102210118), the 111 Project of Sustainable Transportation for Urban Agglomeration in Western China (No.B20035).

References

- [1] Jingwei Liu, Xu Yang, Stephen Lau, Xin Wang, Sang Luo, Vincent Cheng-Siong Lee, and Ling Ding. Automated pavement crack detection and segmentation based on two-step convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, 35(11):1291–1305, 2020. doi:10.1111/mice.12622.
- [2] Jingwei Liu, Xu Yang, Xin Wang, and Jian Wei Yam. A laboratory prototype of automatic pavement crack sealing based on a modified 3D printer. *International Journal of Pavement Engineering*, 23(9):2969–2980, 2022. doi:10.1080/10298436.2021.1875225.
- [3] Jinchao Guan, Xu Yang, Ling Ding, Xiaoyun Cheng, Vincent C.S. Lee, and Can Jin. Automated pixel-level pavement distress detection based on stereo vision and deep learning. *Automation in Construction*, 129: 103788, 2021. doi:10.1016/j.autcon.2021.103788.
- [4] Jinchao Guan, Xu Yang, Pengfei Liu, Markus Oeser, Han Hong, Yi Li, and Shi Dong. Multi-scale asphalt pavement deformation detection and measurement based on machine learning of full field-ofview digital surface data. *Transportation Research Part C: Emerging Technologies*, 152:104177, 2023. doi:10.1016/j.trc.2023.104177.
- [5] Zhihao Pan, Jinchao Guan, Xu Yang, Kang Fan, Jeremy C.H. Ong, Ningqun Guo, and Xin Wang. One-stage 3D profile-based pavement crack detection and quantification. *Automation in Construction*, 153:104946, 2023. doi:10.1016/j.autcon.2023.104946.
- [6] Jianqi Zhang, Xu Yang, Wei Wang, Jinchao Guan, Ling Ding, and Vincent C. S. Lee. Automated guided vehicles and autonomous mobile robots for recognition and tracking in civil engineering. *Automation in Construction*, 146:104699, 2023. doi:10.1016/j.autcon.2022.104699.
- [7] Chengjia Han, Tao Ma, Ju Huyan, Xiaoming Huang, and Yanning Zhang. CrackW-Net: A Novel Pavement Crack Image Segmentation Convolutional Neural Network. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):22135–22144, 2022. doi:10.1109/TITS.2021.3095507.
- [8] Guijie Zhu, Jiacheng Liu, Zhun Fan, Duan Yuan, Peili Ma, Meihua Wang, Weihua Sheng, and Kelvin C. P. Wang. A lightweight encoder–decoder network for automatic pavement crack detection. *Computer-Aided Civil and Infrastructure Engineering*, pages 1–23, 2023. doi:10.1111/mice.13103.

- Machine-filling of cracks in asphalt concrete. Automation in Construction, 141:104463, 2022. doi:10.1016/j.autcon.2022.104463.
- [10] Jianqi Zhang, Xu Yang, Wei Wang, Jinchao Guan, Wenbo Liu, Hainian Wang, Ling Ding, and Vincent C. S. Lee. Cross-entropy-based adaptive fuzzy control for visual tracking of road cracks with unmanned mobile robot. Computer-Aided Civil and Infrastructure Engineering, pages 1-20, 2023. doi:10.1111/mice.13108.
- [11] Firdes Çelik and Markus König. A sigmoidoptimized encoder-decoder network for crack segmentation with copy-edit-paste transfer learning. Computer-Aided Civil and Infrastructure Engineering, 37(14):1875-1890, 2022. doi:10.1111/mice.12844.
- [12] Xinzi Sun, Yuanchang Xie, Liming Jiang, Yu Cao, and Benyuan Liu. DMA-Net: DeepLab With Multi-Scale Attention for Pavement Crack Seg-IEEE Transactions on Intelligent mentation. Transportation Systems, 23(10):18392-18403, 2022. doi:10.1109/TITS.2022.3158670.
- [13] Bo Chen, Hua Zhang, Guijin Wang, Jianwen Huo, Yonglong Li, and Linjing Li. Automatic concrete infrastructure crack semantic segmentation using deep learning. Automation in Construction, 152:104950, 2023. doi:10.1016/j.autcon.2023.104950.
- [14] Honghu Chu and Pang-jo Chun. Fine-grained crack segmentation for high-resolution images via a multiscale cascaded network. Computer-Aided Civil and Infrastructure Engineering, pages 1-20, 2023. doi:10.1111/mice.13111.
- [15] Dongho Kang, Sukhpreet S. Benipal, Dharshan L. Gopal, and Young-Jin Cha. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. Automation in Construction, 118:103291, 2020. doi:10.1016/j.autcon.2020.103291.
- [16] Zhihan Zhang and Di Bai. Optimization of Improved PID Control Strategy Based on Genetic Algorithm. Journal of Physics: Conference Series, 2417(1):012025, 2022. doi:10.1088/1742-6596/2417/1/012025.
- [17] YeFei Kang, ZhiBin Li, and Tao Wang. Application of PID Control and Improved Ant Colony Algorithm in Path Planning of Substation Inspection Robot. Mathematical Problems in Engineering, 2022:1-10, 2022. doi:10.1155/2022/9453219.

- [9] Frank K.A. Awuah and Alvaro Garcia-Hernández. [18] Ultralytics. ultralytics/yolov5: v7.0 YOLOv5 SOTA Realtime Instance Segmentation. https: //github.com/ultralytics/yolov5.com, 2022. URL https://doi.org/10.5281/ zenodo.7347926. Accessed: 7th May, 2023.
 - [19] Yuchuan Du, Shan Zhong, Hongyuan Fang, Niannian Wang, Chenglong Liu, Difei Wu, Yan Sun, and Mang Xiang. Modeling automatic pavement crack object detection and pixel-level segmentation. Automation in Construction, 150:104840, 2023. doi:10.1016/j.autcon.2023.104840.
 - [20] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, Zhaohui Yang, Yiman Zhang, and Dacheng Tao. A Survey on Vision Transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1):87-110, 2023. doi:10.1109/TPAMI.2022.3152247.
 - [21] Yuqing Wang, Zhaoliang Xu, Xinlong Wang, Chunhua Shen, Baoshan Cheng, Hao Shen, and Huaxia Xia. End-to-End Video Instance Segmentation with Transformers. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8737-8746, Virtual, Online, United states, June 2021. IEEE Computer Society. doi:10.1109/CVPR46437.2021.00863.
 - [22] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip H.S. Torr, and Li Zhang. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 6877-6886, Virtual, Online, United states, June 2021. IEEE Computer Society. doi:10.1109/CVPR46437.2021.00681.
 - [23] Wael Farag. Complex Trajectory Tracking Using PID Control for Autonomous Driving. International Journal of Intelligent Transportation Systems Research, 18(2):356-366, 2020. doi:10.1007/s13177-019-00204-2.
 - [24] Salman Bari, Syeda Shabih Zehra Hamdani, Hamza Ullah Khan, Mutte Ur Rehman, and Haroon Khan. Artificial neural network based selftuned PID controller for flight control of quadcopter. In 2019 International Conference on Engineering and Emerging Technologies (ICEET), pages 1-5, Lahore, Pakistan, February 2019. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/CEET1.2019.8711864.

Toward shotcrete process simulation to support robotic operation

Mohammad Reza Yazdi Samadi¹, Ralf Waspe¹ and Christian Schlette¹

¹SDU Center for Large Structure Production, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Denmark

mry@mmmi.sdu.dk, raw@mmmi.sdu.dk, chsch@mmmi.sdu.dk

Abstract -

In this study, we introduce a detailed physics-based particle system for simulation of the shotcrete process. The development of this framework concept is informed by an extensive literature review encompassing diverse modeling and simulation methodologies applied to shotcreting processes, coupled with insights derived from experimental studies on shotcrete. This method can effectively capture key shotcrete characteristics such as adhesion, cohesion, and rebound. Furthermore, simulated shotcrete particles have interaction with different objects, colliding and bouncing off different geometries that represent various construction substrates and conventional reinforcements. The primary objective of the process simulation is to expedite the advancement of robotic systems tailored for executing shotcreting operations, extending beyond mere visualization purposes. Through the integration of shotcrete process simulation into a virtual environment, a simulated representation of a robotic concept can be systematically experimented upon to anticipate and understand its behavior. This approach proves instrumental in refining the design of robotic systems, optimizing robot motion planning, mission planning, and enhancing management and operation practices.

Keywords -

Shotcrete; Process Simulation; Robot; Particle System

1 Introduction

Despite having a pivotal role in the world's economy, construction sector has been one of the last domains to adopt and apply automation technologies. This could be attributed to the irregular and ever-evolving nature of the construction site and the diverse, intricate tasks typically mandated. However, the scarcity of skilled work force, escalating labor costs, accrescent demand for new infrastructure, and the imperative need for maintaining existing ones emphasize the necessity for the integration of automation and robotic technologies in the construction industry [1].

Shotcrete as a concrete compaction method has gained

popularity over the past century due to its exceptional economic efficiency. Even though shotcreting is highly mechanized today, the quality of the process is still heavily reliant on the skill of human operators. Conversely, this method poses significant challenges, subjecting workers to strenuous physical exertion, continuous exposure to hazardous dust, and the constant risk of ground instability.

In a standard shotcrete project, understanding the thickness and distribution of the material adhered to the target surface is fundamentally important. Knowledge of the rebounded material is also essential to minimize waste and operational costs. If a robot is to be deployed to autonomously carry out tasks traditionally performed by humans, profound understanding of the intricacies of the process becomes imperative. Without such knowledge, the robot may be at risk of sustaining damage (e.g., from rebounded shotcrete material) or not be able to carry out its task (e.g., without awareness of shotcrete cohesive failure, the robot may become indefinitely stuck in a loop of spraying the same spot repeatedly).

Therefore, we have developed a particle system integrated with a physics engine, with the aim of improving the accuracy and efficacy of shotcrete process simulation. This conceptual framework is considered an essential facilitator for robot-based shotcreting. The paper is structured into two main sections. Firstly, it provides an overview of the current state of robot-based shotcreting and shotcrete process simulation. Secondly, it presents and discusses the simulation's composition and delves into implementation details.

2 Background

Efforts have been made to employ robotic manipulators for shotcreting in underground structures, including tunnels [2, 3, 4]. Cheng et al. [5] simplified the control system of a semi-automated shotcreting manipulator and validated their model through real-time computer simulation. They also utilized a simulation model for calculating the nozzle path in a fully automated shotcreting robot. Girmscheid and Moser [6] introduced a versatile robotic system capable of manual, semi-automatic, and fully automated shotcreting, demonstrating promising results and emphasizing the potential of automation in construction. These methodologies incorporated shotcrete profile measurement and application control, utilizing laser scanning before and after the shotcreting process, resulting in accurate estimations of the thickness of the accumulated shotcrete. However, one might ask, what if the applied layer of shotcrete is not within the specified tolerances, requiring its removal and negating the economic efficiency of shotcrete.

The noticeable differences in nozzle trajectories observed between manual and automatic shotcreting on a shared robotic platform, as emphasized in the study by Nabulsi et al. [7], suggest divergent approaches taken by robots and humans in the application of shotcrete. In contexts such as tunneling and ground support, ensuring uniform coverage is paramount. Equally crucial is a comprehensive understanding of how different nozzle paths can influence the final shape of the shotcrete on the surface, particularly in applications like infrastructure restoration. In conclusion, despite significant advancements in remotecontrolled robot-based shotcreting, challenges persist in achieving full autonomy. The question of whether current technology can enable robots to autonomously perform shotcreting remains unanswered.

Shotcrete is a complex process of spraying concrete mixture onto a designated surface with high-impact velocity. This involves shooting a multi-phase blend of cementitious material, water, aggregates, and admixtures through a hose. Besides air, an additive, a quick-setting agent, is often introduced at the nozzle. In application, a majority of the sprayed concrete adheres to the target substrate. However, a fraction of the material may exhibit rebound, impinging upon objects and adjacent surfaces or returning to the ground. The efficacy of the shotcrete operation significantly hinges on the proficiency of the operator, who plays a pivotal role in minimizing rebound and preventing the detachment of previously applied shotcrete.

It is common to apply sprayed concrete over steel reinforcements. These structures are affixed to the surface with a specific distance, partially obstructing it. Spraying with the nozzle perpendicular to the surface can lead to the creation of large air pockets or sand lenses behind these elements. A skilled operator directs the nozzle to fill the space between the reinforcement and the wall, ensuring complete encapsulation with concrete. A basic illustration of the shotcrete process is shown in Figure 1.

A realistic simulation ought to incorporate shotcrete sprayability parameters. According to Trussell and Jacobsen [8], the term sprayability is frequently used to describe properties such as:

• Adhesion which is the ability of sprayed concrete to attach to the substrate and avoid falling. It is mostly



Figure 1. Simplified shotcrete process model

affected by the nature and status of the substrate and the composition of the mix.

- Cohesion which is the ability of fresh sprayed concrete to stick to itself and avoid falling and slumping under its own weight. It can be measured in terms of the thickness that can be applied before the build-up material starts to fall under self-weight.
- Rebound which is an unwanted yet unavoidable byproduct of the shotcrete process and occurs when the shotcrete particles fail to adhere to the substrate and instead bounce back from the surface. Armelin and Banthia [9] put an effort to derive a constitutive model of the rebound phenomenon.

There are additional properties associated with the sprayability of shotcrete. Some researchers found strength gained with time to be an intriguing aspect [10, 11]. In a study by Han et al. [12], Artificial Intelligence (AI) was employed to rapidly determine the mix proportion of wetmix shotcrete. Others, such as [13], focused on examining the impact of process parameters and model-based process control.

The absence of any need to prepare concrete forms renders shotcreting an extremely competitive technology. Nevertheless, the skills of the nozzle operator play a vital role in minimizing material waste during application, thus impacting the economic efficiency of the construction process [14]. Therefore, construction companies invest significantly in training their shotcrete personnel. They require workers to undergo repetitive shotcreting drills as part of their training until their skills are fully developed. While this method is not sustainable and not always applicable, researchers have addressed the issue by focusing on the development of real-time simulators for shotcrete training.

Presumably, Börjesson and Thell [15] were the first researchers to experiment with various rendering techniques in developing a virtual environment for shotcrete training. They implemented a particle system and put forward the idea of addition of gravity and dynamic forces to the shotcrete particle, albeit in their implementation they utilized ray casting for their adhesion model. They deliberately limited the particle system to rendering of the spraying effect with no impact on the surfaces. Moreover, their work lacked realistic sprayability parameters of shotcrete. Nevertheless, they managed to commercialize the findings of their research into a real-time training simulator [16].

In their study, Velez et al. [17] developed a real-time shotcrete simulation that introduced gravity to affect rebound and detached concrete. They also employed ray tracing for the adhesion model and introduced a shadow mapping technique to account for cells on the target surface that were shadowed by the steel mesh from the nozzle's perspective. While their approach was commendable, it leans toward oversimplification, especially in the aspect of particle collision with other objects—a complexity that may not be fully captured by their model.

3 Shotcrete process simulation

In this chapter, a novel shotcrete process simulation is introduced to address the limitations identified in previous research. The proposed method enhances critical aspects of the process, addressing adhesion, cohesion, rebound, and simulation of the shotcrete application over additional supporting elements. This contributes to a more thorough and realistic depiction of the process.

The simulation involves parameters that can be categorized into two main groups: those related to the performance of the particle system, such as the number of active particles and simulation time step, and those related to the shotcreting process, such as nozzle spread angle and rebound rate. The latter parameters are adopted and adapted from the current state of the art (i.e., scientific articles and technical reports).

This section is divided into three subsections. The first part explains the functioning of the particle system, including the initialization of particles, and their emission. The next part delves into the implementation of the sprayability parameters of shotcrete—detailing what happens when particles hit an object. Finally, the output of the simulation is described and discussed.

3.1 Particle system

Ballou [18] drew an analogy between shotcrete and the action of throwing balls dipped in paste at a surface. He highlighted that the transfer of paste to the surface occurs upon the ball rebounding, emphasizing the importance of a certain degree of rebound as the primary method for transferring cementitious material to the target surface. Subsequently, each successive ball hitting the surface contributes to and compacts the already deposited concrete into the voids and porosity of the surface. It is not an exaggeration to assert that shotcreting is similar to shooting out particles. Similarly, aggregates, acting as particles, transfer mortar to the target surface, filling cracks and voids, contributing to excellent in-place compaction due to their high kinetic energy.

The foundation of this model is rooted in the intricate functioning of a particle system, which serves as the primary mechanism for simulating and visualizing various aspects of the shotcrete process. The particle system is designed to emulate the behavior of individual particles, representing shotcrete particles, as they interact with surfaces, undergo collisions, and contribute to the build up of material on the target substrate. This sophisticated approach allows for a detailed and dynamic simulation, enabling a closer approximation to real-world shotcrete scenarios.

Firstly, particles are initialized with a randomized direction. Illustrated in Figure 2, the parameter *Spread* denotes the angle of the spray cone, and its value can be adjusted within the simulation environment. In practical scenarios, the spray cone angle depends on factors such as the type of mix (wet or dry), material composition, and air pressure. The data for *Spread* have been sourced from [19]. All particles originate from a circle with a variable diameter, simulating the nozzle aperture, and each is assigned a direction vector. The magnitude of the velocity vector is randomly chosen from a range between a minimum and maximum value, with velocity data derived from experimental studies conducted by Ginouse et al. [20].

To account for the effect of gravity on the particle, a vector expressed as (0, 0, -9.8) ^m/s is added to the defined velocity vector in each second of the simulation. Without gravity, the particle would have followed a straight path along the velocity vector toward the point (x, y, z), as illustrated in Figure 2. However, due to gravity, the particle deviates toward the negative z direction in each time step, eventually hitting a point (x, y, z') slightly lower. Substantial evidence from the work of Ginouse and Jolin



Figure 2. Particle initialization and emission
[19] supports this observation. In their study, high-speed cameras captured images of shotcrete spray, revealing that the shotcrete particles are influenced by gravity even over short distances.

The particle engine handles creating and updating particles throughout the simulation. Essentially, the number of particles to be emitted is determined by the value of this parameter, which is set in the simulation. Each particle is randomly initialized and is updated at each time step during its active lifespan. When a particle hits a surface and adheres to it, it is deactivated and returned to the pool of total number of particles.

The total number of particles and particles emitted per second are two essential parameters in the simulation. Increasing the value of these parameters would enhance the realism of the simulation, but it comes at the cost of computational efficiency. Nonetheless, this provides flexibility to run the simulation on machines with varying computational power.

3.2 Shotcrete sprayability parameters

It is assumed that during a small time-step, each particle travels in a straight line. Hence, the collision problem is simplified to finding the intersection between a line segment and other basic geometries [21]. For instance, when simulating the adhesion of material on a wall, the target surface is treated as a finite plane. The algorithm then seeks the intersection between a line segment and a finite plane. If collision is detected, the intersection point is calculated. A similar approach is applied to handle intersections between a line segment and a cylinder, making it suitable for simulating shotcreting in tunnels and objects along the spray path (i.e., reinforcements).

When an intersection point is found, the algorithm determines whether the particle adheres to the surface or rebounds. This is achieved by calculating the rebound percentage rate of that particle, referencing Melbye curves. In his technical handbook, Melbye [22] identified and categorized the factors affecting rebound, declaring the nozzle angle to the substrate as the decisive one. Therefore, the angle between particle's velocity vector and the surface normal vector at intersection point is calculated. Based on this angle, rebound percentage rate is determined. If the particle adheres to the surface, a value based on the deposition model is added to the corresponding heightfield. Subsequently, the particle is deactivated, and the visualization is updated in the simulation. If the particle rebounds, the collision response involves assigning a new direction, which is the reflection of the velocity vector with respect to the surface normal vector. The magnitude of the reflected velocity is multiplied by the coefficient of restitution of shotcrete (e) derived from [23]. The particle continues to be updated until it adheres to any surface

Table 1. Proposed algorithm for particle collision with surface, and subsequent adhesion or rebound

Pseudo-code:

nowPos = shotcreteParticle -> position;
velocity = shotcreteParticle -> velocity;
nextPos = nowPos + (velocity * diffTime);
<pre>point = compileIntersection(lineSegment, surface);</pre>
if (!contactOnSurface(<i>point</i>))
return;
<pre>chance = generateRandomNumber();</pre>
<i>normal</i> = compileSurfaceNormal(<i>point</i>);
<i>rebound</i> = calculateReboundRate(<i>velocity</i> , <i>normal</i>);
if (chance <= rebound)
deactivate(shotcreteParticle);
updateHeightField();
updateVisualization();
else
<pre>velocity = e* compileReflection(velocity, normal);</pre>

(the target surface, reinforcements, or the ground). The pseudo-code for this algorithm is presented in Table 1.

In the simulation of cohesive failure, the algorithm simplifies the process using a modified seed fill algorithm, akin to the approach in [17], to identify the detachment area. A detachment occurs when the difference between neighboring values exceeds h_{max} , as specified in [24]. The detachment volume is then calculated based on the disparity between the number of adhered particles in the detachment zone and those in neighboring cells. Consequently, an equivalent number of particles detach from the target surface and fall to the ground. This may result in more material being detached, creating a hole-like effect characteristic of shotcrete cohesive failures. For a more in-depth illustration of the proposed simulation process and the interrelationships among its various components, please refer to Figure 3.

3.3 Output of the simulation

The most critical outcome of the process simulation is the amount of concrete deposited at each point on the target substrate. To capture and analyze this information, a data structure capable of storing accumulated material values is essential. A height-field, representing a twodimensional array of integers, emerges as a valuable tool for presenting the simulation results. During the simulation, the height-field is visualized as heat-maps, providing a dynamic representation of the concrete deposition. Subsequently, this data is stored and shared with other software in image file formats, facilitating in-depth analysis or rendering within an environment visualization engine. While realistic rendering of the height-field could enhance the



Figure 3. Proposed system component diagram of the shotcrete process simulation

simulation's visual realism, it is important to note that this research is not primarily focused on training human operators. Therefore, realistic rendering for human perception is not a priority; instead, the emphasis lies in ensuring that the output is well-defined for machine interpretability.

This methodology enables the execution of simulations involving 10 million particles, with 0.1 million particles emitted per second, and a simulation time step of 0.001 second, close to real time. These simulations were conducted on a system equipped with an AMD Ryzen 7 3700X CPU and 16 GB of RAM, using VEROSIM®, a software solution for virtual reality and simulation.

Figures 4 and 5 display the robot shotcreting within the virtual environment, offering insight into particle interaction. Within this scene, two planar obstacles are introduced: one simulating the accumulation of concrete on a hypothetical wall and another positioned on the ground to capture rebounded material. At the nozzle's tip, an emitter releases particles in a cone-shaped spray pattern. The varying colors on each surface represent heat-maps, correlating accumulated material thickness with color. Analogous to topographic maps, this visualization method aids in identifying unevenness, crucial for assessing material deposition uniformity. Furthermore, as extensively discussed in this section, the height-field is concurrently generated behind this layer.



Figure 4. Screenshot of a conceptual robot performing shotcreting in the virtual environment



Figure 5. Detailed screenshot of robot-based shotcrete process within the simulation software

4 Conclusions and future work

This article introduces a concept of a framework for realistic shotcrete process simulation. In contrast to previous methods that relied on ray tracing, our approach utilizes a particle system to track particles from emission until collision with surfaces. These collisions can result in either adhesion to the surfaces or rebound. The incorporation of a physics engine allows particles to be influenced by gravity, air drag, and interactions with obstacles. The algorithm evaluates collisions and estimates rebound at each time step. When a particle adheres to a surface, the simulation increments the corresponding value in the height-field. The output of this simulation is a height-field representing the accumulated shotcrete on the surface, visualized through heat-maps in the simulation. This data can be saved as an image for further analysis or rendering.

Integrating the process simulation with a robot's kinematic simulation establishes a platform for offline optimization for achieving objectives, including optimal surface coverage, uniform material distribution, and effective filling of irregular voids. Furthermore, enhances shotcrete application efficiency and contributes to the versatility of robotic operations across diverse construction scenarios.

Opportunities for refinement persist in enhancing the realism of the process simulation. Continued efforts could be directed toward improving the model's fidelity, with a focus on incorporating nuanced aspects such as the spread of shotcrete material post-particle impact. This may be achieved through the integration of a cellular automata simulation, allowing for a more accurate representation of shotcrete sprayability parameters, including adhesion, compaction, and shrinkage over time. Furthermore, the integration of the process simulation into a Digital Twin of a robot holds promise for conducting high-fidelity, experimentable simulations. Such simulations can contribute significantly to the design, evaluation, and deployment of efficient robots tailored for shotcreting tasks.

5 Acknowledgements

Grateful acknowledgement is made to the European Commission's Horizon program for providing financial support for this study under grant agreement number 101058731.

References

- [1] J. Delgado, L. Oyedele, A. Ajayi, L. Akanbi, O. Akinade, M. Bilal, and H. Owolabi. Robotics and automated systems in construction: Understanding industry-specific challenges for adoption. *Journal of Building Engineering*, 26:100868, 2019. doi:10.1016/j.jobe.2019.100868.
- [2] X. Lin, D. Song, M. Qin, W. Zhang, X. He, and B. Xie. An automatic tunnel shotcrete robot. In

Proceedings of the Chinese Automation Congress (CAC), pages 3858–3863, Hangzhou, China, 2019. doi:10.1109/cac48633.2019.8996350.

- [3] G. Liu, X. Sun, Y. Liu, T. Liu, C. Li, and X. Zhang. Automatic spraying motion planning of a shotcrete manipulator. *Intelligent Service Robotics*, 15(1): 115–128, 2022. doi:10.1007/s11370-021-00405-3.
- [4] G. Moniz and H. Costelha. Path generation and execution for automatic shotcrete in railway tunnels. In Proceedings of the IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), pages 214–219, Tomar, Portugal, 2023. doi:10.1109/ICARSC58346.2023.10129548.
- [5] M. Cheng, Y. Liang, C. Wey, and J. Chen. Technological enhancement and creation of a computeraided construction system for the shotcreting robot. *Automation in construction*, 10(4):517–526, 2001. doi:10.1016/s0926-5805(00)00104-7.
- [6] G. Girmscheid and S. Moser. Fully automated shotcrete robot for rock support. *Computer-Aided Civil and Infrastructure Engineering*, 16(3):200– 215, 2001. doi:10.1111/0885-9507.00226.
- [7] S. Nabulsi, A. Rodriguez, and O. Rio. Robotic machine for high-quality shotcreting process. In Proceedings for the 41st International Symposium on Robotics and 6th German Conference on Robotics, pages 1–8, Munich, Germany, 2010. doi:10.1201/9781482266597-25.
- [8] 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.
- [9] H. Armelin and N. Banthia. Mechanics of aggregate rebound in shotcrete—(part i). *Materials and structures*, 31:91–98, 1998. doi:10.1007/BF02486470.
- [10] L. Malmgren, E. Nordlund, and S. Rolund. Adhesion strength and shrinkage of shotcrete. *Tunnelling and underground space technology*, 20(1):33–48, 2005. doi:doi:10.1016/j.tust.2004.05.002.
- [11] R. Schütz, D. Potts, and L. Zdravkovic. Advanced constitutive modelling of shotcrete: Model formulation and calibration. *Computers and Geotechnics*, 38(6):834–845, 2011. doi:doi:10.1016/j.compgeo.2011.05.006.
- [12] B. Han, K. Ji, B. Singh, J. Qiu, and P. Zhang. An optimization method for mix proportion of wetmix shotcrete: Combining artificial neural network with particle swarm optimization. *Applied Sciences*, 12(3):1698, 2022. doi:doi:10.3390/app12031698.

- [13] B. Schuler and O. Sawodny. Spray pattern analysis using wet-mix concrete for model based process control towards automated construction. In *Proceedings for the 15th International Conference* on Automation Science and Engineering (CASE), pages 661–666, Vancouver, BC, Canada, 2019. doi:doi:10.1109/COASE.2019.8842853.
- [14] N. Dadiani. Use of sprayed concrete in the construction of tunnels. *Power Technology and Engineering*, 52(3):291–297, 2018. doi:10.1007/s10749-018-0947-8.
- [15] P. Börjesson and M. Thell. Shotcrete simulator for education of shotcrete robot operators. M.Sc. thesis, University of Gothenburg, 2010.
- [16] Edvirt training for the mining and tunneling industry. online: https://www.edvirt.com/, Accessed: 17/11/2023.
- [17] G. Velez, L. Matey, A. Amundarain, A. 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.
- [18] M. Ballou. Shotcrete rebound-how much is enough? Shotcrete magazine, American Shotcrete Association, 2003.
- [19] N. Ginouse and M. Jolin. Investigation of spray pattern in shotcrete applications. *Construction and Building Materials*, 93:966–972, 2015. doi:10.1016/j.conbuildmat.2015.05.061.
- [20] N. Ginouse, M. Jolin, and B. Bissonnette. Effect of equipment on spray velocity distribution in shotcrete applications. *Construction and Building Materials*, 70:362–369, 2014. doi:10.1016/j.conbuildmat.2014.07.116.
- [21] D. Eberly. *3D game engine design: a practical approach to real-time computer graphics*. Elsevier, Amsterdam, 2. edition, 2007.
- [22] T. Melbye. Sprayed concrete for rock support. MBT International, 2001.
- [23] H. Armelin and N. Banthia. Mechanics of aggregate rebound in shotcrete—(part i). *Materials and structures*, 31:91–98, 1998. doi:10.1007/bf02486470.
- [24] S. Austin, C. Goodier, and P. Robins. Low-volume wet-process sprayed concrete: pumping and spraying. *Materials and Structures*, 38:229–237, 2005. doi:10.1617/14025.

A Framework of Integrating HBIM and GIS for Automated Fire Risk Assessment of Heritage Buildings

Yutong Qiao¹, Chi Chiu Lam¹, Mun On Wong¹ and Yujin Xu²

¹Department of Civil and Environmental Engineering, University of Macau, Macau SAR, China ²The Third Architectural Design Institute, Guangzhou Design Institute Group Co., Ltd., Guangdong, China <u>yutong.qiao@connect.um.edu.mo, fstccl@um.edu.mo, mowong@um.edu.mo, xuyujin@gzdi.com</u>

Abstract -

Historic buildings face multi-faceted fire risks that threaten their conservation. A comprehensive fire risk assessment is essential to prevent fires and protect cultural heritage. Conventional practices rely on surveys and site visits, which are inefficient in capturing up-to-date information digitally and analyzing the risk levels quantitatively. This paper proposes a framework integrating Historic Building Information Models (HBIM) and Geographic Information Systems (GIS) to enable automated, data-driven fire risk assessment of historic buildings. The framework consists of two key phases: quantitative fire risk modeling and automated risk assessment. The risk modeling defines unified indicators in accordance with fire safety regulations to quantify the risks exposed to the heritage buildings internally and externally. Both inherent building characteristics (e.g. fire resistance rating) and external spatial characteristics (e.g. adjacent access and spatial separation) were assessed. Next, the automated assessment utilizes HBIM and GIS to extract building and surrounding area information, compute the quantitative risks, and develop an interactive visualization platform to facilitate stakeholders in decision-making. The feasibility of this framework is verified through a case study of Mandarin's House in Macau. The results indicated the framework is capable of quantifying the risk related to fire-resistant materials (0.66), external access (0.75) and separation (0.76). The results demonstrate that the proposed framework could contribute a unified fire risk model quantification method and a BIM and GIS-combined mechanism for automated risk assessment to support the proactive conservation of valuable cultural heritage assets.

Keywords -

Historic Building Information Modeling (HBIM); Geographic Information System (GIS); Fire Risk Assessment; Heritage Building

1 Introduction and Related Work

Heritage buildings carry rich historical and cultural values. However, they still face multiple risks of fire. The combustible building materials and lack of fire protection systems make them vulnerable to fire hazards [1]. In addition, electrical, heating and other systems in these historic structures can become ignition sources due to insufficient maintenance [2]. This renders heritage architecture especially susceptible to fire threats. For instance, the original Church of Mater Dei in Macao was destroyed by recurrent fires in history, leaving only ruins. The Zhengjiao Chanlin Buddhist Pavilion at A-Ma Temple also caught fire again after restoration. This indicates that even refurbished heritage buildings remain prone to fire risks [3]. Comprehensive fire risk assessments are one of the proactive measures to identify underlying fire hazards and evaluate the vulnerability of buildings to fire emergencies, and thus are imperative for fire prevention and heritage conservation.

Effective Fire risk assessment can help determine the fire risk level of buildings and provide guidance for fire protection measures and emergency plans. Fire risk assessment for heritage buildings is particularly challenging, as it involves various factors, such as the building materials, structures, occupancies, firefighting systems, surrounding environments, and access routes. Moreover, heritage buildings may have complex and irregular shapes and sometimes undertake restorations, making it difficult to collect accurate and up-to-date data for fire risk assessment.

Traditionally, fire risk assessment for heritage buildings is mainly based on qualitative methods such as questionnaire surveys, site visits, and expert judgments. These methods heavily rely on human efforts to collect and process the fire risk-related information of the building, such as the combustibility of the materials, the presence of ignition sources, the availability of fire protection systems, and the accessibility of fire vehicles. Then, according to relevant criteria and standards, the fire risk level of the building is calculated and classified. However, these methods have several limitations. Firstly, fire risk assessment is time- and labor-consuming, prone to errors and uncertainties, resulting in the inefficient acquisition of up-to-date information and inconsistent assessment results affected by assessors' experience. Secondly, there is a lack of comprehensive risk assessment models grounded in regulations and practices to quantify and evaluate fire risk factors for heritage buildings. Thirdly, assessment results are ambiguous, lacking data visualization and user-friendly interactive tools to support firefighting decision-making and planning.

Therefore, there is a need for a more efficient and reliable method for fire risk assessment of heritage buildings [4]. In recent years, with the development of digital technologies such as Building Information Modeling (BIM) and Geographic Information Systems (GIS), new methods have emerged to facilitate fire risk assessment of heritage buildings. BIM is a digital representation of the physical buildings that contain various attributes and parameters such as materials, structures, fire-resistant ratings and space functions, which can help analyze the building's fire hazard [5,6]. GIS is a data management and analysis system that incorporates various spatial information regarding buildings and road networks, such as locations, orientations, distances, and relations, offering rich geographic data of the external environment to support the analysis of the building's fire vulnerability [7,8].

By integrating BIM and GIS, both building internal and external information can be utilized to assess the fire risk of heritage buildings, further providing suggestions for fire prevention measures and emergency plans [9]. However, previous research on the integration of BIM and GIS has mainly focused on route planning and spatial visualization, while their usage for automated and quantitative fire risk assessment is still in its infancy. Many key factors, such as the intrinsic properties of the building and the surrounding terrain features, have not yet been systematically explored. Therefore, there are still many untapped opportunities for fire risk assessment combining BIM and GIS. Moreover, the mechanism of identifying relevant data in BIM and GIS and utilizing the data for risk quantification deserves further exploration in order to establish automated fire risk assessments for heritage buildings.

This paper proposes and validates an integrated historic BIM (HBIM) and GIS framework for automated fire risk assessment of heritage buildings, which can overcome the limitations of conventional methods by improving the efficiency of data acquisition and reducing manual efforts. The framework combines building inherent features from HBIM models and the external space characteristics from GIS to quantify the risk levels of heritage sites. The framework is applied to a case study of the Mandarin's House, a Chinese heritage building in Macao listed in the UNESCO World Heritage Site [3,10]. The results validate that the proposed framework can effectively quantify and visualize the fire risk levels in multifaceted aspects and provide data-driven analytics for fire prevention and emergency management of heritage buildings.

2 Overview of the Framework

The proposed framework of HBIM and GISintegrated automated fire risk assessment of heritage buildings consists of two phases: quantitative fire risk modeling and automated risk assessment. Figure 1 shows the overview of the framework.

Quantitative fire risk modeling aims to define and quantify the fire risk factors of heritage buildings based on the building parameters from HBIM models and the spatial data sets from GIS. The fire risk factors are divided into two categories: building inherent features and external space characteristics. For each category, a set of indicators is selected and weighted according to the relevant regulations and best practices in the field of fire safety and heritage conservation. Then, risk indices are established as numerical values to calculate and indicate the fire risk level of the heritage building.

The second phase is automated risk assessment, which analyzes and visualizes the fire risk of heritage buildings based on the risk indices. It consists of three steps: HBIM-based internal risk analysis, GIS-based external risk analysis, and automated risk assessment and visualization. The HBIM-based indoor risk analysis leverages HBIM models of heritage buildings to extract data of building components. GIS-based external risk analysis uses GIS data of heritage sites to identify and evaluate the spatial factors that influence the fire vulnerability of the buildings, such as the adjacent structures and firefighting parking routes. Last, automated risk assessment and visualization integrates the HBIM and GIS data and develops an interactive platform to visualize the 3D BIM models, 2D GIS base maps, meshes of neighbor structures, and routes along with the associated risk results. The visualization platform serves as a graphical interface for stakeholders to understand the potential dangers exposed to the heritage buildings and further facilitate fire prevention and emergency management.

3 Quantitative Fire Risk Modeling

In general, factors affecting heritage fire risks can be categorized into four groups, i.e., fire hazards, heritage building characteristics, surrounding environments, and fire safety management [11]. Specifically, heritage building characteristics, including



Figure 1. Overview of the framework

building inherent features, means of egress, portable firefighting devices, integrated fire service installations, and water supplies, are primarily related to the intrinsic attributes of heritages, and thus are highly related to HBIM modeling. Also, the surrounding environments cover the aspects of external space characteristics, road traffic, firefighting resources, external water supplies, and surrounding potential dangers, which fit well with the scope of GIS analysis. As an initial exploration, this study would focus on building inherent features and external space characteristics. To quantify the fire risks of these two factors, relevant regulations and codes in different countries and regions are reviewed and referenced [12–15].

3.1 Fire Risk of Building Inherent Feature

Building inherent features refers to the attributes of building elements that affect the fire spread and propagation. In this study, the concept of fire resistance rating (FRR) is employed as the indicator of building inherent features to measure the duration (in hours) of building elements for withstanding fires passively. According to [13], the FRR can be ranged as 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4 and 6 hours, in which a higher value of FRR represents a better fire-resistant performance. To comprehensively quantify the fire risks of entire buildings, the FRR of various building element classes, including walls, roofs, floors, doors, columns, windows, slabs, and stairs, would be obtained individually and then calculated aggregately to form a united fire risk index. For consistency, the study defines the fire risk index to represent the potential risks of the heritage building suffering from fires, with values ranging from 0 (lowest risk) to 1 (highest risk). Correspondingly, the relation between the fire risk index of different building elements and their FRRs is denoted as in Equation (1):

$$R_{ij} = 1 - FRR_{ij} / FRR_{max} \tag{1}$$

where R_{ij} is the fire risk index of the *i*-th building element in the *j*-th class in terms of building inherent features, FRR_{ij} is the fire-resistance rating (unit: hour) of the building element, and FRR_{max} is the maximum regulated FRR (i.e., 6 hours in [13]).

Next, to holistically assess the fire risks of each building element class and the entire heritage building, the weighted fire risk index is introduced by multiplying the individual fire risk index with the volume of the building element, as shown in Equation (2). The volume of building elements is considered as it can reflect the utilization ratio of elements with different FRRs to approximate their fire-resistant performance to the whole building.

$$\begin{cases} V_j = \sum V_{ij} \\ R_j = (\sum R_{ij} \times V_{ij})/V_j \\ R_{BIF} = (\sum R_j \times V_j)/\sum V_j \end{cases}$$
(2)

where V_{ij} is the volume of the *i*-th building element in the *j*-th class, V_j and R_j is the aggregated volume and fire risk index of the j-th building element class, R_{BIF} is the holistic fire risk index in terms of building inherent features.

Consequently, through decomposition and weighed aggregation of building elements, R_{BIF} can be utilized to represent the fire-resistance capabilities of the holistic heritage buildings.

3.2 Fire Risk of External Space Characteristics

External space characteristics refer to the spatial conditions that determine the exposure and isolation of heritage buildings and affect the spread and severity of fires, which include proximal access and spatial separation. Proximal access is considered as it affects how the way fire trucks can be accessed and parked near the fire-affected heritage buildings. Spatial separation evaluates whether the heritage building has sufficient fire separation distance to its adjoining buildings to prevent fire propagation across buildings.

To quantify the risks of the proximal access, the quadrilateral sides of heritage buildings would be first investigated. More specifically, according to [14], if the side is directly next to a road with a width no less than the minimum width for emergency vehicular access (i.e., 4 meters in [13]), it is suitable for fire response operation. In this manner, all the quadrilateral sides would be examined to verify their suitability. In case all the quadrilateral sides are not suitable, [15] also regulated that the alternative roads within a closely accessible area (i.e., 30 meters in [14]) serve as a secondary option to assist in emergency operations.

Correspondingly, this study introduces the fire risk index of proximal access to comprehensively take the quadrilateral sides and the closely accessible area into account. As indicated in Equations (3) and (4), the fire risk index is related to the number of suitable quadrilateral sides when there exists at least one suitable side (i.e., $\exists w_s \ge W_{EVA}$), otherwise, it would be determined by the suitability of the roads in the closely accessible area (i.e., either $\exists w_r \ge W_{EVA}$ or $\forall w_r < W_{EVA}$). Hence, the values of the fire risk index of proximal access could be 0 (lowest risk), 0.25, 0.5, 0.75, 0.875, and 1 (highest risk).

$$R_{s} = \begin{cases} 0, & \text{when } w_{s} \ge W_{EVA} \\ 1, & \text{when } w_{s} < W_{EVA} \end{cases}$$
(3)

$$R_{pa} = \begin{cases} \sum R_s/4, \text{ when } \exists w_s \ge W_{EVA} \\ 0.875, \text{ when } \forall w_s < W_{EVA} \& \exists w_r \ge W_{EVA} \\ 1, \text{ when } \forall \{w_s, w_r\} < W_{EVA} \end{cases}$$
(4)

where R_s is the fire risk index of an individual quadrilateral side of a heritage building, w_s and w_r are the widths of the quadrilateral side and the roads in the accessible area, W_{EVA} is the regulated minimum road width for emergency vehicular access, and R_{pa} is the fire risk index of proximal access.

For spatial separation, the height and the minimum distance of the surrounding buildings directly adjacent to the heritage building is considered. According to [13], the regulated minimum fire separation distance varies. It depends on whether either the heritage building or its adjoining buildings are high-rise. This is detailed in Equation (5):

$$D_{min,i} = \begin{cases} D_{min,H}, & \text{when } \exists \{H_h, H_{a,i}\} > H_{hr} \\ D_{min,L}, & \text{when } \forall \{H_h, H_{a,i}\} \le H_{hr} \end{cases}$$
(5)

where H_h and $H_{a,i}$ are the building height of the heritage and the *i*-th adjoining building, H_{hr} is the regulated height in the definition of high-rise buildings (i.e., 27 meters for residential buildings in [11]). $D_{min,H}$ and $D_{min,L}$ are the regulated minimum fire separation distances when at least one high-rise building exists and when no high-rise building exists (in [13], $D_{min,H}$ is 9 meters and $D_{min,L}$ is 6 meters for conservative calculation). $D_{min,i}$ is then calculated as the minimum fire separation distance between the heritage and the *i*-th adjoining building.

Correspondingly, the fire risk index of spatial separation of each adjoining building is introduced by considering whether the minimum fire separation distance is satisfied or not, as shown in Equation (6). The holistic fire risk index of spatial separation of the heritage building is eventually calculated by averaging the results of all the adjoining buildings, as indicated in Equation (7):

$$R_{ss,i} = \begin{cases} 0, & \text{when } D_i \ge D_{min,i} \\ 1 - D_i / D_{min,i}, \text{ when } D_i < D_{min,i} \\ R_{ss} = \sum R_{ss,i} / N \end{cases}$$
(6)

where D_i is the actual fire separation distance between the heritage and the *i*-th adjoining building, $R_{ss,i}$ is the fire risk index associated with the adjoining building, Nis the total number of adjoining buildings, R_{ss} is the holistic fire risk index of spatial separation.

Eventually, the total risk of external space characteristics (R_{ESC}) is then calculated as a weighted combination of the risks of the proximal access and the spatial separation using a coefficient φ , as shown in Equation (8). Theoretically, the coefficient φ represents the importance ratio between the proximal access and the spatial separation. However, no quantitative evidence was found in the existing literature for comparing their importance, which in turn suggests a further investigation with firefighting experts to determine the specific values suitable for the local countries or regions. For simplicity, this study adopts $\varphi = 0.5$, assuming that proximal access and spatial separation are equivalently important. As a result, R_{ESC} represents a unified fire risk index that comprehensively considers the potential impacts of proximal access and spatial separation on the heritage building in the surrounding environment.

$$R_{ESC} = \varphi R_{pa} + (1 - \varphi) R_{ss} \tag{8}$$

4 Automated Risk Assessment Using HBIM and GIS

4.1 HBIM-based Internal Risk Analysis

1

After modeling the fire risk of the building's inherent feature, relevant building data such as the fire-resistant

rating and volume of various building elements needs to be obtained for risk analysis. In this study, the HBIM of heritage buildings is utilized as a digital tool to access and manage fire risk-related information. More specifically, Industry Foundation Classes (IFC) [16] is selected for data exchange as it is a standardized opensource data schema widely used in the architecture, engineering, construction, and operations industry. To acquire the fire-resistant rating, the property 'FireRating' in the property set 'Pset FireRating' defined in the IFC 4 schema is extracted [16]. In addition, volumes of various building elements are obtained individually using the property 'Volume' in the property set 'Pset BuildingElementPhysical'. To enable automated data extraction, self-developed scripts are compiled using IfcOpenShell and Pandas libraries [17,18] to iterate all the building elements and retrieve their fireresistant rating and volumes, serving the preprocessed data for further risk quantification.

4.2 GIS-based External Risk Analysis

For the risks of external spatial characteristics, this study adopts OpenStreetMap (OSM) [19] for data acquisition and ArcGIS software [20] for data processing and visualization. OSM contains the critical spatial data regarding adjacent buildings and roads of the heritage building, such as the width of roads, the height and footprint of buildings, and the longitudinal and latitudinal coordinates of common spatial features. Accordingly, the OSM data of the heritage building region is first collected using the Overpass Turbo wizard [21], and then relevant attributes are extracted and compiled into tables. The OSM data and tables are further imported into ArcGIS Pro for spatial analysis and visualization. To calculate the risks of proximal access, the width of roads can be directly retrieved from the OSM data, i.e. 'width' of the 'highway' objects in OSM. Regarding the calculation of spatial separation, the height of the heritage building and adjoining buildings can be also obtained from the OSM data ('height' or 'levels' of the 'building' objects in OSM), and the fire separation distance can be determined by utilizing the geometric boundary of the buildings. Specifically, this study adopts the built-in function named Near in ArcGIS to measure the shortest distance between the heritage and its adjoining building. As a result, all the input data required by the quantitative risk models (i.e., Equations (1) - (8)) can be efficiently collected using HBIM models and GIS systems to support further automated risk assessment.

4.3 Automated Risk Assessment and Visualization

The HBIM model and GIS data are integrated into a

common environment in ArcGIS Pro. To align the HBIM model to its geographical location, built-in functions (Join and Relate) are used to connect the 3D IFC model to the 2D base map. Furthermore, layers regarding the roads and 3D meshes of the adjoining buildings are created by employing the OSM data. To enable automated risk assessment, custom Python scripts are created to access the HBIM and GIS data and calculate the risk index of the building's inherent feature R_{BIF} and the external spatial characteristics R_{ESC} . For visualization, the results are further associated with the HBIM model, the GIS basemaps, the road layers, and the 3D adjoining building meshes, in which users can query the attributes and the results by interacting with the graphical representation. Different colors are set to display the severity of fire risks for more intuitive visualization. Eventually, the analytical HBIM and GIScombined scene are published via ArcGIS Online to allow users to access results and support them in risk assessment and subsequent decision-making.

5 Case Study

5.1 Application to heritage building

To validate the feasibility of the proposed framework, a case study was conducted on a heritage building in Macao. The Mandarin's House is selected, which is a typical Lingnan-style courtyard house built around 1881 [9], with traditional Chinese brick and timber architectural features and intricate decorative details.

A LOD 200 HBIM model of Mandarin's House was created in Revit [19] (as shown in Figure 2). The model included basic architectural elements like walls, columns, doors, windows, and roofs. The HBIM model was exported as an IFC data model, which was further imported into ArcGIS Pro. The OSM data of the surrounding area was also collected and imported into the same geographical scene. Python scripts were developed to extract the relevant building components and properties from the IFC data model and the OSM spatial data. The scripts also computed the fire risk index based on the proposed equations. Specifically, in the calculation of building inherent features, 582 building elements in total were processed to retrieve their FRRs and calculated their elemental fire risk index, e.g., 236 walls (fire risk index: 0.58), 123 doors (0.84), 116 windows (0.92). Regarding the external spatial characteristics, the boundary of the heritage building was simplified as a quadrilateral polygon, in which the width of the four sides was extracted to check if they satisfied the requirements of the minimum width for emergency vehicular access or not. Also, 21 roads within the 30-meter area were also obtained as the secondary options for alternative proximal access. In addition, the fire risk index of the 16 adjoining buildings was investigated to calculate their fire risk indices according to their heights and fire separation distances. To visualize the results, an analytical scene was published into a web scene viewer via ArcGIS Online.



Figure 2. The ground floor of Mandarin's House

5.2 Results and Discussion

The case study demonstrated the application of the proposed framework for a fundamental fire risk assessment utilizing BIM and GIS methodologies. Tables 1 to 3 summarize the quantitative results of the fire risk assessment for both building inherent features and external space characteristics. The results indicate that the fire risk index of the inherent features of Mandarin's House is 0.66, representing a moderate fire risk level. In addition, the fire risk index of diverse building element classes and their volumes are presented in Table 1. The holistic result revealed that walls and slabs notably impacted the overall fire risk, given their substantial volumetric ratios.

Table 1. Fire risk results for building inherent features

Building/	Element	Volumetric	Fire risk
element class	volume (m ³)	ratio	index
1: Column	6.97	0.13%	0.75
2: Door	109.06	2.11%	0.84
3: Ramp	7.56	0.15%	0.75
4: Roof	272.06	5.27%	0.83
5: Slab	1506.62	29.16%	0.78
6: Stair	2.17	0.04%	0.75
7: Wall	3164.26	61.24%	0.58
8: Window	98.03	1.90%	0.92
Mandarin's	5166.74	100.00%	0.66
House	• • • • •		

Table 2. Fire risk results of spatial separation

Building name	Building height (m)	Actual distance (m)	Mini. fire separation distance	Fire risk index
1. Wo Fong Building	27.00	0.00	<u>(III)</u> 6.00	1.00
2: Lei Heng Building	34.40	0.00	9.00	1.00
3: King Yip Building	32.40	0.00	9.00	1.00
4: No.12-16, Barra Street	32.40	0.00	9.00	1.00
5: No.12, António da	23.20	0.00	6.00	1.00

Silva Crossing				
6: Wan Seng Building	20.40	0.00	6.00	1.00
7: Weng On Building	20.00	0.69	6.00	0.89
8: Kun Seng Pavilion	35.70	0.74	9.00	0.92
9: Weng Leng Building	20.40	1.64	6.00	0.73
10: Wang Fu	31.00	3 00	0.00	0.67
Building	51.90	5.00	9.00	0.07
11: Kou Fu Building	31.40	3.00	9.00	0.67
12: No.2, Barra Street	23.20	3.50	6.00	0.42
13: Hou Kio Building	34.00	4.00	9.00	0.56
14: No.5A - 5E, Barra Street	34.00	4.00	9.00	0.56
15: Heng Vo Building	37.40	4.00	9.00	0.56
16: No.4, Barra Street	23.20	4.50	6.00	0.25
Mandarin's House	11.00	N/A	N/A	0.76

Road Name	Road width (m)	W _{EVA} (m)	Fire risk index
Side 1: Barra Street	4.00	4.00	0.00
Side 2: António da Silva Crossing	3.00	4.00	1.00
Side 3: no road	0.00	4.00	1.00
Side 4: no road	0.00	4.00	1.00
Mandarin's House	N/A	4.00	0.75

Next, the fire risk index of the external space characteristics is 0.76 (proximal access: 0.75, spatial separation: 0.76), indicating a high fire risk level. Specifically, the calculation of proximal access showcased that two out of four sides had no roads next to the heritage site, and among the remaining two sides, only one side satisfied the requirement of road width for emergency vehicular access. Given that there existed a suitable side next to Mandarin's House, this study did not further consider the alternative roads within the 30meter accessible area. Regarding the spatial separation, half of the adjacent buildings are less than one meter from Mandarin's House, implicating very high risks of fire propagation among these buildings and the heritage site. From another perspective, it is critically challenging to protect the heritage site from fire emergencies in such a high-density city like Macao.

The results of this case study provided a quantitative assessment of the fire risk for Mandarin's House in terms of building inherent features and external spatial characteristics. The results identify the most vulnerable and critical risks of the heritage building, thereby assisting heritage conservation and fire protection parties in prioritizing fire prevention and mitigation measures. For instance, in Mandarin's House, the window elements get a high level of risk (0.92), which suggests that further proactive measures are required to protect these vulnerable assets. In addition, from the perspective of external characteristics, the adjacent buildings that are extremely close to Mandarin's House are all distributed in the northern and eastern regions of the heritage, implicating that additional fire barriers could be strengthened correspondingly (Figure 3). Hence, the proposed framework serves as a data-driven approach to evaluate the internal and external risks of heritage building in an efficient, automated, and quantitative manner. The proposed framework and the HBIM and GIS-combined visualization platform can be potentially expanded to other heritage buildings and further help stakeholders govern and manage the fire risks of historic centers worldwide.

Limitations also exist in quantitative risk modeling and automated risk assessment. First, the scope of this study only focused on building inherent features and external space characteristics, while many other factors such as ignition sources, emergency exits, fire extinguishers, smoke detection systems, firefighting resources, and road traffic have not been systematically investigated. Site visits and interviews with stakeholders could reveal additional vulnerabilities not captured in the models. Second, the risk models had made certain assumptions and simplifications, and the critical parameters are referenced from regional codes and guidelines. Hence, the risk models need to be carefully interpreted to accommodate the relevant requirements when applying them to other countries or regions. In addition, surveys or questionnaires with stakeholders may provide professional judgement to rationalize the risk formulation based on local conditions. Last, it was found that data from HBIM models and GIS systems was occasionally inaccurate or missing, which would adversely influence the reliability of the quantitative risk results. This highlights the necessity of developing solutions for the semi-automated acquisition of geometric data of buildings to guarantee reliable inputs that do not overly depend on the modeler's manual

inputs. Therefore, future work should construct more standardized and automated data collection updating mechanisms (e.g., reality capture and aerial survey techniques) to improve the reliability of the result. In addition, more heritage buildings with different types (e.g., churches and palaces) should be considered in future studies to verify the generalizability of the proposed framework.

6 Conclusions

This paper proposes and validates a framework that establishes quantitative fire risk models and integrates HBIM and GIS techniques to develop automated fire risk assessments of heritage buildings. Specifically, the fire risk models cover both internal and external factors, including building fire resistance, proximal access for fire trucks, and spatial separation against fire propagation. By referencing fire safety regulations and practical guidelines, the quantitative modeling of fire risks can effectively provide clear and unified indices to help stakeholders better understand the diverse threats facing heritage sites. Therefore, the fire risk models can form a fundamental instrument to support stakeholders in prioritizing their tasks and allocating resources to mitigate the most crucial risks in a targeted and informed manner.

In addition, an HBIM and GIS-combined mechanism is developed in this study to enable automatic data extraction, risk analysis, and interactive visualization. The information requirements of risk quantification are identified and matched with BIM and GIS data sources, where critical information is extracted programmatically to support the subsequent risk computation. Eventually, an interactive visualization platform is developed to



Figure 3. Visualization of spatial fire risk patterns

present the fire risks in different colors to help recognize their spatial distributions.

Through the case study of Mandarin's House, the proposed framework is validated to quantify the multifaceted risks exposed to the heritage site efficiently and effectively. Furthermore, the result discloses that the spatial separation aspect gained a high level of risk and thus deserves more attention for proactive mitigation. Overall, the proposed framework enables data-driven analytics to reveal spatial patterns of fire risk in heritage buildings, thereby forming a solution to facilitate fire prevention and emergency management.

Future work would further investigate other fire riskrelated factors of heritage buildings and incorporate various data collection methods to enhance the reliability of the results.

Acknowledgements

This work was supported by the Research Services and Knowledge Transfer Office of the University of Macau, China (Project No. MYRG2022-00186-FST).

References

- [1] Salazar, L., Romão, X. and Paupério, E. Review of vulnerability indicators for fire risk assessment in cultural heritage. *International Journal of Disaster Risk Reduction*, *60*, 102286, 2021.
- [2] Yuan, C., He, Y., Feng, Y. and Wang, P. Fire hazards in heritage villages: A case study on Dangjia Village in China. *International Journal of Disaster Risk Reduction*, 28, 748–757, 2018.
- [3] Macau Cultural Affairs Bureau. Macao's Historic Buildings—Overview of the World Cultural Heritage Application and Project Navigation. In Macau Cultural Affairs Bureau (Ed.), International Academic Seminar of Protection of Urban Cultural Heritage - Macao Vision, pages 65–120, Review of Culture, 2003.
- [4] Wang, J., You, H., Qi, X. and Yang, N. BIM-based structural health monitoring and early warning for heritage timber structures. *Automation in Construction*, 144(104618), 2022.
- [5] Zou, Y., Kiviniemi, A. and Jones, S. A review of risk management through BIM and BIM-related technologies. *Safety Science*, 97: 88–98, 2017.
- [6] Wong, M. and Lee S. IFC-based information exchange for multi-agency response to indoor fire emergencies. *Automation in Construction*, 144(104623), 2022.
- [7] Shi, J. and Liu, P. An Agent-Based Evacuation Model to Support Fire Safety Design Based on an Integrated 3D GIS and BIM Platform. *Computing in Civil and Building Engineering*, 1893–1900, 2014.
- [8] Danielová, M., Kumke, H. and Peters, S. 3D

Reconstruction and Uncertainty Modelling Using Fuzzy Logic of Archaeological Structures: Applied to the Temple of Diana in Nemi, Italy. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 51(3), 137–146, 2016.

- [9] Meschini, S., Accardo, D., Locatelli, M., Pellegrini, L., Tagliabue, L. and Di Giuda, G. BIM-GIS integration and crowd simulation for fire emergency management in a large diffused university, 2023.
- [10] Chen, Y. Research on the Development of the Style of Lingnan Homestead in Macau — Taking Mandarin's House as an Example. In Proceedings of the International Conference on Architecture: Heritage, Traditions and Innovations, 2019.
- [11] Qiao Y., Lam C. and Wong M. A GIS-based Framework of Fire Risk Assessment and Emergency Route Planning for Heritage Buildings. In International Symposium on Advancement of Construction Management and Real Estate, 2023.
- [12] *GB50352-2019 Uniform Standard for Design of Civil Buildings.* Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2019.
- [13] *REGADM39-2022 Macao Technical Regulations for Fire Safety in Buildings and Premises.* Macao Land and Urban Construction Bureau, 2022.
- [14] JGJ/Tx-2023 Standard for Fire Protection Design of Historic Conservation Areas and Historic Buildings (Draft for Soliciting Opinions). Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2023.
- [15] GB152-2017 Urban Fire Station Construction Standard. Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2017.
- [16] Industry Foundation Classes (IFC). BuildingSMART International Ltd. https://www.buildingsmart.org/standards/bsistandards/industry-foundation-classes/, Accessed: 23/12/2023.
- [17] *IfcOpenShell The open source IFC toolkit and geometry engine*. OS Arch Community. On-line: https://ifcopenshell.org/, Accessed: 23/12/2023.
- [18] Wes McKinney. pandas Python Data Analysis Library. OVHcloud. On-line: https://pandas.pydata.org/, Accessed: 23/12/2023.
- [19] Steve Coast. OpenStreetMap (OSM). OpenStreetMap Foundation. On-line: https://www.arcgis.com/apps/View/index.html?appi d=df7cee38677f479c8697026ebf920431, Accessed: 23/12/2023.
- [20] ArcGIS Mapping and Analytics Software and Services, Esri. On-line: https://www.esri.com/enus/arcgis/about-arcgis/overview, Accessed: 23/12/2023.
- [21] *overpass turbo*. OpenStreetMap. On-line: https://overpass-turbo.eu/, Accessed: 23/12/2023.

3D reconstruction of a bridge with concrete damage classification using deep learning

Christopher Núñez^a, Marck Regalado^a, Luis Huaypar^a, Antonio Bedón^a, Jordan Romaní^a, Oscar Solís^a, Karol Agreda^a and Micaela Cárdenas^b

^aFaculty of Civil Engineering, National University of Engineering, Perú

^bFaculty of Civil Engineering, National University of Trujillo, Perú

E-mail: christopher.nunez.v@uni.pe, marck.regalado.e@uni.pe, luis.huaypar.a@uni.pe, antonio.bedon.r@uni.pe, jordan.romani.c@uni.pe, oscar.solis.g@uni.pe, karol.agreda.e@uni.pe and t534000120@unitru.edu.pe

Abstract

The classification of concrete damage in bridges poses challenges, characterized by time-consuming, hazardous, and often subjective inspection methods. Recognizing the need for efficient damage identification and the creation of 3D models for maintenance purposes, this paper introduces an innovative approach to the inspection of reinforced concrete bridges. The proposed methodology involves 3D reconstruction of a bridge, coupled with a concrete damage classification system based on severity. Notably, the analysis ensures objectivity through the implementation of deep learning for classifying concrete damage in UAV-captured images. A noteworthy aspect of this research is that, in the training models, a precision of over 90% is achieved for each type of concrete damage. This methodology serves as a valuable contribution to automating and streamlining concrete bridge inspections, aiming to reduce costs and enhance efficiency throughout its life cycle.

Keywords –

UAV; Bridge; Concrete damage; SHM; CNN

1 Introduction

In contemporary society, bridges hold significant environmental significance by establishing vital connections between various localities. They play a crucial role in fostering economic and cultural development, among other societal factors [1-3]. Therefore, preserving their structural integrity, safety, and functionality is paramount throughout their lifespan, especially for emergency situations such as resource transportation and rescue operations [3,4].

Currently, numerous countries, particularly those in Europe and North America, confront the deterioration of an extensive infrastructure inventory that has surpassed its originally planned service life [1]. In the United States, the report states that 7.5% of bridges are categorized as structurally deficient, predominantly falling below the established standards [5]. In fact, a minimum of one-third of the over 600,000 bridges in the United States feature a concrete superstructure or wearing surface [6]. In recent years, notable bridge collapses attributed to various factors have occurred, including the Morandi cable-stayed bridge in Genoa, Italy (2018), the Florida International University Pedestrian Bridge (2018), and the Nanfang'ao steel single-arch bridge in Taiwan (2019) [4]. Taiwan, with over 28,000 bridges, faces challenges due to natural disasters like earthquakes, typhoons, and rainstorms, along with its unique geographic features, such as mountainous and stream-filled regions. Over the past decades, several bridge collapses have been attributed to various scour issues, including general scour and local scour [3].

Concrete structures are increasingly experiencing deterioration due to various factors such as aging, increased traffic loads, loadings from earthquakes and extreme weather conditions (wind, ambient vibrations) [4,5,7–9], such deterioration is usually caused by inadequate or untimely maintenance [1] Similarly, the prevalent structural configuration employed for bridges has been the multispan design with simply supported Reinforced Concrete (RC) girders. These girders are interconnected by transverse beams and are supported by piers and abutments at the ends, facilitating construction phases [1]. Consequently, while the emergence of cracks in concrete bridges is a crucial indicator of structural performance, it doesn't necessarily signify imminent collapse; however, it frequently results in diminished structural integrity [8,10]. Additionally, concrete damaged impact various aspects of structures, including bearing capacity, stiffness, energy absorption capacity, and resistance to reinforcement corrosion [8,9,11]. Therefore, timely detection and measurement of damaged elements are crucial for making informed decisions regarding necessary repairs and maintenance [8]. However, conventional inspection methods often prove laborious, time-consuming, and capital-intensive. Especially, in the case of large span bridges, traditional methods are not effective for rapid full-field monitoring and hence a radical monitoring approach is most needed [12]. During recent decades, ensuring life safety and the need to reduce inspection costs have emerged as the top priorities for practicing engineers and researchers [5].

Compared to the traditional contact methods, most of the noncontact Structural Health Monitoring (SHM) methods have improvement in the convenience and efficiency of structural inspection and concrete damage [9]. SHM is a data processing approach that employs technology to offer early signals of disruption and the progression of damages and deterioration to avert potentially hazardous results to a specific structure, which is useful for enhanced decision-making [1,7]. The majority of existing SHM system consists of various sensors and accelerometers [7,13,14].

Consequently, an increasingly number of SHM applications with noncontact means have been developed and applied for the monitoring and inspection of concrete damage in a variety of critical concrete structures [9]. The deployment of Unmanned Aerial Vehicles (UAV) for civil infrastructure monitoring is a relatively recent development, with only a limited number of practical case studies conducted for industries, monuments, and other civil structures. Typically, UAVs are equipped with an image acquisition system, and the captured data are manually processed [1,7,12]. While this approach is convenient, it involves labor-intensive efforts in quantifying and analyzing the acquired data [12].

In this paper, we propose a methodology for the classification of damage in the concrete components of a bridge, employing Deep Learning (DL) based method for image processing. DL serves as an automated solution for predicting and classifying data, with the Convolutional Neural Network (CNN) emerging as a prevalent and straightforward method, particularly wellsuited for image classification tasks. CNN excels in processing and categorizing information presented in image formats, making it a widely adopted approach in the realm of deep learning [15]. The images utilized in this analysis were captured by an Unmanned Aerial Vehicle (UAV). This approach leverages advanced algorithms to categorize and analyze the identified damaged elements, contributing to a nuanced understanding of structural integrity. The integration of UAV technology and CNN method in the proposed methodology offers a sophisticated and efficient means of assessing the condition of concrete elements in bridge infrastructure. This method extracts features of different abstract levels and maps raw pixel intensities of the crack patch into a feature vector through several fully connected layers. All convolutional filter kernel elements are trained from the data in a supervised fashion, learning from the labeled set of examples. This approach not only leverages advanced aerial imaging capabilities but also employs CNN to automatically extract and analyze features from the captured images, enhancing the precision of structural assessment in the

context of concrete damage classification.

The paper is organized as follows: In Section 2, we delve into the Background and Related Studies. Section 3 provides the methodology for obtaining the 3D reconstructed case study with the classification of the damaged concrete using deep learning. The findings of our study are presented in Section 4, which covers the Results. Finally, Section 5 presents the scientific contribution and conclusions of this paper.

2 Background and Related studies

2.1 Concrete Damage classification

During the last three decades, there has been notable expansion in the utilization of high-strength concrete applications in bridge construction [16]. The five most prevalent damage, as outlined in Table 1, encompass cracks, corrosion, efflorescence, spalling, and exposed steel reinforcement [17]. To establish a ranking system for the magnitude of primary damages in reinforced concrete, reference is made to Hüthwohl et al. [18] and Highways England [19] particularly its document "CS 450 Inspection of Highway Structures." This is undertaken with the objective of formulating a model to categorize concrete damages based on their severity.

Hence, we have identified three magnitudes for each concrete damage, serving as benchmarks to evaluate bridge inspection damage classification methods: (1) Not Found (No color), (2) Moderate Damage (Orange color), and (3) Critical Damage (Red color). In instances where multiple damage types coexist, the color corresponding to the highest magnitude will be applied. As such, in the present article, neither the location of the damage nor the combination of types of damage in concrete were taken into consideration when calculating the magnitude. The focus was solely on the damage itself in a 288x288-pixel image. Additionally, due to the variability in image quality, it hinders the analysis of combinations of damage in concrete.

2.2 Relevant studies

The important aspect of the research presented in this paper lies in the utilization of computer vision to identify defects and damages, thereby establishing an objective classification process.

Domogo		Damage Scale (Color reference)	
Damage	No damage (No color)	Moderate Damage (Orange)	Major Damage (Red)
Corrosion	No signs of corrosion attack	Moderate corrosion attack	Major corrosion attack
Craak	No signs of cracks or difficult	Cracks less than 1 mm	Cracks more than 1 mm (easily
Сгаск	to detect visually	(difficult to detect visually)	visible)
Efflorescence	No signs of efflorescence attack	Moderate efflorescence attack	Major efflorescence attack
Exposed Bars	No exposed bars	Moderately exposed bars	Fully exposed bars
Spallation	No spalls	Minor deep spalls exposing	Collapsed

Table 1 Damage scale for reinforced concrete.

Mansuri & Patel [20] devised an automated visual inspection system for defect detection in heritage structures, leveraging artificial intelligence through an R-CNN (Faster Region-based Convolutional Neural Network) object detection model. The inspection accuracy of this model demonstrated optimal detection precision, reaching 91.58%, particularly in identifying three damage types: "spalling," "exposed bricks," and "cracks."

Zhao et al. [21] conducted a three-dimensional reconstruction based on images captured by unmanned aerial vehicles for the monitoring and inspection of dams, focusing on the identification of damages in the obtained images. Subsequently, non-contact optical measurements were performed for disaster prevention.

Wang et al. [1] suggested a UAV-based method to promptly evaluate seismic risk in bridges. Their methodology encompassed the acquisition of aerial photogrammetric data and the automated extraction of geometric features, subsequently integrated into structural models to assess seismic risk in relation to capacity-demand. The feasibility of their approach was substantiated through a case study conducted on an Italian bridge, thus contributing valuable insights to the field of seismic risk assessment for infrastructure.

Nappo et al. [22] proposed the utilization of Unmanned Aerial Vehicles (UAVs) for the semiautomatic detection and classification of damages in asphalt-paved roads affected by landslides. Leveraging 3D models and 2D images derived through UAV-based photogrammetry, the approach aimed to overcome the limitations associated with traditional visual inspections. The developed semi-automatic procedure quantitatively identified and classified longitudinal and transverse cracks in the pavement, presenting a swift, systematic, and objective alternative to conventional field surveys. Applied in the Province of Como, Northern Italy, the results underscored the methodology's utility for road management, providing maps of damage hotspots, pavement damage detectors, criteria based on the International Roughness Index (IRI), and road damage severity maps.

3 Methods and Implementations

3.1 Selected UAV

The aerial survey was performed using a commercial quadrotor, Phantom 4 V2. (DJI, China). This portable UAV has a built-in GPS that is used for way-point navigation and dataset geotagging. The drone was deployed for visual inspections of various structural components across four reinforced concrete bridges. Figure 1 illustrates the implementation of the proposed methodology outlined in this paper for bridge inspection utilizing a drone. The specifications of both the vehicle and the camera, outlined in Table 2, play a crucial role in the image processing carried out by the deep learning model.

Table 2 DJI Phantom 4 V2. technical specification

Unmanned Aerial		DJI Phantom 4 V2
Vehicle (U	JAV)	
Dimension	(mm)	289.5x289.5x196
Weight (kg)	1.375
Max flight	time	Approx. 30 minutes
Hovering flying accuracy (with GPS and vision system)		Vertical: ± 0.1 m Horizontal: ± 0.3 m
Max. speed (mph)		45
Photo resolution		5472x3648 pixels
Camera	Sensor Pixels	1" CMOS 20 Mega
Remote range (km)		8



Fig. 1. UAV used for the data acquisition

3.2 Data acquisition

The task of data acquisition includes (i) site prechecking of the bridge and surrounding area, (ii) flight plan drafting, and (iii) on-site data collection. Firstly, an examination of the area should be conducted to consider multiple factors such as the complexity of the surroundings and the visibility of the structure (e.g., presence of obstacles/vegetation around the structure), the accessibility of the area to identify suitable take-off and landing points, and flight restrictions according to local regulations (e.g., the presence of no-fly zones, the highest permissible altitude for flying, etc.). Then, the planning of the flight path exerts the most significant influence on data quality, as it pertains to factors such as lighting conditions, camera angles, offset distances, the flight pattern chosen, and the level of overlap between images [1]. Additionally, to explain the relationship between camera angle and distance, the terminology ground sampling distance (GSD) is referred to the ratio between the measure of an actual object to a pixel size [1,23,24], which is used here to describe the image quality.

Our work is motivated by a project aiming at the damage classification in concrete bridge elements within a millimeter (1 mm) of accuracy, which is considered adequate following previous research such as Chen et al. [23] who show a relationship between GSD and Working Distance (WD) for the DJI phantom 4 UAV with respect to the tilt angle and offset distance.

The overlap (OV) represents the percentage of the object covered in two consecutive frames, depending on the distance to the center of two consecutive photos. The value of OV is one of the parameters for determining the fly path in the vertical and horizontal directions. Even though overlapping rates are seldom documented and seem to be chosen based on empirical observations, in Chen et al. [23] it is recommended 60% \pm 5% for endlap and 30% \pm 15% for sidelap. In this work, a 60% OV vertical and 30% OV horizontal direction should be considered in the to ensure the survey quality.

After establishing the Ground Sample Distance (GSD), overlap (OV), an appropriate working distance (WD), and tilt angle that align with the research objectives for image acquisition, we then move forward to elucidate the procedure for capturing imagery of the structural elements to create 3D models, as detailed in the subsequent subsection.

3.2.1 Super-structure

Linear trajectories parallel to the longitudinal direction of the bridge and outer beams were employed, adjusting the inclination angle to ensure comprehensive coverage of the structural beam (See Fig. 2).



Fig. 2. Image acquisition process of a concrete bridge structural beam.

3.2.2 Sub-structure

In the case of the pillars (see Fig. 3), linear trajectories are executed, aligning with the lateral elevation profile of the bridge at varying heights and employing different camera pitch and yaw angles (including upward orientation for mapping beneath the deck). Where deemed safe, additional spiral (or point of interest) flights along the height of each pillar are conducted to ensure comprehensive coverage of all surfaces. As for the abutments (see Fig. 4), a parallel survey will be conducted along the primary faces, adjusting both the height and inclination angles for a thorough inspection.



Fig. 3. Image acquisition process of the structural column of a concrete bridge.



Fig. 4. Image acquisition process of concrete bridge abutments.

3.3 Imagen processing algorithms

The images were captured at a resolution of 5472x3648 pixels from a dataset comprising four distinct bridges. The selection of these bridges was based on a non-probability sampling approach, considering their proximity to the designated study area, The training dataset incorporates images from three of these bridges, totaling 848 images. Additionally, to enhance the performance of our training model, we have incorporated another dataset provided by Hüthwohl et al. [18]. As part of the validation dataset, we have selected the Iniche bridge (refer to Fig. 5) as a case study, yielding a total of 142 images.



Fig. 5 Case study: Iniche bridge. Chiclayo, Perú.

This implies that a higher number of neurons require more parameters to be learned, increasing both the learning time and complexity of the model. To mitigate the size of the CNN model, the original image is resized to a smaller image, specifically 228x228 pixels. Subsequently, 384 images, each measuring 228x228 pixels, were generated for every original image of dimensions 5472x3648 pixels. Regarding the image quality, it is important that the images have a good resolution, such as HD or 4K, since the reduced image used to analyze bridge damage (228 x 228 pixels) may present noise that generates incorrect image treatment in the model. Additionally, the normalization of the pixels should be between 0 and 1, following the image binarization criterion based on Elizondo et al. [25].

3.3.1 Data Preparation

In this paper, we employ a deep convolutional neural network based in Zhang et al., [26]. In the data preparation phase, we implemented a code snippet using the TensorFlow Keras API. This code segment focuses on organizing and structuring the input data for model training. Subsequently, the images are reshaped into the appropriate format for TensorFlow, ensuring compatibility by adjusting the dimensions to (number of images, width, height, channels). The reshaped data is then stored in an HDF5 file format using the h5py library, providing an efficient and compressed representation of the dataset.

3.3.2 Data Training

In the data training phase, we utilized the TensorFlow Keras API to construct a Convolutional Neural Network (CNN) architecture for the classification of concrete damage. The input data, stored in an HDF5 file format, is loaded, and pre-processed, including resizing the images to a standardized dimension of 228x228 pixels and normalizing pixel values between 0 and 1. The CNN model is designed as a sequential stack of layers, starting with a convolutional layer with 16 filters, followed by maxbatch normalization, and dropout for pooling, regularization. This pattern is repeated with additional convolutional layers, each increasing the number of filters. The final layer is a dense layer with softmax activation, representing the three categories of damage scale (See Table 1). During training, the model is fed with the pre-processed images, aiming for 30 epochs with a batch size of 64 and a validation split of 20%. The trained model is then saved for subsequent use. This architecture combines convolutional and pooling layers with normalization and dropout techniques, demonstrating its potential for accurate concrete damage classification in structural inspection applications. Regarding the number of images in the training model, this is summarized in the following Table 3. As such, it is important to note that from the set of photos, each image was manually labeled according to the concrete damage.

Table 3 Number of samples used in the training model.

Comonata	Training Samples			
Domogo	No	Moderate	Major	
Damage	damage	Damage	Damage	
Corrosion		544	690	
Cracks		4609	6998	
Efflorescence	2180	1196	1094	
Exposed Bars		289	766	
Spalling		1310	1592	

3.3.3 Data Classification

In this section, a function is developed to manage the loading, resizing, and normalizing each grayscale image to fit the required input dimensions of the model. From the case study, 142 images of 5472x3648 pixels were obtained from which only 89 images were selected for the classification process.

The subsequent iteration through each image file involves making predictions using the loaded model. Additionally, contours are detected in the original image, and based on the predicted class, they are highlighted with semi-transparent colors to emphasize the severity of concrete damage.

Notably, the code incorporates error handling to ensure the successful loading and preparation of images. Furthermore, contours are drawn on images to visually represent the detected damage patterns, contributing to a more comprehensive analysis. The entire process is geared towards automating the classification of concrete damage, making it a valuable tool for efficient structural health assessment.

3.4 3D Reconstruction

The initial step involves the application of colorization to highlight the severity of the damage in each 228x228-pixel image. Subsequently, a meticulous merging process is undertaken to reconstruct these images to their original resolution of 5472x3648 pixels.

This merging process is pivotal, as resizing each image independently would result in the loss of critical georeferencing information. This information, including Latitude, Altitude, Longitude, Focal length, orientation, and other metadata, is embedded in the data provided by the drone for each image. Maintaining the integrity of this georeferencing data is essential for the accuracy of the final 3D spatial reconstruction.

Following the image merging, each reconstructed image is meticulously reassigned its corresponding metadata. This involves a comprehensive analysis and adjustment to ensure that the geospatial information aligns accurately with the reconstructed visual data.

To achieve a technically robust 3D reconstruction, specialized software designed for photogrammetric processing of digital images is employed. Notably, Agisoft Metashape stands out as a prominent example of such software. This software employs advanced algorithms and techniques to process the merged images and generate precise 3D spatial data.

4 Results

4.1 Data Training

The key metrics during the training of the deep learning model are summarized in Table 4.

		Tasiaias	Validation
Concrete	г 1		validation
Damage	Epoch	Loss	Loss
2 4111490		(Accuracy)	(Accuracy)
	1/20	1.0573	4.4956
Compasion	1/30	(0.7799)	(0.6442)
Corrosion	20/20	0.0928	3.3723
	50/50	(0.9766)	(0.5447)
	1/20	1.0725	9.1893
Creatra	1/30	(0.7264)	(0.1653)
Cracks	20/20	0.1149	0.6825
	30/30	(0.9625)	(0.8677)
	1/20	1.5263	1.0956
E.C.	1/30	(0.6063)	(0.3747)
Enforescence	30/30	0.1526	2.6873
		(0.9452)	(0.6756)
	1/30	0.9333	2.0844
E- 1D		(0.8396)	(0.6909)
Exposed Bars	30/30	0.0388	1.1794
		(0.9884)	(0.8779)
	1/20	1.4644	1.9797
G 11:	1/30	(0.6148)	(0.4435)
Spalling	30/30	0.1393	2.9182
		(0.9542)	(0.5280)
		· /	/

Table 4. Loss and accuracy of the training model

4.2 Data Classification

Given the extensive dataset, we present a table displaying the classification of select images featuring damaged concrete along with their corresponding confidence levels.

In Table 5, the classification results of concrete damage for a photograph are illustrated. It can be observed that, in the case of the example, after processing, only corrosion, efflorescence, and spalling damage are noticeable. Similar results are obtained for all other photographs based on the level of damage they exhibit.

4.3 3D Reconstruction

After the classification process, the 228x228 pixels colored images are combined with their counterparts from the original image, which is 5472x3648 pixels (See Fig. 6). Subsequently, the metadata from the original image is transferred to the reconstructed image.

Concrete Damage	Example image	Damage Classification (Level of confidence)
Corrosion		Major (59.74%)
Cracks		No color (100.00%)
Efflorescence	(228x228 pixels)	Moderate (98.06%)
Exposed Bars		No color (88.44%)
Spalling		Major (99.88%)



Fig. 6 Reconstructed image with concrete damage classification

Ultimately, utilizing these 89 reconstructed images from the case study, we proceed to generate a point cloud and perform the 3D reconstruction of the bridge (See Fig. 7).



Fig. 7 3D reconstruction of the Iniche bridge with concrete damage classification

5 Conclusions and discussions

The proposed methodology demonstrates efficiency in bridge inspections by combining threedimensional reconstruction with concrete damage classification using deep learning. This suggests an effective and automated alternative for assessing bridge conditions, minimizing time and costs associated with traditional methods.

The implementation of deep learning in concrete damage classification ensures an objective approach. The model's ability to accurately identify and categorize defects in concrete from UAV-captured images suggests a significant improvement in result objectivity compared to conventional inspection methods.

Upon reviewing the outcomes derived from the 228x228 pixel images, it is evident that a more extensive dataset is essential for refining the training model. Also, due to the use of an external database, the GDS has been variable which affects the quality of the images and therefore affects the training database.

Addressing the loss of georeferenced information after resizing images is achieved through threedimensional reconstruction, where the original metadata is transferred to the reconstructed images. This highlights the importance of integrating geospatial data for a more comprehensive and accurate assessment of infrastructure.

The main limitations of this method were that it does not assess the quality of the images due to their reduction in size. Additionally, the classification of the magnitude of the damage was partially subjective. Other limitations included the acquisition of data, where external conditions to the infrastructure, such as the flow of a river, varied the quality of the images.

References

- Wang X, Demartino C, Narazaki Y, Monti G, Spencer BF. Rapid seismic risk assessment of bridges using UAV aerial photogrammetry. Eng Struct. 2023 Mar 15;279:115589.
- [2] Scattarreggia N, Salomone R, Moratti M, Malomo D, Pinho R, Calvi GM. Collapse analysis of the multi-span reinforced concrete arch bridge of Caprigliola, Italy. Eng Struct. 2022 Jan 15;251:113375.
- [3] Yau NJ, Tsai MK, Wang HL, Hung DM, Chen CS, Hsu WK. Improving bridge collapse detection and on-site emergency alarms: A case study in Taiwan. Saf Sci. 2014 Dec 1;70:133–42.
- [4] Mandirola M, Casarotti C, Peloso S, Lanese I, Brunesi E, Senaldi I. Use of UAS for damage inspection and assessment of bridge infrastructures. International Journal of Disaster Risk Reduction. 2022 Apr 1;72:102824.
- [5] Azimi M, Eslamlou AD, Pekcan G. Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review. Sensors [Internet]. 2020;20(10). Available from: https://www.mdpi.com/1424-8220/20/10/2778
- [6] Dorafshan S, Thomas RJ, Maguire M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Constr Build Mater. 2018 Oct 20;186:1031–45.
- [7] Payawal JMG, Kim DK. Image-Based Structural Health Monitoring: A Systematic Review. Applied Sciences [Internet]. 2023;13(2). Available from: https://www.mdpi.com/2076-3417/13/2/968
- [8] Nyathi MA, Bai J, Wilson ID. Concrete Crack Width Measurement Using a Laser Beam and Image Processing Algorithms. Applied Sciences [Internet]. 2023;13(8). Available from: https://www.mdpi.com/2076-3417/13/8/4981
- [9] Zhang Z, Zhou H, Wang S, Xu C, Lv Y. Design and Research of Low-Cost and Self-Adaptive Terrestrial Laser Scanning for Indoor Measurement Based on Adaptive Indoor Measurement Scanning Strategy and Structural

Characteristics Point Cloud Segmentation. Han J, editor. Advances in Civil Engineering [Internet]. 2022;2022:5681771. Available from: https://doi.org/10.1155/2022/5681771

- [10] Zoubir H, Rguig M, El Aroussi M, Chehri A, Saadane R. Concrete Bridge Crack Image Classification Using Histograms of Oriented Gradients, Uniform Local Binary Patterns, and Kernel Principal Component Analysis. Electronics (Basel) [Internet]. 2022;11(20). Available from: https://www.mdpi.com/2079-9292/11/20/3357
- [11] Li J, Li X, Liu K, Yao Z. Crack Identification for Bridge Structures Using an Unmanned Aerial Vehicle (UAV) Incorporating Image Geometric Correction. Buildings [Internet]. 2022;12(11). Available from: https://www.mdpi.com/2075-5309/12/11/1869
- [12] Sankarasrinivasan S, Balasubramanian E, Karthik K, Chandrasekar U, Gupta R. Health Monitoring of Civil Structures with Integrated UAV and Image Processing System. Procedia Comput Sci. 2015 Jan 1;54:508–15.
- [13] Bao Y, Chen Z, Wei S, Xu Y, Tang Z, Li H. The State of the Art of Data Science and Engineering in Structural Health Monitoring. Engineering. 2019 Apr 1;5(2):234–42.
- [14] Ye XW, Dong CZ, Liu T. A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications. Oddo CM, editor. J Sens [Internet]. 2016;2016:7103039. Available from: ttps://doi.org/10.1155/2016/7103039
- [15] Permana SDH, Saputra G, Arifitama B, Yaddarabullah, Caesarendra W, Rahim R. Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm. Journal of King Saud University - Computer and Information Sciences. 2022 Jul 1;34(7):4345–57.
- [16] Sagara A, Pane I. A Study on Effects of Creep and Shrinkage in High Strength Concrete Bridges. Procedia Eng. 2015 Jan 1;125:1087–93.
- [17] Abubakr M, Rady M, Badran K, Mahfouz SY. Application of deep learning in damage classification of reinforced concrete bridges. Ain Shams Engineering Journal. 2024 Jan 1;15(1):102297.
- [18] Hüthwohl P, Lu R, Brilakis I. Multi-classifier for reinforced concrete bridge defects. Autom Constr. 2019 Sep 1;105:102824.
- [19] Highways England. CS 450 Inspection of highway structures. In: 0.1.0. Design Manual for Roads and Bridges; 2021.
- [20] Mansuri L, Patel D. Artificial intelligence-based automatic visual inspection system for built heritage. Smart and Sustainable Built Environment. 2021 Feb 3;11:622–46.
- [21] Zhao S, Kang F, Li J, Ma C. Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction. Autom Constr. 2021 Oct 1;130:103832.

- [22] Nappo N, Mavrouli O, Nex F, van Westen C, Gambillara R, Michetti AM. Use of UAV-based photogrammetry products for semi-automatic detection and classification of asphalt road damage in landslide-affected areas. Eng Geol. 2021 Dec 5;294:106363.
- [23] Siyuan C, F LD, Eleni M, Iman ZSM, Jonathan B. UAV Bridge Inspection through Evaluated 3D Reconstructions. Journal of Bridge Engineering [Internet]. 2019 Apr 1;24(4):05019001. Available from: https://doi.org/10.1061/(ASCE)BE.1943-5592.0001343
- [24] Atencio E, Plaza-Muñoz F, Muñoz-La Rivera F, Lozano-Galant JA. Calibration of UAV flight parameters for pavement pothole detection using orthogonal arrays. Autom Constr. 2022 Nov 1;143:104545.
- [25] Esqueda Elizondo J, Palafox L. Fundamentos de Procesamiento de Imágenes. 2005.
- [26] Zhang L, Yang F, Daniel Zhang Y, Zhu YJ. Road crack detection using deep convolutional neural network. In: 2016 IEEE International Conference on Image Processing (ICIP). 2016. p. 3708–12.

Point Cloud-based Computer Vision Framework for Detecting Proximity of Trees to Power Distribution Lines

Fardin Bahreini¹, Amin Hammad¹, and Mazdak Nik-Bakht²

¹Concordia Institute for Information Systems Engineering, Concordia University, Canada ²Department of Building, Civil & Environmental Engineering, Concordia University, Canada fardin.bahreini@concordia.ca, amin.hammad@concordia.ca, mazdak.nikbakht@concordia.ca

Abstract -

The maintenance of power lines is challenged by the encroachment of vegetation, posing significant risks to the reliability and safety of power utilities. Traditional methods, based on manual inspection, are not only resource-intensive but also lack the necessary precision for effective and proactive maintenance. This paper aims to develop an automated, accurate, and efficient approach to vegetation management in the vicinity of power lines. It leverages advancements in data collection using LiDAR scanning technology, which despite its potential, faces computational challenges in processing large-scale 3D point clouds to accurately identify power lines and surrounding vegetation. To overcome this challenge, the proposed method deploys the RandLA-Net model for the semantic segmentation of power lines and nearby vegetation in point cloud datasets. Furthermore, the post-processing analysis of the segmented data uses clustering and rule-based thresholding to refine the identification of vegetation. Then, proximity detection is applied using spatial queries based on a KDTree structure. The results of the case study demonstrate the computational efficiency and accuracy of the proposed method, presenting a promising solution for power utilities.

Keywords -

Computer Vision; 3D Point Cloud; Power Lines; Proximity Detection

1 Introduction

Vegetation management is critical for ensuring the safety and reliability of power distribution systems. The encroachment of overgrown vegetation near power lines poses significant risks, potentially leading to power outages, fires, and other hazards. The primary problem in this sector has been the reliance on manual inspection, which is labor-intensive, time-consuming, and often limited in accuracy and frequency. This traditional approach struggles to keep pace with the growing demand for stable energy and the urgent need to mitigate risks associated with overgrown vegetation. Consequently, there is a need for more efficient, automated solutions in vegetation management. Advances in sensing technologies, particularly Light Detection and Ranging (LiDAR) scanning, combined with Machine Learning (ML) algorithms, have led to the development of Automated Vegetation Management (AVM) systems. These systems promise more frequent monitoring, potentially revolutionizing vegetation management [1]. However, a significant challenge lies in processing the vast volumes of 3D point cloud data generated by LiDAR [2], particularly in accurately segmenting and classifying each point to identify vegetation-related risks effectively. Unlike conventional methods relying on visual inspection or 2D imaging, point cloud data provides comprehensive spatial representation, allowing accurate distance measurements and identification of fine details of power lines and surrounding vegetation. This advancement offers a systematic and reliable approach to power line monitoring, supported by recent studies demonstrating its efficacy in automated vegetation management systems and power line inspection [3].

The objectives of this paper are: (1) to accurately detect vegetation and power lines from LiDAR data using Deep Learning (DL), and (2) (2) to conduct detailed post-processing analysis to detect the proximity of trees and power lines. This approach is expected to enhance the reliability of power distribution systems and potentially lead to significant cost savings for utility companies. The results demonstrate the practical application of the proposed method in a real-world urban setting.

2 Literature Review

LiDAR technology has emerged as a powerful AVM tool for power distribution lines. Its ability to provide high-resolution 3D data has made it crucial for detecting and analyzing vegetation in the context of power line management. Gollob et al. [4] investigated the accuracy of estimates for individual trees and forest stand variables using a mobile laser scanning system. Their study highlights the impact of scan variation on tree parameter measurements. Voelsen et al. [5] segmented point cloud data from a Mobile Mapping LiDAR dataset. They used a method combining region growing and random forest classification to distinguish between static and dynamic objects, such as poles and vegetation. Wen et al. [6] emphasized the significance of high-accuracy and highefficiency 3D sensing and associated data processing techniques for various applications, including detecting trees and poles. Lu et al. [7] introduced a localization system for autonomous vehicles using cluster-based methods to extract pole-like objects, including trees and street lights, from 3D LiDAR point clouds. Kutz et al. [8] discussed the application of high-resolution imagery and LiDAR-derived canopy height models in land cover mapping, crucial for resource management and planning. Gaha et al. introduced a new LiDAR-based clustering method for detecting poles and distribution lines, offering improvements in accuracy and efficiency [9]. However, the scope of their study was primarily focused on singlephase lines and had limited effectiveness in occluded environments.

ML has significantly changed AVM around power lines, allowing for rapid data processing and previously unattainable insights. Kyuroson et al. developed an unsupervised ML framework to detect and analyze power lines and surrounding vegetation in Power Line Corridors (PLCs) using various remote data acquisition techniques such as airborne, mobile, and terrestrial laser scanning [10]. Torres de Almeida et al. combined satellite imagery, airborne LiDAR data, and ML algorithms, including Linear Regression, Classification and Regression Trees (CART), and Random Forest (RF) to map vegetation height in PLCs, aiding in management planning [11]. Li et al. employed drone data, airborne LiDAR, and ML algorithms, including RF, and Support Vector Machine (SVM) for classifying tree species in transmission line corridors [12].

Abongo et al. introduced a novel framework for detecting power lines using LiDAR data, utilizing a combination of ML (XGBoost) and geometric methods [13]. However, their approach was primarily limited to the detection aspect, without exploring subsequent data processing and analysis for vegetation management. Haroun et al. reviewed vegetation encroachment detection techniques using satellite images, emphasizing the potential of ML and DL algorithms to enhance detection accuracy and flexibility [14]. Park et al. used feature-enhanced convolutional neural networks (CNNs) including AlexNet, ResNet18, and VGG11 for classifying images from Google Street View into categories related to utility systems and vegetation overgrowth, aiding in vegetation management prioritization [15]. Mohd Rapheal et al. assessed a ML-

based geospatial method for classifying electricity assets using high-density mobile laser scanning data, achieving detection accuracies of 65% for overhead power lines and 63% for electricity poles [16]. Although focused on river management, Rabanaque et al. presented a ML approach (SVM and RF) for analyzing geomorphological characteristics and vegetation density using LiDAR and multispectral satellite images [17]. Horning et al. discussed the challenges and advances in mapping land cover using ultra-high-resolution aerial imagery, including ML algorithms for image processing [18]. Oehmcke et al. utilized DL systems (MSENet14, KPConv, PointNet) to predict wood volume and aboveground biomass directly from airborne LiDAR point clouds [19]. Their method showed significant improvements in accuracy compared to traditional approaches. Gribov and Duri proposed a solution for constructing line features modeling each catenary curve present within a series of points representing multiple catenary curves [20]. This solution can be applied to extract power lines from LiDAR point clouds.

Mahoney et al. utilized a combination of various ML algorithms, including RF, Gradient Boosting Machine, and Artificial Neural Network (ANN), to integrate remote sensing of structural and optical properties of vegetation cover for classifying and mapping shrubland habitats [21]. Furthermore, studies like that of Amani et al., which utilized bathymetric LiDAR data for marine habitat mapping, showcase the versatility of LiDAR and RF algorithms in vegetation classification [22]. Amado et al. presented a method for extracting power lines from LiDAR point cloud data, demonstrating accurate and automatic extraction capabilities [23]. Awrangjeb introduced a power line extracting and modeling approach using LiDAR data, which significantly aids in the detection and modeling of power lines, offering a reliable solution to the challenges faced in power line extraction [24]. Li and Guo discussed the application of LiDAR technology for power line inspection, highlighting its advantages in obtaining high precision 3D spatial information and entire power line corridor data, which is critical for effective inspection and maintenance [25]. Table 1 shows a comparative overview of most related works, outlining key aspects such as methodology, utilized dataset, main research focus, and critical performance metrics across different studies.

3 Proposed Framework

This paper proposes using Random Sampling in Large-scale Point Cloud Analysis Network (RandLA-Net) model [28], designed for the semantic segmentation of large 3D point clouds, for AVM. This model was selected due to its highest overall accuracy level in semantic segmentation of the Toronto-3D dataset [29].

Q"D. 7. %5?.),%#. &0 "&;)D(+#",)(7(';(>;.g)?-+5D.) %,%#. &;)D',?)'-%9. \$'/\$ &(,(\$"#+'&%'#)?.) '#,.#+',2)'/),?.)9'''#,+)(#&/''\$-+'#4)"#),?.),?%.)\$;(++,+) ''/)'#,.%+,U7.4., (,''#5)9'';.+5)(#&)9''D.%;'#.+=)E?'+ %'+B)>(+.&) "#)+9(,'(;)%;(,''' #+?9+)>.,D..#),%.+)(#&) 9"D.%;'#.+)'#)%(#).#7' %#0 .#,+=)b'4-% O9%7' &.+(#) "7.%7'.D)''/),?.)9%9"+.&)%0 .D "%B=

E(>;.)O)! "0	9(%, '7.)"7.%7'.D)"	/)0 "+,)%(,	.&Ɗ"%₿+
--------------	----------	-------------	-------------	---------

!"#'\$" & *"	(")\$	* "+,/ -01	2 3". '4)+)3 "+	*)5%6-&73	8-2 '#\$'&)33''3	9 -: ;)\$)+5<"''
" #\$%&!	+.+/	01 2345\$'6 &77!	>'5(5&;!#!@&A	E:#\$@ 5@5 4\$!	348 &!(\$\$&@)	_ 50,750,75 D#BJ@K.4L5< 5M!
#()!*+,-		8 59!0 44:;5\$#'&!	B4\$@;=B&!	B4::5;4:!	,G)HI	A4B=(X)(48 & !Q+>
		1 ''&\$54\$!	>1C! D5; #:!	@&%<&\$'#'54\$		
		<4;=(&	;##@&	-		
0#\$42	+.+/	CQ3D26&!	>1C! 5<#%&M	C&% & 54\$!#\$;!	C&%#\$54\$F.GGIN	W44;! 5 \$!V&% #85 4\$!
P4(5@x!#()!		A4B=36%4\$!	<#M(#BJ!	K48&!(\$\$&	348 & !(5\$& @HIN	;&&B54\$N(&@@#\$\$!
*+G-		&\$& <r(&!< th=""><th>D5S1TV@&K9!</th><th>@&%<&\$'#'54\$!</th><th>@48 5%:!44< !Ai:!</th><th>; (&; !B4\$'&L'N(#BJ@</th></r(&!<>	D5S1TV@&K9!	@&%<&\$'#'54\$!	@48 5 %:!44< !A i :!	; (&; !B4\$'&L'N(#BJ@
		< <i>&</i> 94;@	:&@='54\$	8 594='!	5< K4V& &	K4L5<5M#\$#(M5@)
				K4L555MA4B=@		
1 R4\$%4&!	+.+/	YWZ44@8 5'9!	P'#\$;#:;!	P4(&A4B=@4\$!	348 &!(5 \$& B)	QAA&BV&5\$!R#@B!
#()!*X/-		R#@ B %&4<&:5B	D5S1T!;#'#@&!	K48&!(5\$&	[+)H\I	; &&B 54\$N(#BJ@!
		#\$#(M 5@)	8 594='!	;&&B54\$		B4 <k&l5m< th=""></k&l5m<>
			@&&B \$ &;!			
			B4 <k&l5m< th=""><th></th><th></th><th></th></k&l5m<>			
] =:!@=;M	+.+H	1 ;V#\$B&;!	E4:4\$'42' SN	S=#(!A4B=@4\$!	E:&&F@) [XI N	^SE:&&4:!K:4L5<5M!
		T#\$;D126&!	K4V55\$%!	R4'9!K48&!(\$\$&	348 &!(5 \$& B)	#\$#(M&&;!
		85'9!@&&B \$ B!	;5V&@&:R#\$!	#\$;!V &‰8# ′54\$!	[G)[/I NB4(&@E	B(#@ @&&B 5A 5® 4> !
		K4@K:4B&@&%	(#\$; @B#K&!	8 5'9!K:4L5<5'M	G\)/,I	@B4&@#\$;!&\$9#\$B\$\$%!
		4K5< 5?# 54\$@	B9#((&\$ % @	#\$#(M 5@)		; &&B54\$!#BB=:#BM

 $\begin{array}{l} (99\%(\$?) (\&\&\&\%++),?.) \$?(;;.#4.) "/)?(\#\&';\#4).V,.#+'7.) \\ 7";-0.+)"/)R8)9"#,)\$;"-\&)\&(j),")(\$-\%,:;2)+40.#,) \\ (\#\&\$;(++'/2)(\$?)9""#,)").//'\$'.#,:2)9'#9'"#,)7.4., (,''#F \\ \%;(,.\&) \%'+B)'#) HCM= Q"D.7. %3,?.) +40.#, (,''#) \\ 9\%(\$.++)'+)"#;2),?.)/'\%,)+,9=)E?.)+>+.G.-#,)\$?(;:,#4.5) \\ (\#\&))\$\%'s(;)(+9.\$,)"/),?'+)9(9.\%'+),?.) 9"+,B9\%s.++'#4) \\ (\#(;2++)"/) +0 (\#,'\$) +40.#, (,''#=) E?'+)'#7";7.+) \\ .0 9;"2'#4) 8.#+',2F (+.&) 19(.'(;) !:-+,.\%'4) "/) \\ H99;'\$(,''#+)D',?)d"'+.)Ks:1 !HdL)NRaP(#&)'4;.F>(+.&) \\ .?\%'f'';&\#4),")+";(,.)">T$,+),?(,0...)+9.$/'$)$%. %(5) \\ \$\%'s'(;) /"\% +9(\%(,'#4) & +;'#$), -\%(#) ::0.#, +(#&) \\ \%\&-\$'#4)#'''+=)H&\&'''#(;;25)D.'09;.0.#,)()9\%V'0',2) \\ \&.,\$,''#) -+'#4)] 8 E\%) 0..?"&),").7(;-(,.)9",.#,'(;) \end{array}$

567 8*- "&\$\$68*:- *&44%

E?.) J (#&IHEd. ,)0 "&;)NSP).//'\$'.#,)#)9%5.++'#4) ?'4?F&.#+,2) 9'''#,) \$;''-&+5) D(+) -+.&) /''%)+.0 (#,'\$) +40 .#, (,''#5)''\$-+'#4)''#),%+59''D.%&+,%>-, ''#);'#.+5) (#&)9'';.+ \Rightarrow J(#&IH Fd. , +,(#& \Rightarrow)''-,)&-.) ,') ',+)-#' G-.) (99%(\$?5 .0 9;'2'#4) %(#&''0) 9'''#,) +(0 9;'#4) /''% &"D#+(0 9;'#4=E?'++'4#'/'\$(#,;2)%&-\$.+\$''0 9-, (,''#(;) \$''0 9;.V',2)D?';.) 0 ('#,('##4),?.) '#,.4%,2)''/),?.) 9'''#,) \$;''-&)4.''0 .,%\$)&.,(';+ \Rightarrow ,)'#\$''9''%,.+)();"\$(;)/.(, -%) (44%4(, ''#) 0 "&-;.5) D?'\$?) '#\$;-&.+) I "\$(;) 19(,'(;) 3#\$''(#&H,,#, '7.)`'''; '#45,'')\$(9,-%)'#,%\$(.);"\$(;)



b'4-% Of 7.%7'.D)"/),?.)9%9"+.&)%0.D"%B

structures effectively [28]. The network's architecture, featuring shared Multilayer Perceptrons (MLPs) and dilated residual blocks, enhances its processing speed, allowing it to handle up to one million points in a single pass with notable accuracy. This efficient and innovative approach makes RandLA-Net particularly suitable for large-scale point cloud analyses, demonstrating superior performance in both speed and accuracy compared to other methods, such as PointNet++.

To assess the performance of the RandLA-Net model in this context, we used the Intersection over Union (IoU) metric. This is a standard metric in evaluating segmentation models, as it quantifies the accuracy of the model in classifying each point. It works by measuring the overlap between the model's predicted classifications and the actual, ground truth classifications. The IoU provides a comprehensive overview of the model's performance across all classes, offering insights into its precision and effectiveness in segmenting different urban elements. A higher IoU value indicates better model performance and a more accurate representation of the real-world scenario.

3.2 Clustering and Rule-Based Thresholding

In order to refine the segmented point cloud data in the initial stage of post-processing, partial clustering algorithm, DBSCAN, was selected for its ability to identify clusters of various shapes and densities within the data without the need for predefined number of clusters. This feature made DBSCAN particularly suitable for handling the complex and varied structure of the urban dataset. Key parameters like the epsilon values and minimum samples were carefully adjusted to align with the dataset's unique features, ensuring sensitivity to the varied densities and distributions of urban elements. In addition to DBSCAN, rule-based thresholding was implemented, setting specific height and point count thresholds for each urban feature class, such as trees and poles. This approach effectively isolated significant urban objects within each category while reducing noise and irrelevant data, thereby enhancing the clustering results' overall quality and accuracy.

3.3 Proximity Detection between Trees and Power Lines

To assess the risk associated with trees near power lines, a K-Dimensional Tree (KDTree) structure was adopted for streamlined spatial querying, aiming to effectively evaluate potential risks. The KDTree, known for its ability to rapidly query points in a multi-dimensional space, proved ideal for analyzing spatial relationships within the point cloud data. The process involved utilizing the KDTree structure to efficiently identify the nearest power line to each tree. Once these proximities are calculated, they are compared against a safety threshold. Tree areas falling within this threshold are identified as potential hazards. Considering the dataset's large scale, ensuring computational efficiency was a critical concern. This approach optimizes performance and resource utilization in large-scale, complex data operations.

4 Implementation and Case Study

In the implementation phase, post-processing techniques including clustering, rule-based thresholding, and proximity detection were deployed using Python.

4.1 Data Acquisition and Preparation

Toronto-3D dataset, developed by Tan et al. [2] was used in the case study. This dataset was collected along a 1-kilometer section of Avenue Road in Toronto, Canada. It is a large dataset containing about 78.3 million data points. The dataset is notable for its high point density, with an average of 1000 points per square meter on the ground. This high density is crucial for capturing detailed features of the urban environment, which is vital for this study. The data was collected using a LiDAR sensor on a Mobile Laser Scanning (MLS) system. The LiDAR sensor captured up to 700,000 points per second, with a vertical field of view from -10 to +30 degrees, and an accuracy of better than 3 cm. Each point in the dataset has several attributes including the position in meters (XYZ coordinates), the color reflectance (RGB), LiDAR intensity, GPS time of collection, scan angle, and the object class label. The object class labels cover a range of urban features, making the dataset useful for semantic segmentation. These labels include roads, road markings, natural elements (trees, shrubs), building parts, power distribution lines, poles (utility poles, traffic signs), vehicles, and vertical barriers (fences, walls).

Data preparation involved loading the point cloud data and performing grid subsampling with a grid size of 0.06 meters to reduce data volume while preserving key features. For validation datasets, projection indexes were created to map model results back to the original dataset, ensuring a structured and efficient dataset ready for semantic segmentation and analysis. A projection index is a reference that maps each point in a subsampled point cloud back to its original location in the full dense point cloud, ensuring that any analysis or modifications applied to the reduced dataset can be accurately reflected in the original, larger dataset. The raw data from the dataset underwent preprocessing to convert the .ply files into a suitable text format for the semantic segmentation process. Figure 2 shows a sample area of point cloud data.



Figure 2. A sample area of point cloud data

4.2 Semantic Segmentation

The Toronto-3D dataset was divided into four sections, each covering about 250 meters of the road. Sections 1, 3, and 4 were used for training and Section 2 was used for testing. The dataset underwent two separate training processes. Initially, it was trained without considering RGB and intensity data to focus on the geometric features (X, Y, Z). Then, it was trained again, this time incorporating the RGB and intensity information (X, Y, Z, R, G, B, intensity) to assess the impact of these attributes on model performance.

Using this dataset, the RandLA-Net model underwent 100 epochs of training to enhance its accuracy in segmenting point cloud data. The Adam optimizer, known for its efficiency with large-scale data, was utilized. An initial learning rate of 0.01 was set, gradually reduced by 5% per epoch to refine model adjustments and convergence. A batch size of 4 was maintained to balance computational resources and effective learning during training sessions. Throughout this segmentation process, the model accurately assigned a class label to each point, enabling the differentiation of various urban elements. The training time for the model was 124 hours and 33 minutes on a LAMBDA workstation with one NVIDIA RTX A6000 GPU, 48 GB RAM/GPU, and an AMD Ryzer Threadripper 3960×48-core CPU. The model without considering RGB and intensity achieved an overall accuracy of 93.08%, representing the ratio of correctly labeled points to the total number of points across all classes. The model considering RGB and intensity achieved an overall accuracy of 95.42%.

During the testing phase, each epoch was composed of 25 steps, with every step processing a batch of test data. In the testing, a step is a single iteration over a batch of data and an epoch represents a single pass through the entire test dataset. For individual classes, the model without considering RGB and intensity showed very good performance with the classes of trees, power distribution lines, and poles achieving IoU of 95.76%, 87.61%, and 76.37%, respectively. The model considering RGB and Intensity showed improved performance with the classes of trees, power distribution lines, and poles achieving IoU of 96.81%, 87.83%, and 79.36%, respectively. The observed improvement of over 3% in the poles class in the model considering RGB and intensity may be due to the distinctive and unique colors of the poles. However, the lower accuracy in pole detection in both models could be attributed to the inherent complexity of urban environments where poles are located. Factors such as occlusion by other objects, varying lighting conditions, and the poles' similarity to other vertical structures might make them harder to distinguish accurately. Compared to the work of Abongo et al. [13] focusing only on power distribution line detection, which achieved an IoU of 82.49%, our method notably surpasses this performance, achieving an IoU of 87.83% for power distribution lines.

4.3 Clustering and Rule-Based Thresholding

In the subsequent stage of our analysis, we used a combination of DBSCAN, height, and point count thresholding techniques to identify and isolate objects meeting our predefined criteria. This method involved adjustments of DBSCAN parameters, such as epsilon (i.e. maximum distance between samples) and the minimum number of samples, to match the unique attributes of the Toronto-3D dataset. Following the clustering process, we retained clusters that surpassed the designated height threshold (e.g., 8 meters for poles) and fulfilled the

minimum point count threshold (e.g., 500 points for poles). This step resulted in a refined dataset, distinctly differentiating and emphasizing significant urban features from less relevant objects. Table 2 presents the clustering values and thresholds for each class. Figure 3 shows segmentation results before and after clustering and rule-based thresholding.

Table 2. Clustering values and thresholds for each class

	Clustering values		Defined thresholds	
Class	Epsilon	Minimum	Minimum	Minimum
	(m)	samples	point count	height (m)
Trees	0.5	20	12000	6.5
Poles	0.5	10	500	8
Power lines	0.3	3	-	-

RGB and intensity data analysis for precise urban vegetation management. This is further complemented by novel post-processing techniques, including DBSCAN clustering and rule-based thresholding, which collectively refine risk assessment and provide a detailed understanding of vegetation's proximity to power lines. This paper presents a comprehensive approach for urban vegetation management in proximity to power lines using point cloud data in conjunction with the RandLA-Net model. The approach is further enhanced by postprocessing techniques such as clustering and rule-based thresholding considering the specific needs of the application. Moreover, the incorporation of proximity detection for risk assessment added a practical dimension to the proposed framework. The RandLA-Net model considering RGB and intensity showed improved performance with various classes, including trees, power





Figure 3. Comparing segmentation results before and after clustering and rule-based thresholding

4.4 Proximity Detection between Trees and Power Lines

The next step aims to compare the spatial data of trees and power lines to enable proximity detection for risk assessment. Leveraging KDTree's queries, we efficiently determined the closest power line point to each tree point in a multidimensional space. Assessing this against a predetermined safety threshold of 1 meter allowed us to identify parts of trees posing risks due to their proximity to power lines, which were flagged as potential hazards. Figure 4 shows the post-processed semantic segmentation result, highlighting hazardous tree areas (red points) within the safety perimeter of power distribution lines.

5 Conclusions and Future Work

The contribution of this paper lies in its integrated approach, utilizing the RandLA-Net model enhanced by distribution lines, and poles, achieving IoU of 96.81%, 87.83%, and 79.36%, respectively. The effectiveness of DBSCAN clustering and rule-based thresholding was apparent in the clarity and distinction of the isolated objects. Additionally, the proximity detection analysis efficiently pinpointed significant tree areas where trees can pose threats to power lines.

The study's findings rely on the Toronto-3D dataset, which might not fully represent all urban, and/or suburban landscapes. While the results of RandLA-Net exhibited high accuracy, its performance in different or more complex environments requires further exploration. Moreover, the computational demands of these methods could limit their feasibility in resource-constrained settings. Future research should prioritize testing the proposed framework across diverse environments, refining the algorithms for broader applicability, and integrating additional data sources for a more comprehensive approach.



Figure 4. Post-processed result of semantic segmentation, highlighting hazardous tree areas

References

- Cao, W., Wu, J., Shi, Y. and Chen, D., Restoration of Individual Tree Missing Point Cloud Based on Local Features of Point Cloud, *Remote Sensing*, vol. 14, no. 4, p. 1346.
- [2] Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K. and Li, J., "oronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Seattle, WA, USA, 2020.
- [3] Mongus, D., Brumen, M., Žlaus, D., Kohek, Š., Tomažič, R., Kerin, U. and Kolmanič, S., A Complete Environmental Intelligence System for LiDAR-Based Vegetation Management in Power-Line Corridors. *Remote Sensing*, vol. 13, no. 24, p.5159, 2021.
- [4] Gollob, C., Krassnitzer, R., Ritter, T., Tockner, A., Erber, G., Kühmaier, M., Hönigsberger, F., Varch, T., Holzinger, A., Stampfer, K. and Nothdurft, A., Measurement of Individual Tree Parameters with Carriage-Based Laser Scanning in Cable Yarding Operations, *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, vol. 44, no. 22, pp. 401-407, 2023.

- [5] Voelsen, M., Schachtschneider, J. and Brenner, C., Classification and change detection in mobile mapping LiDAR point clouds, *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 89, no. 3, pp. 195-207, 2021.
- [6] Wen, C., Habib, A.F., Li, J., Toth, C.K., Wang, C. and Fan, H., Special issue on 3D sensing in intelligent transportation, *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 1974-1949, 2021.
- [7] Lu, F., Chen, G., Dong, J., Yuan, X., Gu, S. and Knoll, A., Pole-based localization for autonomous vehicles in urban scenarios using local grid mapbased method, in Proceedings of 5th International Conference on Advanced Robotics and Mechatronics (ICARM), Shenzhen, China, 2020.
- [8] Kutz, K., Cook, Z. and Linderman, M., Object based classification of a riparian environment using ultra high-resolution imagery, hierarchical landcover structures, and image texture, *Scientific Reports*, vol. 12, no. 1, p. 11291, 2022.
- [9] Gaha, M., Jaafar, W., Fakhfekh, J., Houle, G., Abderrazak, J.B. and Bourgeois, M., Anew lidarbased approach for poles and distribution lines detection and modelling, *Comput. Sci. Inf. Technol*, vol. 11, no. 1, pp. 85-97, 2021.
- [10] Kyuroson, A., Koval, A. and Nikolakopoulos, G., Autonomous Point Cloud Segmentation for Power Lines Inspection in Smart Grid, *IFAC*-

PapersOnLine, vol. 56, no. 2, pp. 11754-11761, 2023.

- [11] Torres de Almeida, C., Gerente, J., Rodrigo dos Prazeres Campos, J., Caruso Gomes Junior, F., Providelo, L.A., Marchiori, G. and Chen, X., Canopy Height Mapping by Sentinel 1 and 2 Satellite Images, Airborne LiDAR Data, and Machine Learning, *Remote Sensing*, vol. 14, no. 6, p. 4112, 2022.
- [12] Li, X., Wang, R., Chen, X., Li, Y. and Duan, Y., Classification of Transmission Line Corridor Tree Species Based on Drone Data and Machine Learning, *Sustainability*, vol. 14, no. 4, p. 8273, 2022.
- [13] Abongo, D.A., Gaha, M., Cherif, S., Jaafar, W., Houle, G. and Buteau, C., A novel framework for distribution power lines detection, in Proceedings of IEEE Symposium on Computers and Communications (ISCC), Gammarth, Tunisia, 2023.
- [14] Haroun, F.M.E., Deros, S.N.M. and Din, N.M., A review of vegetation encroachment detection in power transmission lines using optical sensing satellite imagery, *arXiv preprint arXiv:2010.01757*, 2020.
- [15] Park, A., Rajabi, F. and Weber, R., Slash or burn: Power line and vegetation classification for wildfire prevention, arXiv preprint arXiv:2105.03804, 2021.
- [16] Mohd Rapheal, M.S.A., Farhana, A., Mohd Salleh, M.R., Abd Rahman, M.Z., Majid, Z., Musliman, I.A., Abdullah, A.F. and Abd Latif, Z., Machine Learning Approach for Tenaga Nasional Berhad (TNB) Overhead Powerline and Electricity Pole Inventory Using Mobile Laser Scanning Data, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 46, no. 1, pp. 239-246, 2022.
- [17] Rabanaque, M.P., Martínez-Fernández, V., Calle, M. and Benito, G., Basin-wide hydromorphological analysis of ephemeral streams using machine learning algorithms, *Earth Surface Processes and Landforms*, vol. 47, no. 1, pp. 328-344, 2022.
- [18] Horning, N., Land cover mapping with ultra-highresolution aerial imager, *Remote Sensing in Ecology* and Conservation, vol. 6, no. 4, pp. 429-430, 2020.
- [19] Oehmcke, S., Li, L., Revenga, J.C., Nord-Larsen, T., Trepekli, K., Gieseke, F. and Igel, C., Deep learning-based 3D point cloud regression for estimating forest biomass, in Proceedings of the 30th International Conference on Advances in Geographic Information Systems, Seattle, USA, 2022.
- [20] Gribov, A. and Duri, K., Reconstruction of power lines from point clouds, *in Proceedings of International Conference on Document Analysis and Recognition*, Cham: Springer Nature

Switzerland, 2023.

- [21] Mahoney, M.J., Johnson, L.K., Guinan, A.Z. and Beier, C.M., Classification and mapping of lowstatured shrubland cover types in post-agricultural landscapes of the US Northeast, *International Journal of Remote Sensing*, vol. 43, no. 19-24, pp. 7117-7138, 2022.
- [22] Amani, M., Macdonald, C., Salehi, A., Mahdavi, S. and Gullage, M., Marine Habitat Mapping Using Bathymetric LiDAR Data: A Case Study from Bonne Bay, Newfoundland, *Water*, vol. 14, no. 23, p. 3809, 2022.
- [23] Amado, M., Lopes, F., Dias, A. and Martins, A., LiDAR-based power assets extraction based on point cloud data, in Proceedings of the IEEE International Conference on Autonomous Robot Systems and Competitions, Santa Maria da Feira, Portugal, 2021.
- [24] Awrangjeb, M., Extraction of power line pylons and wires using airborne lidar data at different height levels, *Remote Sensing*, vol. 11, no. 15, p.1798, 2019.
- [25] Li, X. and Guo, Y., 2018, July. Application of LiDAR technology in power line inspection, *IOP Conference Series: Materials Science and Engineering*, vol. 382, no. 5, p. 052025, 2018.
- [26] Cano-Solis, M., Ballesteros, J.R. and Sanchez-Torres, G., VEPL-Net: A Deep Learning Ensemble for Automatic Segmentation of Vegetation Encroachment in Power Line Corridors Using UAV Imagery, *ISPRS International Journal of Geo-Information*, vol. 12, no. 11, p. 454, 2023.
- [27] Wang, G., Wang, L., Wu, S., Zu, S. and Song, B., Semantic Segmentation of Transmission Corridor 3D Point Clouds Based on CA-PointNet++, *Electronics*, vol. 12, no.13, p.2829, 2023.
- [28] Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N. and Markham, A., RandLA-Net: Efficient semantic segmentation of large-scale point clouds, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020.
- [29] Toronto-3D GitHub repository. On-line: https://github.com/WeikaiTan/Toronto-3D, Accessed: 09/08/2023.
- [30] Ahmed, K.N. and Razak, T.A., A comparative study of different density based spatial clustering algorithms, *International Journal of Computer Applications*, vol. 99, no. 8, p. 8887, 2014.

Reinforcement Learning for Smart Mobile Factory Operation in Linear Infrastructure projects

Jianpeng Cao¹, Irfan Čustović¹, Ranjith Soman², and Daniel Hall¹

¹Faculty of Architecture & the Built Environment, Delft University of Technology, Netherlands ²Faculty of Civil Engineering and Geosciences, Delft University of Technology, Netherlands <u>C.J.P.Cao@tudelft.nl, I.Custovic@tudelft.nl, R.Soman@tudelft.nl, D.M.Hall@tudelft.nl</u>

Abstract

Mobile factories promise an increased project efficiency with on-demand production and Just-in-Time delivery of prefabricated elements. However, traditional scheduling methods predominantly focus on either factory or site and neglect the factory often leading mobility, to suboptimal synchronization. To address this gap, this paper introduces a novel reinforcement learning (RL)-based model for optimizing the operational policy of mobile factories in infrastructure projects. The developed model simultaneously schedules on-site and off-site operations, effectively integrating the performance metrics at the project level. Utilizing RL, the factory's production management system continuously learns and adjusts in response to real-time project developments, ensuring optimal decision-making regarding scheduling and resource allocation.

Keywords –

Reinforcement Learning; Mobile Factory; Scheduling;

1 Introduction

1.1 Mobile Factories for Infrastructure Projects

The evolution of on-site and near-site prefabrication factories in construction and architecture is marked by notable milestones. Early 20th-century pioneers like Walter Gropius, Martin Wagner, and Adolf Meyer introduced systems such as the "Occident System" and the "Frankfurt Assembly Method," which emphasized standardization [1]. The latter part of the century witnessed unique projects, including Moshe Safdie's Habitat '67 [2] in Montreal and Thomas Herzog's EXPO 2000 timber roof in Hannover, highlighting local production for specialized architecture [3]. Additionally, SKANSKA AB's "Flying Factories" [4] and LiWood's "Field Factories" [5] for near-site modular timber prefabrication represent systematic attempts to bring prefabrication closer to construction sites, focusing on flexibility through low levels of automation.

More recent research has shifted focus towards enhancing the mobility of mobile factories. Alix et al. [6] introduced a reconfigurable manufacturing system designed for frequent relocations, adept at accommodating fluctuating demand. Following this, Wagner et al. [7] unveiled a transportable and adaptable timber construction platform, specifically for carpentry. This innovation was validated through the construction of an intricate wooden pavilion, demonstrating its potential to elevate both the quality and efficiency of carpentry work.

Benefits of mobile factories include efficient manufacturing and pre-assembly operations near the building site, safer and cleaner working environments, and reduction in the number of transport kilometers between the factory and the building site [8]. Particularly, this concept of a mobile factory is suitable for situations with long distances and high logistics costs like the fabrication of components on the construction site.

Despite the benefits of mobile factories, existing research underscores the necessity for broader industrial testing across various domains, as noted by Alix et al [6]. Specifically, the application of mobile factories in large infrastructure projects like rail and road construction remains limited. This gap is noteworthy given the alignment between the intrinsic benefits of mobile factories and the demands of infrastructure projects. Therefore, it is crucial to urgently develop operational policies and decision support systems for scheduling mobile factories in infrastructure construction.

1.2 Integrated Project Scheduling

Project scheduling is a crucial aspect of project management, especially in dynamic and complex environments like factory production and site construction. However, these two areas are typically addressed separately [9]. This separation overlooks the potential efficiencies that could be realized through an integrated scheduling approach. In the realm of industrialized construction, this integration becomes increasingly important. Industrialized construction requires a more streamlined and coherent workflow, ensuring that the prefabrication process in factories aligns precisely with the timelines and demands of site construction, thereby optimizing resource utilization and reducing project delays.

Most researchers in this field have adopted a strategy of integrating site scheduling with storage, delivery, and other logistics processes. For example, Ahn et al. [9] streamline the synchronization of factory output with site demands by optimizing truck-dispatching schedules, and enhancing resource utilization and project timelines. Wang and Hu [10] integrate site scheduling into production scheduling by adding element storing and transportation processes to the traditional production model. This modification allows for simultaneous storage of different elements in the stockyard postproduction, with the timing of the storing and transportation processes being closely aligned with site requirements and schedules. However, both works do not consider factory mobility, which is essential for infrastructure projects, where the factory is transportable in alignment with the project progression.

To address the identified limitations in current research on mobile factories, this paper proposes an innovative approach using a reinforcement learningbased model to optimize operational policies in infrastructure projects. Unlike traditional project scheduling methods, which typically segregate factory production from site construction, our approach focuses on integrating these two critical components. Consequently, this approach not only promises improved project efficiency but also marks a significant step in adaptive project management. Building upon this foundation, the following section reviews existing research in RL-based scheduling methodologies, setting the stage for a deeper understanding of the approach's context and significance.

2 Literature Review

Reinforcement learning (RL) has emerged as a powerful tool in this domain, offering adaptive and efficient solutions. The current literature on RL in project scheduling demonstrates significant advancements in site and factory production scheduling.

2.1 Site Scheduling

The application of RL in site scheduling is characterized by a variety of approaches aimed at addressing the dynamic and complex nature of construction environments. Kedir et al. [11] and Lee et al. [12] showcase how RL can be used to simulate and adapt to changing conditions on construction sites. The hybrid reinforcement learning-graph embedding network model proposed in [11] exemplifies an innovative approach to simulating complex construction planning environments. It shows the potential of RL in reducing computational burdens while establishing effective activity sequences and work breakdown structures. Similarly, [12] applies a digital twin-driven RL method for adaptive task allocation, indicating RL's capability to enhance realtime decision-making and efficiency in dynamic construction environments. This emphasis on adaptability and prompt responsiveness is similarly reflected in [13], which presents a novel method for generating Look-Ahead Schedules using RL. This method addresses the challenges of manual planning by offering a faster, more efficient approach to scheduling construction site activities.

2.2 Factory Production Scheduling

In the realm of factory production scheduling, RL is utilized to address the challenges of variability and the need for adaptability in manufacturing processes. Several studies highlight various aspects of how RL can improve efficiency and adaptability in factory settings [14-17]. Du and Li focus on automated assembly planning for robot-based construction, employing Deep Reinforcement Learning (DRL) in a re-configurable simulator to enhance assembly planning processes [14]. This approach aligns with [15] and [16], which also explore the dynamic nature of factory environments and how RL can be used to respond to changes in orders and resources. The comprehensive review in [17] of RL applications in production planning and control further underscores the versatility of RL in managing diverse aspects of manufacturing, including production scheduling, capacity planning, inventory and management.

2.3 **Point of departure**

While the reviewed literature on RL in project scheduling offers significant insights, it reveals a notable limitation: the lack of integration between site scheduling and factory production planning. To enhance overall operational efficiency and achieve the promised benefits of on-demand production and Just-in-time delivery, it is crucial to achieve a seamless integration of mobile factories and construction sites. By integrating these two components, RL can drive the evolution of project management towards more streamlined, efficient, and sustainable practices.

achieved.

3 Methodology

The research methodology followed in this research is centered around the development and validation of a RL algorithm. RL is a branch of machine learning that draws inspiration from the natural learning process. In RL, the behavior of an entity, known as an agent, is shaped by the outcomes of its previous actions. Positive outcomes reinforce certain behaviors, lending greater importance to those actions and the decisions leading up to them. RL builds upon the foundations of Markov Decision Processes [18] and stands apart from supervised learning in that it does not rely on labeled input-output pairs for learning [19]. A RL model formally comprises:

- A discrete set of states, S;
- A discrete set of possible actions for the agent, A; and
- A set of scalar rewards.

In this framework (shown in Figure 1), an agent selects actions based on the rewards previously received in similar states. The ultimate objective is to devise a policy π that maps states to actions in a way that maximizes the overall reward.



Figure 1. Reinforcement Learning Feedback Loop

The algorithm proposed in this research work aims to optimize operational policies in infrastructure projects that employ mobile factories for construction supply. The methodology uniquely combines logistical mobility with the complexities of production and assembly processes within the context of a mobile factory. This presents a novel and challenging environment for the application and exploration of RL techniques. In the following subsection, we will explain how the RL problem was formulated. These include, definition of environment, agent's action space, reward and penalty.

3.1 Environment

The simulation divides a construction project into sections of uniform length, with on-site assembly of building components produced by the mobile factory. The assembly unfolds in a linear fashion, advancing from section to section until the project's completion is

3.2 Agent

The agent operates as the mobile factory, commencing at the first section and advancing toward the terminal section. It continuously tracks its position relative to the project and the quantities of both produced and assembled components.

3.3 Actions

The agent's operational choices are determined on a daily basis, introducing an element of strategic timing to the simulation. The agent has a repertoire of actions that directly influence the environmental state:

- **Production**: Engaging in this action, the factory commits to the fabrication of building elements at a set rate. While in production mode, the factory's status is updated to reflect the new production count, and its location remains unchanged.
- **Movement**: Opting to move prompts the factory to transition to the next section. This phase is characterized by a cessation of production, which realistically simulates operational downtime during relocation.
- Idle: By choosing to idle, the factory does not produce or move. This inaction provides an opportunity for strategic timing, potentially waiting for more favorable conditions or to better align with other segments of the project.

3.4 Reward

The reward function in this project environment is designed to incentivize optimal scheduling and resource allocation. It includes the following components:

- **Project Completion Reward**: This substantial reward is granted upon the successful completion of the entire construction project, i.e., when all sections have met their assembly requirements and the project has reached its final stage. This reward reflects the ultimate goal of project completion.
- Milestone Reward: Awarded each time the project successfully meets the assembly requirements for a current section and progresses to the next. This reward is a key driver for phasewise project execution, encouraging the timely accomplishment of individual project segments. The milestone bonus not only acknowledges the completion of specific sections but also promotes a steady pace, ensuring that the project advances methodically from one stage to the next without unnecessary delays.

3.5 Penalty

The penalty function encapsulates various operational costs and risks, promoting efficient and strategic planning:

- **Duration Cost:** Emphasizes project time management, where shorter durations align with industry objectives. This cost accrues daily and escalates with extended project timelines.
- Factory Running Cost: Emphasizes the operational expenses associated with the daily functioning of the factory. This cost accrues continuously, reflecting the resource utilization and maintenance required to keep the factory operational.
- Factory Movement Cost: Underscores the expenditures associated with relocating the factory within the construction project area. This cost is incurred when the factory needs to be moved to a different location within the project site, often to align with the construction progress.
- Shipment Cost: Reflects the logistical complexity of material transportation from the factory to the construction site. This cost is quantified by the distance to the current project section and is enhanced by a predetermined factor, underscoring the value of logistical efficiency.
- **Inventory Cost**: Signals potential inventory management inefficiencies. This cost is activated when production outperforms assembly. The incurred cost, proportional to the imbalance and multiplied by a coefficient, advocates for a balance between production and assembly.
- Underproduction Cost: Underproduction Cost addresses the potential consequences of producing fewer components than required for the construction project. To mitigate this risk, additional resources may be needed, such as sourcing components from external suppliers or resorting to on-site production, often under urgent circumstances.

The penalty function complements the reward function by creating a balanced and comprehensive system of disincentives and incentives. This system encourages behaviors that are conducive to the overarching objectives of efficiency, cost-effectiveness, and timely delivery in construction project management.

4 Algorithm

Proximal Policy Optimization (PPO) [20] is selected for the mobile factory simulation. The PPO algorithm combines ideas from A2C (Advantage Actor-Critic) and TRPO (Trust Region Policy Optimization). It is wellregarded for its effective balance between exploration and exploitation, ensuring gradual improvements in decision-making. It operates by making incremental adjustments to its policies, which prevents drastic changes that could destabilize the learning process. This characteristic of PPO makes it particularly suitable for the mobile factory simulation, where decisions have a direct and significant impact on operational efficiency and project cost. The algorithm's ability to handle complex decision spaces and maintain steady progress is aligned with the requirements of coordinating production, assembly, and movement in the simulated environment.

5 Implementation

The implementation for the RL problem described utilizes the OpenAI Gym framework to create a custom environment, **FactoryEnv**, which simulates a mobile factory moving through different sections of a construction project. It is important to note that the values of parameters used in this setup are for illustrative purposes only. In a real-world project setting, users have the flexibility to customize these values according to specific project requirements. This customization capability ensures that **FactoryEnv** can be adapted to various construction scenarios, allowing for more accurate simulations and effective training of RL models tailored to the unique dynamics of each project. The environment is characterized by parameters:

Table 1. Environment parameters

Parameter	Description	Value
num_stops	The total number of	20
	sections the	
	infrastructure is divided	
	in	
parts_per_stop	The number of parts	10
	required at each section	
assembly_rate	The rate at which the	5 parts
	site assembles parts per	per day
	day	
production_rate	The rate at which the	8 parts
	factory produces parts	per day
	per day	
movement_time	The time it takes to	3 days
	move from one section	
	to the next	

Along with these operational parameters, the environment's behavior and agent's performance are influenced by a set of reward and penalty parameters, defined as follows:

Table 2 Reward and penalty parameters

DAILY_COST	Cost incurred for each day	-1
	of the project duration.	
FACTORY_R	Cost incurred for daily	-10
UNNING_COS	functioning of the factory	
T_PER_DAY		
MOVE COST	Additional cost for each	-20
PER DAY	day the relocation of the	
	factory	
SHIPMENT_C	Cost per unit of distance	-0.5
OST	between factory and	
	assembly location	
INVENTORY	Cost for parts produced in	-1
COST	excess of the assembly	
	requirements	
UNDERPROD	Cost incurred for parts	-5
UCTION_COS	underproduced relative to	
Т	the assembly requirements	
COMPLETIO	Reward granted upon the	1000
N_REWARD	successful completion of	
	the entire project.	
MILESTONE_	Reward for each project	30
REWARD	section completed	

Employing PPO via the stable baselines3 library, this study utilizes a multi-layer perceptron for simultaneous policy and value function approximation within a custom-defined FactoryEnv environment. The model's architecture and hyperparameters are meticulously calibrated: a linearly scheduled learning rate commencing at 1e-4, a discount factor at 0.99, a GAE (generalized advantage estimate) lambda at 0.95, and an entropy coefficient of 0.005. Additionally, the network architecture comprises dual-layered structures with 128 neurons each for both policy and value functions. Batch processing is implemented with 2048 steps per batch, balancing computational efficiency with learning efficacy. The model's initialization incorporates these parameters, while TensorBoard integration facilitates detailed progress monitoring.

6 Results

In the presented results, we observe the performance metrics of a RL model over the course of training, measured across one million timesteps. Figure 2 delineates the trajectory of the training loss, a key indicator of the model's prediction accuracy regarding future rewards. The plot reveals an initial phase with a high variance in loss, indicative of the model's exploratory learning and parameter optimization. As training progresses, a clear downward trend emerges, culminating in a stable, low loss value, which suggests that the model's predictions have become more reliable and consistent. Figure 3 showcases the evolution of the average reward during the model's evaluation phase. The initial negative values represent suboptimal decisionmaking by the model. However, an enhancement is noted as the average reward increases, eventually reaching a plateau, demonstrating significant learning and policy improvement throughout the training process.



Figure 2. Training Loss Over Time



Figure 3. Evaluation of Average Reward

7 Validation

7.1 Single Environment Validation

In assessing the performance of our PPO-based RL model, we employed a quantitative validation strategy that entailed a comprehensive analysis of reward distributions. This strategy involved executing a random policy across 100,000 episodes within a consistent environmental setting of **FactoryEnv**. The objective was to establish a baseline distribution of rewards that could be leveraged as a comparative measure against the deterministic output of our trained RL model.

The histogram depicted in Figure 4 illustrates the frequency of total rewards obtained from the random policy across the 100,000 episodes. A dashed black line represents the reward achieved by our trained RL model, and a dashed red line denotes the maximum reward attained by the random policy throughout its trials. The RL model's reward, markedly higher than the random policy's mean and maximum reward, underscores the learning algorithm's success in optimizing decision-making to enhance reward outcomes.



Figure 4. Reward Distribution of Random Policy and Comparison with RL Policy over the Same Setting

7.2 Diverse Environments Validation

To validate the robustness of the trained RL model, a comparative study was conducted against a baseline random policy. The comparison was done across a set of 100 diversified scenarios within a simulated environment, specifically designed to mimic a factory setting (FactoryEnv). Each scenario presented a unique configuration by varying the assembly rate, a critical parameter influencing the environment's dynamics. The assembly rates for each scenario were sampled from a normal distribution with a mean of 5 and a standard deviation of 1, ensuring a spectrum of operating conditions to challenge the robustness of the model. The trained RL model, developed using the PPO algorithm, was compared against the random policy in these scenarios to assess its adaptability and performance. The key metric for comparison was the total cumulative reward achieved by the end of each episode, serving as a proxy for the model's decision-making quality and efficiency.

The resulting performance, as shown in Figure 5, indicates a significant and consistent outperformance of the trained RL model over the random policy across all tested scenarios. The RL model achieved higher cumulative rewards in each individual case, demonstrating not only the ability to generalize across different settings but also the robustness of its learned policy.



Figure 5. Comparison of RL Model versus Random Policy over Different Settings

8 Conclusion

This study contributes a novel RL-based scheduling model for optimizing the operation of mobile factories in infrastructure projects. It encompasses a comprehensive method for considering an array of performance indicators at the project level, including production and inventory costs, project duration, and shipping expenditures. As such, the reward and penalty parameters are designed to encourage cost-effectiveness and timely delivery of prefabricated elements. This aligns with the very motivation for applying mobile factories in construction projects - on-demand production and justin-time delivery. Moreover, our approach exhibits remarkable flexibility, effectively adapting to a wide spectrum of production environments characterized by varying rates, the mobility of production facilities, and differing operational states such as idleness. Thus, the proposed method presents a holistic decision-making tool that can empower factory managers to optimize project execution strategies.

However, this research has some limitations. The environmental and reward parameters employed within the simulated setting may not entirely capture the complexity of real-world projects. For example, the COMPLETION_REWARD value could include various dimensions like the effort required, the time to completion, resources needed, and the complexity of tasks. As a next step, the practical application and validation of our approach in a real-world project setting will be imperative to ascertain its effectiveness and to fine-tune the model parameters for enhanced realism and applicability. By bridging the gap between theoretical modeling and practical implementation, we anticipate that our RL-based approach will offer tangible benefits in the management of factory and construction operations.

Acknowledgements

This work was carried out within the research project "Smart Mobile Factory for Infrastructure Projects (SMF4INFRA)" and supported by the Swiss National Science Foundation [grant no. 204852].

References

- [1] Seelow A. The Construction Kit and the Assembly Line—Walter Gropius' Concepts for Rationalizing Architecture. In Arts, pages 95, 2018.
- [2] Safdie M. Beyond Habitat, volume 978-0262690362. MIT Press, 1973.
- [3] Herzog T. Expodach: Roof Structure at the World Exhibition, Hanover 2000, volume 978-3791323824. Prestel Pub, 2000.
- [4] Haukka S. and Lindqvist M. Modern Flying Factories in the Construction Industry, Master's Thesis, Lulea University of Technology, Lulea, Sweden. 2015.
- [5] Alvarez M. et al. The BUGA Wood Pavilion -Integrative Interdisciplinary Advancements of Digital Timber Architecture. In Proceedings of the 39th ACADIA Conference 2019, pages 490–499, Austin, USA, 2019.
- [6] Alix T., Benama Y., and Perry N. A framework for the design of a Reconfigurable and Mobile Manufacturing System. Procedia manufacturing, 35:304–309, 2019.
- [7] Wagner H. J., Alvarez M., Kyjanek O., Bhiri Z., Buck M., and Menges A. Flexible and transportable robotic timber construction platform–TIM. Automation in Construction, 120:103400, 2020.
- [8] Martínez S., Jardón A., Victores J. G., and Balaguer C. Flexible field factory for construction industry. Assembly Automation, 33(2):175–183, 2013.
- [9] Ahn S. J. et al. Integrating off-site and on-site panelized construction schedules using fleet dispatching. Automation in Construction, 137:104201, 2022.
- [10] Wang Z. and Hu H. Improved precast productionscheduling model considering the whole supply chain. Journal of Computing in Civil Engineering, 31(4):04017013, 2017.
- [11] Kedir N. S., Somi S., Robinson Fayek A., and Nguyen P. H. D. Hybridization of reinforcement learning and agent-based modeling to optimize construction planning and scheduling. Automation in Construction, 142:104498, 2022.
- [12] Lee D., Lee S., Masoud N., Krishnan M. S., and Li V. C. Digital twin-driven deep reinforcement learning for adaptive task allocation in robotic construction. Advanced Engineering Informatics, 53:101710, 2022.
- [13] Soman R. K., and Molina-Solana M. Automating look-ahead schedule generation for construction using linked-data based constraint checking and reinforcement learning. Automation in Construction, 134:104069, 2022.

- [14] Du Y., and Li J.-q. A deep reinforcement learning based algorithm for a distributed precast concrete production scheduling. International Journal of Production Economics, (2023): 109102.
- [15] Shiue Y.-R., Lee K.-C., and Su C.-T. Real-time scheduling for a smart factory using a reinforcement learning approach. Computers & Industrial Engineering, 125:604-614, 2018.
- [16] Shi D., Fan W., Xiao Y., Lin T., and Xing C. Intelligent scheduling of discrete automated production line via deep reinforcement learning. International Journal of Production Research, 58(11):3362-3380, 2020.
- [17] Esteso A., Peidro D., Mula J., and Díaz-Madroñero M. Reinforcement learning applied to production planning and control. International Journal of Production Research, 61(16):5772-5789, 2023.
- [18] Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction, 2nd ed. MIT Press, Cambridge, Massachusetts, 2018.
- [19] Kaelbling L. P., Littman M. L., and Moore A. W. Reinforcement learning: A survey. Journal of Artificial Intelligence Research, 4:237–285, 1996.
- [20] Schulman J., Wolski F., Dhariwal P., Radford A., and Klimov O. Proximal policy optimization algorithms. arXiv preprint:1707.06347, 2017

Semantic annotation of images from outdoor construction sites

Layan Farahat¹ and Ehsan Rezazadeh Azar¹

¹ Faculty of Engineering and Architectural Science, Toronto Metropolitan University, Canada <u>layan.farahat@torontomu.ca</u>, <u>ehsan.azar@torontomu.ca</u>

Abstract –

Valuable information is embedded in construction images which can be used for different construction engineering and management purposes. The availability of low-cost cameras and robust artificial intelligence methods has increased the use of imaging technology in construction sites. However, these rich data sources are not often used to their full potential due to subjective documentation, leading to potentially overlooking valuable content. This study proposes an ensemble approach that utilizes deep learning techniques for object recognition, pixel-level segmentation, and text classification to annotate images from outdoor construction scenes at medium (ongoing activities) and high (project type) levels. Experimental results demonstrate the potential of this approach by achieving a 70% overall recall rate.

Keywords -

Annotation; construction images; construction management; deep learning; object detection; semantic segmentation.

1 Introduction

The construction industry has vastly employed image and video recording technologies, and a growing number of research projects investigate methods for better utilization of this valuable data [1]. This trend has been facilitated by the emergence of low-cost capturing systems, robust computer vision methods, and the flexibility in imaging offered by the UAV systems [1], [2]. The resulting image databases, typically organized through manual labels and metadata, could serve various project management purposes, such as progress tracking, quality inspection, safety audits, and training [1]. However, the unrestrained accumulation of these visual data poses challenges in annotation and retrieval, which potentially results in underutilizing valuable information [3]. Unlike captioning of generic images, the technical complexities of construction images demand expert knowledge for a practical annotation [4]. Past research focused on feature extraction and object detection to

enhance image annotation [5]. Some explored equipment poses [6] and interactions [7], while others provided semantic annotation for construction videos through spatiotemporal data interpretation of equipment motion [4]. However, current annotations mainly cover appearing resources and their interactions, lacking identification of ongoing activities without visible actors. For examples, methods were developed to caption an image as "a dozer is pushing the soil on the ground" [7] or "the excavator is loading dirt to the truck". [14]. But there is a gap to provide useful annotation where there is no actor, i.e. equipment and workers, in the image.

This research introduces an innovative ensemble method utilizing deep learning for low-level, mediumlevel, and high-level annotations of outdoor construction images, defining objects and materials as low-level, activities as medium-level, and project types as highlevel annotations. The proposed approach integrates deep learning-based object detection, semantic segmentation, and text classification, focusing on outdoor construction images to demonstrate the potential advancements of this approach. Since indoor and outdoor construction scenery includes distinct elements and resources, this research only focuses on outdoor settings.

2 Literature Review

The construction industry has experienced widespread application of digital imaging since the 1990s, leading to a substantial increase in image-making rates [8]. Initial efforts focused on feature extraction for image retrieval based on material patterns and shapes, incorporating metadata like location and date [9]. Feature-based object detection techniques were used to detect construction equipment and workers, but they had mediocre performance and faced limitations in multiclass detections [10]. Deep convolutional neural networks (DCNN) addressed these limitations, offering multiclass recognition, automated feature extraction, and improved detection performance [11]. Some of the DCNN models, like Region-Based CNN (R-CNN) [12] and You Only Look Once (YOLO) [13], were employed in construction contexts for different tasks like progress monitoring [14]
and safety management [15]. DCNN models were also employed for semantic segmentation at pixel-level recognition. Methods like Mask R-CNN [16] and DeepLab V3+ were used for progress estimation [17], and understanding of construction visual data [18]. Efforts were made to produce semantic annotations by combining object recognition and Bayesian belief networks [4]. Long Short-Term Memory (LSTM) architecture [19] could be employed in conjunction with a CNN model for extracting semantic information from images. This approach contributed to safety management [20] and descriptive caption generation for equipment activities and interactions [7] by analyzing the interaction between two objects. Recent developments to caption construction images focus on actors and their action(s), and generally produce a caption as "an equipment/worker is doing something" [7, 14, 28]. Many construction images, however, do not contain equipment/workers and only capture a snapshot of the progressing construction work. This research seeks to address this gap by integrating detected objects and materials/surfaces to annotate construction images with ongoing activities and project types, even in cases where the main actors, such as equipment and workers, are not present in the images.

3 Research Methods

This study proposed an ensemble model, integrating three machine-learning models to predict general construction activities and project types in outdoor construction site images. The approach involved two deep-learning models for object detection and surface segmentation, and a third text classifier using a neural network for predicting construction activities and project type/phase. In this approach, the results of object detection and semantic segmentation, i.e. low-level annotations, are fed to the text classifier to predict medium-level and high-level annotations. Each model was trained and tested on separate datasets. Figure 1 shows a flowchart of the developed approach with detailed steps outlined in the subsequent sections, including results, discussion, and conclusion.



Figure 1. Flowchart of the Proposed Framework

3.1 Data Collection

This study used images of diverse construction projects collected from open online platforms like Google Images, YouTube videos, Pexels, and Pixabay to train and test the developed models. Some of the images were taken by the authors from construction sites using smartphones. Separate datasets were created for each model, tailored to their specific objectives. For the object detection model, 5,260 images featuring various construction equipment were collected. A dataset of 321 images was collected for the semantic segmentation model. The text classifier dataset comprised 545 images representing different types and phases of outdoor construction projects.

3.2 Object Detection Module

The object detection module was trained using 4,990 images (95% training and 5% validation) containing various construction equipment and workers. The training process involved preparing raw data, resizing images, and labelling objects using the Image Labeler tool by MATLAB [24]. Eleven classes were manually labelled with bounding boxes, including nine construction equipment types (excavator, bulldozer, scraper, off-road truck, truck, loader, compactor, grader, concrete mixer), in addition to two generic classes of humans and regular cars. The YOLO v4 with DarkNet53 as the backbone, pre-trained on the COCO public dataset, was chosen for its performance and processing time. The training was conducted on a desktop computer with 32 GB RAM, a 4.7 GHz Intel Core CPU, and an NVIDIA GeForce RTX 3060 GPU. Hyperparameters included a gradient decay factor of 0.9, a learning rate 0.001, and data augmentation with random horizontal reflection and scaling. Batches of 8 images were processed over 85 epochs using the MATLAB® 2022a software Deep Learning toolbox [24] for training and testing.

3.3 Semantic Segmentation Module

The second DCNN model employed semantic segmentation to classify image elements, like materials and construction surfaces, that might be impractical to be recognized by the object detection. This model classifies objects at the pixel level, with training and test datasets manually annotated using the same Image Labeler tool utilized for the object detection module. Twenty-one classes, including concrete, formwork, glass, tower crane, human, bin, dirt, equipment, lumber, asphalt, steel, rebar, scaffold, bitumen, aggregate, rail, waterproofing, pipe, curing blanket, brick, and other (such as sky, mountains, trees), were defined for labeling major elements in construction sites. The DeepLab V3+ ResNet50 [25] model was retrained using a labeled dataset of 220 images. The Deep Learning toolbox of MATLAB®

2022a software was used for training [24], adjusting stochastic gradient descent with a learning rate of 0.01 and a momentum value of 0.9. Batches of six images and 100 epochs were utilized, with data augmentation involving random X and Y translations and right/left pixel reflections. The training process was carried out on the same desktop computer mentioned earlier.

3.4 Text Classification Module

The last model in the system is the text classifier, which analyzes the outcomes of the object detection and semantic segmentation modules to predict medium-level (activities) and high-level (project types) annotations. This module was trained using a dataset of 385 images from a diverse set of construction projects. The objects, associated activities, and project phases/types were manually extracted from these images based on the authors' construction expertise. Then, these extracted textual data were manually converted into binary format, with 1 representing the presence and 0 indicating the absence of objects, surfaces, activities, or project types in the images. The dataset encompassed seven types of construction projects: building sub-structure, building super-structure, bridge construction, road construction, heavy construction (i.e., tunnel construction and subway construction), railway construction, and pipeline construction, and fifteen activities, including rebar installation, steel erection (structural), formwork shuttering and removal, concrete work, lumber work, earth hauling, paving work, material lifting, earthwork (i.e., loading, compacting, and removing dirt), excavation, masonry work, glazing, rail work, waterproofing, and pipework.

Since the neural network (NN) method has shown promising results in various construction-related analysis, such forecasting labor productivity [21], analyzing accidents [22], and project delay risks [23], it was also used in this research. NN models with multiple hidden layers and varying neuron numbers were developed using RapidMiner Studio [27], which offers various operators for data retrieval, model evaluation, and algorithms. The study trained and tested the model across nine configurations, with the most effective performance observed using two hidden layers, each with five neurons, and a Rectifier activation function over 100 epochs. The module training was conducted on a laptop with 16 GB RAM, 2.3 GHz Intel Core i7-11800H, and an NVIDIA GeForce RTX 3060 GPU.

3.5 Ensemble Model

The ensemble model combines all the mentioned modules, in which the trained object detection and semantic segmentation models extract low-level data, i.e. detected objects and material/surfaces, and pass them as input to the text classifier for medium-level and high-level annotations.

4 Experimental Results

The developed modules and the ensemble model were assessed in four phases: 1) the object detection module evaluation, 2) the semantic segmentation module evaluation, 3) the text classifier evaluation, and 4) the ensemble model evaluation, which encompasses all the modules and possible propagation of errors.

4.1 Object Detection Module Results

The object detection module was evaluated with 270 images from various construction site scenery gathered from the same sources as the training dataset. Model evaluation metrics included precision and recall rates. Precision is denoted as the proportion of correctly predicted positive instances (true positives) out of all predicted positives (true positives + false positives). Recall as the proportion of correctly predicted positives) out of all actual positive instances (true positives) out of all actual positive instances in the dataset. The 11 object classes were evaluated individually in addition to the overall performance. Table 1 presents the performance of each class by the trained YOLO V4 – DarkNet50 classifier.

Table 1. Performance metrics for each class

Classes	Recall	Precision
Excavator	80.28%	93.44%
Bulldozer	91.43%	87.67%
Scraper	76.92%	78.95%
Off truck	87.93%	72.86%
Truck	71.83%	77.27%
Loader	61.36%	57.45%
Human	54.10%	94.29%
Compactor	76.36%	66.67%
Grader	91.43%	78.05%
Car	84.00%	72.41%
Concrete Mixer	78.57%	61.11%

The precision rates for the human and excavator classes exceeded 90%, while other equipment classes like grader, truck, scraper, and bulldozer achieved precision rates higher than 75%. The object detection classifier demonstrated an overall recall rate of 77.7% and an overall precision rate of 76.4% in detecting construction equipment and workers.

4.2 Semantic Segmentation Module Results

The semantic segmentation model's performance was assessed using a test dataset of 101 images collected from open online sources for an outdoor construction setting. The test images were manually labeled with the 21 previously mentioned classes to develop the ground truth file for model evaluation. The mean Intersection over Union (mIoU) metric was the evaluation metric used, which calculates the overlap between the ground truth and the classifier output, divided by the area of their union. The Deeplab V3+ with ResNet50 has achieved a weighted mIoU of 54.3%. Table 2 presents the results for the 21 classes in test images. Some classes, such as tower crane, dirt, human, steel, curing blanket, and equipment, achieved mIoU rates higher than or equal to 50%. Others, including concrete, rebar, lumber, and brick, had mIoU rates ranging from 40% to 50%. Classes like formwork, glass, asphalt, pipe, scaffold, bitumen, aggregate, and bin exhibited mIoU values between 17% and 39%. However, the model exhibited poor performance in classes like rail and waterproofing, potentially due to their low number of sample pixels in the training dataset. These results show the model's strengths and limitations, which are further detailed in the discussion section.

Table 2. Semantic segmentation classes' mIoU

Classes	mIoU	Classes	mIoU
Concrete	45%	Steel	51.3%
Formwork	27.6%	Rebar	46.6%
Glass	26.8%	Scaffold	16.8%
Tower crane	56.2%	Bitumen	13.1%
Human	50%	Aggregate	25%
Other	69.3%	Rail	10.6%
Bin	20.7%	Brick	49.1%
Dirt	67.5%	Waterproofing	2.9%
Equipment	60%	Pipe	17%
Lumber	43.7%	Curing_Blanket	68.2%
Asphalt	39%	Steel	51.3%

Despite the varying mIoU rates across the 21 classes, the semantic segmentation model mainly aimed to identify major objects/surfaces in construction images for input to the text classifier. In this study, the top 7 detected classes were selected for the text classifier based on their pixels counts. A "commonality percentage" assessment method determined the accuracy of the top 7 segmented classes in representing the actual classes in test images. Figure 2 illustrates how the commonality percentage of actual classes was calculated among the top 7 detected classes. The average commonality percentage across 101 test images was 85.9%.



Ground truth classes: Concrete, Formwork, Rebar,

Top 7 detected classes: Concrete, Formwork, Other, Dirt, Curing_Blanket, Human, Lumber Rebar, Curing_Blanket, Human

This image shows 5 out of 6 actual classes among the top 7 detected classes. Therefore, the percentage = $\frac{5 \times 100\%}{5}$ = 83.3%

Figure 2. Commonality percentage of existing actual classes

4.3 **Text Classifier Module Results**

The text classifier model was evaluated on 160 test images using precision and recall rates as performance metrics. The aim of this test setup was to assess the performance of the text classifier alone to assess how well it can annotate activities and project phase/type based on the correct appearing objects and surfaces. Thus, the ground truth objects and surfaces were given to the models. In other words, the object detector and semantic segmentation classifiers were not used to feed the text classifier. The text classifier achieved an overall precision rate of 92.1% and an overall recall rate of 86.7%. Table 3 and Table 4 present the results for medium-level and high-level predictions of the NN text classifier. Most activity predictions achieved high precision and recall rates exceeding 90%. Road, rail, and pipe construction showed higher results than other project types among the high-level predictions.

Table 3. Medium-level (activities) prediction results

Class	Dragision	Dagall
Class	FIECISION	Recall
Rebar Installation	98.1%	100%
Steel Erection	88.2%	93.8%
Formwork Shuttering and	98.2%	100%
Removal		
Concrete Work	95.3%	99%
Lumber Work	75%	81.8%
Earth hauling	100%	86%
Paving work	75%	60%
Material Lifting	100%	100%
Earthwork	100%	86.5%
Excavation	100%	100%
Masonry Work	66.7%	100%
Glazing	80%	57.1%
Rail Work	100%	86.7%
Waterproofing	100%	87.5%
Pipework	92.9%	100%
Overall Medium-Level	95.9%	94.6%

Table 4 High-level (project type) prediction results

Class	Precision	Recall
Building Sub-structure	51.6%	53.3%
Building Super-structure	86.5%	59.3%
Bridge Construction	33.3%	8.3%
Road Construction	90.3%	96.6%
Heavy Construction	83%	50%
Railway Construction	100%	93.3%
Pipeline Construction	100%	90%
Overall High-Level	80.2%	65.6%

4.4 Ensemble Model Results

The same 160 images were used to test the NN text classifier to evaluate the entire system (ensemble model). The object detection and segmentation classifiers processed the images, and their results were passed to the text classifier. The ensemble model achieved an overall precision of 60.23% and a recall of 70%. Table 5 and Table 6 detail the medium-level and high-level prediction results. Figure 3 shows a sample image from the test dataset. It encompasses a building superstructure project, formwork activity, concrete work, lumber, earth hauling, and material lifting. The model correctly predicted the project type and four activities. However, rebar installation was also classified, while lumber work was missed, resulting in 5 true positive instances, 1 false positive instance, and 1 false negative. Thus, the ensemble model had a precision and recall rates of 83.33% in this sample.



removal, concrete work, lumber work, earth hauling, material lifting **Predicted project type:** Building Superstructure **Predicted activities:** Rebar work, formwork shuttering and removal, concrete work, earth hauling, material lifting

Figure 3. Ensemble model test image sample

Among the medium-level classifications, there were different performance trends. Concrete work, material lifting, and earthwork exhibited high-performance rates. Some activities, such as rail work, paving work, pipework, and earth hauling, demonstrated higher precision than recall, meaning that the model made fewer predictions than the activities in the ground truth dataset, although most were correct. On the contrary, rebar installation, steel erection, formwork shuttering and removal, and lumber exhibited recall rates surpassing precision, which means that the model made many predictions for these activities; however, not all were correct. Waterproofing and glazing showed lower performance, with precision and recall rates below 33%, and masonry work was not detected. Among high-level classifications, railway and pipeline constructions demonstrated precision values of 100%, yet their recall values were only 33.3% and 10%, respectively. Road construction and building super-structure followed, achieving precision rates of 95.65% and 63.89%, respectively. Building sub-structure, bridge, and heavy construction obtained less than 36% precision values, indicating a lower performance than other project types.

Table 5. Medium-level (activities) prediction results

Class	Precision	Recall
Rebar Installation	60.98%	94.34%
Steel Erection	40%	75%
Formwork Shuttering and	51.85%	100%
Removal		
Concrete Work	78.57%	97.1%
Lumber Work	30.95%	59.1%
Earth hauling	76.92%	47.62%
Paving work	60%	30%
Material Lifting	80.85%	95%
Earthwork	87.1%	72.97%
Excavation	66.67%	51.28%
Masonry Work	0.00%	0.00%
Glazing	28.57%	28.57%
Rail Work	100%	46.67%
Waterproofing	33%	25%
Pipework	85.71%	46.15%
Overall Medium-Level	63.77%	78.2%

Table 6 High-level (project type) prediction results

Class	Precision	Recall
Building Sub-structure	35.59%	70%
Building Super-structure	63.89%	42.59%
Bridge Construction	11.54%	25%
Road Construction	95.65%	75.9%
Heavy Construction	12.5%	10%
Railway Construction	100%	33.33%
Pipeline Construction	100%	10%
Overall High-Level	48.1%	47.5%

The ensemble model exhibited a noticeable performance reduction due to errors from the object detector and semantic segmentation compared to standalone text classifier. A sensitivity analysis was conducted in two scenarios to assess the impact of each DCNN module. In the first scenario, only the object detection classifier was used with the ground truth segmentations, while in the second scenario, only the semantic segmentation classifier processed images. Based on the ground truth dataset for the text classifier, the analysis revealed that semantic segmentation had a more pronounced effect on the performance of the overall system than the object detection. Scenario two exhibited a 6% drop in precision and a 19.38% drop in recall compared to scenario one, as shown in Figure 4.



Figure 4. The difference between the two scenarios and the base scenario

5 Discussion

The outcomes obtained from the three modules showed promising results in improving the annotation of outdoor construction images, by providing appearing objects, ongoing activities, and project types. However, these results also showed challenges and limitations encountered by the classifiers in specific instances. The subsequent four subsections discuss these challenges and propose potential solutions to enhance the practical application of the proposed approach.

5.1 Object Detection Module

The object detection classifier resulted in recall and precision rates of 77.7% and 76.6% across 11 object classes, respectively. Despite this promising performance, there were instances of missed or misclassified objects. For example, truck, off-truck, and concrete mixer classes exhibited lower precision rates than major equipment like excavators and bulldozers. Some misclassifications included regular dump trucks identified as off-road dump trucks and concrete mixers classified as regular trucks. Similarities in the front view of this equipment from different manufacturers contributed to such misclassifications. Misclassifications were observed for loaders, roller compactors, and graders due to visual similarities, specifically from their rearview. Moreover, image quality factors, such as resolution and occlusion, which are common in busy construction sites, contributed to increased false negatives, notably in the human class. Enhancing the training dataset with more images featuring diverse equipment types and poses could potentially improve recall and precision.

5.2 Semantic Segmentation

The mIoU metric was utilized to evaluate the performance of the semantic segmentation classifier in detecting elements in outdoor construction sites. Across 101 test images, the model achieved an mIoU rate of 54.3% in segmenting 21 classes. However, misclassifications occurred, especially when elements shared similar visual features. The model confused lumber, formwork, and wooden scaffolding systems due to material and texture similarities. Waterproofing was misclassified as asphalt or bitumen, and rail was often detected as steel due to similarities in their material composition. Additional misclassifications included aggregate as dirt and bitumen as asphalt. Pipe was detected as dirt or equipment. This module's performance relies highly on the training dataset; thus, it is suggested to expand the training dataset and enhance image quality.

5.3 Text Classifier

The standalone NN classifier achieved an overall precision rate of 92.1% and a recall rate of 86.7%. However, among high-level classifications, heavy construction, building sub-structure, and bridge construction exhibited lower precision and recall rates than other project types. These misclassifications are due to visual element similarities with other project types, as illustrated in Figure 5 (a) and Figure 5 (b), where a heavy construction project (road underpass) and a sub-structure construction of a high-rise shared visual elements like excavators, concrete, and dirt, leading to confusion during classification. The text classifier faced challenges in differentiating instances of these classes, leading to higher classification error rates. Heavy construction was frequently misclassified as building sub-structure, building super-structure, or road construction. Similarly, building sub-structure was often confused with bridge construction, heavy construction, building superstructure, and road construction. Moreover, bridge construction was mixed up with building super-structure and building sub-structure classes.



Figure 5. (a) Heavy Construction (underpass construction) and (b) Building sub-structure

The model struggled to predict paving, masonry, and lumber work in the medium-level classification. For example, the classifier tended to overpredict lumber work, particularly when lumber was stored on the construction site (e.g., for formwork or lagging) without apparent lumber work in the image. Paving work was mainly classified when bitumen or asphalt was present in road project images. However, when a road made of asphalt appears in the background of a project, as seen in Figure 6 in a building sub-structure project, the classifier predicts paving work. Furthermore, the presence of common road construction equipment, including an excavator and dump trucks, in this image further confused the classifier. Similarly, some images included adjacent buildings made of bricks (and other building elements) in the project's background, which is not part of the project. The classifier could mistakenly identify masonry work in such cases, leading to more false positives and reducing the precision rate.



Figure 6. Asphalt roads beside a building substructure project

5.4 Ensemble Model

The integration of the three modules had impacted the overall performance of the ensemble model. For example, it struggled to predict pipeline construction projects, achieving a 10% recall rate, due to the poor performance of semantic segmentation in classifying "pipes" which achieved an mIoU of 16.9%. Consequently, the text classifier could not classify a pipeline project type due to the inaccurate low-level classification. Similarly, masonry work predictions were affected as the text classifier predicted masonry work where the "brick" class was present among the data, but buildings made of bricks in the project surroundings led to such misclassifications. Thus, the precision rate for masonry work was consequently impacted. The semantic segmentation modules also achieved mIoU of 49.1% for "brick" detection, further contributing to the ensemble model's overall low performance.

The sensitivity analysis highlighted the substantial impact of semantic segmentation on the ensemble model's performance; however, developing a robust semantic segmentation model was not the primary goal of this study. It should be mentioned that a properly trained model can result in mIoUs of up to 65%, as shown in recent studies [18]. Inherent issues in annotating project types in images exist, as an image may include limited visual information, making it challenging even for experts to accurately determine the project type without additional context, as shown in Figure 7.



Figure 7. A retaining wall under construction with an unclear project type

6 Conclusion

This study introduces an ensemble model utilizing computer vision and machine learning to annotate outdoor construction images with activities and project types. Two DCNN classifiers for object detection and semantic segmentation were trained to detect key elements and surfaces in outdoor construction scenes. These classifiers initially process construction images to detect various construction elements, materials, and equipment. The identified objects and surfaces are then fed to a trained NN text classifier to predict construction activities and project phases/types. The results showed promising performance, achieving a precision rate of 60.23% and a recall rate of 70% for predicting 15 construction activities and seven project types. The ensemble model offers potential improvements in automating image documentation and retrieval in the construction industry, with suggestions to enhance performance by increasing training datasets. However, one of the limitations of this study is to predict detailed activities due to the absence of a temporal dimension, relying on still images only. Future research may explore dynamic datasets, such as videos, to address this limitation. Additionally, a multitask vision language pretraining approach, such as Bootstrapping Language-Image Pre-training (BLIP) [27], could be explored to develop image and text classifiers simultaneously in future work.

7 References

[1] Paneru, S., & Jeelani, I. Computer vision applications in construction: Current state, opportunities & challenges. *Autom. Constr.*, 132, 103940, 2021.

- [2] Lopes Amaral Loures, L., & Rezazadeh Azar, E. Condition Assessment of Unpaved Roads Using Low-Cost Computer Vision–Based Solutions. *Journal of Transportation Engineering, Part B: Pavements*, 149(1), 04022066, 2023.
- [3] Ma, J. W., Czerniawski, T., & Leite, F. An application of metadata-based image retrieval system for facility management. *Adv. Eng. Inf.*, 50, 101417, 2021.
- [4] Rezazadeh Azar, E. Semantic annotation of videos from equipment-intensive construction operations by shot recognition and probabilistic reasoning. J. Comput. Civ. Eng., 31(5), 04017042, 2017.
- [5] Xiao, B., & Kang, S. C. Development of an image data set of construction machines for deep learning object detection. J. Comput. Civ. Eng., 35(2), 05020005, 2021.
- [6] Kim, J., Chi, S., & Kim, J. 3D pose estimation and localization of construction equipment from single camera images by virtual model integration. *Adv. Eng. Inf.*, 57, 102092, 2023.
- [7] Wang, Y., Xiao, B., Bouferguene, A., Al-Hussein, M., & Li, H. Vision-based method for semantic information extraction in construction by integrating deep learning object detection and image captioning. *Adv. Eng. Inf.*, 53, 101699, 2022.
- [8] Brilakis, I., & Soibelman, L. Content-based search engines for construction image databases. *Automation in Construction*, 14(4), 537-550, 2005.
- [9] Brilakis, I., & Soibelman, L. Multimodal image retrieval from construction databases and modelbased systems. J. Constr. Eng. Manage., 132(7), 777-785, 2006.
- [10] Rezazadeh Azar, E., & McCabe, B. Automated visual recognition of dump trucks in construction videos. J. Comput. Civ. Eng., 26(6), 769-781, 2012.
- [11] Pal, A., & Hsieh, S. H. Deep-learning-based visual data analytics for smart construction management. *Automation in Construction*, 131, 103892, 2021.
- [12] Girshick, R., Donahue, J., Darrell, T., & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE CVPR conference*, pages 580-587, 2014.
- [13] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE CVPR conference*, pages 779-788, 2016.
- [14] Chen, C., Xiao, B., Zhang, Y., & Zhu, Z. Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition. *Autom. Constr.*, 146, 104702, 2023.
- [15] Wu, S., Hou, L., Zhang, G. K., & Chen, H. Realtime mixed reality-based visual warning for

construction workforce safety. Autom. Constr., 139, 104252, 2022.

- [16] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961-2969, 2017.
- [17] Wang, Z., Zhang, Q., Yang, B., Wu, T., Lei, K., Zhang, B., & Fang, T. Vision-based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase. J. Comput. Civ. Eng., 35(1), 04020056, 2021.
- [18] Wang, Z., Zhang, Y., Mosalam, K. M., Gao, Y., & Huang, S. L. Deep semantic segmentation for visual understanding on construction sites. *Computer -Aided Civil and Infrastructure Engineering*, 37(2), 145-162, 2022.
- [19] Hochreiter, S., & Schmidhuber, J. Long short-term memory. Neural computation, 9(8), 1735-1780. 1997.
- [20] Ding, L., Fang, W., Luo, H., Love, P. E., Zhong, B., & Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in construction* 86, 118-124, 2018.
- [21] Heravi, G., & Eslamdoost, E. Applying artificial neural networks for measuring and predicting construction-labor productivity. J. Constr. Eng. Manage., 141(10), 04015032, 2015.
- [22] Gerassis, S., Martín, J. E., García, J. T., Saavedra, A., & Taboada, J. Bayesian decision tool for the analysis of occupational accidents in the construction of embankments. *J. Constr. Eng. Manage.*, 143(2), 04016093, 2017.
- [23] Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. Machine learning algorithms for construction projects delay risk prediction. J. Constr. Eng. Manage., 146(1), 04019085, 2020.
- [24] The MathWorks Inc. MATLAB R2022b: 9.13, Natick, Massachusetts: *The MathWorks Inc.* https://www.mathworks.com, 2022.
- [25] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In *Proceedings of* the IEEE CVPR conference, pages 770-778, 2016.
- [26] Mierswa, I, and Klinkenberg, R. "RapidMiner." *RapidMiner Inc.*, https://rapidminer.com, 2018.
- [27] Li, J., Li, D., Xiong, C., & Hoi, S. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *Proceedings of International Conference on Machine Learning*, pages 12888-12900, 2022.
- [28] Xiao, B., Wang, Y., & Kang, S. C. Deep learning image captioning in construction management: a feasibility study. J. Constr. Eng. Manage., 148(7), 04022049, 2022.

Structural design in the era of digital twins. A case study

Héctor Posada¹, Rolando Chacón¹ and Carlos Ramonell¹

¹Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, Spain <u>hector.posada@upc.edu, rolando.chacon@upc.edu, carlos.ramonell@upc.edu</u>

Abstract -

Digital Twins (DT) represent a technology that provides a way to encompass useful information of built assets for the sake of productivity enhancement in the AECO sector. Its adoption in the industry is still at early developmental stages. Significant strides have been made in establishing frameworks and workflows for effective DT implementation. This paper analyzes a case study in which a DT of a built bridge is developed during the particular episode of a load test, to understand how structural design can benefit (and potentially adapt) to the specific demands this innovative technology may generate. The bridge under twinning was already built and its design was neither determined nor influenced by the fact that load testing procedures were meant to be twinned. By identifying the link between the development of the DT of the bridge and its design, a conceptual diagram illustrates the key components to consider during the structural design phase. The main objective of this diagram is to present an encompassed vision of design that already accounts for the future existence of the DT of the asset, beyond the delivery of a BIM model. The implementation of the proposed diagram for the case study is described. As a result, it is observed that the proactive integration of structural engineers' specialized knowledge holds promise for enhancing the efficiency of DTs for construction and maintenance tasks.

Keywords -

Digital Twins; Structural Design; FEM; BIM; IFC Sensor-based data

1 Introduction

The AECO (Architecture, Engineering, Construction and Operation) sector has presented lower productivity indexes compared with other industries [1]. This can be attributed to the reluctance of a traditional sector to adopt technologies that can genuinely help enhancing capabilities. The involvement of diverse stakeholders in infrastructure projects and the dynamic environment of the construction sites with constant variations in site layout, equipment, materials, and other components, contribute to the complexity of delivering projects avoiding unforeseen issues, delays, or cost overruns [2].

Recent years have witnessed the extensive adoption of Building Information Modeling (BIM), elevating the efficiency of data generation and exchange within collaborative environments. Nevertheless, as digitalization and process automation, fueled by technologies like Internet of Things (IoT) and Artificial Intelligence, continue to advance, the BIM path encounters new hurdles in adapting to a broader interconnected context for delivering more sophisticated construction services [3][4]. Digital Twins (DTs), defined as a sophisticated information construct that enables a timely and useful connection between real and virtual assets, hold the potential to effectively address digitization and smart management needs within the sector [5].

Recent research concerning DTs has been focused on proposing frameworks and workflows for a successful application of the technology [6][7][8]. It can be observed that during the last years, several assets in construction and maintenance stages have been fully (but mostly partially) twinned. However, these assets may not have been designed consciously with the knowledge of the potential of their twinning. A gap regarding the role of the structural design phase in meeting the specific requirements of a DT in subsequent stages is identified. In addition, structural design is not directly connected to long term assessment of the asset (for instance, a direct connection with risk assessment involving structural components or structures). Structural engineers have the specialized knowledge to identify critical construction or maintenance processes that may need monitoring to ensure quality or enhance the management of an infrastructure. During the design phase, engineers consistently grapple with uncertainties. It is of interest to track these uncertainties during subsequent stages. Numerical models require calibration, and this holds particular significance in assessing the structural performance of infrastructures. Over the design stage, defining expected physical values at critical locations within a monitoring plan, and seamlessly integrating them, along with sensors and simulations data into a centralized digital system, can boost the potential of a DT, simultaneously reducing working hours for validation and verification. Risk assessment procedures used on the

long run can also benefit from early identification of crucial aspects of the structure.

In this paper, an example of a twinned infrastructure (a bridge) is provided. In this example, the bridge is twinned after the asset is designed and constructed. It represents a common procedure for the new generation of digitally twinned assets. However, structural engineering design offices may contribute to more sophisticated twinning by providing crucial insights at early design stages. Through a conceptual diagram, four key components are depicted: (1) Monitoring plan, (2) Sensors-data processing (3) Verification and validation, and (4) Simulations data exchange. Afterward, the applications of the conceptual diagram are showcased for the case study.

The practices presented in this research presumes the existence of a CDE (Common Data Environment) framework that enables access and updates of simulations for validation and verification during the lifecycle of an asset within a user-friendly interface, supporting risk management and decision-making. Section 6 outlines the work-in-progress associated with this topic.

2 Structural Analysis and Digital Twins

A DT as a virtual representation of a physical asset requires constant updates to the state and conditions of that real infrastructure. Therefore, varied simulations and predictions must be performed to match with on-site collected data. Ideally, this process can be infused into centralized digital systems where the collection of information of structural models and sensors data allows the validation and verification at both, construction and maintenance stages. This provides valuable information in the form of performance indicators that can support data-driven decision-making [9].

The key challenge lies in effectively bridging the gap between, BIM models, on-site physical data, assessment and structural analysis simulations. This integration should be completed within a flexible framework ensuring the unified use of information while upholding interoperability and maintainability over the infrastructure lifecycle [10].

Literature on the integration of BIM and structural simulations is increasingly growing [11][12][13]. Primarily, the existing research concentrates on facilitating interoperability between structural analysis and BIM software, and on enhancing the workflow for generating a structural analysis model from an architectural model, thereby minimizing the need for rework during initial design cycles.

For instance, Zhao Xu et al. [14] have proposed a new method to transform a BIM model into a connection model using the Brep graphic representation. Then, a discretization of filaments is done to obtain a refined finite element analysis. Additionally, they achieved the visualization of FEM (Finite Element Method) results on an OpenBIM environment following the IFC (Industry Foundation Classes) schema.

The primary goal of this paper is to align the structural design phase as a way to provide crucial information for subsequent DTs. Thus, facilitating the gathering of BIM models, measurements, processing, assessment, and structural analysis simulations in a centralized digital system.

3 The Case Study

This research is performed within the frame of a H2020 European project called Ashvin [15]. Its main goal is to pull out methodologies and cohesive solutions for many kinds of data flows to implement DTs during the design, construction, and maintenance phases. One of the demo cases of the project is La Plata viaduct (see Fig. 1), which belongs to a high-speed train network in Extremadura, Spain. The bridge is a 4-spanned posttensioned slab with a total length of 114 meters and represents the case study of this research.



Figure 1. Viaduct La Plata during the load test

An IFC-based BIM model was generated from existing 2D drawings. The administration provided access during the load test of the bridge. Accelerations, displacements, and strains were acquired for dynamic and static tests. The collected data was synced and uploaded to an IoT (Internet of Things) platform. Mechanical properties of the materials were defined from the original design reports. Information regarding the construction process was not provided. Structural analysis models were generated to match expected against measured values, validating the structural performance of the bridge. The DT paradigm was approach through the integration of multiple layers of information. BIM, FEM models, measures, and related standards (see Fig. 2), were knitted together to generate a virtual asset within a CDE, encompassing information pipelines from on-site measurements to risk assessment applications. Extended details about the procedure are described in [16].



Figure 2. DT information layers of the case study

From the lessons learned in the digital twinning of the case study, which involved dealing with an existing asset without construction procedure data and the participation of the bridge structural designers, the proposed conceptual diagram displays key components for the design stage as a way to anticipate the development of its DT. As a result, singularities about the structural behavior of an infrastructure can be considered *a priori* incorporating the technical knowledge of the designers. In section 5, the application of the conceptual diagram is described for the case study, guiding future implementation.

4 The Conceptual Diagram

Anticipating a DT during the structural design stage of an infrastructure is crucial for meeting technical requirements and effectively monitoring key construction or maintenance tasks.

To address this procedure, a conceptual diagram is presented in Fig. 3. Within the diagram, beyond the deliver of a BIM model, four main components should be integrated into the DT: 1) A comprehensive Monitoring Plan delineating critical construction or maintenance activities, defining physical values for measurement, specifying the location and scale of the measures, recommending the type of sensors, and generating a schedule for data collection. 2) Code-based Sensors Data Processing techniques for each physical magnitude, to convert raw sensor-based data into meaningful information. 3) Verification and Validation processes identifying Key Performance Indicators (KPIs) aligned with standards or quality control requirements. These KPIs must be converted into machine-readable formats such as JSON or XML for seamless integration into a DT. 4) A Simulation Data Exchange to the target CDE, meeting its framework requirements for exporting structural analysis models facilitating accessibility and updates of simulations.

The integration of these components into a CDE will ensure a robust foundation for the implementation of DTs over the construction and maintenance phases.



Figure 3. The conceptual diagram

5 Application to the Case Study

As the access to the bridge was provided during the load test, the actions executed for developing the DT of the bridge were directed on digitizing the components and processes for checking structural performance [16]. From the results and the lessons learned over the developments, in this section a knowledge transfer is done by applying the proposed conceptual diagram, considering a scenario where the structural design anticipates the digital twinning of the load test. Procedures were carried out following the Spanish national standard NAP 2-4-2.0 [17], which regulates the load testing of railway bridges.

5.1 Monitoring Plan

For the case study, the critical construction procedure to monitor was the validation of the structural performance to approve the start of the operation. The physical magnitudes to measure were accelerations, displacements and strains at mid-span and supports. The sensors and their description were added to the IFC-BIM model at the required locations employing the *IfcSensor* entity. The schedule of the measures was generated within the IFC-BIM model as metadata using the entities *IfcWorkPlan*, *IfcWorkSchedule*, *IfcWorkCalendar*, and *IfcTask*. The entities and relations of the IFC schema to perform this procedure are depicted in Fig. 4.



Figure 4. IFC entities and relationships to generate a Monitoring Plan

By generating an IFC-based Monitoring Plan, the digitization of the information is achieved, and its integration into a CDE is facilitated, given that the IFC schema is a well-known open standard.

5.2 Sensors Data Processing

To calculate the values of interest for displacements and strains, it was required to develop Python scripts for processing the collected data from LVDTs and strain gauges, considering the characteristics of the sensors. On the other hand, for accelerometers, an application able to identify vibration modes of a structure has also been implemented. This Python-based application estimates the vibration modes of a structure based on synced accelerometers data [18].

Furthermore, scripts were developed following the cloud computing framework of an IoT platform [19] to upload and retrieve sensors data.

5.3 Verification and Validation

To decide whether the structural performance of the bridge is satisfactory to initiate operation, the NAP 2-4-

2.0 standard defines parameters and limits that must be within acceptable ranges. In this sense, the results of the load test should be available in the form of indicators.

To achieve the digitization of the validation and verification for the case study, the formulas to estimate the parameters established by the national standard, were integrated into the CDE using python scripts. Moreover, the results of the load test were parametrized and consigned into JSON machine-readable dictionaries to enable the display of indicators for end-users. Fig. 5 presents the JSON defined to verify displacements. These procedures allow the automation of verification by digitizing the national standard. This information can be retrieved at any moment during the lifespan of the asset, reducing working hours and the waiting period to get the results of validations.

```
{
"Task":"Static Analysis",
"Date":11-11-24,
"Responsible": "Hector Posada",
"Standard": "NAP 2-4-2.0",
"Results":
[{
    "Hypothesis": "Hyp 1",
    "Span": "Span 1",
    "SensorID": "5clea221-f748-40dc-bc58-6b32ee0427d0",
    "Span deflection(mm)": "2.21",
    "Measured/Predicted deflection (%)": 64.0",
    "Residual deflection(mm)": 0.057",
    "Recovery rate (%)": "95.43"
}]
```

Figure 5. JSON schema for displacements validation

5.4 Simulation Data Exchange

One of the main challenges of digital twinning is embedding simulations of various kinds within a CDE. Usually, the commercial software for performing structural analysis require development of APIs and expertise to package information, which increases the complexity of exchanging data.

For the case study, JSON dictionaries were generated to include the results of a proprietary license Software package called MIDAS [20] within the target CDE, as shown in Fig. 6.

However, due to the absence of automation, the process was time-consuming. In addition, it would be inefficient in enabling future updates. In section 6, a work-in-progress research project that aims to develop an open-source data model for exchanging structural analysis models is described (O-SAM). The data model will facilitate the delivery of structural analysis simulations to target CDEs, helping to perform a seamless Simulation Data Exchange facilitating accessibility and updates.



Figure 6. Structural Analysis results within the target CDE

6 Work-In-Progress: The O-SAM Data Model

A DT that pretends to hold various simulations of a real asset must be capable of flexibly exchanging structural analysis models to enable access and updates of these simulations. On the other hand, to perform structural analysis calculations there is a considerable variety of options available in the market. Each of these software generates simulation models and results in its format, hindering interoperability.

The IFC 4x3 schema [21] have tried to cover this issue by its *IfcStructuralAnalysisDomain*. Nevertheless, the standard is limited as it is not possible to properly define structural analysis results and entities such as finite element meshes and 3D solids.

Authors' current research is focused on developing O-SAM data model. It is possible to break down this open-source data model into two parts: Firstly, a unified JSON-based structural simulation model schema. The schema is accompanied by a set of converters that allows multiple FEM software to upload simulation information to the DT. Secondly, the incorporation of a graph-based representation into the existing model will seamlessly unite simulations with other data and models in the DT using knowledge graphs. These graphs streamline the integration and contextualization of information from disparate systems into a cohesive and semantically rich model, characterized by a flexible and intuitive structure that aligns seamlessly with the demands of a DT [22].

7 Conclusions

Through a conceptual diagram, this paper presents the components to consider over the structural design phase for anticipating to a DT. The development of the proposed diagram was executed by transferring the knowledge of digital twinning the load test of a railway bridge. The purpose of the study was to include the specialized knowledge of structural engineers in planning and preparing the framework for monitoring construction or maintenance procedures within a DT. If these procedures are digitized during the design stage, costs and working hours are reduced. For instance, to check the structural performance of the case study it was necessary to generate new structural models to define the expected physical values. Furthermore, the time gap between data collection and the validation processes.

On the other hand, the information provided from the collaborator to develop the digital twin of the case study was limited and delivered as isolated silos. A centralized system which gathers the structural design information will facilitates the execution of construction or monitoring tasks.

There are still obstacles to overcome for DT adoption in the structural engineering domain. There is a lack of civil engineers with expertise in Information Technologies. Moreover, structural design offices may hesitate to share project information, concerned about overexposing their design procedures. Finally, owners or public administrations must include in the infrastructure design contracts the digitization of Monitoring Plans and Verification and Validation procedures, the development of Sensors Data Processing techniques, and the delivery of simulation models and results into a target CDE.

References

- Office for National Statistics. Productivity in the construction industry, UK: 2021. On-line: https://www.ons.gov.uk/economy/economicoutput andproductivity/productivitymeasures/articles/prod uctivityintheconstructionindustryuk2021/2021-10-19, Accessed: 05/12/2023.
- [2] Kumar, S.S., Cheng, J.C.P. A BIM-based automated site layout planning framework for congested construction sites. *Automation in*

Construction, Vol. 59, 2015, Pages 24-37, ISSN 0926-5805,

https://doi.org/10.1016/j.autcon.2015.07.008.

- [3] Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction*, 114, 2020, 103179. ISSN 0926-5805. https://doi.org/10.1016/j.autcon.2020.103179.
- [4] Sacks, R., Girolami, M., & Brilakis, I. Building Information Modelling, Artificial Intelligence and Construction Tech. *Developments in the Built Environment*, 4, 2020, 100011. https://doi.org/10.1016/j.dibe.2020.100011.
- Zhang, J., Cheng, J. C. P., Chen, W., & Chen, K. Digital Twins for Construction Sites: Concepts, LoD Definition, and Applications. *Journal of Management in Engineering*, 38(2), 2022, 04021094. Doi:10.1061/(ASCE)ME.1943-5479.0000948.
- [6] Pregnolato, M., Gunner, S., Voyagaki, E., De Risi, R., Carhart, N., Gavriel, G., Tully, P., Tryfonas, T., Macdonald, J., & Taylor, C. Towards Civil Engineering 4.0: Concept, workflow and application of Digital Twins for existing infrastructure. *Automation in Construction*, 141, 2022, 104421. ISSN 0926-5805. https://doi.org/10.1016/j.autcon.2022.104421.
- [7] 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.
- [8] Schlenger, J., Yeung, T., Vilgertshofer, S., Martinez, J., Sacks, R., & Borrmann, A. A Comprehensive Data Schema for Digital Twin Construction. In Proceedings of the 29th International Workshop on Intelligent Computing in Engineering (EG-ICE), June 2022, pp. 34-44. doi:10.7146/aul.455.c194.
- [9] Chacón, R., Ramonell, C., Posada, H., Tomar, R., de la Rosa, C. M., Stipanovic, I. (2023). Measurements, Simulation, Analysis and Geolocation in a Digital Twin tool for bridge management. In *Proceedings of the conference Eurostructure*, September 2023, 6(5), 474-482. John Wiley & Sons, Ltd. https://doi.org/10.1002/cepa.2017.
- [10] Boje, C., Kubicki, S., Guerriero, A., Rezgui, Y., Zarli, A. (2022). Digital twins for the built environment. *CRC Press.* (pp. 179-199). ISBN: 9781003204381. doi: 10.1201/9781003204381-10.
- [11] Ramaji, I.J., Memari, A.M. Interpretation of structural analytical models from the coordination view in building information models. *Automation in Construction*, 90, 117–133, 2018. https://doi.org/10.1016/J.AUTCON.2018.02.025.
- [12] Park S.I., Lee S.H., Almasi A, Song J.H. Extended

IFC-based strong form meshfree collocation analysis of a bridge structure. *Automation in Construction*, 119, 103364–103364, 2020. https://doi.org/10.1016/J.AUTCON.2020.103364.

- [13] Posada H., Chacón R., Ungureanu L.-C., and García D. Closing the Gap Between Concrete Maturity Monitoring and Nonlinear Time-dependent FEM Analysis through a Digital Twin. Case Study: Posttensioned Concrete Slab of an Office Building, Barcelona, Spain. In Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC), Bogotá, Colombia, 2022.
- [14] Xu, Z., Rao, Z., Gan, V. J. L., Ding, Y., Wan, C., Liu, X. (2019). Developing an Extended IFC Data Schema and Mesh Generation Framework for Finite Element Modeling. *Advances in Civil Engineering*, 2019, 1434093. Hindawi. https://doi.org/10.1155/2019/1434093.
- [15] Ashvin. Ashvin Project Website. On-line: https://www.ashvin.eu/, Accessed: 07/12/2023.
- [16] Chacón, R., Ramonell, C., Posada, H., Sierra, P., Tomar, R., Martínez de la Rosa, C., Rodriguez, A., Koulalis, I., Ioannidis, K., Wagmeister, S. (2023). Digital twinning during load tests of railway bridges
 case study: the high-speed railway network, Extremadura, Spain. *Structure and Infrastructure Engineering*. doi:10.1080/15732479.2023.2264840.
- [17] ADIF. Normativas Técnicas. On-line: http://descargas.adif.es/ade/u18/GCN/NormativaT ecnica.nsf, Accessed: 12/12/2023.
- [18] Sierra, P., Chacon, R., Martinez, X. Iberoeka 'E-TESTING': 'Herramienta numérica-experimental para la determinación del estado de integridad de las estructuras'. A technical report of the national project E-testing. 2020.
- [19] Teodorović, M. Deliverable 1.1 "Launch Version of ASHVIN Platform". The ASHVIN project. Zenodo, 2021. https://doi.org/10.5281/zenodo.4556836.
- [20] Midasoft. Midas Software Structural Engineering Software. On-line: https://www.midasoft.com/, Accessed: 12/12/2023.
- [21] Technical BuildingSmart. IFC Schema Specifications. On-line: https://technical.buildingsmart.org/standards/ifc/ifc -schema-specifications/, Accessed: 12/12/2023.
- [22] Ramonell C., Chacón R., Posada H. Knowledge graph-based data integration system for digital twins of built assets. *Automation in Construction*, 156:105109, 2023. DOI: https://doi.org/10.1016/j.autcon.2023.105109.

Removal of Construction Machinery Occlusion using Image Segmentation and Inpainting for Automated Progress Tracking

Ahmet Bahaddin Ersoz¹ and Onur Pekcan¹

¹Department of Civil Engineering, Middle East Technical University, TÜRKİYE abersoz@metu.edu.tr, opekcan@metu.edu.tr

Abstract -

This study introduces an innovative method for enhancing digital modeling accuracy in construction site monitoring by integrating UAV imaging with advanced machine learning and computer vision algorithms. It focuses on removing temporary elements like construction machinery from images. The method involves two steps: first, using deep learning algorithms, for instance, segmentation to detect and segment construction machinery from UAV images trained on the Aerial Image Dataset for Construction (AIDCON); second, employing image inpainting techniques, utilizing the Places2 dataset and the LaMa algorithm, to fill in the areas left vacant by the removed machinery. Demonstrated on a parking garage construction site in Ankara, Türkiye, the results show high accuracy in machinery segmentation and effective image inpainting, as evidenced by metrics like Normalized Root Mean Square Error (NRMSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). This approach contributes significantly to the field of construction site monitoring by refining digital models and shows potential for broader application in the industry. Future research directions include developing a specialized image inpainting dataset for construction scenarios and extending the methodology to encompass more types of temporary site elements, paving the way for more efficient and accurate project management in construction. Keywords -

Image segmentation, image inpainting, UAV, point cloud, progress tracking

1 Introduction

In the construction industry, digital modeling of job sites is essential for efficient project management and execution. Automated monitoring systems, frequently incorporating advanced technologies such as UAVs, LiDAR, and machine learning, require accurate data to monitor and compare ongoing construction activities with planned ones. These systems also play a crucial role by enabling the detection of deviations or potential delays. Such systems offer a proactive approach to project management, allowing project managers to address issues promptly and keep the project on track.

In practice, challenges arise due to temporary objects such as machinery, equipment, and materials on the construction site. Such objects can create occlusion in the digital model, obscuring the actual progress of the project. Their presence in the digital representation can lead to inaccuracies in assessing the extent of the completed structure, potentially resulting in misguided decisions and inefficiencies. Therefore, they need to be extracted from the digital models in an automated manner.

Researchers in the field of UAV-based photogrammetry pointed out significant challenges in accurate mapping and calculations due to moving objects such as cars, construction equipment, and temporary facilities [1,2]. These obstacles notably affect computations, leading to the erroneous generation of height differential maps. Their findings underline the necessity for enhanced methods to overcome the inaccuracies introduced by non-terrain objects. Automated monitoring systems can gain a clearer and more accurate view of the construction progress by eliminating these objects from the digital model. This practice is crucial as it allows for a more precise comparison between the current site condition and the project plan. It ensures that the progress tracking is focused solely on the permanent structural developments rather than being skewed by temporary site elements. This leads to enhanced overall efficiency and productivity of the construction project, ensuring it is completed on time and within budget.

Building on the critical need for accurate digital site representation, particularly in light of the challenges posed by temporary objects, it becomes clear that visionbased technological solutions are essential. Image segmentation and image inpainting emerge as crucial techniques in this context. Image segmentation involves dividing a digital image into different segments to distinguish between various elements, such as separating temporary objects like machinery and equipment from permanent structural components [3]. On the other hand, image inpainting is reconstructing missing or obscured parts of images [4]. This technique becomes particularly valuable in construction for filling areas from which temporary objects have been removed, thereby providing a more precise and accurate representation of the actual site conditions.

This study proposes a method that combines image segmentation and inpainting to produce well-represented digital models of construction sites such as point clouds. Initially, image segmentation is utilized to identify and isolate temporary objects within the site images. Following their removal, image inpainting is applied to fill in the resultant gaps, effectively recreating the obscured parts of the construction site. The inpainted images serve as an accurate base for the 3D reconstruction process. By utilizing refined images, point clouds can be generated to accurately reflect the actual state of the construction site, free from distortions caused by temporary objects. This approach significantly benefits the automated monitoring systems, providing them with a more reliable data source for tracking the project's progress. It enables precise tracking and assessment of construction activities, leading to better resource allocation, decision-making, and, ultimately, more efficient and timely completion of construction projects.

2 Background

Eliminating occlusion is notably challenging due to the variable geometry of construction surfaces. Various traditional terrain filtering methods have been developed for digital terrain generation, which can also be applied to point clouds of construction. They can be classified based on geometric principles: slope-based, morphologybased, and surface-based methods.

Slope-based approaches [5,6] focus on evaluating the slope in a localized area and categorizing points as ground and non-ground based on a predefined slope threshold. Morphology-based methods [7,8] employ mathematical morphology techniques to effectively identify and remove points that do not correspond to ground surfaces. Surface-based methods take a different approach by gradually selecting points from raw point clouds to construct a ground surface model. This is commonly achieved through the Triangulated Irregular Network (TIN) [9]. Other notable research methods in recent years are the Simple Morphological Filter (SMRF) [10] and the Cloth Simulation Filter (CSF) [11].

While these algorithms have proven effective in various scenarios, they also share certain limitations.

Their successful application requires users to thoroughly understand the algorithms and the specific characteristics of the sampled regions. This necessity for specialized knowledge makes these methods more challenging to apply. Moreover, in cases where the sampled region is extensive and features complex terrain relief, the parameters chosen for one site may not be applicable across the entire area without leading to errors in classification. Numerous classical machine learning algorithms have been introduced to enhance the robustness and level of automation in terrain generation. These algorithms aim to provide more adaptable solutions for terrain filtering in varied and complex construction environments.

In computer vision, deep learning has risen significantly in recent years. Deep learning techniques in image inpainting are adept at extracting semantic details from images, making more accurate predictions about missing content. Techniques such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have proven highly effective in capturing nuanced image data. Many studies have successfully employed CNNs to refine image inpainting processes, leading to notable progress [12, 13]. Among the notable variations of CNNs are Fully Convolutional Networks (FCN) [14] and U-Nets [15]. Additionally, the introduction of GANs has been influential, as they are particularly well-suited for image inpainting tasks due to their strong data generation capabilities [16].

To support these advancements, researchers have developed a variety of image inpainting datasets and applications, each specifically designed for different types of images, scenarios, or inpainting challenges. Key datasets such as ImageNet [17], Places2 [18], Paris StreetView [19], and CelebA-HQ [20] have been instrumental in the progress and assessment of image inpainting algorithms. These datasets have provided the necessary diversity and complexity for refining and evaluating inpainting techniques.

Advances in deep learning-based inpainting methods have been significant and widespread, impacting areas including urban modeling, shadow manipulation, construction management, and infrastructure planning. For instance, Kapoor et al. [21] utilized these techniques to create Nostalgin, a tool designed for reconstructing 3D city models from historical photographs by filling in missing data, thus offering a reasonable representation of the past. Similarly, Wei et al. [22] developed a dual-stage GAN method specifically for shadow inpainting and removal, notably improving color retention in shaded areas. In the realm of construction, Bang et al. [23] applied GANs for enhanced detection and reconstruction of construction resources in UAV imagery. Further, Angah and Chen [24] proposed a context inpainting method to eliminate obstructions in construction site

images, facilitating the creation of Building Information Models. W. J. Kim et al. [2] enhanced the detection of moving objects by improving background details at a pixel level.

J. Park et al. [25] also introduced a technique to generate vehicle-free ortho-mosaics from UAV images, thereby improving transportation infrastructure management. These diverse applications highlight the adaptability and efficiency of inpainting in tackling complex challenges across various fields. In this study, the objective is to utilize deep learning-based image inpainting techniques to identify and remove construction machinery from images. This approach aims to represent the construction field accurately, which is crucial for the subsequent steps in 3D reconstruction.

3 Method

The study introduces a comprehensive method for enhancing the accuracy of digital models in construction sites by removing temporary elements, such as construction machinery, from UAV-captured images. This methodology unfolds the parts mainly: segmentation of construction machinery through instance segmentation, followed by applying image inpainting techniques to refill the absence of these objects from the images (Figure 1).



Figure 1. Flowchart of the method

The first stage aims to detect and segment prominent construction machines like excavators, bulldozers, and trucks from the aerial images. Deep learning algorithms, tailored explicitly for instance segmentation, are deployed for this purpose. Instance segmentation offers a more detailed mapping of an image compared to traditional methods. It partitions an image into regions or pixels corresponding to individual objects, producing an exact "mask" for each object. This heightened level of detail in segmentation is essential in accurately identifying and subsequently removing objects from images.

The deep learning model trained by the AIDCON -Aerial Image Dataset for Construction [26] was used to facilitate this process. The AIDCON dataset includes 2155 images captured by UAVs. It provides bird's-eye views of various construction environments annotated at the pixel level, featuring nine categories of construction machinery like dump trucks, excavators, loaders, and dozers. This model smoothly detects and segments construction machinery in the UAV imagery. Subsequently, the identified objects are converted into binary image masks, which are used in the image inpainting process to eliminate these objects from the images.

The second stage involves the application of image inpainting methods. It utilizes the masks generated in the previous stage. The inpainting algorithm effectively fills the pixels previously occupied by the machinery with pixels suitable for the construction site environment. A pre-trained model on the Places2 dataset [18], renowned for its vast collection of diverse images across numerous unique scene categories, is employed. This dataset provides a robust and varied training environment for the model, significantly improving its capability to detect and remove construction machinery in many scenarios. After evaluating various inpainting techniques mentioned in the literature, the Large Mask Inpainting (LaMa) algorithm [27] is used. This algorithm is particularly adept at handling large missing regions, complex geometric structures, and high-resolution images, making it an ideal choice for the study. Combining the pre-trained Places2 dataset and the LaMa algorithm allows us to achieve robust and visually consistent inpainting results.

By merging the strengths of instance segmentation and advanced image inpainting techniques, the method is designed to produce digital models that accurately reflect the actual state of construction sites, devoid of distortions caused by temporary construction machinery. Building on the success of removing temporary construction machinery from UAV images, the study advances into the phase of 3D reconstruction of the current construction sites. Structure from Motion (SfM) [28] and Multi-View Stereo (MVS) [29] techniques are employed to create detailed point clouds. This integration of advanced image processing with 3D reconstruction technologies ensures that the final digital models accurately represent the actual state of the construction sites, significantly enhancing project management and planning capabilities.

4 Results

This section outlines the experiments conducted to assess the effectiveness of the proposed system in

construction environments. It offers a detailed examination of how image segmentation, inpainting, and 3D reconstruction techniques can be applied to track construction progress in real-world scenarios.

4.1 Study Area

The field study was conducted at a job site in Ankara, Türkiye, where a parking garage covering more than 9,000 square meters is being constructed beneath a courtyard. This site was selected for its suitability for progress monitoring due to the variety of machinery present.

4.2 Data Collection

This study conducted UAV imaging during two site visits, T1 and T2, offering a detailed overview of the construction progress. Aerial views of the construction site are illustrated in Figure 2 and Figure 3. To enhance the quality of the 3D point cloud and ensure accurate positioning in the 3D reconstruction process, Ground Control Points (GCPs) were measured around the perimeter of the construction site.



Figure 2. Aerial View of Construction Site at T1 Site Visit

The DJI Mavic Pro drone was chosen for the imaging process due to its several beneficial features. Its compact size, prolonged flight capability, and precise positioning make it well-suited for such tasks. The drone flights were maintained at a consistent altitude of 40 meters, which was crucial for ensuring data uniformity and enabling a comparison between different flights. The imaging strategy involved maintaining an 80% overlap for both front and side images, greatly enhancing the data quality and enabling accurate data analysis. This standardized approach to data collection was crucial in facilitating reliable comparisons and drawing meaningful conclusions from the data gathered.



Figure 3. Aerial View of Construction Site at T2 Site Visit

4.3 Data Processing

During the T1 and T2 timeframes, the construction site featured a variety of equipment, including dump trucks, excavators, backhoe loaders, and cars. Additionally, a category termed "other" encompassed drilling machines, anchor installation machines, and concrete mixers. The deep learning model previously mentioned was employed to segment construction machinery in both T1 and T2 datasets. The model's performance, measured by the mean Average Precision (mAP) COCO Metrics [30], is detailed for both datasets in Table 1. Additionally, it presents a breakdown of the Average Precision (AP) results, categorized by each type of equipment. AP of the backhoe loader was not present, as it was not visible in the images during the T2 timeframe.

Table 1. Segmentation Results

Dataset	mAP	mAP ₅₀	mAP ₇₅
T1	67.6	87.3	82.3
T2	64.5	87.2	78.7

Table 2. Classwise AP Results (IoU=50%) (D.T: Dump Truck, Exc: Excavator, B.L.: Backhoe Loader)

Data	D.T.	Exc.	B.L.	Car	Other
T1	97.8	98.9	96.3	97.5	45.9
T2	92.8	96.4	-	96.5	63.1

The outcomes of the segmentation stage lay the groundwork for advancing to the subsequent phases, which involve the removal of construction machinery and the creation of 3D surface models. Following the segmentation stage, image masks were created (Figure 4). These masks accurately define the boundaries of each piece of machinery detected in the images, preparing them for the next phase of the process. The LaMa inpainting algorithm, which has been trained using the Places2 dataset, was then employed on these images.



Figure 4. Eliminating Occlusion from the Images

The effectiveness of the inpainting process was assessed using several established image comparison metrics. These include the Normalized Root Mean Square Error (NRMSE) [31], Peak Signal-to-Noise Ratio (PSNR) [31], and Structural Similarity Index (SSIM) [32]. The results of these evaluations are detailed in Table 3.

Table	3.1	npainting	Results
ruore	J. 1	inpunning	results

Dataset	NRMSE	PSNR	SSIM
T1	0.019	37.628	0.900
T2	0.018	32.595	0.901

Table 4. 3D Reconstruction Results

	T1	T2
Processed Images	230 of 230	230 of 230
Sparse P. Cloud	314,299	313,065
Dense P. Cloud	32,491,933	32,012,587
GSD	1.3 cm	1.3 cm
GCP Error	16.7 cm	14.4 cm

Once these construction machines were digitally eliminated from the images, the next step involved converting the 2D images into a 3D point cloud. This transformation required using the SfM-MVS technique to achieve a three-dimensional site representation. SiteEye, a robust photogrammetry software [33], was chosen to manage UAV photogrammetry in this study. SiteEye was selected from a range of available software known for its comprehensive capabilities in photogrammetry. The results of the photogrammetric process using SiteEye are visualized in Figure 5 and summarized in Table 4. The table indicates that all 230 images from each image set were processed successfully, demonstrating the effectiveness of this approach in creating accurate 3D models of the construction site.

5 Discussion

The current research introduces a novel approach for monitoring construction site progress, fusing UAV imaging with advanced machine learning and computer vision algorithms. This innovative method addresses several limitations in traditional techniques for generating point clouds at construction sites. The discussion will highlight the significant contributions of the study, the challenges faced, and potential avenues for future research.

Performance of the Construction Machine Segmentation Model: A pivotal success of the proposed approach is the performance of the deep learning model in segmenting construction machinery. This model exhibited reasonable accuracy, especially for frequently encountered construction vehicles like dump trucks, excavators, backhoe loaders, and cars. The AP scores, often surpassing 90%, attest to the efficacy of deep learning algorithms in machinery segmentation within construction site imagery. This achievement is vital to integrating machine learning technologies into construction site monitoring. It is essential to acknowledge that flight parameters such as height, overlap, and camera angle substantially impact the resolution, coverage, and geometric accuracy of UAV images. These factors significantly influence

the detectability of construction machinery and the quality of the areas inpainted subsequently. For example, flying at a higher altitude may lead to lowresolution images, complicating fine-grained segmentation tasks. In the study, a flight altitude of 40 meters resulted in a ground sampling distance of 1.3 cm, sufficiently identifying the construction machinery targeted in the images.

- Advancements in the Automation of Construction Machine Removal: The study marked a significant stride in the automated removal of construction machinery using image inpainting techniques. The encouraging results from the evaluation metrics-NRMSE, PSNR, and SSIM-underscore this success. For instance, dataset T1 showed an NRMSE of just 0.019 and an SSIM score of 0.900, indicating good structural similarity in the inpainting process. These results were closely mirrored in dataset T2. These metrics validate the effective implementation of image inpainting algorithms for removing construction machinery, thus aiding in the precise generation of digital terrain models.
- Interoperability and Versatility of the Method: A noteworthy aspect of the method is its compatibility with various photogrammetry software. The construction machine-removed images are designed to be georeferenced and processed independently, making the output of the proposed method adaptable for integration with different third-party photogrammetry applications. This flexibility enhances the method's potential for widespread adoption, particularly in automated progress monitoring for construction sites.
- Potential for Future Research Specialized **Image Inpainting Dataset and Removal of More** Types of Site Occlusion: Looking ahead, developing a dedicated image inpainting dataset tailored to scenarios commonly encountered in automated construction site monitoring presents a research opportunity. Such a specialized dataset could enhance the performance of image inpainting algorithms, a critical component of the proposed method. Improved algorithms will further refine automated progress monitoring, making it more efficient and reliable. Future research also offers the potential to expand current methodologies by removing various site occlusion types. While the current focus is on construction machinery, extending this to elements like workers, unused materials, and temporary structures could greatly enhance site management. This would improve the digital representation of construction sites, providing a clearer view of progress and conditions, thereby facilitating more efficient and accurate



(a) 3D Reconstruction with Original Images



(b) 3D Reconstruction with Inpainted Images

Figure 5. Resultant Point Clouds

project management. Developing comprehensive models and datasets is crucial to advancing automated monitoring in the construction industry.

6 Conclusions and Future Work

This study introduces a method combining UAV imaging, machine learning, and computer vision algorithms to improve the digital modeling of construction sites. The focus was on removing temporary elements like construction machinery from images to enhance the accuracy of these models. The approach was tested in a real-world setting on a construction site, demonstrating its practical application. Key findings include the effective use of deep learning for segmenting construction machinery, resulting in high AP scores. This success illustrates the potential of deep learning in construction site monitoring. Additionally, image inpainting proved valuable in creating accurate digital representations of the site after removing temporary objects.

The method's compatibility with various photogrammetry software suggests its potential for broader application in the construction industry. Future research directions could involve developing a specialized image inpainting dataset for construction scenarios and extending the methodology to include more types of temporary site elements. In conclusion, this research contributes to the field of construction site monitoring by offering an innovative method for improving the precision of digital models, with implications for more efficient and accurate project management in the construction industry.

Acknowledgments

This research is funded by a grant from the Scientific and Technological Research Council of Türkiye (TUBITAK), Grant No. 122M055. TUBITAK's support is gratefully acknowledged.

References

- [1] Cho, J. W., Lee, J.-K., and Park, J., Large-scale earthwork progress digitalization practices using series of 3D models generated from UAS images, Drones, vol. 5, no. 4, 2021.
- [2] Kim, W. J., Hwang, S., Lee, J., Woo, S., and Lee, S., AIBM: Accurate and instant background modeling for moving object detection, IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 9021–9036, 2021.
- [3] Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L., and Zhiwei, L., Dataset and benchmark for detecting moving objects in construction sites, Automation in Construction, vol. 122, no. October 2020, p. 103482, 2021.
- [4] Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C., Image inpainting, in Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000, pp. 417–424.
- [5] Vosselman, G., Slope based filtering of laser altimetry data, International Archives of Photogrammetry and Remote Sensing, Vol. 33, Part B3/2, vol. 33, no. Part B3/2, pp. 678–684, 2000.
- [6] Wang, C. K. and Tseng, Y. H., Dem generation from airborne lidar data by an adaptive dualdirectional slope filter, International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, vol. XXXVIII, pp. 628–632, 2010..
- [7] Zhang, K., Chen, S. C., Whitman, D., Shyu, M. L., Yan, J., and Zhang, C., A progressive morphological filter for removing nonground measurements from airborne LIDAR data, IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 4 PART I, pp. 872–882, 2003.
- [8] Chen, Q., Gong, P., Baldocchi, D., and Xie, G., Filtering airborne laser scanning data with morphological methods, Photogrammetric Engineering and Remote Sensing, vol. 73, no. 2, pp. 175–185, 2007.
- [9] Axelsson, P., DEM generation from laser scanner data using adaptive TIN models, International archives of photogrammetry and remote sensing, vol. 33, no. 4, pp. 110–117, 2000..
- [10] Pingel, T. J., Clarke, K. C., and McBride, W. A., An improved simple morphological filter for the terrain classification of airborne lidar data, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 77, no. October 2017, pp. 21–30, 2013.
- [11] Zhang, W. et al., An easy-to-use airborne lidar data filtering method based on cloth simulation, Remote Sensing, vol. 8, no. 6, p. 501, 2016.
- [12] Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., and Catanzaro, B., Image inpainting for irregular holes using partial convolutions, in

Proceedings of the European conference on computer vision (ECCV), 2018, pp. 85–100.

- [13] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T., Free-form image inpainting with gated convolution, Proceedings of the IEEE/CVF international conference on computer vision, vol. 2019-Octob, pp. 4470–4479, 2019.
- [14] Long, J., Shelhamer, E., and Darrell, T., Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431– 3440.
- [15] Ronneberger, O., Fischer, P., and Brox, T., 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, 2015, pp. 234–241.
- [16] Goodfellow, I. et al., Generative adversarial networks, Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.
- [17] Russakovsky, O. et al., Imagenet large scale visual recognition challenge, International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
- [18] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A., Learning deep features for scene recognition using places database - supplementary materials, Advances in neural information processing systems, vol. 27, 2014..
- [19] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A., Context encoders: Feature learning by inpainting, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2536–2544.
- [20] Karras, T., Aila, T., Laine, S., and Lehtinen, J., Progressive growing of GANs for improved quality, stability, and variation, in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018, pp. 1– 26.
- [21] Kapoor, A., Larco, H., and Kiveris, R., Nostalgin: Extracting 3D city models from historical image data, in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019, pp. 2565–2575.
- [22] Wei, J., Long, C., Zou, H., and Xiao, C., Shadow inpainting and removal using generative adversarial networks with slice convolutions, in Computer Graphics Forum, 2019, pp. 381–392.
- [23] Bang, S., Baek, F., Park, S., Kim, W., and Kim, H., Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation

techniques, Automation in Construction, vol. 115, no. September 2019, p. 103198, 2020.

- [24] Angah, O. and Chen, A. Y., Removal of occluding construction workers in job site image data using U-Net based context encoders, Automation in Construction, vol. 119, no. April 2019, p. 103332, 2020.
- [25] Park, J., Cho, Y. K., and Kim, S., Deep learning-based uav image segmentation and inpainting for generating vehicle-free orthomosaic, International Journal of Applied Earth Observation and Geoinformation, vol. 115, no. June, p. 103111, 2022.
- [26] AIDCON Authors, Aerial Image Dataset for Construction (AIDCON). On-line: https://ai2lab.org/aidcon/ Accessed:09/04/2024
- [27] Suvorov, R. et al., Resolution-robust large mask inpainting with Fourier convolutions, in Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2022, pp. 2149– 2159.
- [28] Wu, C., Towards linear-time incremental structure from motion, in Proceedings - 2013 International Conference on 3D Vision, 3DV 2013, 2013, pp. 127–134.
- [29] Furukawa, Y. and Ponce, J., Accurate, dense, and robust multiview stereopsis, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 8, pp. 1362–1376, 2010.
- [30] Lin, T. Y. et al., Microsoft COCO: Common objects in context, in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, 2014, pp. 740–755.
- [31] Eskicioglu, A. M. and Fisher, P. S., Image quality measures and their performance, IEEE Transactions on Communications, vol. 43, no. 12, pp. 2959–2965, 1995.
- [32] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.
- [33] SiteEye Authors, SiteEye: Cloud-based Photogrammetry and Project Management. On-line: https://www.siteeye.co, Accessed: 09/04/2024.

Context-Adaptive CCTV Pan-Tilt-Zoom method for Personal Protective Equipment Detection

Seokhwan Kim¹, Minwoo Jeong¹, Minkyu Koo¹, Taegeon Kim¹, and Hongjo Kim¹

¹Department of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea

yyksh2019@yonsei.ac.kr, minuj9855@yonsei.ac.kr, kmk0119804@yonsei.ac.kr, ktg9655@yonsei.ac.kr, hongjo@yonsei.ac.kr (Corresponding Author)

Abstract -

PPE items, including hardhats, hooks, harnesses, and straps, are critical for fall prevention. Ongoing research in construction safety has focused on using deep learning models to detect Personal Protective Equipment (PPE) worn by high-altitude workers. Despite efforts using computer visionbased models for safety monitoring, small object detection, such as hooks and straps, remains challenging due to image resolution issues. This study introduces a novel technique using mobile CCTV cameras controlled by an automated Pan-Tilt-Zoom (PTZ) algorithm to enhance the detection of small-sized PPE. The method leverages the size gap between worker and PPE. In a zoomed-out state with a short focal length, the system identifies the worker's bounding box (bbox), then zooms in with a longer focal length for precise PPE detection. When encountering multiple workers, the system applies predetermined zoom-in rules. Experimental results demonstrated a significant increase in detection accuracy for the small PPE: hook detection improved from 39.8% to 88.3%, and strap detection from 49.4% to 71.8%, as measured by an mAP of 50. This encouraging performance improvement suggests that automated PTZ control technology could enhance the effectiveness of safety monitoring

Keywords -

Construction safety; PTZ CCTV control; monitoring; PPE detection; Small object detection

1 Introduction

Construction industry, notorious for the highest number of accident victims, experiences most fatalities due to falls, according to the U.S. Bureau of Labor Statistics [1]. In South Korea, the most common type of industrial accidents is also falls, leading to a significant number of injuries [2]. Falls typically occur from high workspaces due to slipping or structural failures. Despite attempts to install safety nets and personal airbags, their high costs and spatial constraints limit widespread usage in construction sites. Therefore, proper attachment and usage of Safety Hooks and Safety Straps to fixed structures such as scaffolding are crucial to prevent falling from height position. To increase the use of PPE, construction companies globally, including in South Korea, are employing construction site safety managers to monitor workers' compliance with PPE usage, with some regions enforcing this as a legal requirement. However, this manual method is labor-intensive, costly, and prone to human error. Recent advancements have seen the integration of CCTV and cutting-edge deep learning for automated safety monitoring in construction sites. These systems use deep learningbased object detection or instance segmentation models to identify workers and PPE from video frame, automatically determining proper PPE usage. This automated safety monitoring tries to aid or replace human managers, reducing errors, cutting costs, and increasing efficiency in construction projects.



Figure 1. Video captured by CCTV installed at an actual construction site.

While automated safety monitoring technology has been continually developing, achieving significant success in certain areas such as recognizing workers' behaviors, detecting hardhats, or differentiating between highaltitude and ground-level workers, it still faces substantial challenges in recognizing small PPE items such as safety hooks and straps. This difficulty arises because these PPE items are often captured as only a few to several pixels in size, making it hard for object recognition models to identify their features. For example, Figure 1 shows this challenge. It is easy to recognize five workers in ground and high-altitude, but hard to find their hooks and straps, even if human eyes.



Figure 2. Results of applying Super-Resolution (SR) to construction site image. (a,b) show before SR application, and (c,d) are after applying SR. (b) and (d) specifically provide enlarged views of a worker in the top right area.

Several software-based solutions have been proposed to address low-resolution issues in images. For example, Super Resolution (SR) technology, aimed at enhancing image resolution software-wise, has been evolving for decades [3][4] and now focuses on deep learning-based methods [5]. Despite its advancements, SR's application in identifying small PPE at construction sites often doesn't meet expectations. Figure 2 shows the use of a contemporary deep learning-based SR method [6] on construction videos. Overall image clarity improves from (a) before SR to (c) after SR application. However, there are limitations when comparing (b) and (d), where despite clearer worker visibility, finding small PPE remains problematic. Thus, the approach of only using post-processing on already captured video frame has limitations in effectively detecting small PPE.

This study tackles the issue of small object detection, such as PPE detection in far-field monitoring. The goal is to overcome problems with low-resolution images by first taking larger pictures of PPE. The study leverages PTZ CCTV cameras with remote control capabilities, employing a worker-centric approach for zooming in to fill video frames, thus enlarging small PPE for better visibility. This research is potentially pioneering in proposing an automated system for adjusting PTZ cameras specifically for PPE detection in construction workers. It utilizes the size different between a typical 180cm worker and 20cm PPE, prioritizing worker detection in a zoomedout state, then calculating center coordinates for targeted zoom-in, enhancing focused on the PPE detection. To evaluate the effectiveness of the suggested system, experiments were carried out in a laboratory environment, capturing both zoomed-in and zoomed-out videos targeting a worker. Performance of the PPE detection model was compared between two scenarios. The feasibility of implementing automatic PTZ control was examined in an indoor setting. The integration of automated PTZ CCTV control with safety monitoring in this study is expected to demonstrate utility and facilitate precise safety monitoring. Moreover, the system's design to operate without the need for personnel to control the CCTV leads to a more efficient monitoring system, reduces human error in surveillance, and is anticipated to improve the accuracy of monitoring.

2 Related works

2.1 Construction safety monitoring with computer vision

In recent years, continuous research has utilized advanced computer vision technology for safety monitoring in construction sites [7]. Studies include training Region-based fully convolutional networks to recognize construction equipment [8], installing CCTV on cranes for worker safety monitoring in complex construction environments [9], and combining computer vision models with IoT sensors for accurate fall hazard detection [10]. Additionally, depth estimation on single-lens captured images has been proposed for improved safety monitoring [11], and optimizing loss functions in deep learning models has been shown to enhance PPE detection [12]. These studies indicate widespread use of computer vision in construction safety monitoring and support the suitability of instance segmentation and object detection for PPE detection. However, they primarily focus on recognition performance in already captured footage, with less discussion on the methods of capturing the footage itself.

2.2 Autonomous PTZ control

Automating PTZ control has been extensively researched. Maximizing PTZ CCTV's object tracking and zooming capabilities is key. Attempts include using classical computer vision methods such as KLT feature trackers for PTZ operation [13]. Efforts to reduce or mitigate delays between video and PTZ control have been made for successful zooming and tracking [14]. Studies on optimally operating multiple PTZ cameras in a space, considering field of view and concurrent tracking, have been conducted [15]. Research on inspection robots for continuous object monitoring [16] and integrating PTZ control with neural networks for end-to-end solutions [17] have also been explored. Previous research on PTZ cameras typically involved wired connections, enabling detection without significant delays, which differs from this study. Here, the CCTV is wirelessly connected, resulting in a delay of about 3 seconds for video reception and motor control. This delay was factored into the design of the automated PTZ control system. The interval between PTZ control commands had to be longer than this delay to prevent malfunctioning.

3 Methodology



Figure 3. Simplified flowchart of the automated PTZ CCTV system



Figure 4. Top view of the PTZ CCTV's PAN coordinates

Figure 3 illustrates the automated PTZ CCTV system, which consists of a hardware control unit for operating the PTZ and an analyzer unit that processes footage from the CCTV. The CCTV continuously captures frames and broadcasts it using the Real-Time Streaming Protocol (RTSP), a standard protocol commonly used in CCTV systems for transmitting live video. The analyzer receives this video via RTSP for analysis. Deep learning models em-



Figure 5. Side view of the PTZ CCTV's TILT coordinates

ployed for analysis include object detection and instance segmentation models such as YOLOv8m-seg [18]. These models are trained to identify workers and PPE, with their training process described later in 3.2. The system calculates the actual coordinates on the CCTV camera from the pixel coordinates of objects detected by the model Figure 4 and Figure 5.

$$\phi_{target} = \frac{x - \frac{W}{2}}{\frac{W}{2}} * \frac{FOV_w}{2} + \phi_{center} \tag{1}$$

$$\theta_{target} = \frac{y - \frac{H}{2}}{\frac{H}{2}} * \frac{FOV_h}{2} + \theta_{center}$$
(2)

The detailed method for coordinate calculation follows equations 1 and 2. *W* and *H* represent the width and height pixel values of the image, also *x* and *y* mean the pixel coordinates of target object in the image. FOV_w and FOV_h refer to the Field of View (FOV) of the camera in the width direction and height direction. ϕ_{target} and θ_{target} represent the Pan and tilt coordinates of target object. ϕ_{center} and θ_{center} represent the center coordinates of image.

Based on the estimated coordinate information, the PTZ motor is activated through a Open Network Video Interface Forum(ONVIF) protocol which is a global standard for the interface of IP-based physical security products, such as network cameras. This processes are repeated, continuously capturing and analyzing footage and operating the PTZ control. The performance of the proposed system is validated by the improved mask-AP(Average Precision) of the deep learning model.

3.1 States and transitions

A Finite State Machine (FSM) is a design method where a device can only exist in one of a finite number of states at a time [19]. The FSM allows a system to operate within predictable states, enabling stable control. The proposed system's PTZ control algorithm is designed as an FSM, as shown in Figure 6. The diamond in the figure represents the starting state, while the circles represent other



Figure 6. Flow chart of proposed auto PTZ control system using finite state machine

states. The lines between states indicate possible transitions. The system transitions from one state to another based on predefined procedures. Details on actions and transition definitions for each state will be discussed in the subsequent sections.

3.1.1 initCCTV

The 'initCCTV' state activates when the system starts. It has no inward transition since it's the initial state. In this state, the system checks the CCTV power, initiates the ON-VIF protocol, positions Pan-Tilt-Zoom to zero, and creates shared memory between states. These operations prepare the system for action, initialize hardware, and stabilize the system. It also checks communication status, attempting reconnection if issues arise. Transition to the next state is based on predefined user instructions: it transitions to 'Find site' if commanded, or 'Do zoom out' otherwise.

3.1.2 Site finding

The 'Site Finding' state enables a PTZ CCTV to automatically detect and orient towards the direction of ongoing construction work, allowing it to start filming independently without remote assistance via human. Figure 7 illustrates how this feature operates. Initially, it performs 'Heading to zero position,' returning pan and tilt to positions 0 and 1. To anticipate network delays, a refresh function clears any buffer backlog. Then, it captures a single frame from the CCTV. 'n' represents the number of captures, dividing 360° by 'n' to determine the pan angle per capture. The deep learning model identifies workers in each frame, storing their locations. This process repeats until a full rotation is completed. Afterwards, the number of detected workers at each rotation point is averaged to identify the current work site, and the camera is oriented accordingly. Once complete, the system transitions to 'Zoom out' state.

3.1.3 Zoom out

In the 'Zoom out' state, the system receives RTSP and searches for workers or, if possible, PPE in the video. This state manages the overall schedule, alternating every 5 seconds to the 'Heading adjustment' state or switching to the 'Zoom in' state every 30 seconds.

3.1.4 Heading adjustment

The 'Heading adjustment' state involves receiving realtime RTSP video to locate workers and automatically adjust the camera's direction towards them. Once this action is completed, the system reverts back to the 'Zoom out' state.

3.1.5 Zoom in

The 'Zoom in' state uses the PTZ's zoom feature to select a worker for closer observation. After detecting a worker and calculating their PTZ coordinates, it zooms in on a certain worker based on a pre-chosen policy: (1) smallest area worker, (2) no PPE, or (3) from left to right sides. The extent of zooming is until the worker's bounding box (b-box) fills the video frame. If the setting for 'Tracking a target' is enabled, it activates the 'Tracking a target' state every 3 seconds. The 'Zoom in' state operates for 15 seconds, after which it returns to the 'Zoom out' state.

3.1.6 Tracking a target

In the 'Tracking a target' state, the system continuously follows the magnified individual. It calculates the necessary pan and tilt adjustments considering the focal length changes due to zooming. This process lasts for 5 seconds, after which the system reverts back to the 'Zoom in' state.

3.2 PPE detection model

Accurately and rapidly detecting workers and PPE is crucial for effective safety monitoring. For this, YOLOv8,



Figure 7. Algorithm of the Site Finding

known for its accuracy and detection speed, was employed. YOLOv8 introduced an anchor-free detection system, enhancing performance with faster computation and better accuracy. Mosaic augmentation, used until 10 epochs before training completion, prevented overfitting, ensuring general detection capabilities. YOLOv8's versatility allows easy modification or addition of features to its head, if needed. Thus, the study adopted YOLOv8m-seg, using transfer learning on data labeled in instance segmentation format from a construction site video collected during 2022-2023 in Korea.

4 Experiment

4.1 Experimental Settings



Figure 8. The PTZ CCTV is mounted on a module equipped with a router and battery.

The PTZ CCTV camera used in this study is the Hikvision 'DS-2DE4A225IW-DE 2MP' model, capable of up to 25x zoom and providing a 57.6° FOV at 1x zoom shown in Figure 8. The RTSP video streams at a resolution of 1280x720 at 10 frames per second (FPS). The computer used for training and inference employs an RTX3090 GPU.

4.2 Image dataset for PPE detection model



Figure 9. Example: (Left) Images and (Right) GT masks

The dataset used for training the model consists of videos collected from 65 construction sites, including apartment and road projects in South Korea, labeled for instance segmentation with four classes: worker, hardhat, strap, and hook (as shown in Figure 9). It comprises a total of 6,523 images, divided into training, validation, and testing subsets in a ratio of 5,877:600:46 for use in training.

4.3 Train the model

The training of the YOLOv8m-seg model followed the default settings suggested in [18], with the only modification being an increase in the maximum epochs to 300. This model pretrained on the MS COCO dataset, and utilized transfer learning in this research to develop a fast and high-performing model. The dataset used was the one introduced in 4.2, focusing on learning and locating features of workers, hardhats, straps, and hooks.

4.4 Evaluation matric

The evaluation metric used was the mask_AP. It assesses instance segmentation by calculating the Intersection over Union (IoU) between the predicted mask and the

true mask, considering instances with IoU over 50% as True Positives (TP). This method of evaluating instance segmentation performance centered on masks incorporates both recall and precision of predictions, offering a comprehensive evaluation of the segmentation's accuracy.

4.5 Lab test of developted PTZ control



Figure 10. (Left) Zoom-out view and (Right) Zoom-in view

Video data of workers and PPE was collected from temporary structures at Yonsei University. This data comprises 230 images, simulating zoomed-in and zoomed-out states as shown in Figure 10. The images were polygon labeled for the same four classes as mentioned in 4.2. This dataset is utilized to assess the practical effectiveness of the zoom-in.

5 Results and Discussion

5.1 Performance of the PPE instance segmentation model

The training results on the dataset from 4.2 showed outcomes as in Table1. High mask_mAP performances of 97.1% for 'Worker' and 95.2% for 'Hardhat' were achieved, whereas 'Strap' and 'Hook' exhibited lower performances at 60% and 48.2%, respectively. This reaffirms the difficulty in recognizing small PPE in far field situations.

Ta	ble 1. Perfo	ormance of the mode
	Class	mask_mAP@50
	Worker	97.1
	Hardhat	95.2
	Strap	60.0
	Hook	48.2

The developed model was applied to the small-scale dataset of zoomed-in and zoomed-out images in 4.5, and its performance was evaluated. As seen in Table2, significant performance improvements were noted for 'Strap' and 'Hook'. In zoomed-out situations, 'Strap' and 'Hook' showed lower performances of 49.4% and 39.8%, respectively, while zoomed-in, they exhibited remarkable improvements with 71.8% and 88.3%.

Table 2.	. Performanc	e of the m	nodel bet	ween zooi	m-in
and zoo	om-out				

Class	mask_mAP@50			
Class	Zoom-out	Zoom-in		
Worker	99.5	99.5		
Hardhat	99.5	99.5		
Strap	49.4	71.8		
Hook	39.8	88.3		

5.2 Qualitative results of automated PTZ system

It was observed that according to the pre-determined rules of the finite state machine, the transition from the Zoom-out state to Zoom-in and Tracking a target occurs as illustrated in Figure 11. With each activation of the zoom, making small PPE more detectable by eye becomes apparent. In contrast, without PTZ control, detection is limited to workers or hardhats only.



Figure 11. Comparison of situations with and without automated PTZ peration

Applying these experimental findings to real construction sites could greatly improve the accuracy of detecting whether PPE is worn, thereby significantly boosting site safety. Moreover, the capability to automatically track workers from a distance allows for effective monitoring of the site, regardless of CCTV camera placement.

6 Conclusion

This study presented a comprehensive examination of the implementation and efficacy of an automated PTZ CCTV system for enhancing safety monitoring on construction sites. Our research found that zooming in on small objects such as hooks or straps significantly enhances detection capabilities. Laboratory experiments with zoomed-in and zoomed-out footage, analyzed using the same model, indicated a substantial improvement in performance—by approximately 1.6 to 2 times. Moreover, the zoom-in and tracking states proved effective in detecting small PPE items previously undetectable.

Still, there are limitations to be addressed for further advancement of the proposed PTZ control method:

- 1. **Need for Field Data Validation:** Testing in real construction environments is necessary to validate the PTZ control system's efficacy, thereby revealing unknown issues that hamper the reliable monitoring system.
- 2. Improvement in Coordinate Calculation at Lower Tilt Values: Future study should focus on enhancing the accuracy of coordinate calculations, especially at lower Tilt angles which are the error sources reducing the current system's precision.
- 3. Delays in Video Transmission Over Wireless Networks: Future study should explore the integration of edge computing and the PTZ CCTV. These efforts are intended to reduce the data transmission time, improving the system's responsiveness.
- 4. Efficient Zoom-In Target Selection: Additionally, identifying a systematic and efficient method for determining zoom-in targets will be essential. This will ensure the PTZ control system can focus on relevant areas quickly and accurately, enhancing its utility in monitoring safety equipment on construction sites.

By addressing these issues and possible solutions, future study aims to significantly improve the PTZ control method's reliability and effectiveness.

The research contributes valuable insights into the field of construction safety and lays the groundwork for future innovations that could potentially automate and improve safety measures, thereby reducing the risk of accidents and enhancing worker protection in construction environments.

Acknowledgment

This research was conducted by the support of the "2023 Yonsei University Future-Leading Research Initiative (No. 2023-22-0114)" and the "National R&D Project for Smart Construction Technology (No. RS-2020-KA156488)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

References

- U.S. BUREAU OF LABOR STATISTICS. Construction deaths due to falls, slips, and trips increased 5.9 percent in 2021. On-line: https:// www.bls.gov/opub/ted/2023/constructiondeaths-due-to-falls-slips-and-tripsincreased-5-9-percent-in-2021.htm, Accessed: 25/12/2023.
- [2] Ministry of Employment and Labor. 2022 industrial accident statistics: Results of 'investigation of fatal accidents' released. On-line: https://www.moel.go.kr/news/enews/report/ enewsView.do?news_seq=14546, Accessed: 25/12/2023.
- [3] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3): 21–36, 2003. doi:10.1109/MSP.2003.1203207.
- [4] Daniel Glasner, Shai Bagon, and Michal Irani. Superresolution from a single image. In 2009 IEEE 12th International Conference on Computer Vision, pages 349–356, 2009. doi:10.1109/ICCV.2009.5459271.
- [5] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, and Vishal Goyal. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, 91:230–260, 2023. doi:https://doi.org/10.1016/j.inffus.2022.10.007.
- [6] Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang, and Chao Dong. Rethinking alignment in video super-resolution transformers. In Advances in Neural Information Processing Systems, volume 35, pages 36081–36093, 2022. doi:https://doi.org/10.48550/arXiv.2207.08494.
- [7] Weili Fang, Lieyun Ding, Peter ED Love, Hanbin Luo, Heng Li, Feniosky Pena-Mora, Botao Zhong, and Cheng Zhou. Computer vision applications in construction safety assurance. *Automation in Construction*, 110:103013, 2020. doi:https://doi.org/10.1016/j.autcon.2019.103013.
- [8] Hongjo Kim, Hyoungkwan Kim, Yong Won Hong, and Hyeran Byun. Detecting construction equipment using a region-based fully convolutional network and transfer learning. *Journal of computing in Civil Engineering*, 32(2):04017082, 2018. doi:https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731.
- [9] Gelayol Golcarenarenji, Ignacio Martinez-Alpiste, Qi Wang, and Jose Maria Alcaraz-Calero. Machinelearning-based top-view safety monitoring of ground workforce on complex industrial sites. *Neural Computing and Applications*, pages 1–14, 2022. doi:https://doi.org/10.1007/s00521-021-06489-3.
- [10] Muhammad Khan, Rabia Khalid, Sharjeel Anjum,

Si Van-Tien Tran, and Chansik Park. Fall prevention from scaffolding using computer vision and iot-based monitoring. *Journal of Construction Engineering and Management*, 148(7):04022051, 2022. doi:https://doi.org/10.1061/(ASCE)CO.1943-7862.0002278.

- [11] Wei-Chih Chern, Jeongho Hyeon, Tam V Nguyen, Vijayan K Asari, and Hongjo Kim. Context-aware safety assessment system for far-field monitoring. *Automation in Construction*, 149:104779, 2023. doi:https://doi.org/10.1016/j.autcon.2023.104779.
- [12] Wei-Chih Chern, Tam V Nguyen, Vijayan K Asari, and Hongjo Kim. Impact of loss functions on semantic segmentation in far-field monitoring. *Computer-Aided Civil and Infrastructure Engineering*, 38(3):372–390, 2023. doi:https://doi.org/10.1111/mice.12832.
- [13] Keni Bernardin, Florian van de Camp, and Rainer Stiefelhagen. Automatic person detection and tracking using fuzzy controlled active cameras. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, 2007. doi:10.1109/CVPR.2007.383502.
- [14] Gengjie Chen, Pierre-Luc St-Charles, Wassim Bouachir, Guillaume-Alexandre Bilodeau, and Robert Bergevin. Reproducible evaluation of pantilt-zoom tracking. In 2015 IEEE International Conference on Image Processing (ICIP), pages 2055– 2059, 2015. doi:10.1109/ICIP.2015.7351162.
- [15] Samer Hanoun, James Zhang, Vu Le, Burhan Khan, Michael Johnstone, Michael Fielding, Asim Bhatti, Doug Creighton, and Saeid Nahavandi. A framework for designing active pan-tilt-zoom (ptz) camera networks for surveillance applications. In 2017 Annual IEEE International Systems Conference (SysCon), pages 1–6, 2017. doi:10.1109/SYSCON.2017.7934744.
- [16] Yong Li, Liang Pan, and Tao Cheng. A camera ptz control algorithm for autonomous mobile inspection robot. In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), pages 962–967, 2021. doi:10.1109/ICBAIE52039.2021.9389970.
- [17] Sandeep Singh Sandha, Bharathan Balaji, Luis Garcia, and Mani Srivastava. Eagle: End-to-end deep reinforcement learning based autonomous control of ptz cameras. arXiv preprint arXiv:2304.04356, 2023. doi:https://doi.org/10.48550/arXiv.2304.04356.
- [18] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics. On-line: https: //github.com/ultralytics/ultralytics, Accessed: 25/12/2023.
- [19] Ferdinand Wagner, Ruedi Schmuki, Thomas Wagner,

and Peter Wolstenholme. *Modeling software with finite state machines: a practical approach*, volume 1. CRC Press, 2006.

Smart Automatic Mixed Reality-Based Construction Inspection Framework

Boan Tao¹, Jiajun Li¹ and Frédéric Bosché¹

¹School of Engineering, University of Edinburgh, Edinburgh, Scotland, UK boan.tao@ed.ac.uk, Jiajun.Li@ed.ac.uk, f.bosche@ed.ac.uk,

Abstract -

With increasingly complex construction projects, improving inspection efficiency and accuracy is an important challenge. This paper proposes a novel MR-based construction inspection framework that integrates BIM, MR, and AI technologies to achieve automatic inspection tasks. The framework comprises object detection, 2D to 3D projection, and digital twin-based object recognition and MRbased visualisation to provide an efficient inspection process. The framework is evaluated in an indoor construction environment with common elements like electrical sockets and switches as a typical example to validate our approach in real-world applications.

Keywords -

Mixed reality; BIM; Digital Twin; Construction Inspection; Camera Projection; Object detection; Deep Learning

1 Introduction

In the construction industry, the importance of efficiency and precision in construction inspection processes cannot be overstated. Traditional inspection approaches, predominantly manual and reliant on 2D drawings and physical presence, are increasingly challenged by the complexity and scale of modern construction projects [1]. Therefore, exploration of digital technologies for enhancing efficiency in these steps is necessary.

Building Information Modeling (BIM), and now Digital Twinning (DT), have emerged as a foundational element in the evolution of construction technologies, offering detailed 3D representations and facilitating effective planning and management. Mixed Reality (MR) blends digital information with the physical environment, which offers an immersive platform that enhances the visualisation of BIM model directly on construction sites. Concurrently, computer Vision (CV) technologies leveraging Artificial Intelligence (AI) are emerging as transformative tools for automating the detection and analysis of site elements and anomalies. Thus, an integrated approach that synergises the detailed visualisation of BIM, the immersive experience of MR, and the analytical capabilities of CV could address the current limitations of traditional inspection methods, including issues with accuracy, efficiency, and safety.

This paper proposes a smart and automatic construction inspection framework that integrates the strengths of BIM, MR, and AI. In the framework, construction inspectors use MR glasses that autonomously perform inspection tasks based on the inspector's location. This system is uniquely optimised to work in a automatic way and with computational efficiency, ensuring effective performance with minimal power consumption during site inspections.

The rest of the paper is organised as follows. Section 2 reviews the relevant literature on BIM, MR and AI in the context of construction inspection applications. Section 3 introduces our automatic inspection framework, detailing its design specifically for integration with BIM, MR and AI technologies. Section 4 illustrates and evaluates the performance of this framework. Section 5 discusses performance and limitations of our method. Section 6 proposes current challenges and future developments. The paper concludes in section 7.

2 Related work

The potential of combining BIM and MR for real-time data processing in construction site inspections is exemplified in Feng and Chen [2]. They propose a system combining BIM and MR, specifically using the headmounted MR device HoloLens. This system allows construction engineers to visualise the BIM model overlaid at the actual construction site, facilitating real-time comparison between planned and actual work, and enabling efficient inspection. Riedlinger et al. [3] demonstrate the potential benefits of the combination of BIM and MR for bridge inspection, including increased precision in locating damages and time-saving potential in damage recording. Ammari and Hammad [4] further extend this integration to multisource facilities information, BIM models, and feature-based tracking in an MR-based setting to enhance collaboration and visual communication between field workers and managers. Similarly, Nguyen et al. [5] design a MR-based system for bridge inspection and maintenance. The system is designed to overlay relevant data and information directly onto the physical bridge structure as viewed through MR devices. This feature enables inspectors to see and assess real-time information about the bridge's condition, maintenance requirements, and other critical data in situ.

The incorporation of AI into MR marks a significant step towards automating inspection processes. Karaaslan et al. [6] and Zakaria et al. [7] discuss the integration of MR and real-time machine learning to enhance structural inspections, particularly for concrete infrastructures like bridges. They use deep learning models that can localise and quantify concrete defects in real-time using MR device. These studies underscore AI's role in analysing BIM data to detect defects and predict maintenance needs, showcasing the potential for more intelligent and proactive construction management.

The existing research primarily focuses on the pairwise combination of these technologies, such as BIM with MR or MR with AI, without fully harnessing the synergistic potential of combining all three. Moreover, current systems still largely depend on manual user input for tasks like locating specific areas or activating the system, which undermines efficiency. There is a need to develop a more autonomous MR system, empowered by BIM and AI, that can independently, automatically and passively identify and process construction site data without extensive user intervention.

3 Method

3.1 Method overview

Our proposed system architecture encompasses two primary components: MR device, specifically chosen as the Hololens2 (HL2), and a Computation Centre (CC), which can be either a local computer or a cloud-based platform. This framework is notably effective in two key use cases within a fully developed BIM context: Facilities Management (FM) inventory and construction project progress and quality monitoring. Firstly, for FM inventory, it enables dynamic interaction with the facility's digital twin, allowing managers to visualise, track, and manage assets efficiently. Secondly, for project progress and quality monitoring, it provides a real-time inspection tool for ensuring construction adheres to planned works. This aids in identifying and rectifying deviations, thus maintaining project integrity and facilitating quality control.

The comprehensive workflow of our proposed framework is depicted in fig. 1. In operation, users equipped with HL2 navigate the construction site. The HL2 (red rectangle) maintains real-time communication with the computation centre, continuously transmitting spatial data regarding the user's position and orientation. Upon receiving this spatial data, the CC (blue rectangle) initiates a series of processes, and send result back to the HL2. Key stages include:

1. Detection zone analysis. The system first evaluates

whether the user is situated within a specially predefined detection zone for each element in the BIM model that needs to be controlled, thereby facilitating a focused and efficient inspection process. The design of the zone is discussed in Section 3.2. It is completed in an offline setting, with the zones stored in the database of CC.

- 2. Camera activation and data acquisition. If the user is within the detection zone, the computation centre sends an activation command to the HL2, which then starts capturing video frames in real-time and transmits them and the camera's intrinsic parameters back to the centre.
- 3. Object detection. The object detector runs in realtime on camera frames to detect target objects (e.g. building components or defects) within that detection zone.
- 4. Orientation validation. The system ensures that user faces the target objects and incident angles between user orientation and wall are within acceptable thresholds, to increase the accuracy of the subsequent camera projection and matching calculations (see next step).
- 5. 2D to 3D Projection. Utilising the 2D detection boxes coordinates, the system computes their projected coordinates in the BIM model (or Digital Twin), through 3D projection using the pinhole camera model.
- 6. Deviation assessment. The projected 3D coordinates are then compared against the as-planned object positions. Compliance is determined based on predefined deviation thresholds, and the results recorded and linked to the project BIM model.
- 7. Visualisation. The inspection results are simultaneously reported to the user visually, highlighting non-compliant from compliant objects, providing an intuitive and immediate visual cue for inspection outcomes.

The following sub-sections provide mode details about the whole process.

3.2 Detection zone analysis

The detection zones are created to focus on specific areas that need inspection or monitoring. When setting up these zones for electrical elements like switches and sockets (which are the focus of the validation presented later), walls are used as primary reference points, with the zones defined as bounding boxes extending from the walls. Parameters for each detection zone are established based on the inspection requirements. Here, the



Figure 1. Real time workflow of the system

primary parameter is the distance between the inspector and the target element. We set the maximum distance at 2 meters, aligning with the optimal range of the HL2 front camera. This distance ensures that the camera captures images of sufficient quality for the computer vision algorithm to perform reliable object detection.

Each detection zone stores their essential information, including: the precise locations and categories of target elements and relevant geometrical data, such as targets' surface normal lines. By pre-storing this data, the system can rapidly process and analyse the images captured by the inspector, significantly speeding up the inspection process.

3.3 Camera activation and data acquisition

HL2 is equipped with an array of sensors that capture spatial and visual data [8]. This includes a depth sensor, an RGB camera, and sensors dedicated to tracking head, hand, and eye movements. Spatial sensors capture spatial information like position, orientation, and movement of the user's head and hands. The front-facing RGB camera captures conventional colour imagery. This can be used for applications requiring visual data from the user's perspective.

Zaccardi et al. [9] provides insights into using Unity's Barracuda on HoloLens 2 for real-time medical AR systems. They found that simpler models like Lenet5 can achieve over 30 fps. In contrast, more complex models

like EfficientNetB0 result in a much lower frame rate, highlighting the balance between model complexity and performance. Therefore, in theory, the computational capabilities of current MR hardware are sufficient to support the execution of deep learning models, including the projection of 3D objects. However, for more effective communication with digital twins and to assess the framework's performance more accurately, we perform both the detection and projection processes in CC. Dibene and Dunn [10] propose a HL2 server application to facilitate the real-time streaming of sensor data over TCP (Transmission Control Protocol). This protocol ensures reliable, ordered, and error-checked delivery of a stream of bytes between applications running on hosts communicating via an IP network. In this project, we implement a multiprocessing approach to efficiently direct the streams of front camera and spatial input data towards a centralised computational hub. This approach facilitates the concurrent processing of diverse data inputs, enhancing the overall efficiency and throughput of the system.

3.4 Object detection

In this study, the overall system is illustrated using the inspection of sockets and switches as an example. But, the method is naturally adaptable to other objects (e.g. fire safety equipment [11]). To detect sockets and switches in images captured by the HL2 camera, a deep learning model is developed, based on YOLOv5m [12], noted for its rapid and precise performance. The pre-trained YOLOv5m model is then retrained (transfer learning) using a dataset comprising 2,026 indoor images featuring sockets and switches, enhanced through various augmentation techniques such as rotation, shearing, and mosaic effects to mimic lens distortion and complex indoor scenarios. The evaluation of the system involved the analysis of 73 images, incorporating 163 instances, and yielded a precision rate of 95% and a recall rate of 86.6%. The system has an inference time of 8.4 milliseconds, and a Non-Maximum Suppression (NMS) time of 2.5 milliseconds per image for an image dimension of (32, 3, 640, 640). This processing speed is particularly advantageous for real-time applications in construction inspection, highlighting the system's capability in both accuracy and efficiency in object detection tasks.

3.5 Real-time position and orientation

In the HL2, image and video streams undergo distortion correction within the image-processing framework prior to application accessibility [13]. Thus we assume that the transmitted image frames conform to a perfect pinhole camera model without distortion. It satisfies the perspective projection equation [14]:

$$p_i = K[\mathbf{R}|\mathbf{t}]P_i \tag{1}$$

The value of camera's intrinsic matrix K, which encapsulates the camera's focal length and the principal point offset, is computed in real-time by the HL2 auto focus-length system and communicated to the computation centre. The extrinsic matrix $\mathbf{E} = [\mathbf{R}|\mathbf{t}]$, encapsulating the rotation and translation vectors of the camera, represents the camera's pose relative to the world coordinates. It undergoes real-time updates to reflect the changes in the camera's position and orientation as the user navigates through the site.

3.5.1 Initialisation

The camera's initial pose $\mathbf{E}_0 = [\mathbf{R}_0|\mathbf{t}_0]$ can be measured by various methods, including QR code scanning [15], or visual analysis of recognisable structures or features [16]. In this study, the initialisation of the camera's pose is conducted through the scanning of a QR code, strategically affixed to a predetermined location (a wall in the case of the experiments reported below).

The QR code is identified, and the coordinates of its corners are extracted, denoted as q_i in the image coordinates. Their corresponding 3D coordinates in a local world coordinate system, designated as Q_i , are known from the pose of the matching twin QR code in the BIM model.

Using the 2D-3D point correspondences $(q_i \text{ and } Q_i)$, the rotation vector (**R**) and translation vector (**t**) of the camera coordinate relative to the world coordinate is calculated. This computation is grounded in the principles outlined in eq. (1).

3.5.2 Real-time updating

HL2 transmits real-time orientation ($\Delta \mathbf{R}$) and position ($\Delta \mathbf{T}$) changes relative to the initial pose. This data is used to update the user's pose and the camera's extrinsic matrix.

Rotation update: The new orientation matrix \mathbf{R}_{new} is computed by multiplying the initial orientation \mathbf{R}_0 with the change in orientation $\Delta \mathbf{R}$:

$$\mathbf{R}_{\text{new}} = \mathbf{R}_0 \cdot \Delta \mathbf{R} \tag{2}$$

Position update: The new position vector \mathbf{P}_{new} is updated by applying the change in position $\Delta \mathbf{T}$ relative to the initial orientation \mathbf{R}_0 , and adding it to the initial position \mathbf{P}_0 :

$$\mathbf{P}_{\text{new}} = \mathbf{R}_0 \cdot \Delta \mathbf{T} + \mathbf{P}_0 \tag{3}$$

Extrinsic matrix update: The extrinsic matrix \mathbf{E}_{new} of the camera, which transforms points from the world coordinates to the camera coordinates, is updated using the new orientation and position:

$$\mathbf{E}_{\text{new}} = \begin{bmatrix} \mathbf{R}_{\text{new}} & | & -\mathbf{R}_{\text{new}} \cdot \mathbf{P}_{\text{new}} \end{bmatrix}$$
(4)

3.6 2D to 3D projection

Given the 2D image coordinates set (u, v) of the vertices of the bounding box enclosing the detected object from section 3.4, the first step is to normalise these coordinates to the camera's coordinate system. The normalised camera coordinates (x, y) are obtained by:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \tag{7}$$

In each frame, the detection target is identified based on the camera-object angle, defined as the angle formed between the camera's line of sight and the normal to the object's surface. This process involves measuring the camera-object angle for every object within the designated detection zone. The object exhibiting the smallest such angle is then selected as the primary detection target for that specific frame. The orthogonal distance, represented as d, between this selected object and the camera, is effectively the z-coordinate value of the object within the camera's coordinate system.

Subsequently, the camera coordinates are transformed by applying a scaling factor equal to d. This step translates the 2D coordinates into 3D camera coordinates (X_c, Y_c, Z_c) :

$$X_c = x \cdot d, \quad Y_c = y \cdot d, \quad Z_c = d.$$
(8)

The final step involves transforming these camera coordinates into 3D world coordinates. This transformation is accomplished using the camera's extrinsic matrix \mathbf{E}_{new} obtained in section 3.5.2:

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{E}_{\text{new}} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}.$$
 (9)

3.7 Deviation assessment and Visualisation

Section 3.6 calculates in real-time the projection of 3D bounding boxes that captures the 'as-is' location of elements within the detection zone. For each detected 'as-is' element, we compute the centroid of its 3D bounding box. This centroid serves as a representative point for comparing the 'as-is' element with corresponding 'as-designed' elements of the same category within the detection zone. The comparison process involves identifying the 'as-designed' element. This proximity-based selection aims to match each 'as-is' element to the most relevant 'as-designed' counterpart.

Given the dynamic and continuous operation of the camera, multiple 3D bounding boxes are projected for the same target over time. These projections may exhibit variations due to factors such as noise, distortion, and limitations inherent to the sensing equipment. To account for these variations, we compute an average centroid for the 'as-is' element across all captured frames. This averaged centroid is then compared to the centroid of the closest 'as-designed' element.

The spatial deviation between the averaged 'as-is' centroid and the 'as-designed' centroid is quantitatively assessed against a predefined threshold. This assessment determines whether the 'as-is' element conforms to the planned design specifications.

The detection and conformance checking results are recorded in the Digital Twin as the average of the projected bounding boxes.

Finally, the result is sent back to the HL2 where the detected bounding box are shown coloured in:

- green, if the element is matched and found conforming;
- *red*, if the element is matched and found non-conforming;
- grey, if the element is not matched.

4 Experimental result

4.1 Result visualisation

Figure 2 and fig. 3 show the digital twin as updated in real time in the CC. The grey mesh is the BIM of room. Within this virtual representation of the room, four different coloured squares are observable; these are designated as detection zones. The HL2 in its current pose (updated in real time) is shown in black. As introduced in section 3.1, the front camera on the HL2 is only activated when the HL2 is situated within these coloured detection zones. If the target object is detected and checked as conforming, the target object is shown with a small green sphere, representing the 'as-is' position. In the digital twin screenshot in fig. 2, three green spheres can be seen on the wall next to the blue detection zone, representing three detected and conforming objects.

The HL2 screen interface, shown in fig. 4 and fig. 5, reports essential information to the user during the inspection process. It reports when the user enters a detection zone and the designated targets for inspection. Objects that align with the as-planned design are explicitly listed on the screen, and for enhanced visual clarity, these compliant objects are highlighted within green bounding boxes. Conversely, objects detected but found to deviate from the as-planned design are enclosed within grey boxes, indicating that their projected 3D positions do not match any element's as-planned position.

4.2 Performance analysis

4.2.1 Initialisation

Using scanning QR codes for determining camera position and orientation is a cost-effective and accessible method. However, this approach has its limitations. The accuracy can be significantly affected by factors such as poor lighting, low camera resolution, and environmental interference. To enhance the accuracy of the initialisation of the camera's position and orientation, we continuous scan the QR code for a duration of 5 seconds while remaining stationary. Then we calculate the mean value of the position and orientation collected during this period. Therefore, transient errors caused by sudden changes in the environment or by the initial positioning of the camera can be averaged out.

In our experiment, a comparative evaluation is conducted between the computed camera position derived from the pin hole model and the position obtained through manual measurements. This comparison revealed that the average position deviation in this initialisation step is approximately 3.49 cm.

This discrepancy can be attributed to two significant factors. Firstly, lens distortion, particularly in the form of radial and tangential distortions, can alter the perceived geometry of the scanned QR code, leading to inaccuracies in the calculation of the camera's position and orientation. Secondly, during the process of breathing, subtle but impactful body movements occur, which can inadvertently shift the camera's position, albeit slightly.

4.3 Real-time projection

During 2D to 3D projection, the method casts rays from the camera's origin through the image plane and into the 3D world. The precision of the projection process is subject to variation due to several factors, including the camera-object angle, the distance between the camera and the object, and the camera's incidence angle, which is defined as the angle between the camera's optical axis and the normal of the surface. To elucidate the correlation between these factors and projection errors, we conducted an experimental study using a single socket target. The experiment is initialised by scanning QR code and then detection and projection are performed at varying angles and distances

We define deviation as the spatial distance calculated from the centre point of the 'as-designed' socket to the centroid of the 3D projected bounding box. In total 12,507 data points are acquired for analysis. In the analysis, the controlled variable method is utilised to ensure rigour and accuracy in the interpretation of the data.

Initially, we fix the camera incidence angles at 0° or 5° , given that the majority of the data fall within this range. Additionally, these angles are chosen due to their



Figure 2. Screenshot of the Digital Twin (switch#1, socket#1, socket#2)



Figure 4. HL2 Screen Interface 1 (switch#1, socket#1)

minimal distortion impact on the projection, ensuring they did not significantly affect the analysis of other parameters. Employing the set parameter of camera incidence angle to select the test subdataset(4,586 data), we analyse the relationship between the camera-object distance and the observed deviations. The results are summarised in the 2D scatter plot shown in fig. 6. Our findings indicate that the deviation maintains a consistent level of stability, remaining below 0.25 m, up to a camera-object distance of 1.1 m. Beyond this threshold, the deviation increases significantly and becomes more erratic. This phenomenon can be attributed primarily to two factors: (1) the amplification of errors in preceding stages, such as sensor measurement or QR code initialisation, due to longer distances; and (2) the inherent limitations of the camera's capabilities adversely affecting detection at extended ranges.

Setting the specified range, where the camera-object distance is less than 1.1 m, result in minimal deviation, as evidenced by prior findings. Then, we investigate the relationship between camera-object incidence angle and deviation, as illustrated in Figure 7. The analysis demonstrates that there is a direct correlation between the deviation and camera-object incidence angle within a range of less than 10° . As this angle surpasses 10° , we observe that the deviation becomes both unstable and significantly higher. Within the angle range of 0° to 5° , deviation remains below 0.22 m, with an average



Figure 3. Screenshot of the Digital Twin 2 (socket#4)



Figure 5. HL2 Screen Interface 2 (socket#4)



Figure 6. Relationship between the camera-object distance and 3D projection deviation.

deviation of 0.16 m. This can be attributed to two possible reasons: (1) camera-object incidence angle affects image distortion and perspective projection, leading to greater deviations at wider angles; and (2) the Inertial Measurement Unit (IMU) sensor measurement inside HL2 is not accurate and stable and thus accumulates errors during calculations.

5 Discussion

In light of the aforementioned findings, it can be deduced that optimal system performance is attained when the camera-object incidence angle is less than 5° and camera-object distance is under 1.1 m. Under these spe-


Figure 7. Relationship between the camera-object incidence angle and 3D projection deviation.

cific conditions, the system demonstrates enhanced efficacy, as evidenced by a mean deviation of approximately 16 cm. That deviation can be ascribed to the following factors.

First, there is an inherent error in the process of initialising the camera's location and orientation using QR code scanning. As discussed in section 4.2.1, this error results in a positional deviation of approximately 3.5 cm. Additionally, a deviation in orientation has been identified, further investigation into which is considered for future research endeavours.

Second, several types of distortions can affect the outcome. Firstly, perspective and lens distortions impact how the sizes and shapes in an image are seen, which can lead to errors in the final 3D model. Then, the way lighting and shadows appear in the image can also change how accurately objects are detected and represented. Additionally, sensor errors, particularly from devices like inertial measurement units (IMUs), introduce further errors. These sensors sometimes struggle to track the exact position and movement of the camera, especially during quick motions.

Considering the various challenges inherent in the process of 2D to 3D projection, and the technological capability of HL2, it appears that using that system, construction positioning conformance can only be confirmed with a threshold of 16 cm. To improve the accuracy of our object detection and projection, two main strategies can be employed. First, we can train our object detection model with images taken in different lighting conditions. This approach would make the model more versatile and accurate in varying lighting environments. Second, we can use additional tools like external sensors to support and enhance the initialisation of camera position and orientation.

6 Future development

It is important to note that our methodology currently assumes a singular detection object per frame. In scenarios involving multiple objects, the projection outcomes for objects other than the primary target are prone to deviations. To address this issue, our future research will develop and integrate an algorithm capable of filtering outliers and averaging projection results.

In construction site management, accurately identifying complex elements like multifunctional media sockets is challenging due to their diverse designs and the need to distinguish their specific types and orientations. A strategy to address this would be to utilise sophisticated object detection technologies, trained on an extensive array of socket designs and configurations.

Besides, construction sites often involve situations where materials and equipment that partially occlude crucial elements. The compact placement of items on sites complicates the identification process. To overcome these obstacles, applying data augmentation methods such as cutout and mosaic in the training phase can enhance the model's ability to handle occlusions. Additionally, enhancing the network design with attention mechanisms enables the model to pinpoint more nuanced features, boosting its detection performance.

The proposed system is designed to automate the process of (progress and) quality control in construction projects, ensuring that all installed components, such as sockets, switches, and structural elements, adhere to the project's specifications and quality standards. This application can significantly reduce human error and increase the efficiency of the inspection process. The system holds potential for other applications, such as monitoring and ensuring compliance with safety regulations on construction sites. By detecting potential hazards or non-compliance with safety standards (e.g., improper installation of safety equipment, obstruction of emergency exits), the system can contribute to a safer work environment.

7 Conclusion

This paper presents a novel MR-based construction inspection framework. The framework integrates AIbased object detection with 2D to 3D projection techniques and matching against the facility's DT to achieve automatic and passive inspection work, facilitated by the communication system between the MR device and computation centre. The results are stored in the DT and can be reviewed in an interactive, and user-friendly way by the MR user on site. The framework's practicality and effectiveness were evaluated in an indoor construction environment. The results from these tests demonstrate the system's feasibility in real-world inspection processes, albeit with limitations on the quality of the results that can reasonably be achieved.

References

- Y. Li and P. Gu. Free-form surface inspection techniques state of the art review. *Computer-Aided Design*, 36(13):1395–1417, 2004. doi:10.1016/j.cad.2004.02.009.
- [2] C.-W. Feng and C.-W. Chen. Using bim and mr to improve the process of job site construction and inspection. WIT Transactions on the Built Environment, 192:21–32, 2019. doi:10.2495/BIM190031.
- [3] U. Riedlinger, F. Klein, M. Hill, C. Lambracht, S. Nieborowski, R. Holst, S. Bahlau, and L. Oppermann. Evaluation of mixed reality support for bridge inspectors using bim data: Digital prototype for a manual task with a long-lasting tradition. *i-com*, 21(2):253–267, 2022. doi:10.1515/icom-2022-0019.
- [4] K. El Ammari and A. Hammad. Remote interactive collaboration in facilities management using bim-based mixed reality. *Automation in Construction*, 107:102940, 2019. doi:10.1016/j.autcon.2019.102940.
- [5] D. C. Nguyen, R. Jin, and C. H. Jeon. Developing a mixed-reality based application for bridge inspection and maintenance. In *The 20th International Conference on Construction Applications of Virtual Reality (CONVR 2020)*, 2020. doi:10.1108/CI-04-2021-0069.
- [6] E. Karaaslan, M. Zakaria, and F. N. Catbas. Mixed reality-assisted smart bridge inspection for future smart cities. In *The Rise of Smart Cities*, pages 261–280. 2022. doi:10.1016/B978-0-12-817784-6.00002-3.
- [7] M. Zakaria, E. Karaaslan, and F. N. Catbas. Realtime ai-based bridge inspection using mixed reality platform. In *Structures Congress 2023*, pages 120– 131, 2023. doi:10.1061/9780784484777.012.
- [8] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. Stühmer, T. J. Cashman, B. Tekin, J. L. Schönberger, et al. Hololens 2 research mode as a tool for computer vision research. arXiv preprint arXiv:2008.11239, 2020. doi:10.48550/arXiv.2008.11239.
- [9] S. Zaccardi, T. Frantz, D. Beckwée, E. Swinnen, and B. Jansen. On-device execution of deep learning models on hololens2 for real-time augmented reality medical applications. *Sensors*, 23 (21):8698, 2023. doi:10.3390/s23218698.
- [10] J. C. Dibene and E. Dunn. Hololens 2 sensor streaming. arXiv preprint arXiv:2211.02648, 2022. doi:10.48550/arXiv.2211.02648.

- [11] A. Corneli, B. Naticchia, M. Vaccarini, F. Bosché, and A. Carbonari. Training of yolo neural network for the detection of fire emergency assets. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 37, pages 836–843, 2020. doi:10.22260/ISARC2020/0115.
- [12] G. Jocher, A. Stoken, J. Borovec, L. Changyu, A. Hogan, L. Diaconu, J. Poznanski, L. Yu, P. Rai, R. Ferriday, et al. ultralytics/yolov5: v3. 0. *Zenodo*, 2020. doi:10.5281/zenodo.3983579.
- [13] Microsoft. Locatable camera in mixed reality. https://learn.microsoft. com/en-us/windows/mixed-reality/ develop/advanced-concepts/ locatable-camera-overview, 2023. Accessed: 2023-12-01.
- [14] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the Seventh IEEE International Conference* on Computer Vision, pages 666–673 vol.1, 1999. doi:10.1109/ICCV.1999.791289.
- [15] J.-I. Kim, H.-S. Gang, J.-Y. Pyun, and G.-R. Kwon. Implementation of qr code recognition technology using smartphone camera for indoor positioning. *Energies*, 14(10):2759, 2021. doi:10.3390/en14102759.
- [16] P.-E. Sarlin, A. Unagar, M. Larsson, H. Germain, C. Toft, V. Larsson, M. Pollefeys, V. Lepetit, L. Hammarstrand, F. Kahl, and T. Sattler. Back to the feature: Learning robust camera localization from pixels to pose. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3246–3256, 2021. doi:10.1109/CVPR46437.2021.00326.

Unveiling Building Façade Deterioration: A Drone-Powered Deep Learning Approach for Seamless Tile Peeling Detection

Ngoc-Mai Nguyen¹, Minh-Tu Cao² and Wei-Chih Wang²

¹Department of Civil Engineering and Environmental Informatics, Minghsin University of Science and Technology, Taiwan

²Department of Civil Engineering, National Yang Ming Chiao Tong University, Taiwan <u>nnmai@must.edu.tw, mtcao@nycu.edu.tw, weichih@nycu.edu.tw</u>

Abstract

The accurate segmentation of tile peeling on building facades holds considerable significance for effective building maintenance, particularly in regions like Taiwan, where tiles are the predominant facade protection. This research introduces YOLOM, a novel deep-learning-based segmentation model designed to address this challenge. YOLOM harnesses the capabilities of You Only Look Once version 7 (YOLOv7) and incorporates the BlendMask-based segmentation technique, further augmented by the Efficient Layer Aggregation Network (ELAN) to enhance feature discrimination and extraction capabilities specifically tailored for scenarios involving tile peeling. Employing a dataset comprising 400 images featuring 758 instances of peeling and 525 instances of sealed tiles observed during on-site surveys of public buildings, YOLOM exhibits outstanding segmentation performance. It outperforms the Resnet-BlendMask50 FPN with improvements of 7.1% of mean average percentage (mAP) and 0.4% of the average precision (AP) at the intersection over union (IoU) of 50%. Remarkably, YOLOM consistently surpasses other models, showcasing a 19.5% and 2.2% lead in AP for small and large objects, respectively. In a noteworthy advancement, YOLOM seamlessly integrates with drone technology, enhancing its capabilities for aerial surveying of building facades. This integrated approach proves invaluable for building maintenance teams, enabling proactive and cost-effective interventions. The study introduces a distinctive framework seamlessly integrating cutting-edge neck backbone and modules, particularly emphasizing the ALAN. The innovative YOLOM model establishes a new standard in artificial intelligence (AI) techniques for building maintenance, contributing significantly to academic discussions surrounding AI-enhanced image segmentation.

Keywords -

Tile peeling; Building façade; Building maintenance; Computer vision; YOLOv7; Deep learning; BlendMask technique; ALAN.

1 Introduction

While tile peeling may initially seem like a cosmetic concern in aging buildings, its ramifications are farreaching. Beyond aesthetics, the detachment of tiles poses a direct threat to residents, risking falling accidents and compromising structural stability. The erosion of safety extends to the building's core, disrupting waterproofing and insulation capabilities. Exposed areas become susceptible to rainwater and humidity, accelerating material degradation and jeopardizing the adhesion between tiles and the structure. Recognizing tile peeling as more than a visual issue is crucial; it's a fundamental step in safeguarding both the safety and longevity of the built environment.

Routine inspections of building facades are imperative to address these risks. However, traditional inspection methods, involving manual surveys, photographic documentation, and physical condition recording, are labor-intensive and expensive and pose safety hazards for surveyors—particularly when navigating precarious sections of buildings such as highrise rooftops and sidewalls [1]. Given the many buildings necessitating inspection, there is an urgent need to refine traditional methods to mitigate prolonged risks to structures and residents.

In response to the challenges inherent in infrastructure inspection, researchers and industry professionals are collaborating to explore innovative solutions. They are turning to advanced image analysis techniques, driven by the synergy of computer vision and artificial intelligence (AI), as a promising avenue for automating the assessment of damaged components in buildings and infrastructure [2-5]. Various models, including Faster R-CNN, SSD, SSD_Lite, and different iterations of You Only Look Once (YOLO), have proven successful in detecting and categorizing damage on surfaces such as concrete structures, metro tunnels, bridges [6, 7] and road surfaces [8].

Despite numerous studies on defect survey work, a literature gap exists regarding advanced deep learning (DL) methods for identifying damage to architectural components, such as tile peeling on building facades. Existing methodologies lack practical modifications, potentially compromising accuracy or processing time. Additionally, the intricate context of captured images and constraints in drone-to-building surface access pose challenges in feature extraction for recognizing areas with peeling. The synergy between potent segmentation models and drone-powered technologies holds immense revolutionizing potential for building facade maintenance practices.

Addressing these literature gaps in problem-solving and methodology, this study introduces YOLOM, a pioneering segmentation model meticulously crafted to address the challenge of segmenting tile peeling areas on building facades. YOLOM leverages the strengths of YOLO version 7 [9] and integrates the segmentation framework of the BlendMask technique [10, augmented by efficient layer aggregation network (ELAN) blocks [10]. These ELAN blocks enhance feature discrimination and counteract gradual convergence deterioration, bolstering the model's performance in identifying and delineating tile peeling instances. Operating within a onestage framework for pixel-wise segmentation, YOLOM capitalizes on the BlendMask-based segmentation mechanism, offering a robust solution to overcome identified limitations in the literature. By synergistically combining YOLOv7 with BlendMask, our aim is to establish a resilient segmentation model that significantly enhances the effectiveness and comprehensiveness of tile peeling inspection on building facades.

2 Literature review

In contemporary scholarly discourse, a discernible focus exists on harnessing AI and computer vision methodologies to facilitate scrutinizing structural components within buildings. These advancements have yielded significant benefits by furnishing tools identifying nuanced features such as subtle cracks, deformations, and structural irregularities. These imperceptible nuances might elude the human eye or escape manual inspections, making technological interventions indispensable for transforming the efficacy, precision, and inclusivity of inspections of building structures [11].

Despite the predominant emphasis on structural elements, a noticeable lack of attention has been directed towards architectural components, specifically facades and exterior wall cladding. This oversight is significant considering that, similar to their structural counterparts, architectural elements are vulnerable to wear, damage, and degradation as time progresses. The consequences of their decline extend beyond aesthetic considerations, influencing the overall functionality of a building and contributing to heightened maintenance expenses [12]. It becomes imperative to customize AI and computer vision techniques for architectural inspections, presenting a more comprehensive strategy to ensure the optimal condition of every aspect of a building, encompassing both structural and architectural facets.

Within the domain of computer vision applications for building inspections, especially in examining architectural components, enduring challenges persist despite recent advancements. A notable example is illustrated in the study undertaken by Kung, Pan [13], where a VGG-16 classifier [14] successfully classified damage on exterior wall tiles, attaining commendable accuracy. However, practical apprehensions regarding the viability of such a system emerge, particularly concerning capturing images at elevated heights and acquiring detailed images encompassing entire wall spans.

Expanding upon the initial research efforts, Guo, Wang [15] delved into applying a semi-supervised convolutional neural network (CNN) to classify façade damage, particularly under constraints of limited training datasets. Subsequent advancements were realized by Guo, Wang [11], who employed the Mask Region-based Convolutional Neural Network (Mask R-CNN) model to delineate plastered and painted façades. This application exhibited promising segmentation accuracy, with a mean average precision (mAP) of 58.4%. In a more comprehensive inquiry, Lee, Hong [16] scrutinized the efficacy of a Faster R-CNN architecture in the segmentation and categorization of defects on building facades. Notably, an average precision (AP) of 62.7% was achieved across all trained defects, employing an intersection over union (IoU) threshold of 0.5. Despite the laudable predictive performance, it is pertinent to acknowledge a significant limitation inherent in the Mask R-CNN model-its protracted inference time.

In recent research endeavors, Junior, Ferreira [17] made notable contributions by introducing the U-net, coupled with diverse Residual networks as the backbone architectures, to track crack lines in ceramic tiles. Extending the application of computer vision to address issues related to building facades, scholars have employed Faster R-CNN and Mask R-CNN models to identify and segment scratches on building glass panels [18]. The experimental results presented by Dais, Bal [19] compellingly support the effectiveness of DL on the crack segmentation on masonry surfaces. These findings, in conjunction with the previously mentioned studies,

underscore the growing potential of computer vision and DL in advancing methodologies for assessing facade defects.

The knowledge extracted from existing literature emphasizes that DL methods for detecting or segmenting defects on building facades largely align with conventional approaches. This observation underscores a notable gap, indicating an urgent necessity to innovate and enhance advanced DL models tailored to the intricate challenges of detecting building facade defects, including issues like tile peeling. Models prioritizing speed, robustness, and user-friendliness are essential to address practical concerns faced by building maintenance agencies. Beyond the immediacy of pragmatic considerations, the prospective trajectory and widespread integration of sophisticated DL-based computer vision models hold transformative potential for the building engineering sector. This paradigm shift can endow professionals with heightened levels of precision, efficiency, and comprehensiveness in the realm of facade inspections and the formulation of intervention strategies.

3 BlendMask-based YOLOv7 model

3.1 BlendMask-based image segmentation procedure

Introduced by Chen, Sun [20], BlendMask is a onestage instance segmentation model within the Fully Convolutional One-Stage Object Detection (FCOS) framework. Its departure from the pre-defined anchor boxes employed in YOLO family models or region proposals like Mask R-CNN sets it apart, contributing to BlendMask's notably swift inference time. The architecture of the BlendMask model is composed of a feature extraction network and a mask prediction branch. The feature extraction network integrates a fusion of a residual network (Resnet) and a feature pyramid network (FPN). Concurrently, the mask branch incorporates three crucial components: 1) a bottom module determining the relative position of object instances, denoted as score maps; 2) a top layer generating specific attention maps for a detected region, concentrating on relationships between pixel pairs within the same instance by learning an embedding space, and 3) a blender module aligning the score maps with attention tensors (refer to Figure 1).

The bottom module, known as a "score map," predicts the location of a target object. Consequently, the output of the bottom module comprises bases (B) with a shape of $b \times n \times H/_S \times W/_S$, where b represents the batch size, n is the number of bases, and s is the output stride. The feature pyramid network output, including P3, P4, and P5, serves as the input for the bottom module. P4 and P5 undergo interpolation using the DeepLabV3+ decoder to match the size of P3, followed by concatenation through stacking. The loss function in this phase, termed semantic segmentation loss, is computed using the cross-entropy function.

Obtaining the feature pyramid network output (P3 \sim P7) involves applying a convolutional layer to the tower's output. The tower is then expanded with a solitary convolutional layer, responsible for producing the bounding box size, center coordinates of the bounding box (center-nest), and determining the object class confined within the bounding box. Additionally, attention A is provided as the bounding box score, where the shape of this attention is $n \times M \times M$, with $M \times M$ denoting the resolution set at 14×14 in this study, and n=4. To finalize the bounding boxes for subsequent steps, the post-processing technique of FCOS [21] is applied to refine the bounding boxes $P = \{p_d \in \mathbb{R}^0_+ | d = 1, ..., D|\}$ with the highest scores $A = \{a_d \in \mathbb{R}^{K \times M \times M}_+ | d =$ 1, ..., D. Two components contribute to the loss function in this phase, namely focal loss (L_{cls}) and IoU regression loss (L_{reg}) .

$$L(\{p_{x,y}\},\{t_{x,y}\}) = \frac{1}{N_{pos}} \sum_{x,y} L_{cls}(p_{x,y}, c_{x,y}^{*}) + \frac{\lambda}{N_{pos}} \sum_{x,y} \mathbb{I}_{\{c_{x,y}^{*}>0\}} L_{reg}(t_{x,y}, t_{x,y}^{*})$$
(1)

where N_{pos} is positive samples, λ is the weight of regression loss (L_{reg}) term, $\mathbb{I} = 1$ if $c_{x,y}^* > 0$, and $\mathbb{I} = 0$ if $c_{x,y}^* \le 0$



Figure 1. Content of tower block

The pivotal element within BlendMask is the blender module, which plays a crucial role in assimilating information from the bottom-level bases (*B*) and region proposals accompanied by corresponding top attention (*A*). During the training process, ground truth boxes serve as the region proposals, while in the prediction process, the bounding box is deduced. Given the varying sizes of the proposals, the Blender module employs the RoIPooler function to extract the area of the *K* bases associated with each proposal (p_d). Subsequently, this area is resized to a fixed size ($R \times R$) with the feature shape denoted as r_d , as outlined in Equation (2). In executing this task, the RoIAlign technique was adopted, implementing bilinear poolers.

$$r_d = RoIPool_{R \times R}(B, p_d), \quad \forall d \in \{1, \dots, D\}$$
(2)

The attention resolution, denoted as $M \times M$, undergoes interpolation to match the size of the proposals $(R \times R)$, forming a shape set $R = \{r_d | d = 1, ..., D\}$. Subsequently, a'_d is subjected to normalization using the SoftMax function across the K bases, yielding the score map set $S = \{s_d | d = 1, ..., D\}$. The next step involves the element-wise product between each entity r_d of the region proposal set R and the corresponding s_d of the score set S. This operation is performed for each of the K bases, and the results are summed to determine the mask logit (m_d) , as outlined in Equation (3). The parameter K is consistently set at a value of 4, while R assumes values of 28 and 56, as proposed by Chen, Sun [20].

$$\begin{aligned} a'_{d} &= interpolate_{M \times M \to R \times R}(a_{d}), \\ &\forall d \in \{1, \dots, D\} \end{aligned} \tag{3}$$

$$s_d = softmax(a'_d), \ \forall d \in \{1, \dots, D\}$$
(4)

$$m_d = \sum_{k=1} s_d^k \circ r_d^k, \quad \forall d \in \{1, \dots, D\}$$
(5)

3.2 ELAN-backbone and CSP-SPP + ELAN-PAN integration

Effective extraction and processing of features hold a central role in the analysis of image data. The advent of big data and advancements in convolutional neural networks (CNNs) and high-performance computers have facilitated the practicality of analyzing extensive image datasets. Using random trials is deemed impractical for developing efficient CNN networks tailored to extract specific task-related features. Therefore, a meticulous analysis of the particular task, incorporating intricate adjustments, becomes imperative.



Figure 2. BlendMask with ELAN-base backbone and CSP-SPP+ELAN-PAN

After the YOLOv7 operations are completed, applying the SiLU activation function follows all batch normalizations. In a departure from conventional methods, this study introduces a unique network based on the ELAN [10]. ELAN, designed to counteract gradient deterioration in larger networks, combines elements from VoVNet [22] and CSPNet [23] to optimize the gradient length within a computational block. Notably, the layer aggregation network is trainable independently of the backbone network, facilitating faster training and experimentation, rendering it highly efficient for real-time operations.

Concerning the construction of the backbone, the ELAN is strategically incorporated between the downsampling blocks, as depicted in Figure 2. Diverging from YOLOv7, a deviation occurs by splitting two adjacent Cross-Stage Partial (CSP) blocks, and the output of all six CSP blocks in ELAN is directly concatenated, forming a structure referred to as ELAN-6. This ELAN-6 output, spanning C2 to C5, serves as the input at the forefront of the network. Moreover, ELAN-6 is employed at the head of the network to enhance feature extraction in CSP-SPP + ELAN-PAN, with PAN representing the Path Aggregation Network.

In this investigation, the Feature Pyramid Network (FPN) is substituted with the Pyramid Attention Network based on the Efficient Layer Aggregation Network (ELAN-PAN). ELAN is integrated into the layer scaling of PAN, processing backbone features before entering ELAN-PAN. In contrast to BlendMask, ELAN-PAN's scaling progresses from P2 to P7, as opposed to P7 to P2. ELAN is introduced into the transformation between the layers of ELAN-PAN. Due to a significant increase in the number of parameters in the P6 and P7 generations without substantially enhancing the segmentation model's inference power, ELAN avoids the downscaling task for P6 and P7. To align with the foundational structure of the BlendMask operating system, the channel number for each layer in ELAN-PAN is standardized. As a result, this study designates the proposed model as YOLOM.

4 Data collection and processing

Emphasizing the concern for public safety posed by peeling tiles, the surveyed buildings were strategically selected in high pedestrian-traffic zones. These encompassed various structures such as university campus buildings, apartment complexes, hospitals, government offices, and activity centers. Image data collection employed a Nikon D3200 digital camera, Autel Robotic EVO Lite+ unmanned aerial vehicle (UAV), Canon EOS M10, and iPhone 12 Pro, capturing photos across different seasons, times of day, and lighting conditions (*e.g.*, cloudy days, shadows, high and low light, and reflected light). Over a year, the survey team conducted fieldwork, ensuring diverse images with complex backgrounds to enhance the model's applicability in real-world scenarios. Each object was documented from various angles and within randomly sized rooms, contributing to the model's adaptability. The survey team utilized maximum zoom settings, especially for images capturing tile peeling at elevated heights. Following model training, an Autel Robotic EVO Lite+ unmanned aerial vehicle (UAV) was deployed to survey tile peeling in high-rise buildings.

A dataset comprising 400 surveyed images was employed in developing the tile peeling segmentation model, encompassing 758 instances of peeling and 529 instances of sealed tiles (refer to Figure 3 for representative samples). Upholding the quality of the dataset was a meticulous process involving labeling and verification by two additional members of the research team to annotate object instances. This rigorous approach ensured that the DL models were trained on a dataset of superior quality. Table 1 details the number of images and instances at each survey location for a comprehensive overview of the dataset. The primary survey locations across northern Taiwan are visually represented in Figure 4, accompanied by image samples collected from these regions.



Figure 3. Surveying locations in the northern Taiwan



Figure 4. Surveying locations in northern Taiwan

Location	Imagas	Dhaga	instances		
	mages	Fliase	Peeling	Sealed	
8	400	Training	638	410	
	400	Test	120	119	

Table 1. Experimental dataset

5 Experimental results and discussions

In object segmentation, the assessment of model performance extends to metrics such as AP across varying IoU thresholds and mAP considering different object sizes. These metrics are pivotal for gaining a nuanced comprehension of the capabilities inherent in an object segmentation model. IoU is a critical gauge, elucidating the accuracy with which the predicted bounding box aligns with the actual object. The exploration of diverse thresholds, including AP50 (IoU > 0.5) and AP75 (IoU > 0.75), for small (APs), medium (APm), and large (API) objects enables a nuanced evaluation of the model's precision across distinct levels.

The models in this comparative analysis underwent training on a sophisticated computing setup featuring an NVIDIA GeForce RTX 4090 24G GDDR6, RAM DDR5 5600MHz (2x32G), SSD Samsung 970 EVO Plus NVMe M.2, and an Intel CPU i7-13700-Core Processor. Hardware selection plays a substantial role in influencing training outcomes, particularly in DL models where GPU capabilities are paramount. A standardized image input size of 640x640 was maintained, and batch sizes of six were employed during the training process. Each model underwent an extensive training regimen encompassing 10,000 iterations, with meticulous fine-tuning to ensure optimal segmentation precision. These values were

empirically chosen to yield the most favorable outcomes in the experiments, enhancing the model's proficiency in accurately segmenting tile peeling.

The experimental results of comparative models are presented in Table 2. Regarding segmentation, YOLOM stands out as the dominant model because it achieves the greatest values of all evaluation metrics.

The YOLOM obtained at least a 7.1% improvement compared with the second-best model, BlendMask– Resnet50 FPN, in terms of increasing the strictest AP value. The segmentation difference between the proposed model and other models incrementally increases as the complex challenge increases from large to small objects. As seen in Table 2, the YOLOM attained 2.2%, 11,9%, and 19.5% improvements in segmenting small, medium, and large objects compared with the remaining models, respectively.

YOLOM dominates the YOLOv7 mask that was published along with the YOLOv7 detection model by Wang et al. [25], boosting the AP50 and mAP values by 4.6% and 14.0% improvements. YOLOMASK is established by integrating CSP-SPP + ELAN-PAN and substituting the Resnet backbone with the YOLOv7 backbone while preserving the FPN found in BlendMask. This model structure was not streamlined, which is proven by the large segmentation accuracy drop of 16.7% and 14.0% of AP50 and mAP values. However, there is still an appraisal for its performance in segmenting small objects compared with BlendMask-Resnet50 FPN because it is supported by ELAN blocks. This study also uses BlendMask-CSPDarknet FPN to compare with the proposed model by substituting the YOLOv7 backbone with CSP Darknet. This model does not work efficiently, as proven by yielding much lower AP values than YOLOM and BlendMask-Resnet50 FPN with different IoU challenges.

	Model	AP50	AP75	mAP	AP ^s	AP^m	AP^{l}
tion	BlendMask–Resnet50 FPN	0.799	0.537	0.496	0.181	0.399	0.652
	BlendMask–CSPDarknet FPN	0.639	0.386	0.367	0.107	0.385	0.471
nenta esult	YOLOv7 mask by Wang, Bochkovskiy [24]	0.757	NA	0.427	NA	NA	NA
Segn	YOLOMASK	0.636	0.348	0.352	0.125	0.307	0.473
01	YOLOM	0.803	0.699	0.567	0.376	0.518	0.674
3	BlendMask–Resnet50 FPN	0.811	0.611	0.543	0.334	0.475	0.662
sounding box results	BlendMask–CSPDarknet FPN	0.650	0.433	0.402	0.187	0.431	0.475
	YOLOv7 mask by Wang, Bochkovskiy [24]	NA	NA	NA	NA	NA	NA
	YOLOMASK	0.650	0.420	0.387	0.162	0.366	0.476
щ	YOLOM	0.807	0.736	0.622	0.540	0.583	0.702

Table 2. Experimental results of comparative models

6 Conclusions

In the realm of building façade recognition, the early identification and precise delineation of issues emerge as pivotal factors. This study introduces a cutting-edge deep-learning-based segmentation tool named the YOLOM model, meticulously crafted to discern instances of tile peeling on building exteriors. Its application extends valuable support to building owners, aiding in proactive maintenance and resource conservation endeavors.

The YOLOM model has set a performance benchmark by incorporating state-of-the-art backbone and neck modules. A comparison with the BlendMask-Resnet50 FPN underscores its superiority, boasting a remarkable 7.1% increase in superior mean (mAP) and a notable 16.2% enhancement in AP75 values. Furthermore, it exhibits substantial leads in various precision metrics—APs, APm, and APl-with improvements of 19.5%, 11.9%, and 2.2%, respectively. These outcomes stem from a diligently curated dataset featuring 400 building façade images containing 1287 instances of peeling and sealed tiles. These validations affirm the model's robustness and propel academic and practical advancements.

Future endeavors will focus on integrating global building visuals and collecting tile-peeling images with diverse resolutions, enhancing the adaptability of YOLOM for improved training dynamics without compromising accuracy. Subsequent research will involve an ablation analysis to comprehensively assess the impact of various model components on the performance of the YOLOM, offering valuable insights for future enhancements.

Future studies could explore how variations in photographic datasets, including facade-distance, camera angle, lighting conditions, and image resolution, affect the performance of YOLOM, with a focus on strategies to mitigate challenges and optimize model performance across diverse real-world scenarios. Integration of augmented and synthetic data could augment dataset diversity and size, enhancing the model's generalization capabilities. Additionally, investigating transfer learning and domain adaptation techniques could improve YOLOM's adaptability to different datasets and mitigate domain shift issues. Evaluation of real-world deployment challenges should also be conducted to ensure successful implementation and adoption of YOLOM in building facade maintenance workflows. advancing its effectiveness and impact in practical applications.

Beyond its role as a mere model, this research contributes a robust framework to the academic community, accentuated by the ELAN-based structure. This foundation encourages researchers to explore advanced segmentation models. In conclusion, YOLOM is a guiding light in academic exploration, symbolizing AI's transformative potential in practical spheres, particularly building façade maintenance. As the journey forward unfolds with an array of improvements and expansions, collective efforts hold the potential to reshape façade maintenance narratives.

Acknowledgments

The authors would like to thank the National Science and Technology Council, Taiwan for financially supporting this research under contracts 111-2221-E-A49-189- and 112-2221-E-159-001-.

References

- [1] Kim B, Yuvaraj N, Sri Preethaa KR, Arun Pandian R. Surface crack detection using deep learning with shallow CNN architecture for enhanced computation. *Neural Computing and Applications*, 33:9289-305, 2021.
- [2] Dabetwar S, Padhye R, Kulkarni NN, Niezrecki C, Sabato A. Performance evaluation of deep learning algorithms for heat loss damage classification in buildings from UAV-borne infrared images. *Journal of Building Engineering*, 75:106948, 2023.
- [3] Katsigiannis S, Seyedzadeh S, Agapiou A, Ramzan N. Deep learning for crack detection on masonry façades using limited data and transfer learning. *Journal of Building Engineering*, 76:107105, 2023.
- [4] Xu Y, Qian W, Li N, Li H. Typical advances of artificial intelligence in civil engineering. *Advances in Structural Engineering*, 25:3405-24, 2022.
- [5] Qiao W, Zhao Y, Xu Y, Lei Y, Wang Y, Yu S, et al. Deep learning-based pixel-level rock fragment recognition during tunnel excavation using instance segmentation model. *Tunnelling and Underground Space Technology*, 115:104072, 2021.
- [6] Li D, Xie Q, Gong X, Yu Z, Xu J, Sun Y, et al. Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. *Advanced Engineering Informatics*, 47:101206, 2021.
- [7] Deng J, Lu Y, Lee VC-S. Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, 35:373-88, 2020.
- [8] Huyan J, Li W, Tighe S, Zhai J, Xu Z, Chen Y. Detection of sealed and unsealed cracks with complex backgrounds using deep convolutional neural network. *Automation in Construction*, 107:102946, 2019.
- [9] Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:220702696*, 2022.

- [10] Wang C-Y, Liao H-YM, Yeh I-H. Designing Network Design Strategies Through Gradient Path Analysis. arXiv preprint arXiv:221104800, 2022.
- [11] Guo J, Wang Q, Li Y. Evaluation-oriented façade defects detection using rule-based deep learning method. *Automation in Construction*, 131:103910, 2021.
- [12] Li J, Wang Q, Ma J, Guo J. Multi-defect segmentation from façade images using balanced copy-paste method. *Computer-Aided Civil and Infrastructure Engineering*, 37:1434-49, 2022.
- [13] Kung R-Y, Pan N-H, Wang CCN, Lee P-C. Application of Deep Learning and Unmanned Aerial Vehicle on Building Maintenance. *Advances in Civil Engineering*, 2021:5598690, 2021.
- [14] He K, Zang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*:770-8, 2016.
- [15] Guo J, Wang Q, Li Y. Semi-supervised learning based on convolutional neural network and uncertainty filter for façade defects classification. *Computer-Aided Civil and Infrastructure Engineering*, 36:302-17, 2021.
- [16] Lee K, Hong G, Sael L, Lee S, Kim HY. MultiDefectNet: Multi-Class Defect Detection of Building Façade Based on Deep Convolutional Neural Network. *Sustainability*, 12, 2020.
- [17] Junior GS, Ferreira J, Millán-Arias C, Daniel R, Junior AC, Fernandes BJT. Ceramic Cracks Segmentation with Deep Learning. *Applied Sciences*, 11, 2021.
- [18] Wang N, Zhao X, Zou Z, Zao P, Qi F. Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 35:277-91, 2020.
- [19] Dais D, Bal İE, Smyrou E, Sarhosis V. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. *Automation in Construction*, 125:103606, 2021.
- [20] Chen H, Sun K, Tian Z, Shen C, Huang Y, Yan Y. Blendmask: Top-down meets bottom-up for instance segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition:8573-81, 2020.
- [21] Tian Z, Shen C, Chen H, He T. Fcos: Fully convolutional one-stage object detection. *Proceedings of the IEEE/CVF international conference on computer vision*:9627-36, 2019.
- [22] Lee Y, Hwang J-w, Lee S, Bae Y, Park J. An energy and GPU-computation efficient backbone network for real-time object detection. *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition workshops:0-, 2019.

- [23] Wang C-Y, Liao H-YM, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H. CSPNet: A new backbone that can enhance learning capability of CNN. *Proceedings* of the IEEE/CVF conference on computer vision and pattern recognition workshops:390-1, 2020.
- [24] Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*:7464-75, 2023.

Accelerating Indoor Construction Progress Monitoring with Synthetic Data-Powered Deep Learning

Mathis Baubriaud^{1,2}, Stéphane Derrode¹, René Chalon¹ and Kevin Kernn²

¹Centrale Lyon, CNRS, Claude Bernard Lyon 1, INSA Lyon, Lumière Lyon 2, LIRIS, UMR5205, 69130 Ecully, France ²SPIE Building Solutions, F-69320 Feyzin, France

(mathis.baubriaud, stephane.derrode, rene.chalon)@ec-lyon.fr, kevin.kernn@spie.com

Abstract -

In recent years, there has been a growing interest in automated indoor construction progress monitoring (ICPM) to maximize precision and reduce human intervention. Computer vision approaches, especially based on deep learning (DL) methods, have shown great potential in this task. However, training DL models require large-scale datasets, which are often costly and laborious to obtain, specifically for indoor construction environments. This study proposes an automated approach to generate real-world-like synthetic data of indoor construction by combining building information modeling (BIM) and a photorealistic graphics engine. The approach was validated by efficiently producing annotated synthetic datasets of mechanical, electrical, and plumbing components from various BIM models. A state-of-theart instance segmentation network was trained using those datasets alongside real manually annotated data and transfer learning methods to assess the results. Preliminary experiments using an on-site augmented reality device demonstrate the promising efficiency of DL for ICPM.

Keywords -

Building Information Modeling; Computer Vision; Deep Learning; Progress Monitoring; Indoor Construction; Augmented Reality.

1 Introduction

Progress monitoring of construction sites is essential because it gives managers the information needed to act quickly and wisely. Ineffective progress monitoring leads to a loss of control, causing time and cost overruns. The manual data entry required by conventional progress monitoring techniques is laborious, time-consuming and prone to human error [1]. Inspections of interior work — for example, mechanical, electrical, and plumbing (MEP) installation — can be even more difficult for inspectors due to the level of detail and interdependence of tasks [2]. Therefore, it becomes desirable to automate these tasks.

The use of computer vision (CV) has been studied on a wide range of construction applications, such as site safety inspection [3], localization, navigation [4], and 3D recon-

struction [5, 6], among others. Several studies on visionbased construction progress monitoring have also been reported [7]. Deep learning (DL) techniques excel in numerous CV tasks and is increasingly used in this field [8]. Although DL models are capable of powerful feature representation, they rely heavily on large-scale, high-quality training datasets. Three methods can be found commonly in the literature to acquire such datasets (1) using preexisting datasets, (2) using web crawling techniques, and (3) by capturing the data manually [9].

Regarding (1), the CV community has produced a number of publicly available datasets, including ImageNet [10], S3DIS [11], MS COCO [12], and ADE20K [13]. Although a handful of datasets such as MOCS (moving objects in construction sites) [14] and CIS (construction instance segmentation) [15], targets the construction domain, they mainly focus on outdoor environments, leaving indoor environments underexplored.

The second and third data collection techniques present a unique set of challenges that hinder the creation of largescale, high-quality real-world datasets. On one hand, data collection through web crawling requires manual review for quality, privacy regulations, intellectual property rights, and consistency. On the other hand, capturing data in the real world involves tedious work, access to construction sites, and specialized expertise [16]. The data annotation step, arguably the most time-consuming task due to the dataset's large scale, follows the data collection process. Crowdsourcing services like Amazon SageMaker and Google Cloud Vertex AI can significantly reduce the manual labor involved, but they remain costly and prone to error [17].

Because the acquisition of real-world datasets is a challenging and resource-intensive endeavor, researchers have also explored the generation of synthetic data from controllable and computable virtual environments as a costeffective and efficient alternative [18]. In the realm of synthetic image data generation in construction, there have been a number of virtual environments constructed using 3D modeling and computer graphics software such as Revit, Blender, Unreal and Unity [19, 20, 21]. DL models trained on synthetic or mixed datasets have been shown to outperform or achieve comparable performance to models trained solely on real images [22].

One emerging source of data over the past decade comes from the building information modeling (BIM). BIM is a set of interacting policies, processes and technologies aimed at managing the essential data of a construction site, in digital form, throughout its life cycle [23]. BIM models have become an appropriate data source for generating synthetic images of indoor building scenes due to the inclusion of accurate geometry and sometimes appearance information (*i.e.*, material and texture). BIM authoring tools can provide virtual cameras and several studies have leveraged their capabilities to generate synthetic data [24, 25]. However, the potential of BIM models combined with computer graphics software has not been fully explored and there is no universal approach to generate data automatically.

This paper proposes a procedural approach to generate synthetic datasets from BIM models to address the data collection and availability drawbacks mentioned above. To test this approach, three synthetic datasets were generated from three different BIM models (*i.e.* Figure 1) and used to train and test a state-of-the-art instance segmentation model. Lastly, a preliminary approach is presented that compares the prediction of the trained neural network model with the BIM model view through an augmented reality (AR) device.



Figure 1. Snapshots of three BIM projects imported in the graphic engine.

This article is structured as follows: Section 2 delves into the technical aspects of synthetic image generation and annotation. Section 3 presents an experimental procedure to evaluate the framework performance, followed by a concise experimental study showcasing its practical application. Finally, Section 4 summarizes the findings and outline directions for future research.

2 Methodology

The objective of this study is to develop an automated approach that generates photorealistic synthetic RGB images of indoor building scenes with comprehensive annotations using BIM and a computer graphics engine. Figure 2 illustrates the workflow of the approach, each step is further described in the following.



Figure 2. Framework to generate annotated photorealistic images from a BIM model.

2.1 Synthetic image generation

Since the study focuses on indoor construction sites, the first step is to acquire BIM models of such projects. A BIM model for a building is typically created by a team of professionals from various trades, each contributing their expertise to the development of the model. These trades can be broadly categorized into the following groups: architecture, structural engineering, MEP and interior design. The MEP trade, representing the construction phase with the most dynamic and complex changes throughout a project's life cycle, is a critical area for progress monitoring automation. Therefore, we focus on generating a synthetic dataset in this field.

NVIDIA Isaac Sim is a robotics simulation toolkit for the NVIDIA Omniverse platform that provides researchers and practitioners with the tools and workflows they need to create robust, physically accurate simulations and synthetic datasets. With the use of a Revit plugin, one can directly import construction projects into Isaac Sim, making it a viable tool for synthetic dataset generation using BIM. For each physical object, the following information is obtained: object class, instance ID, and triangular mesh. However, the projects lack materials and textures, which we need to add through Isaac Sim to create a photorealistic environment.

The application of true-to-life textures to each object class within the BIM environment can be achieved through the utilization of the application programming interface (API) and the Omniverse material library. By leveraging the unique IDs associated with each object instance, it is possible to efficiently apply a variety of textures to specific object classes, such as pipes, for enhanced realism. For

example, red and blue metal textures can be applied to represent warm and cold currents, respectively, while black foam textures can be used to depict insulation. Additionally, the ability to create or import custom texture packs from external platforms, along with the incorporation of 3D models of specific elements, such as fan coils, further enhances the realism of the BIM environment. Moreover, the accurate representation of lighting, including both indoor and natural illumination, is crucial for achieving a realistic virtual representation. Artificial lighting should be meticulously tailored to match the lighting fixtures specified within the BIM model, while additional light sources may need to be strategically placed to compensate for the absence or non-functional state of lighting fixtures at the actual construction site. Various rendering parameters can be adjusted to optimize the depiction of light behavior, its interaction with objects, their reflectivity, colors, and transparency/absorption of materials. To achieve the highest level of fidelity, we use the interactive path tracing mode instead of approximation methods that prioritize performance while sacrificing accuracy.

The camera sensor simulation provides granular control over parameters like lens properties, aperture, shutter, clipping, and fisheye distortion, replicating real-world camera behavior. Key parameters include focal length, field of view, output resolution, and focus distances. It is important to note that digital cameras may produce radial and tangential distortions due to manufacturing imperfections in their lenses. The virtual camera supports various distortion models, with popular options including rational polynomial, brown conrady, and fisheye. Post-processing enhancements, including exposure adjustment, tone mapping, color grading, color correction, depth of field, motion blur, bloom, and others, further refine the simulated camera feed. We target a low-end camera simulation mimicking a laptop webcam for embedding compatibility.

After configuring the virtual camera, the next step is to determine its possible positions within each scene. This involves defining routes that mimic the movement of a worker inspecting the construction site. The API provides a tool for manually creating these routes. While grid-based viewpoints could be used, manually drawn routes provide a more natural representation of an inspector's movement.

2.2 Instance segmentation annotation

We use the semantics schema API to associate semantic data to each object class in our dataset. The synthetic data recorder options are RGB, depth, semantic and instance segmentation, 2D tight and loose bounding box. For this dataset, we only select RGB and instance segmentation data, capturing one image per half second along the predefined route.

So far, valid photorealistic synthetic images have been

generated. We now need to translate the semantic data to a DL model annotation format. We chose to use the Darknet text format compatible with the YOLOv8 instance segmentation model, one of the latest iteration of the popular YOLO model [26] released in January 2023 by Ultralytics. The architecture of YOLOv8 combines a convolutional neural network (CNN) backbone with a selfattention mechanism to achieve high accuracy and speed for object detection. This CNN model's high-speed inference makes it ideal for real-time applications on edge devices. Additionally, a segmentation head is added to predict the binary segmentation masks for each object detected.

Each line in a Darknet text file corresponds to a single object annotation. We create a Python script to attribute to each pair of RGB and semantic image a text file containing each object's class IDs, bounding box coordinates, and the mask coordinates, determined using the Ramer–Douglas–Peucker (RDP) algorithm. The RDP algorithm is a curve-fitting algorithm that simplifies the contours of an object by reducing the number of points that describe it while preserving its shape. By configuring the spatial resolution, minimum surface, and the number of points describing a contour, we can eliminate objects that are too small.

3 Experimental study

3.1 The generated synthetic dataset

We perform a series of experiments to evaluate the proposed approach. All experiments are run on a laptop equipped with an Intel Core i7-10750H CPU, 32 GB of RAM, and a NVIDIA Quadro RTX 3000 GPU. SPIE Building Solutions, a subsidiary of SPIE, provides access to three construction projects for the experiment (see Figure 1). The first project is an eight-story tower that will be used as an office, the second is an extension to a scientific university laboratory, and the third is a completely new site for a business school, making them an excellent sample of tertiary construction projects. Prior to importing the geometry into Isaac Sim, filters are applied to the BIM model in Revit to eliminate all interior design elements and architectural components that obstruct the MEP components. A route is created within each building, traversing each floor and alternating the viewpoints.

The dataset generated in this study is named MEP-SEG. By capturing one image every half a second of the created route, we collected 8,751 samples from the BIM projects, with a rendering time of approximately 9 hours. The distribution of assets for each of the 13 classes is shown in Table 1. Examples of the generated samples are depicted in Figure 3. As shown, some classes are severely under-represented compared to others. This

Table 1. Distribution of assets among the 13 classes in the generated dataset.

Class	No. of assets
Wall	90801
Pipe	44998
Floor	44266
Circular duct	34973
Rectangular duct	26227
Framework	11627
Air vent	8585
Pole	5131
Fan coil	4286
Radiant panel	3031
Ceiling	2431
Pipe accessory	1449
Climatic equipment	1309

will have to be taken into account when using this dataset. The entierety of the data is publicly available at : https://datasets.liris.cnrs.fr/mep-seg-dataset-synthetic-images-generated-building-information-modeling-bim-v1.



Figure 3. Three synthetic images generated on the left column and their corresponding semantic segmentation image on the right column.

3.2 Evaluation procedure

3.2.1 Evaluation of synthetic pre-trained models on small real datasets

For this initial evaluation, we compare transfer learning (TL) using two different pre-trained models of the same YOLOv8 architecture. The first model is pre-trained on the COCO image dataset [12], the second model on our MEP-SEG dataset. The objective is to determine whether

using a synthetic dataset in the targeted domain can yield better performance compared to a generic but more comprehensive public dataset. We focused on detecting only three classes: circular ducts, rectangular ducts, and pipes. These objects are among the most challenging to detect in images of the MEP domain due to their various shapes, textures, and sizes. We extracted these three classes from our MEP-SEG dataset, resulting in a duct and pipe dataset named D&P-SEG. A blank YOLOv8 neural network was trained using the API of YOLOv8. All available images are distributed 70% for training, 20% for validation and 10% for testing. The training strategy included a batch size of 6, learning momentum of 0.937, weight decay of 0.0005, learning rate (LR) of 0.01, width and height of 640 for 1000 epochs. Training took approximately 8 hours on an RTX 3000 GPU.

Real-world images were collected from inside five construction sites: the three used for the MEP-SEG dataset and two new projects: a residential building for Olympic athletes and a new corporate headquarters. Two distinct devices were used for data collection: a smartphone and the Microsoft HoloLens 2 AR glasses. A total of 217 images were acquired and manually labeled. Then, two datasets of varying dimensions are created: (1) the first small (S) dataset included 45 images for training and 19 images for validation. (2) The second medium (M) dataset contained 131 images for training and 43 images for validation. Both pre-trained models are fine-tuned on those real-world datasets with the same strategy, adding a patience of 50 for early stopping. The two models obtained their lowest validation loss in around 200 epochs, for a training time of 20 minutes. For testing, 40 additional images are selected. Table 2 summarizes the test results.

Table 2. Performances on COCO vs. synthetic TL on small (S) and medium (M) real datasets.

Metrics	COCO TL		Synthetic TL		
S dataset	box	mask	box	mask	
Precision (%)	43	51	66	64	
Recall (%)	43	34	47	46	
mAP50 (%)	42	38	53	49	
mAP50_95 ⁽ %)	26	19	37	30	
M dataset	box	mask	box	mask	
Precision (%)	52	52	69	63	
Recall (%)	45	42	40	38	
mAP50 (%)	43	41	47	43	
mAP50 95 (%)	29	24	30	24	

The performance metrics used on both the predicted boxes and masks are precision, recall, mean average precision at an intersection-over-union (IoU) threshold of 0.5 (mAP50), and mean average precision at IoU thresholds ranging from 0.5 to 0.95 (mAP50_95). As we can see on the S dataset, the pre-trained model on our synthetic data outperforms the one pre-trained on COCO in every metric. Additionally, the gap between box and mask on precision and recall is lower using the synthetic TL method, suggesting that the model better detects the edges of objects. Looking at the results on the M dataset, we observe a less significant difference between the two methods, indicating that the more real data available, the less relevant is using a synthetic dataset for a pre-trained model. In conclusion, using a synthetic pre-trained model showed promising results for transferring knowledge onto a small real dataset.

3.2.2 Evaluation of augmented datasets for MEP object detection

For the second evaluation, we constructed two sets of training datasets, one comprising 142 real images and the other 500 synthetic images. Secondly, we prepared a real test dataset containing 50 real images. This time, we focused on five classes: duct (encompassing circular and rectangular ducts), pipe, radiant panels, fan coil, and diffuser. The real training images were carefully selected from the MEP-SEG dataset to ensure they were the most representative. Thirdly, to explore the effectiveness of augmenting synthetic data with a small amount of real data, we prepared two additional training datasets:

- A mixed training dataset with 550 images, containing 90% synthetic images and 10% real images.
- A mixed training dataset with 600 images, containing 80% synthetic images and 20% real images.

In each training dataset, 80% of the images were used for training, and the remaining 20% were used for validation. The YOLOv8 model pretrained on the COCO image dataset served as the foundation and the same training strategy as in the previous evaluation was employed. The evaluation on the 50 real test images of the precision metric across all training datasets is provided in Table 3.

Table 3. Performance of the YOLOv8 model trained on different datasets.

Training dataset	Precision (%)			
	box	mask		
Real	77	75		
Synthetic	30	29		
Synthetic +10% real	71	69		
Synthetic +20% real	80	79		

The purely synthetic dataset exhibited inferior performance compared to the same dataset enhanced with 10% real images, which itself performed less effectively than that augmented with 20% real images. To achieve performance that is comparable or even superior to that of the real image dataset, it appears that at least 20% of real images are needed. This indicates that augmenting a dataset of real images with synthetic images holds promise for effectiveness and efficiency.

3.3 Preliminary validation of effectiveness.

This section presents a preliminary work towards the automation of progress monitoring in the construction industry using the Hololens 2 AR glasses. The work involves the use of an YOLOv8 model, pre-trained on our MEP-SEG dataset and fine-tuned on a real dataset to accurately detect and segment diverse MEP objects within images. The NEXT-BIM application serves as a foundation for this work, which already enables the visualization of BIM models using AR glasses as well as a semiautomatic alignment of the user view. Post-alignment, the BIM view is superimposed onto the real-world view in real-time, enabling a seamless comparison between the two. A prototype tool for visual progress monitoring has been developed on top of NEXT-BIM's application. The tool is capable of assessing the MEP work progress in a real on-site scenario.

Figure 4 provides an example of the method used to assess MEP work progress.



Figure 4. Captured versus BIM comparison.

The process starts with the capture of the real-world scene (1) using the integrated webcam. Subsequently, the corresponding view within the BIM model (2) is extracted based on the webcam sensor's coordinates and orientation. The captured on-site image is run through our instance segmentation model (3), resulting in a prediction mask that categorizes and labels each pixel with a unique color corresponding to its respective class. For instance, ventilation ducts are represented in cyan, while piping is depicted in green. In parallel, semantic extraction is applied to the BIM model image (4) to generate the ground truth mask. This involves transforming each material component within the view frustum to its corresponding class, followed by color filtering and shape estimation. The two generated segmentation masks, namely the prediction mask and the ground truth mask, are then superimposed (5), with adjustments made to ensure alignment. Finally, a comparison of the superimposed masks yields the resulting discrepancy mask (6), where red represents the ground truth, blue represents the prediction, and green represents the overlapping pixels.

To ensure the robustness of our AR-based progress monitoring system, a rigorous validation methodology will be employed. Success criteria for detection, segmentation, alignment, and processing speed will be defined first. The next step involves using the Hololens glasses to collect diverse pairs of real-world images and ground truth segmentation masks, deliberately targeting challenging edge cases. Iterative evaluation metrics will be utilized to pinpoint weaknesses, which will guide the refinement of our model and data. The final stage involves designing a reallife progress assessment scenario, where the efficiency and user experience of our method will be compared to traditional alternatives.

This innovative approach, enabled by AR technology, will empowers inspectors to visualize and interact with the BIM model in real time, enhancing their understanding of the physical environment and enlighten informed decisionmaking. This automatic segmentation makes it easier to assess the presence or absence of MEP objects, calculate the possible difference between the as-planned and asbuilt, and facilitate ICPM in the end.

4 Conclusion and future work

We presented a promising solution towards the automation of indoor construction progress monitoring (ICPM) using synthetic data and deep learning (DL). We proposed a procedural building information modeling (BIM) based synthetic image generation approach to address the data scarcity of real-world data, particularly for indoor construction environments. Our experimental study showcases the successful implementation of the proposed approach in three real-world construction projects, and the preliminary validation of effectiveness using augmented reality (AR) glasses further demonstrates the potential practical applications of the developed tool.

However, our current method has limitations related to the differences between the real and synthetic worlds. Building scenes represented by BIM models are often cleaner or more well organized than real-world scenes, where random objects can be found, and the locations of movable objects can be arbitrary. Also, it relies heavily on the quality of the BIM models, which frequently contain labeling errors, geometric clashes, or an absence of object geometry.

Future work will address these limitations to make the captured images more scene-realistic. Furthermore, we now focus on exploiting instance segmentation models for ICPM. We investigate comparison methods to compute more information between as-planned and as-built segmentation masks. Additionally, to improve the model performance, we consider running it through a live video feed instead of a still image and adding temporal and tracking methods to reduce the imprecision of the model.

5 Acknowledgements

The research leading to these results is part of a thesis co-supervised between École Centrale Lyon and SPIE Building Solutions. NEXT-BIM is thanked for its support in implementing the solution in their working environment and the use of their development platform.

References

- Teizer and Jochen. 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, 2015. doi:10.1016/j.aei.2015.03.006.
- [2] Koo B. and Fischer M. Feasibility Study of 4D CAD in Commercial Construction. *Jour*nal of Construction Engineering and Management, 126(4):251–260, 2000. doi:10.1061/(ASCE)0733-9364(2000)126:4(251).
- [3] Tang S., Roberts D., and Golparvar-Fard M. Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction*, 120:103356, 2020. doi:10.1016/j.autcon.2020.103356.
- [4] Asadi K., Ramshankar H., Noghabaei M., and Han K. Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction. J. Comput. Civ. Eng., 33 (5):04019031, 2019. doi:10.1061/(ASCE)CP.1943-5487.0000847.
- [5] Fathi H., Dai F., and Lourakis M. Automated asbuilt 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. *Advanced Engineering Informatics*, 29 (2):149–161, 2015. doi:10.1016/j.aei.2015.01.012.

- [6] Wang Boyu, Wang Q., Cheng J.C.P., Song C., and Yin C. Vision-assisted BIM reconstruction from 3D LiDAR point clouds for MEP scenes. *Automation in Construction*, 133:103997, 2022. doi:10.1016/j.autcon.2021.103997.
- [7] Reja V.K., Varghese K., and Ha Q.P. Computer vision-based construction progress monitoring. *Automation in Construction*, 138:104245, 2022. doi:10.1016/j.autcon.2022.104245.
- [8] Pal A. and Hsieh S. Deep-learning-based visual data analytics for smart construction management. *Automation in Construction*, 131:103892, 2021. doi:10.1016/j.autcon.2021.103892.
- [9] Wei W., Lu Y., Zhong T., Li P., and Liu B. Integrated vision-based automated progress monitoring of indoor construction using mask regionbased convolutional neural networks and BIM. *Automation in Construction*, 140:104327, 2022. doi:10.1016/j.autcon.2022.104327.
- [10] Deng J., Dong W., Socher R., Li L., Li K., and Fei-Fei L. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision* and Pattern Recognition (CVPR), pages 248–255, 2009. doi:10.1109/CVPR.2009.5206848.
- [11] Armeni I., Sener O., Zamir A.R., Jiang H., Brilakis I., Fischer M., and Savarese S. 3D Semantic Parsing of Large-Scale Indoor Spaces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1534–1543, Las Vegas, NV, USA, 2016. IEEE. doi:10.1109/CVPR.2016.170.
- [12] Lin T., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., and Zitnick C.L. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision (ECCV)*, pages 740– 755, 2014. doi:10.1007/978-3-319-10602-1_48.
- [13] Zhou B., Zhao H., Puig X., Fidler S., Barriuso A., and Torralba A. Scene Parsing through ADE20K Dataset. In *IEEE Conference on Computer Vision* and Pattern Recognition (CVPR), pages 5122–5130, 2017. doi:10.1109/CVPR.2017.544.
- [14] Xuehui A., Li Z., Zuguang L., Chengzhi W., Pengfei L., and Zhiwei L. Dataset and benchmark for detecting moving objects in construction sites. *Automation in Construction*, 122:103482, 2021. doi:10.1016/j.autcon.2020.103482.
- [15] Yan X., Zhang H., Wu Y., Lin C., and Liu S. Construction instance segmentation (CIS) dataset for deep learning-based computer vision.

Automation in Construction, 156:105083, 2023. doi:10.1016/j.autcon.2023.105083.

- [16] Soltani M.M., Zhu Z., and Hammad A. Automated annotation for visual recognition of construction resources using synthetic images. *Automation in Construction*, 62:14–23, 2016. doi:10.1016/j.autcon.2015.10.002.
- [17] Northcutt C.G., Athalye A., and Mueller J. Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2021. doi:10.48550/arXiv.2103.14749.
- [18] Handa A., Patraucean V., Badrinarayanan V., Stent S., and Cipolla R. Understanding Real-World Indoor Scenes with Synthetic Data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4077–4085, 2016. doi:10.1109/CVPR.2016.442.
- [19] Ha I., Kim H., Park S., and Kim H. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, 140:23–31, 2018. doi:10.1016/j.buildenv.2018.05.026.
- [20] Acharya D., Khoshelham K., and Winter S. BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:245–258, 2019. doi:10.1016/j.isprsjprs.2019.02.020.
- [21] Lee J.G., Hwang J., Chi S., and Seo J. Synthetic Image Dataset Development for Vision-Based Construction Equipment Detection. *Journal of Computing in Civil Engineering*, 36(5):04022020, 2022. doi:10.1061/(ASCE)CP.1943-5487.0001035.
- [22] Acharya D., Singha Roy S., Khoshelham K., and Winter S. A Recurrent Deep Network for Estimating the Pose of Real Indoor Images from Synthetic Image Sequences. *Sensors*, 20(19):5492, 2020. doi:10.3390/s20195492.
- [23] Sacks R., Eastman C., Lee G., and Teicholz P. BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers. Wiley, 2018. doi:10.1002/9781119287568.
- [24] Hong Y., Park S., Kim H., and Kim H. Synthetic data generation using building information models. *Automation in Construction*, 130:103871, 2021. doi:10.1016/j.autcon.2021.103871.

- [25] Ying H., Sacks R., and Degani A. Synthetic image data generation using BIM and computer graphics for building scene understanding. *Automation in Construction*, 154:105016, 2023. doi:10.1016/j.autcon.2023.105016.
- [26] Redmon J., Divvala S., Girshick R., and Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. In *IEEE Conference on Computer Vision* and Pattern Recognition (CVPR), pages 779–788, 2016. doi:10.1109/CVPR.2016.91.

Single-Stage Spatiotemporal Activity Recognition of Excavators: A Case Study

Ali Ghelmani¹, Ghazaleh Torabi², Amin Hammad^{1*}, Chen Chen³

¹Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada ²ZestyAI, Montreal, Canada

³School of Civil Engineering and Architecture, Zhejiang University of Science and Technology, Hangzhou, China <u>ali.ghelmanirashidabad@concordia.ca</u>, <u>ghazalehtrb@gmail.com</u>, <u>amin.hammad@concordia.ca</u>*,

chen.c@zust.edu.cn

Abstract -

Manual excavator activity monitoring to evaluate their performance and productivity is laborious, time-consuming, and error-prone. To address these problems, many automated computer vision-based frameworks have been developed for the detection of excavators and the classification of their activities. Most of the current methods consist of several separately optimized modules that are applied to the input video sequentially. Recently, single-stage spatiotemporal activity recognition methods are gaining more popularity in the construction community. You Only Watch Once (YOWO) network and its variation (i.e., YOWO53) have proved to be superior to the three-stage approaches for activity recognition of construction workers. This paper investigates the benefits of using YOWO and YOWO53 over the three-stage methods for the activity recognition of excavators, by utilizing a large custom dataset of 1,060 video clips collected from both local construction sites and YouTube, with different camera angles, illuminations, occlusions, weather conditions, and video resolutions. The results demonstrate 88.9 and 88.7% classification accuracy and F1-score, respectively for the YOWO method compared to 70.4% and 69.8% classification accuracy and F1-score for the three-stage method. This indicates the feasibility and benefits of deploying single-stage methods to near real-time the applications.

Keywords -

Site monitoring, Activity recognition, Computer vision

1 Introduction

Nowadays, the growing demand for the completion of construction projects within schedule and under budget has resulted in the development of automated, continuous monitoring routines to provide project managers with vital productivity and safety information [2]. Traditionally, monitoring the activities of excavators and other earthmoving equipment is performed manually by superintendents on the site. However, such methods can be very time-consuming, labor-intensive, and errorprone especially on large construction sites [1, 2]. Considering that excavators are at the core of earthmoving operations [3], monitoring their activities can provide information about work cycle duration and consequently productivity. This information in turn enables site managers to make more informed projectrelated decisions, such as adjusting resource allocations. Considering that excavators are at the core of earthmoving operations [3], monitoring their activities can provide productivity and work cycle duration information, which in turn enables site managers to make more informed project-related decisions, such as resource allocations and scheduling [4, 5]. Videos can provide detailed information about the visual features and physical motions of equipment, and therefore increase the interpretability of the results and their shortcomings by viewing the recorded video and the detected activities [2, 6]. Before the rise of deep learning, vision-based methods generally relied on hand-crafted features to extract useful information for activity recognition from images and videos [7]. However, advances in deep learning methods demonstrated their superiority over traditional hand-crafted methods in different applications such as object detection [8] and activity recognition [9], which resulted in a corresponding change in the use of vision-based methods in the construction domain.

Convolutional Neural Networks (CNN) are the main building blocks in all vision-based deep learning methods, and in the past few years, many 2D CNN-based construction equipment activity recognition methods have been proposed. For instance, Roberts et al. [1] used a combination of 2D CNN with Hidden Markov Models to detect, track, and identify the activities of excavators and dump trucks. Luo et al. [10] used a combination of 2D CNNs and relevance networks for detecting various construction-related objects and their associated set of interactive activities by exploiting the two-dimensional pixel proximity of the detected objects. Kim and Chi [11] also performed interaction analysis to identify the activities and operation cycles of excavators and dump trucks by combining 2D CNN and Long Short-Term Memory (LSTM) architectures. Similar combinations of 2D CNNs and LSTMs, were also used by Slaton et al. [12] to detect the routine tasks of excavators and roller compactors, and by Kim et al. [6] to detect excavator activities via exploiting their sequential working patterns for automatic productivity analysis.

While 2D CNN-based methods try to combine the spatial and temporal information using different methods, 3D CNN-based methods incorporate the spatiotemporal data extraction into a single architecture, which allows the deep learning models to extract relevant spatiotemporal data. Chen et al. [2, 13] proposed a threestage method in which excavators are first detected in the input frames. Then, the detected excavators are fed into a tracking algorithm, and finally, the tracked results are input to a 3D CNN network to classify the activities. A similar three-stage framework was also proposed by Lou et al. [14] in which workers were first detected using the You Only Look Once (YOLOv3) network. The detected workers were then tracked, and the activities performed by them were classified using a 3D CNN architecture. Although these frameworks can potentially extract more informative spatiotemporal features using 3D CNN architectures, their three-stage approach still limits their accuracy. The main limitations of three-stage methods are: (1) not being fully optimized, and (2) the propagation of errors from earlier stages to the later ones, which results in the degradation of the performance of the entire framework [14,15]. For example, if an equipment is not detected in a few frames or if it is not tracked properly through the frames in which an activity is occurring, the final 3D CNN stage cannot correctly classify the underlying activity.

The benefits of using a single-stage method over the three-stage methods were studied for the case of detecting activities of construction workers by Torabi et al. [16]. They proposed a network called You Only Watch Once 53 (YOWO53) to jointly detect construction workers that appeared small in the video frames and classify their activities. YOWO53 is based on a general human activity recognition network called YOWO [17]. The results showed YOWO53 improves the detection recall of YOWO for small objects (e.g., workers) by at least 2%, and both single-stage networks (i.e., YOWO and YOWO53) improved the activity classification accuracy of one of the state-of-the-art three-stage methods [2] by at least 16%. Jung et al. [15] also

proposed a single-stage architecture for detecting the activities of multiple construction equipment simultaneously. This framework uses a 3D CNN architecture and performs equipment detection and activity recognition in one stage to alleviate the limitations of the three-stage methods. However, 39% of the video clips in the reported dataset of seven activities correspond to the idling state of the equipment. Such a dataset, in addition to being unbalanced, limits the real-world applicability of the trained model.

The aim of this paper is to investigate whether the same improvement achieved by single-stage YOWO53 method for workers [16] is achievable for the case of excavators. Furthermore, another important factor in the final performance and applicability of a developed activity recognition model is the size and variability of the data included in a dataset. To this end, a large balanced dataset of excavator activities with more than 1,060 video clips, collected both from local construction sites and YouTube, has been gathered. The prepared dataset contains the three common excavator activities of digging, swinging, and loading the trucks under various camera angles, illuminations, occlusions, weather conditions, and video resolutions. Thus, enabling a thorough evaluation of the YOWO and YOWO53 methods under various real-world conditions.

2 Methods Used in The Case Study

The single-stage YOWO [17] and YOWO53 [16] methods are compared in this study with a state-of-theart three-stage method proposed by Chen et al. [2]. The general architectures of the selected methods are shown in Figure 1, and a more detailed description of these methods is presented in the following sections.

2.1 YOWO

YOWO [17] is a spatiotemporal activity recognition method, which uses two branches in its architecture. One branch extracts 2D features from the current frame while the second branch extracts 3D features from a stack of successive frames. Afterwards, the outputs of the two branches are combined using a channel fusion and attention mechanism (CFAM), which provides the essential performance boost.

The 3D CNN branch is utilized for extracting the spatiotemporal features. In this work, the ShuffleNetV2_2x [18] 3D CNN is chosen for this branch for comparison with the activity recognition results of workers [16]. The input to this network is a video clip comprised of a sequence of frames with the dimension of $[C \times D \times H \times W]$, with C being equal to 3 (RGB channels), D representing the number of input frames, and H and w representing the height and width of the



Figure 1. The general architecture of the selected methods

frames, respectively. The shape of the output is $[C' \times D' \times H' \times W']$, with C' being the number of output channels, D' = 1, H' = H/32, and W' = W/32. By default, the output of the 3D CNN branch is 4-dimensional, while the output of the 2D CNN branch is 3-dimensional. To make the output of these two branches compatible before combining, the output of the 3D CNN branch is designed to have a reduced depth component (D' = 1), which can be dropped, hence becoming three-dimensional in effect.

While some studies only rely on a single 3D CNN network for simultaneous activity recognition and localization (e.g., [15]), the YOWO model also uses a 2D CNN branch in parallel to provide more accurate spatial information. YOWO uses the Darknet19 [19] network in the 2D CNN branch for object detection, which takes images of the form $[C \times H \times W]$ as input, while the shape of the output feature map is of the form of $[C'' \times H' \times W']$, where C'' is the number of output channels.

Afterwards, the output of the two branches is input into the CFAM module, which fuses the two 2D CNN and 3D CNN extracted information. Since the two branches are designed to have the same shape for their output feature maps, they can be easily concatenated along the channel dimension. The CFAM module then utilizes the fused feature maps to generate a combination of both motion and appearance data. Finally, the loss function used to train the YOWO model, is inspired by the losses defined in YOLO [19] and Fast R-CNN [20] models, and is comprised of the activity classification loss, and the localization loss between the bounding box predicted by the model and the ground truth bounding box.

2.2 YOW053

The YOWO method has a modular architecture in which the networks in the 2D CNN branch and the 3D CNN branch can be replaced by other corresponding networks. The YOWO53 model exploits this architecture by applying the following modifications. As stated in Section 2.1, the YOWO method uses the Darknet19 network in the 2D CNN branch for extracting spatial features from the current frame. While the Darknet19 network is fast, its detection accuracy for small objects is not very high. To address this problem, the Darknet53 network [21] is utilized in the YOWO53 method, which although slower, provides more accuracy for detecting small objects and hence is more suitable for applications in the construction domain. Particularly, considering that most of the surveillance cameras on construction sites are installed at high altitudes and consequently, workers and construction equipment at a distance can appear very small in the recorded site videos. Following the above change in the 2D CNN branch, to keep the output shape of the 2D and the 3D branches consistent for fusion in CFAM, a single max-pooling layer is removed from the architecture of the network used in the 3D branch. Thus, doubling the size of the output feature maps of the two branches (H' = H/16, and W' = W/16). This modification not only allows the concatenation of the two feature maps but also decreases the receptive field of YOWO53, which helps with the detection of small objects.

The receptive field of a particular feature in the output feature map of a CNN is the region in the input image that this feature encodes. The size of this region depends on the depth of the feature as well as the combination of the previous layers. Usually as the size of the output feature map is reduced, its receptive field increases. For example, if an image is reduced to a single feature by a CNN, this feature encodes the important information of the entire input image. In object detection, the size of the object that can be detected by the network depends on the receptive field of the last layer (detection layer). If the size of the object is larger than the receptive field of a layer, it may not be correctly detected using the output feature map of that layer. Larger feature maps have smaller receptive fields and can be used to detect smaller objects.

2.3 Three-stage method

To investigate the benefits of using single-stage methods over the three-stage method for the activity recognition of excavators, YOWO and YOWO53 methods are compared with the state-of-the-art threestage method proposed by Chen et al. [2]. This method is composed of detection using the YOLOv3 method, tracking using the Simple Online and Real-time Tracking (Deep SORT) method [22], and activity recognition using the 3D-ResNext-101 [23] network, with each stage optimized separately. The previous studies (e.g., [2,15]) did not fine-tune the Deep SORT module since it is one of the state-of-the-art methods capable of tracking multiple objects at the same time. The detection module can be trained using simple frame-level bounding box annotations. However, the activity recognition module used in the three-stage method requires the input video clip to contain only a single excavator performing a single activity. Thus, the detected and tracked excavators should be cropped before being input into the 3D CNN network for activity recognition.

3 Dataset description

Considering that the majority of publicly available datasets for the task of activity recognition are focused on human activities in various environments [24, 25], the first step in excavator activity recognition is to create a proper dataset. The video clips used in creating the dataset were manually collected from various sources including local construction sites and videos posted online on websites such as YouTube. Each video clip contains one or more excavators performing three types of activities: digging, swinging, and loading the trucks. To add to the diversity of the collected dataset and enable a thorough analysis of the selected methods, the videos are collected from 25 different construction sites, incorporating various site conditions, such as different camera angles, illuminations, occlusions, weather conditions, and video resolutions. The detailed statistics of the collected dataset are presented in Table 3. For labeling the collected dataset, the Computer Vision Annotation Toolbox (CVAT) [35], which is a free webbased video and image annotation toolbox [35], was used in this paper. The ground truth for each labeled frame includes the type of the occurring activity and the top left and bottom right coordinates of the encompassing bounding box.

4 Implementation details

All of the models are trained on three RTX A6000 GPUs in Ubuntu 20.04 and Python 3.7 environment and PyTorch 1.8, with 80% of the video clips randomly selected for training and the remaining 20% used for testing. The ShuffleNetV2_2x 3D CNN network was pre-trained on the large-scale Kinetics-600 [26] dataset. Only the last layer of this network is fine-tuned on the collected excavator dataset. The 2D CNN networks, i.e., Darknet19 and Darknet53, are pre-trained on the COCO [27] dataset and only their last two layers are fine-tuned in this work. The models are trained for 20 epochs using the Adam optimizer [28] and their best results are saved.

5 Experimental results

Table 1 provides a comparative performance of the YOWO, YOWO53, and the three-stage methods. It can be seen that both YOWO and YOWO53 methods significantly outperform the three-stage method. The results in Table 1 show that the YOWO53 method obtains 15% improvement in classification accuracy and 15.6% improvement in F1-score over the three-stage method, which aligns closely with the results reported in [16]. However, in contrast to the performance improvements of YOWO53 over YOWO for worker activity recognition [16], Table 1 shows that the YOWO

method obtains 3.5% improvement in classification accuracy and 3.3% improvement in F1-score over the YOWO53 method for excavator activity recognition.

Table 1. Comparing YOWO, YOWO53, and the threestage method

	Classification accuracy (%)	F1-score (%)
YOWO	88.9	88.7
YOW053	85.4	85.4
Three-stage method	70.4	69.8

To further examine the difference in the performance of the YOWO and YOWO53 methods, Table 2 shows the classification accuracy, localization recall, overall precision, overall recall, and F1-score obtained by training and evaluating YOWO and YOWO53 methods on three different input sizes. The results for the smallest frame size (i.e., 128×128) agree with what was reported for workers in [16], with YOWO53 achieving better performance than YOWO in all of the metrics. However, as the size of the input frame increases, the performance of YOWO53 drops lower than YOWO. Additionally, after the second smallest input size (i.e., 256×256), the performance of YOWO53 drops by around 1% in all metrics. This indicates that the smaller receptive field of the YOWO53 method (as stated in Section 2.2), while efficient for detecting workers, is not capable of covering big equipment such as excavators and consequently, adversely affects the model performance. Finally, the overall best result is obtained by the YOWO method using the largest input size (i.e., 448×448), with 88.9% classification accuracy, and 88.7% F1-score.

To evaluate the real-world applicability of the YOWO and YOWO53 methods, the number of parameters of both methods, as well as the speed of processing each input size along with the highest batch size that can fit in the three RTX A6000 GPUs are presented in Table 2. It can be seen that the YOWO method processes 448×448 frames at 10.7 FPS, which is comparable to the processing speed of the YOWO53 methods for the same frame size. However, it should be noted that YOWO is a smaller network compared to YOWO53 with 79 million parameters compared to 90 million parameters, allowing higher batches to be processed at the same time by the network.

5.1 Sensitivity analysis

To further investigate the performance of the best YOWO model under various conditions, a sensitivity analysis was performed with different video conditions such as camera angles, illumination, occlusion, weather conditions, and video resolution. The results are presented in Table 3, demonstrating the high performance and applicability of the model for various real-world conditions.

As mentioned in Section 4, 80% of the total number of video clips in the dataset was selected randomly for training, while the remaining 20% was used for testing the performance of the model. A consequence of this division strategy is the varying ratio between training and testing data for each of the considered sensitivity analysis cases, especially for cases in which the total amount of data in the dataset is relatively small. For example, for the high occlusion category there are 455 frames for testing compared to 1,493 frames for training, a ratio of almost 1:3, while for the snowfield category, the amount of training data is even less than that of testing, with 421 frames available for training while 780 frames are used for testing. However, it should be noted that even in the snowfield category with such a train/test data imbalance, the model still achieves 94.7% activity classification accuracy and recall, showing its high accuracy even for cases with low training data, as long as the image quality is not severely degraded (e.g., high occlusion category).

Another interesting example is the "Below ground level" category, with only 143 frames from one video clip for testing, for which due to the high quality of the input video clip, the performance is still high (82.6% activity classification accuracy and 82.5% recall). Another effect of the low amount of training data for some cases can be seen in the apparent performance contradiction for the low-resolution video clips only with 13,056 training and 3,226 testing frames, which seems to outperform the high-resolution cases with 117,173 training and 29,840 testing frames (about nine times more frames). However, after further investigation, the lower performance for high resolution video clips is found to be due to the inclusion of most other difficult cases, which resulted in almost the same performance for this category as the full dataset reported in Table 2.

Finally, it can be seen from Table 3 that the worst performance of the model (66.4% activity classification accuracy) is in the high occlusion category, which is generally one of the biggest limitations of single-camera CV-based methods. However, considering that in these cases more than half of the excavator is not visible due to occlusion by other equipment, self-occlusion, or not being fully in the camera's field of view, the results show the impressive performance of the model.

Model	Input size	Classification	Precision	Recall	F1-score	Speed (FPS)	Batch	#param
		accuracy (%)	(%)	(%)	(%)	(FPS)	size	
YOWO	448×448	88.9	88.7	88.7	88.7	10.7	64	
YOWO	256×256	88.0	87.2	87.4	87.3	12	256	~79M
YOWO	128×128	84.3	82.5	82.8	82.7	12.6	256	
YOW053	448×448	85.4	85.4	85.4	85.4	10.4	32	
YOW053	256×256	86.8	86.3	86.3	86.3	10.9	32	~90M
YOW053	128×128	86.3	85.2	85.5	85.3	11.0	256	

Table 2. Performance comparison for variants of YOWO and YOWO53 methods

Video clip condition		Training		Testing		Classification	Precision	Recall	F1-score
		#clips	#frames	#clips	#frames	accuracy (%)	(%)	(%)	(%)
	High altitude	41	5,872	12	1,288	76.5	75.8	75.8	75.8
Camera	Mid altitude	520	80,571	130	21,601	89.3	89.2	89.2	89.2
angle	Ground level	284	43,296	68	10,034	76.5	75.8	75.8	75.8
	Below ground level	4	490	1	143	82.6	82.5	82.5	82.5
Illumination	Low	54	7,160	20	2,471	78.3	77.7	77.7	77.7
	Mid	562	91,784	133	22,469	93.6	93.6	93.6	93.6
	High	233	31,285	58	8,126	83.3	83.0	83.0	83.0
	Contre-jour	15	2,399	5	691	92.9	92.9	92.9	92.0
	Low	221	31,875	62	10,445	87.6	87.6	87.6	87.6
Occlusion	Mid	58	8,995	13	1,891	75.8	75.8	75.8	75.8
	High	11	1,493	2	455	66.4	66.4	66.4	66.4
Weather condition	Cloudy	117	12,870	35	4,316	80.9	80.3	80.3	80.3
	Sunny	714	114,539	170	27,279	90.2	90.1	90.1	90.1
	Snow field	3	421	1	780	94.7	94.7	94.7	94.7
Deselution	Low	96	13,056	22	3,226	96.4	96.4	96.4	96.4
Resolution	High	753	117,173	189	29,840	88.2	87.9	87.9	87.9

Table 3. Results of the sensitivity analysis for the YOWO model

6 Conclusions and future work

This paper investigates the benefits of using YOWO and YOWO53 methods over the state-of-the-art threestage method for the activity recognition of excavators. The performance is evaluated using a custom dataset of 1060 videos collected from local construction sites and YouTube videos. The obtained results show that the joint optimization of single-stage methods (i.e., YOWO, YOW053), provides significant performance improvement over the three-stage method, in which each stage is optimized separately. In particular, the YOWO model achieved an activity classification accuracy of 88.9% and an F1-score of 88.7%. In comparison, the YOWO53 model recorded slightly lower metrics, with both activity classification accuracy and F1-score at 85.4%. However, the best performance of the three-stage method was the activity classification accuracy, and F1score of 70.4% and 69.8%, respectively.

Although both single-stage methods proved to be superior to the three-stage method, however, in contrast to the results obtained in a previous study for workers [16], the performance of the YOWO53 method was lower than that of the YOWO method when increasing the input size. Considering that the YOWO53 method was developed to improve the detection performance for small objects (i.e., workers), the performance gain over the YOWO method is only for the cases where either the object of interest or the input size is small, while the opposite behavior is seen for excavators. Hence, indicating to a possible shortcoming of the current singlestage methods and a possible future approach which is able to simultaneously recognize the activities of equipment and workers at different scales, especially for the interactive activities, which involve both workers and equipment. Therefore, a single network should be able to recognize both small (e.g., workers) as well as large (e.g., excavators) objects.

References

[1] Roberts D. and Golparvar-Fard M., "End-to-end vision-based detection, tracking and activity

analysis of earthmoving equipment filmed at ground level," *Automation in Construction*, vol. 105, p. 102811, Sep. 2019, doi: 10.1016/j.autcon.2019.04.006.

- [2] Chen C., Zhu Z., and Hammad A., "Automated excavators activity recognition and productivity analysis from construction site surveillance videos," *Automation in Construction*, vol. 110, p. 103045, Feb. 2020, doi: 10.1016/j.autcon.2019.103045.
- [3] Rezazadeh Azar E. and McCabe B., "Part based model and spatial-temporal reasoning to recognize hydraulic excavators in construction images and videos," *Automation in Construction*, vol. 24, pp. 194–202, Jul. 2012, doi: 10.1016/j.autcon.2012.03.003.
- [4] Bohn J. S. and Teizer J., "Benefits and Barriers of Construction Project Monitoring Using High-Resolution Automated Cameras," *Journal of Construction Engineering and Management*, vol. 136, no. 6, pp. 632–640, Jun. 2010, doi: 10.1061/(ASCE)CO.1943-7862.0000164.
- [5] Kim J., Chi S., and Seo J., "Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks," *Automation in Construction*, vol. 87, pp. 297–308, Mar. 2018, doi: 10.1016/j.autcon.2017.12.016.
- [6] Kim J. and Chi S., "Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles," *Automation in Construction*, vol. 104, pp. 255–264, Aug. 2019, doi: 10.1016/j.autcon.2019.03.025.
- [7] Chen C., Zhu Z., and Hammad A., "Critical Review and Road Map of Automated Methods for Earthmoving Equipment Productivity Monitoring," *Journal of Computing in Civil Engineering*, vol. 36, no. 3, p. 03122001, May 2022, doi: 10.1061/(ASCE)CP.1943-5487.0001017.
- [8] Redmon J., Divvala S., Girshick R., and Farhadi A., "You Only Look Once: Unified, Real-Time Object Detection," In *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pages 779–788, Jun. 2016, doi: 10.1109/CVPR.2016.91.
- [9] Donahue J., Hendricks L. A., Guadarrama S., Rohrbach M., Venugopalan S., Darrell T., et al., "Long-term recurrent convolutional networks for visual recognition and description," In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2625–2634, Jun. 2015, doi: 10.1109/CVPR.2015.7298878.
- [10] Luo X., Li H., Cao D., Dai F., Seo J., and Lee S., "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*, vol. 32, no. 3, p. 04018012, May 2018, doi: 10.1061/(ASCE)CP.1943-5487.0000756.

- [11] Kim J. and Chi S., "Multi-camera vision-based productivity monitoring of earthmoving operations," *Automation in Construction*, vol. 112, p. 103121, Apr. 2020, doi: 10.1016/j.autcon.2020.103121.
- [12] Slaton T., Hernandez C., and Akhavian R., "Construction activity recognition with convolutional recurrent networks," *Automation in Construction*, vol. 113, p. 103138, May 2020, doi: 10.1016/j.autcon.2020.103138.
- [13] Chen C., Zhu Z., Hammad A., and Akbarzadeh M., "Automatic Identification of Idling Reasons in Excavation Operations Based on Excavator–Truck Relationships," *Journal of Computing in Civil Engineering*, vol. 35, no. 5, p. 04021015, Sep. 2021, doi: 10.1061/(ASCE)CP.1943-5487.0000981.
- [14] Luo X., Li H., Yu Y., Zhou C., and Cao D., "Combining deep features and activity context to improve recognition of activities of workers in groups," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 9, pp. 965–978, 2020, doi: 10.1111/mice.12538.
- [15] Jung S., Jeoung J., Kang H., and Hong T., "3D convolutional neural network-based one-stage model for real-time action detection in video of construction equipment," *Computer-Aided Civil* and Infrastructure Engineering, vol. 37, no. 1, pp. 126–142, Jan. 2022, doi: 10.1111/mice.12695.
- [16] Torabi G., Hammad A., and Bouguila N., "Two-Dimensional and Three-Dimensional CNN-Based Simultaneous Detection and Activity Classification of Construction Workers," *Journal* of Computing in Civil Engineering, vol. 36, no. 4, p. 04022009, Jul. 2022, doi: 10.1061/(ASCE)CP.1943-5487.0001024.
- [17] Köpüklü O., Wei X., and Rigoll G., "You Only Watch Once: A Unified CNN Architecture for Real-Time Spatiotemporal Action Localization." arXiv, Oct. 18, 2021. doi: 10.48550/arXiv.1911.06644.
- [18] Köpüklü O., Kose N., Gunduz A., and Rigoll G., "Resource Efficient 3D Convolutional Neural Networks," In *IEEE/CVF International Conference on Computer Vision Workshop* (*ICCVW*), pages 1910–1919, Oct. 2019, doi: 10.1109/ICCVW.2019.00240.
- [19] Redmon J. and Farhadi A., "YOLO9000: Better, Faster, Stronger," In *IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), pages 6517–6525, Jul. 2017, doi: 10.1109/CVPR.2017.690.

- [20] Girshick R., "Fast R-CNN," In IEEE International Conference on Computer Vision (ICCV), Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.
- [21] Redmon J. and Farhadi A., "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [22] Wojke N., Bewley A., and Paulus D., "Simple online and realtime tracking with a deep association metric," In 2017 IEEE International Conference on Image Processing (ICIP), pages 3645–3649, Sep. 2017, doi: 10.1109/ICIP.2017.8296962.
- [23] Hara K., Kataoka H., and Satoh Y., "Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?," In *Proceedings of the IEEE* conference on Computer Vision and Pattern Recognition, pages 6546–6555, 2018.
- [24] Soomro K., Zamir A. R., and Shah M., "UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild," *arXiv*:1212.0402, Dec. 2012, http://arxiv.org/abs/1212.0402
- [25] Gu C., Sun C., Ross D. A., Vondrick C., Pantofaru C., Li Y., et al., "Ava: A video dataset of spatiotemporally localized atomic visual actions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6047–6056, 2018.
- [26] Carreira J., Noland E., Banki-Horvath A., Hillier C., and Zisserman A., "A Short Note about Kinetics-600," arXiv:1808.01340, Aug. 2018, http://arxiv.org/abs/1808.01340
- [27] Lin T.-Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., et al., "Microsoft COCO: Common Objects in Context," arXiv:1405.0312, Feb. 2015, http://arxiv.org/abs/1405.0312
- [28] Kingma D. and Ba J., "Adam: A Method for Stochastic Optimization," *ICLR*, 2014.

Development of an immersive digital twin framework to support infrastructure management: a case study of bridge asset health monitoring.

Muhammad Fawad^{1,2}, Marek Salamak¹, Qian Chen³ and Kalman Koris²

¹ Department of Mechanics and Bridges, Faculty of Civil Engineering, Silesian University of Technology, Poland, ² Department of Structural Engineering, Budapest University of Technology and Economics, Hungary, ³ Faculty of Applied Sciences, University of British Columbia, Canada.

muhammad.fawad@posl.pl, marek.salamak@polsl.pl, qian.chen@ubc.ca, kalman.koris@emk.bme.hu

Abstract -

Digital Twin (DT) has been widely adopted in construction and infrastructure projects to focus on the management and control of physical systems and assets responsively to meet the users' requirements. When integrating immersion functions with the digital and wireless transformation models, the DT applications can further allow users to interact with and experience the model information in a more engaging and intuitive way. To add new perspectives to DT applications, this research gave reality functions to develop an immersive DT framework and presented a case study of VR- and AR-based bridge monitoring and twining systems health to demonstrate the functionalities. The case study was focused on the Structural Health Monitoring (SHM) processes for a newly constructed extradosed bridge, facilitated by the integration of Building Information Modeling (BIM), Virtual Reality (VR), Augmented Reality, 3D game engines, and Internet of Things (IoT) technologies. The bridge's geometric and property information was modeled in BIM parametric design software as a basis for VR model development and was used to develop the BIM-based sensory model in the 3D game engine according to the on-site conditions of the physically installed system. The integrated virtual model was further deployed to the AR device via a VR-to-AR workflow and was presented through the immersive DT dashboard. Project stakeholders can perform bridge damage detection using the information available from the dashboard. In this dashboard the IoT tools were used for the multi-source data integration process by storing, processing, and transforming the monitoring data, lending opportunities for predictive simulations of bridge condition states.

Keywords -

Augmented Reality (AR), Structural Health Monitoring (SHM), Virtual Reality (VR), Digital Twin (DT), data fusion.

1 Introduction

Technological advancement has continuously challenged the way of managing, monitoring, and maintaining the infrastructure assets' structural health [1]. The application of "digital twin" systems has been a game-changer that substantially revamped the processes of bridge health monitoring and management [2][3]. DT applications assist stakeholders in making reliable and quick decisions by automatically diagnosing the faults which leads to improved bridge performance and enhanced service life, which in turn reduces the maintenance and operation costs of the infrastructure [3][4].

Developing a DT prototype requires a complex and comprehensive approach to address the problem of inadequate resources at the network level while appropriately taking a bridge's health data into account [5]. A DT of a bridge comprises a connectivity module that enables synchronization of the physical and virtual assets along the asset's life cycle stages [6], as well as a virtual duplication of an actual bridge [7]. To deal with this complexity, this research proposes the Immersive DT framework ('Figure 1'). A structural twin can be built using the Finite Element (FE) model [8], whereas the bridge's 3D geometry and properties can be produced using the BIM technique [9]. Though the FE model is not perfectly interoperable with BIM data schema, data regarding environmental conditions, loads, and the structure's response to those loads can be obtained from sensors installed during a SHM procedure and be semiautomatically integrated with BIM designs [10]. The virtual assets produced as a result of the BIM model, can be used as a replica of the physical system and simulate



Figure 1 – The proposed immersive DT framework focusing on information flows between virtual and reality models.

the actual bridge's condition electronically which can be remotely visualized by the applications of AR [1][11]. This integration facilitates the remote and on-site monitoring of bridge health by visualizing and interpreting the data under both "normal" and "damaged" conditions. This data is then used to train AI data-driven models, which can instantly identify damages [12]. By using the information produced, the bridge management team and other interested parties can maximize the resources at their disposal for predictive decision-making [13].

The proposed immersive DT based on AR/VR technology not only visualizes the bridge health data but also helps to manage and process this data for predictive decision-making [14]. This way the wireless link between the digital and physical SHM system can be developed which controls the interface for the physical model including AR capability and represents real-time data fusion on the digital model (3D) of the bridge [15].

Considering these benefits, this study aims to develop an Integrated Bridge Digital Twin immersed in the AR environment to assist the infrastructure management and monitoring. This immersive DT explains the conception, development, and implementation of an Integrated Bridge Digital Twin, and how predictive analytics can be performed for intervention planning. This research explores the many facets of the implementation of the proposed framework, from the designing, initial data collection, and visualization of bridge SHM data using the applications of 3D game engines and simulation approaches. It also emphasizes how this digital twin

supports predictive decision-making, allowing

authorities to anticipate possible structural problems, maximize maintenance plans, and schedule interventions ahead of time.

2 Description of the bridge project case study and bridge SHM modeling

To test the usefulness of the proposed DT framework, the case study of a newly constructed bridge on Poland's National Highway 75 was chosen. This bridge has a continuous, four-span (100.0, 200.0, 200.0, and 100.0 m) extradosed structure with a C60/75 concrete box girder superstructure. The Load Model (LM1) according to Eurocode (EC)-1 standard [16], was utilized in the bridge's design, with adaptation coefficients of $\alpha = 1.00$. For class A, the Polish Norms (PN) 85/S standard [17] was followed. The span cross-section ('Figure 2') measures 17.68 m in width, while the intermediate supports are 23.0 m wide. The bridge is traversed by a dual carriageway road that is 8.60 m wide (between curbs) and has a 4.0 m wide pedestrian and bicycle lane.

To identify the possible damages and the expected location of monitoring devices, the Finite Element Analysis of the bridge was carried out. Based on that the bridge monitoring system that reflects the real-time condition states of the bridge assets was designed [18]. The FEA parameters of investigation include internal forces and span displacement. The possible cracking of the bridge surfaces was observed from this analysis but no prominent abnormalities were found. The bridge was verified for ULS and SLS as required by EC, and it was



Figure 2 – The BIM design of the bridge case study

found to be satisfactory. The deflected shape of the cables was considered to evaluate their deformations for the identification of monitoring points. For numerical analysis, only the parameters that have a direct impact on the bridge health and whose monitoring is critical were modeled including the stresses, strains, internal forces, crack widths, temperature, and humidity variations. Based on the FEA parameters, we proposed an integrated SHM system equipped with various IoT sensors, including Wired Strain Gauges (WSG), Liquid Levelling Sensors (LLS), MEMS accelerometers, and a Weather Monitoring Station (WMS) to monitor concrete deformations, vertical displacements, structural vibrations, and weather conditions. As a basis of the Digital Twin of the SHM model, the BIM design of the bridge was developed to include geometry and properties information ('Figure 2'). All the families of structural elements were developed by importing actual details of materials and accessory elements.

The developed BIM model is further used in the 3D game engine to develop the DT model which can perform automated bridge health monitoring.

3 Case study demonstration of the immersive DT

Using the bridge case study, the immersive DT was demonstrated on the bridge health monitoring processes by automatically identifying the damages. The whole process of digital twinning involves three different states, Physical State (PS) (reality state), Virtual State (VS), and Hybrid State (HS) for simulations.

The physical or reality state involves the physical model of the bridge, where the real sensors are installed on the bridge along with the network of monitoring systems. The PS defines the base of the DT framework where the SHM system collects the data and sends it to web servers or web platforms which then communicate with the VS. The VS involves the development of the BIM design and DT dashboard which replicates the physical model and SHM system in the virtual environment and each element of the bridge and SHM system is designed as an individual entity to communicate with the PS. The HS of DT involves the data streaming from the physically installed sensors to the virtual system to provide accurate information about bridge health. The DT dashboard was developed based on the integration of the gamification approach of the 3D game engine (UNITY 3D). All the parameters are automated by the relevant C# codes to formulate a datacentric mode of bridge SHM system. The novelty of this research provides the framework of the bridge SHM system, its integration with AR devices, and automated data-processing workflows which help the authorities in predictive decision-making.

The first step of DT generation involves the development of the BIM model using any of the BIM software (e.g. Autodesk Revit). This model is then

imported to the 3D game engine using the Revit plugin directly. It can also be done by using the Industry Foundation Class (IFC) file format but importing models with all the native properties makes them heavy and inefficient thus requiring some extra manual work to import all the design data. Therefore, direct import is preferred and recommended in such cases.

After importing the BIM model, the virtual replicas of all the sensors are developed at their exact installation locations in the, where they are installed in the PS. All these replicas virtually represent the actual sensors giving birth to the sensory model of the bridge in the VS.

The next workflow lies under the DT where each sensor is provided with several functions. For embedding any function in the DT dashboard, canvases are used to mesh the virtual model which develops the regenerative algorithms and links them with the sensory model. These meshes control the automation of the generated virtual elements and directly call the web platform connected to the sensory model. To automate the connection between the sensory model and the canvas, a proprietary C# script is developed in the visual studio and then embedded in the canvas. This automatically starts the communication between the PS and VS of the SHM system. Inside the canvas, a special button is generated which can give clickable features. This button is then embedded into the virtual replicas of the sensors, enabling communication with the actual sensors.

The data processing is automatically done using the AI tools on the web platform of the provided SHM system. As the SHM system is provided, installed, and developed by our industrial partners under a Non-Disclosure Agreement, detailed decision information about the smart sensors, and their Internet of Things (IoT)-based web platform are not discussed in this research. The dashboard of the DT model and the different states of the process are shown in 'Figure 3'.

4 VR-to-AR flow of bridge SHM data using Immersive DT

To demonstrate the VR-to-AR workflows for the case study, the VS of the model is transformed from VR to AR using the proprietary AR application. The development of this application is supported by the Mixed Reality Tool Kit (MRKT) and Universal Window Platform (UWP) in the UNITY game engine. These computing platforms help to develop applications that can run on the Windows system and AR platforms. DT model is used as the base 3D model for this application and transformed to immersive DT by importing all assets to this application. The real-time functionality of the application is developed using the virtual buttons as a connection



Figure 3 – Defining the three states in the proposed immersive DT framework.

between the PS and VS. To import the reality, an AR

development plugin is used (Vuforia) [19][20]. This plugin enables its own camera to convert the static DT model into an immersive DT model. So, when the game mode of UNITY is turned on, the VR system connects to the AR system in real time. Thus, the DT model can be used as an immersive AR application in the gaming environment as shown in 'Figure 4'. After this, the application is deployed to an AR headset (Microsoft HoloLens (HL)) using Visual Studio (ViS). Debugging is initiated in ViS, and the DT model is visualized in the HL app menu, allowing the immersive DT to function in an AR environment.

The successful testing of the deployed DT involved data integration and onsite bridge damage detection in the AR environment. Data integration in the developed DT application uses the IoT platform to gather, consolidate, and transform data from the installed IoT sensors, and systems into the web platform. The data on the web platform is available in graphical and tabulated formats so it can be transferred from the system to the users even when working on AR devices. As the data from different sensors is collected with different frequencies, data fusion methodology is adopted to integrate multiple sensor's data and produce a more consistent, accurate, and useful dataset for meaningful analysis. The outcomes of the measured data are then compared with the outcomes of numerical analysis (Finite Element Analysis) and the damages are identified in the system.



Figure 4 – VR to AR workflow for the proposed immersive DT framework

5 Conclusions

The use of Digital Twin applications for bridge monitoring processes has become a growing trend in construction and infrastructure projects. This research implemented the DT concept on the bridge health monitoring system by designing an immersive DT framework, which took the major step to tackle the research challenges in the integration of the DT model with VR/AR technologies to provide a holistic technologically available framework.

proposed immersive DT The model was demonstrated using the existing bridge project in Poland. As a basis of the DT model, the Physical State (PS), the Virtual State, and the Hybrid State of the bridge assets were defined for the SHM processes. The Virtual State (VS) is in the form of a BIM model and virtual replica of installed sensors and involves the data streaming from the physical system to the virtual system to provide accurate information about bridge health. Using such state definitions, the DT model of the bridge was developed in the UNITY game engine and deployed to the AR device (Microsoft HoloLens). The developed DT model is then successfully tested in its immersive nature and successful integration of measured data is observed in the AR environment. The case study shows that the ARenhanced DT of bridge SHM systems not only helps the real-time monitoring of bridge health but also performs predictive decision-making regarding bridge health. This integrated system has turned the conceptual designs into a matured BIM framework that is taking the BIM implementation to a new dimension.

6 Practical implications

The developed framework is practical in nature as it achieved the integration of the conceptual designs of BIM workflows in the AR/VR domain to digitize the physically installed SHM system. Therefore, the application of the developed methodology on this bridge is already planned. The authors have already contacted the authorities to practically implement the research results so in case of the approval, the research findings will be practically implemented on the subject bridge, which could also involve the professional training of the working staff. Also, once the research is implemented in the real life, it could be possible to calculate the cost details and the comparison of installation and maintenance costs.

Moreover, the core information extracted from the framework can help to define the future development of bridge immersive DT applications to aid the monitoring strategies of infrastructure assets. This way a detailed comparison of the performance of the immersive DT of SHM system with the use of traditional methods can be performed.

Further, this paper reinforces the understanding of immersive digital twins and contributes to their practical adaptability, showcasing the DT applications as an extension of BIM tools integrated with monitoring technologies for the management and monitoring of infrastructural assets.

Acknowledgment

This publication is supported by the Rector's pro-quality grant, Silesian University of Technology, grant no. 32/014/RGJ24/2023. Also, the partial funding of this publication was supported by Budapest University of Technology and Economics, Hungary.

References

- Palma V., et al., "Innovative technologies for structural health monitoring of SFTs: proposal of combination of infrared thermography with mixed reality," J. Civ. Struct. Heal. Monit., no. 0123456789, 2023, doi: 10.1007/s13349-023-00698-1.
- [2] Scianna A., Gaglio G. F., and Guardia La M. "Structure Monitoring with BIM and IoT: The Case Study of a Bridge Beam Model," *ISPRS Int. J. Geo-Information*, vol. 11, no. 3, 2022, doi: 10.3390/ijgi11030173.
- [3] Jiménez A., Plevris V., and Nogal M., "Bridge management through digital twin-based anomaly detection systems: A systematic review," *Front. Built Environ.*, vol. 9, no. April, pp. 1–18, 2023, doi: 10.3389/fbuil.2023.1176621.
- [4] Ye C., "A digital twin of bridges for structural health monitoring, in: Structural Health Monitoring 2019: Enabling Intelligent Life-Cycle Health Management for Industry Internet of Things (IIOT) -," in Proceedings of the 12th International Workshop on Structural Health Monitoring, 2020, pp. 1619–1626, doi: https://doi.org/10.12783/shm2019/32287.
- [5] Shabani A., Skamantzari M., Tapinaki S., Georgopoulos A., Plevris V., and Kioumarsi M., "3D simulation models for developing digital twins of heritage structures: Challenges and strategies," *Procedia Struct. Integr.*, vol. 37, no. C, pp. 314–320, 2021, doi: 10.1016/j.prostr.2022.01.090.
- [6] Sacks R., Brilakis I., Pikas E., Xie H. S., and Girolami M., "Construction with digital twin information systems," *Data-Centric Eng.*, vol. 1, no. 6, 2020, doi: 10.1017/dce.2020.16.
- [7] Honghong S., Gang Y., Haijiang L., Tian Z., and Annan J., "Digital twin enhanced BIM to shape full life cycle digital transformation for bridge engineering," *Autom. Constr.*, vol. 147, no. January, p. 104736, 2023, doi: 10.1016/j.autcon.2022.104736.
- [8] Fawad M., et al., "Automation of structural health monitoring (SHM) system of a bridge using BIMification approach and BIM-based finite element model development," *Sci. Rep.*, vol. 13, no. 1, pp. 1–18, 2023, doi: 10.1038/s41598-023-40355-

7.

- [9] Xiong B., Jin Y., Li F., Chen Y., Zou Y., and Zhou Z., "Knowledge-driven inference for automatic reconstruction of indoor detailed as-built BIMs from laser scanning data," *Autom. Constr.*, vol. 156, no. February, p. 105097, 2023, doi: 10.1016/j.autcon.2023.105097.
- [10] Cai H., Abudayyeh O., Abdel-Qader I., Attanayake U., Barbera J., and Almaita E., "Bridge deck load testing using sensors and optical survey equipment," *Adv. Civ. Eng.*, vol. 2012, 2012, doi: 10.1155/2012/493983.
- [11] Hasan S. M., Lee K., Moon D., Kwon S., Jinwoo S., and Lee S., "Augmented reality and digital twin system for interaction with construction machinery," *J. Asian Archit. Build. Eng.*, vol. 21, no. 2, pp. 564– 574, 2022, doi: 10.1080/13467581.2020.1869557.
- [12] Awadallah O. and Sadhu A., "Automated multiclass structural damage detection and quantification using augmented reality," *J. Infrastruct. Intell. Resil.*, vol. 2, no. 1, p. 100024, 2023, doi: 10.1016/j.iintel.2022.100024.
- [13] Boddupalli C., Sadhu A., and Rezazadeh Azar E., "An integrated structural health monitoring tool using building information modeling," 6th Int. Struct. Spec. Conf. 2018, Held as Part Can. Soc. Civ. Eng. Annu. Conf. 2018, pp. 87–95, 2018.
- [14] Catbas N. and Avci O., "A review of latest trends in bridge health monitoring," *Proc. Inst. Civ. Eng. Bridg. Eng.*, vol. 176, no. 2, pp. 76–91, 2022, doi: 10.1680/jbren.21.00093.
- [15] K K., "A BIM-Based Visualization Tool for Facilities Management: Fault Detection through Integrating Real-Time Sensor Data into BIM," J. Archit. Eng. Technol., vol. 9, no. 1, pp. 1–12, 2020, [Online]. Available: https://www.omicsonline.org/open-access/abimbased-visualization-tool-for-facilitiesmanagement-fault-detection-through-integratingrealtime-sensor-data-into-bim-111564.html%0Ahttps://www.omicsonline.org/pee r-reviewed/a-bimbased-visualization-tool-for-fac.
- [16] PN-EN, "PN-EN 1991-2: 2007 Eurocode 1: Actions on structures. Part 2. Moving loads on bridges.," *Eurocode 1*. 2007.
- [17] Polish Committee for Standardization, "PN 85/S-10030: 1985, Bridge structural loads," *PN 85/S-10030: 1985.* 1985.
- [18] Fawad M., Koris K., Salamak M., Gerges M., Bednarski L., and Sienko R., "Nonlinear modelling of a bridge: A case study-based damage evaluation and proposal of Structural Health Monitoring (SHM) system," *Arch. Civ. Eng.*, vol. 68, no. 3, pp. 569– 584, 2022, doi: 10.24425/ace.2022.141903.
- [19] Vuforia, "Vuforia Engine in Unity," Vuforia

devloper library, 2023. https://library.vuforia.com/getting-started/gettingstarted-vuforia-engine-unity.

 [20] Yogeeswaran K., Chen Q., and de Soto B. G., "Utilizing Augmented Reality for the Assembly and Disassembly of Panelized Construction," *J. Inf. Technol. Constr.*, vol. 28, no. April, pp. 571–584, 2023, doi: 10.36680/j.itcon.2023.030.

A Graph Neural Network Approach to Conceptual Cost Estimation

Hao Liu¹, Jack C.P. Cheng¹ and Chimay J. Anumba²

¹Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong

²College of Design, Construction and Planning, University of Florida, United States of America <u>hliuci@connect.ust.hk</u>, <u>cejcheng@ust.hk</u>, <u>anumba@ufl.edu</u>

Abstract –

Conceptual cost estimation plays an important role in construction projects since it is the basis for stakeholders to produce financial plans (e.g., establishing project budgets). The current practice, heavily dependent on cost engineers' subjective judgment and manual work, tends to be error-prone and labor-intensive. In response, this paper introduces a Graph Neural Network (GNN) approach to accurate and efficient conceptual cost estimation. Firstly, cost factors impacting construction costs, as well as their relationships, are identified based on literature review to form a graph representation. Afterwards, a GNN model is deployed to predict the construction cost. A real-world dataset from school projects is used for validation. The results show that the proposed approach achieved high accuracy, demonstrating the potential of graph neural networks in conceptual cost estimation.

Keywords -

Conceptual cost estimation; Deep learning; Graph neural network

1 Introduction

Conceptual cost estimation predicts the construction cost at the early stages of the project (e.g., conceptual design, budget setup) [1]. The estimation result is vital for the success of the project since stakeholders rely on it to set project budgets and make cost management plans before and during construction [2]. However, design information at project early stages is limited and full of uncertainties, which leads to low level of confidence on the estimation [3]. Traditional methods are heavily dependent on the experience and subjective assessments of cost engineers [4]. Such subjective evaluations, however, can vary and be unreliable, often resulting in inaccurate estimations and potential financial losses in the project [5]. Moreover, the dependence on the expertise of cost engineers makes the estimation process laborious and time-consuming, which is problematic for construction projects that typically operate on tight timelines [6].

To this end, deep learning, as a data-driven approach, presents an appealing alternative. It refines its accuracy autonomously by learning from historical data, and has demonstrated its reliability in data analysis and prediction within the construction sector [7]. Yet, standard deep learning models fall short in expressing the nuances of construction cost estimation. Specifically, construction projects exhibit intricate interrelations among cost factors (e.g., influence of contract type on project duration), highlighting their complex dependencies [8]. Traditional deep learning models, which typically use isolated factors to form tabular inputs, fail to account for these real-world characteristics of construction cost factors [9], missing out on capturing the interactive effects crucial for accurate construction cost estimation.

Graph deep learning, a branch of deep learning adept at representing intricate interrelations among input variables, offers a solution for encapsulating the complexity of construction cost factors. In this method, data is represented in a graph format where the edges in the graph provide linkage between the nodes, thereby depicting the impact relationships between them [9]. A number of studies have recognized the effectiveness of graph deep learning in representing input data in an expressive way for accurate predictions in the construction domain [10]. Nevertheless, its application in construction cost estimation is still in its infancy.

Thus, this research targets at developing a graph deep learning-based approach for accurate and efficient conceptual cost estimation for construction projects. A graph representation that integrates various construction cost factors and their relationships will be established, enabling expressive input data modeling. Then, a Graph Neural Network (GNN) will be developed and deployed using the graph representation to conduct data-driven conceptual cost estimation.

2 Related Work

Numerous studies have been conducted for accurate and efficient conceptual cost estimation. Early research leveraged statistical models to fit historical project data. For instance, Williams employed a univariate linear regression model to predict construction costs of highway projects [11]. In contrast, Stoy et al. used regression analysis with various factors (e.g., floor height, project type) to estimate construction costs for residential buildings [12]. The traditional statistical models have highlighted the dependency between the construction costs and their influential factors. However, they can be complex to implement and may lack robustness to intricate nonlinearity [13].

As a result, researchers began exploring the use of machine learning, known for its automatic pattern recognition and good prediction capabilities, to facilitate conceptual cost estimation. Fang et al. developed a Support Vector Machine (SVM) model with design information as the inputs to predict construction costs of building projects [14]. Based on the gradient boosting model, Chakraborty et al. used information structure and material design for construction cost estimation at the value engineering stage [15]. Although the majority of studies using conventional machine learning models for conceptual cost estimation yielded credible outcomes, they encounter limitations associated with shallow learning [16]. This limits their ability to discern more intricate patterns that could enhance estimation accuracy.

To address the limitations of traditional machine learning models, deep learning, a branch of machine learning renowned for its advanced capability to autonomously extract features and provide state-of-theart accuracy, has been investigated in recent years. Saeidloua and Ghadiminia implemented a Deep Neural Network (DNN) to estimate construction costs of buildings and found that DNN outperformed traditional machine learning models, such as SVM, in terms of estimation accuracy [17]. Similarly, Kim and Cha applied a DNN model to predict construction costs of data-scarce renovation projects, with the consideration of probability distributions of cost factors [18]. The deep learning methods have advanced conceptual cost estimation models regarding accuracy and efficiency. However, they rely on structured tabular inputs and tend to overlook the intricate relationships between cost factors in construction projects, which can lead to unreliability in real-world applications [9]. Therefore, to bridge the research gap, this paper introduces a graph deep learning approach, including a graph representation for modeling interrelated cost factors, and a GNN model for accurate conceptual cost estimation in a data-driven manner.

3 Proposed Method

Figure 1 provides an overview of the proposed graph neural network-based approach. Firstly, a thorough literature review is undertaken to identify factors influencing construction costs and their relationships. The identified results are validated by professionals construction cost specializing in estimation. Subsequently, a graph representation is established to effectively express the cost factors and relationships. A GNN model is then developed to conduct model training based on the formulated graph data representation. With the trained GNN model, the construction cost is predicted in an end-to-end manner. The main focus of this research is building projects, given their applicability to a diverse range of stakeholders, including public/private developers and contractors of various sizes. Detailed explanations of the proposed method are presented in subsequent subsections.



Figure 1. Overview of the proposed GNN-based approach

3.1 Graph Formulation

A comprehensive review of existing literature on construction cost estimation was carried out to pinpoint various factors that influence the construction costs, along with how these factors interrelate. Professionals in construction cost estimation then validated the findings. For inclusion in the graph deep learning model, these factors should meet specific criteria relevant to the focus and approach of this research, as follows:

- The factors should be those influencing the overall construction cost, rather than cost indices or unit rates.
- The factors should be relevant and easily obtainable in conceptual cost estimation (i.e., project early stages).
- The factors should be quantifiable for inputting into deep learning models.

Cost f	actor	Description		
Design information	Construction floor area	Construction floor area. A numerical variable in m^2 .		
	Building height	Height of the building. A numerical variable in <i>m</i> .		
	Soil condition	Geology condition type based on the geological map. A categorical variable.		
Project characteristics	Project type	Project type. A categorical variable.		
	Project duration	Planned duration of the construction works. A numerical variable in <i>month</i> .		
	Project location	Location of the project by region. A categorical variable.		
	Contract type	Contract type. A categorical variable.		

Table 1 Description of the cost factors

Figure 2 displays the identified cost factors and their relationships, with Table 1 providing the descriptions of the factors. As shown in Figure 2, there are three levels of the identified hierarchy of cost factors. The first level is the target of the estimation, i.e., the construction cost. The second level represents the influential aspects. The third level consists of specific cost factors in each aspect. In general, design information and project characteristics are two major aspects that impact construction costs. The design information aspect refers to construction floor area [19], building height [20], and soil condition [21]. As for the project characteristics aspect, it covers project type [1], project duration [1], project location [22], and contract type [23].

Moreover, there are relationships between these cost

factors. The dotted lines among the specific cost factors represent their impact relationships. In terms of design information, the soil condition plays a pivotal role in determining the building height [24]. Besides, there is a notable relationship between the height of a building and its construction floor area, since taller buildings may have larger floor areas, and vice versa. Regarding project characteristics, the project type has impacts on the choice of contract type [8]. Furthermore, both the contract type and project location can influence the duration of the project [25]. In addition, the project location often decides the soil condition according to geological maps.

Drawing from the identified cost factors and the relationships between them, Figure 3 illustrated the formulated graph representation. The nodes in the graph denote the identified cost factors, while the arrows describe the impact relationships between them (i.e., the dotted lines shown in Figure 2). The unidirectional arrows represent the certain impact from one factor to the other, while the bidirectional arrows denote the fuzzy interactions between two factors. Consequently, the graph representation encapsulates the cost factors and their correlations. This forms the foundational input for the GNN model in the next step.



Figure 2. The identified cost factors and their relationships



Figure 3. The formulated graph representation
3.2 Graph Neural Network Development

After establishing the graph representation of cost factors, a GNN model is developed to take the formulated graph as the input, utilize the graph structure in the model learning, and predict the construction cost. Figure 4 describes the architecture of the GNN model.



Figure 4. The developed GNN model

As shown in Figure 4, the developed model features graph convolutional layers to take advantage of the input graph structure in the neural network model training. The graph convolutional layer is a core component of Graph Convolutional Network (GCN), which is a mainstream GNN model and has demonstrated state-of-the-art accuracy performance in various graph-based prediction applications [26]. The graph convolution mechanism utilizes the graph structure to capture both node features and their topological relationships [26]. Equation (1) describes the operation in a graph convolutional layer as follows:

$$H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$
(1)

Where:

- $H^{(l)}$ stands for the node feature matrix at the l^{th} layer.
- $\hat{A} = A + I$ denotes the adjacency matrix of the graph, A, with added self-connections represented by the identity matrix I.
- \widehat{D} is the diagonal degree matrix of \widehat{A}
- $W^{(l)}$ means the trainable weight matrix at the l^{th} layer.
- σ is the non-linear activation function

Firstly, each node aggregates features from its neighbors, which learns contextual information in the graph. This is achieved by the multiplication of the adjacency matrix with the node feature matrix. The symmetric normalization, applied via $\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}$ considers different node degrees to ensure balanced influences from each node's neighbors. Then, a linear transformation is performed for the aggregated features through the trainable weight matrix. By doing so, the features are projected to a higher-level feature space, which facilitates the learning of complex patterns in the graph. Afterwards, a non-linear activation function, such as Rectified Linear Unit (ReLU), is used to introduce non-linearity in the neural network model for capturing complex data patterns. Multiple graph more convolutional layers are stacked as the essential parts of the developed GNN model. An average graph readout operation is conducted to compile the features of all the nodes into a single feature vector as a graph-level representation. The operation is defined in Equation (2), where N is the number of nodes in the graph, h_i stands for the feature vector of node i, and h_G denotes the graph readout feature vector to represent the entire graph. After producing the graph-level representation, fully connected layers are integrated to output the predicted construction costs.

$$h_G = \frac{1}{N} \sum_{i=1}^{N} h_i \tag{2}$$

Upon training the developed GNN model, a prediction model is established for conceptual cost estimation. The trained model can autonomously generate construction cost predictions for new projects that have the same graph input information. Such an end-to-end process helps to reduce the need for manual labor.

4 Experiments and Results

4.1 Experimental Design

The proposed method was implemented on a personal computer equipped with the Windows 10 operating system, an Intel(R) Core(TM) i7-11700KF @ 3.60GHz processor, an NVIDIA GTX 3060Ti GPU, and 32GB of RAM. The experiments were conducted with the Python 3.9.7 [27] programming language. PyTorch 1.10.2 [28] and scikit-learn 1.0.2 [29] libraries are adopted as the development platforms.

This study employs a dataset from school construction projects for validation. The dataset is from the Development Bureau of the Government of the Hong Kong Special Administrative Region, an authoritative organization overseeing building and infrastructure projects in Hong Kong. The dataset encompasses 50 school projects in Hong Kong, including their actual construction costs and the cost factor values required for the graph representation described in Section 3.1. The project scope includes various school types: primary, secondary, secondary-cum-primary, and special schools. Secondary-cum-primary through secondary levels, while special schools cater to students with physical or intellectual disabilities.

The dataset undergoes min-max normalization as part of its preprocessing, a technique aimed at enhancing model performance and expediting convergence speed [30]. It is randomly split into training and testing sets in an 8:2 ratio. The training set is used to develop and train the GNN model as detailed in Section 3.2. For the training process, Adam optimizer, which is recognized for its superior performance over other common optimizers such as SGDNesterov and RMSprop [31], is used for model optimization. After training, the model is evaluated using the testing set to assess the model's prediction accuracy.

4.2 Model Prediction Results

Regarding the model evaluation, the widely used metric for regression problems, Mean Absolute Percentage Error (MAPE), is adopted to assess the performance of the model in predicting construction costs. MAPE is a straightforward and scale-invariant metric that measures the discrepancy between predicted and actual values. The calculation is defined in Equation (3), where *n* represents the total number of projects being tested, A_i and P_i denote the actual and predicted construction costs, respectively. The MAPE indicates the percentage variance between predicted and actual costs, with a range from 0% to 100%. A low MAPE value means that the prediction is close to the actual cost.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - P_i}{A_i} \right|$$
(3)

Table 2 summarizes the prediction accuracy results of the developed GNN model and the comparison with common statistical, machine learning, and deep learning models in previous literature on conceptual cost estimation. The 5-fold cross-validation is conducted to thoroughly evaluate the performance of the GNN and other baseline models. The dataset is split into 5 folds, each of them is used for validation iteratively. The final performance is obtained by averaging the 5-fold crossvalidation results to reduce the risk of overfitting in a relatively small dataset and provide a more reliable performance comparison.

The developed GNN model demonstrated superior accuracy in predicting construction costs when measured against other typical baseline models. More specifically, it achieved better accuracy results than the typical statistical model, Least Absolute Shrinkage and Selection Operator (LASSO), highlighting its enhanced performance over traditional statistical techniques. A comparison between the GNN model and a conventional machine learning model, Support Vector Regression (SVR), was conducted, and the results showed the advantage of the GNN model in improving the prediction accuracy. Both the DNN and GNN models belong to deep learning methods. The DNN refers to an artificial neural network with an input layer (i.e., the features), hidden layers, and an output layer (i.e., predicted cost). For a fair comparison, the DNN adopts the same number of hidden layers (i.e., 5) as the GNN. The grid search is used to decide other optimal hyperparameters (e.g., learning rate, batch size) for the DNN and GNN, respectively. The performance comparison between the formulated graph representation incorporating interrelationships in the GNN and normal flat tabular inputs in the DNN verifies whether the graph representation is more effective or not. The higher accuracy provided by the GNN model highlights the superiority of the proposed approach compared with typical deep learning models that adopt tabular input formats. This indicates the importance of considering the relationships between cost factors within the deep learning model.

Table 2. Prediction performance of the developed GNN model and its comparison with other baseline models

	LASSO	SVR	DNN	The GNN
				model
MAPE	25.5%	21.0%	22.4%	15.2%

5 Conclusions

Conceptual cost estimation is pivotal in establishing

the financial scale for ensuring the success of the project. This study introduces a graph neural network approach to predict construction costs accurately and efficiently. The contributions are twofold. Firstly, an expressive graph representation is established for cost data modeling based on identification of cost factors and their complex relationships. Secondly, a novel GNN-based deep learning model is developed to predict construction costs, which demonstrates the effectiveness and potential of graph deep learning in conceptual cost estimation.

Of note is that while the proposed GNN-based method is more complex than simple techniques (e.g., regression, decision tress), the complexity is a response to the intricate nature of construction cost estimation where multiple cost factors are interrelated. Simple models fail to capture these interactions, which can lead to significant errors, as shown in the results. We argue that the proposed GNN-based approach provides a structured and systematic method for encapsulating the complex relationships while maintaining operational simplicity. The model's input requirements and the process flow remain straightforward in an end-to-end manner. In addition, the ability to handle non-structured data allows the proposed method to handle irregularities and maintain performance where simpler models might falter

Although the outcomes of this study are promising, several limitations exist. The formulated graph representation reflects the impact relationships among the specific cost factors, but does not comprehensively capture the hierarchical relationships in the identified two-level cost factor hierarchy shown in Figure 2. In the future, such relationships can be incorporated for more expressive modeling of cost factors. Besides, the blackbox nature of the neural networks raises concerns on the explainability of the cost estimation results. Future research is suggested to investigate and integrate advanced explainable artificial intelligence (XAI) techniques to reveal the hidden cost patterns learnt by the model to facilitate a trustable and reliable decisionmaking process.

Acknowledgement

The authors would like to thank The Government of the Hong Kong Special Administrative Region -Development Bureau for providing partial support to this research, including dataset preparation. Any opinions and findings are those of the authors and do not necessarily reflect the views of The Government of the Hong Kong Special Administrative Region - Development Bureau.

References

[1] He X., Liu R. and Anumba C.J. Data-Driven

Insights on the Knowledge Gaps of Conceptual Cost Estimation Modeling. *Journal of Construction Engineering and Management*, 147 (2): 04020165, 2021.

- [2] Zhang S., Migliaccio G.C., Zandbergen P.A. and Guindani M. Empirical Assessment of Geographically Based Surface Interpolation Methods for Adjusting Construction Cost Estimates by Project Location. *Journal of Construction Engineering and Management*. 140 (6): 04014015, 2014.
- [3] Elmousalami H.H. Artificial Intelligence and Parametric Construction Cost Estimate Modeling: State-of-the-Art Review. *Journal of Construction Engineering and Management*, 146 (1): 03119008, 2020.
- [4] Liu H., Cheng J.C.P., Gan V.J.L. and Zhou S. A novel Data-Driven framework based on BIM and knowledge graph for automatic model auditing and Quantity Take-off. *Advanced Engineering Informatics*, 54: 101757, 2022.
- [5] Akintoye A. Analysis of factors influencing project cost estimating practice. *Construction Management and Economics*, 18 (1): 77–89, 2000.
- [6] Zhang Y., Minchin R.E., Flood I. and Ries R.J. Preliminary Cost Estimation of Highway Projects Using Statistical Learning Methods. *Journal of Construction Engineering and Management*, 149 (5): 04023026, 2023
- [7] Akinosho T.D., Oyedele L.O., Bilal M., Ajayi A.O., Delgado M.D., Akinade O.O. and Ahmed A.A. Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, 32: 101827, 2020.
- [8] Chen Q., Xia B., Jin Z., Wu P. and Hu Y. Choosing Appropriate Contract Methods for Design-Build Projects. *Journal of Management in Engineering*, 32 (1): 04015029, 2016.
- [9] Mostofi F., Toğan V., Başağa H.B., Çıtıpıtıoğlu A. and Tokdemir O.B. Multiedge Graph Convolutional Network for House Price Prediction. *Journal of Construction Engineering and Management*, 149 (11): 04023112, 2023.
- [10] Li M., Liu Y., Wong B.C.L., Gan V.J.L. and Cheng J.C.P. Automated structural design optimization of steel reinforcement using graph neural network and exploratory genetic algorithms. *Automation in Construction*, 146: 104677, 2023.
- [11] Williams T.P. Predicting final cost for competitively bid construction projects using regression models. *International Journal of Project Management*, 21 (8): 593–599, 2003.
- [12] Stoy C., Pollalis S. and Schalcher H.-R. Drivers for Cost Estimating in Early Design: Case Study of Residential Construction. *Journal of Construction*

Engineering and Management, 134 (1): 32–39, 2008.

- [13] Wang R., Asghari V., Cheung C.M., Hsu S.-C. and Lee C.-J. Assessing effects of economic factors on construction cost estimation using deep neural networks. *Automation in Construction*, 134: 104080, 2022.
- [14] Fang S., Zhao T. and Zhang Y. Prediction of construction projects' costs based on fusion method. *Engineering Computations*, 34 (7): 2396–2408, 2017.
- [15] Chakraborty D., Elhegazy H., Elzarka H. and Gutierrez L. A novel construction cost prediction model using hybrid natural and light gradient boosting. *Advanced Engineering Informatics*, 46: 101201, 2020.
- [16] Pan Y., Fu X. and Zhang L. Data-driven multioutput prediction for TBM performance during tunnel excavation: An attention-based graph convolutional network approach. *Automation in Construction*, 141: 104386, 2022.
- [17] Saeidlou S. and Ghadiminia N. A construction cost estimation framework using DNN and validation unit. *Building Research & Information*, 1–11, 2023.
- [18] Kim J. and Cha H.S. Expediting the Cost Estimation Process for Aged-Housing Renovation Projects Using a Probabilistic Deep Learning Approach. *Sustainability*, 14 (1): 564, 2022.
- [19] Jafarzadeh R., Ingham J.M., Wilkinson S., González V. and Aghakouchak A.A. Application of Artificial Neural Network Methodology for Predicting Seismic Retrofit Construction Costs. *Journal of Construction Engineering and Management*, 140 (2): 04013044, 2014.
- [20] Elmousalami H.H. Comparison of Artificial Intelligence Techniques for Project Conceptual Cost Prediction: A Case Study and Comparative Analysis. *IEEE Transactions on Engineering Management*, 68 (1): 183–196, 2021.
- [21] Cheng M.-Y., Tsai H.-C. and Sudjono E. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Systems with Applications*, 37 (6): 4224–4231, 2020.
- [22] Lowe D.J., Emsley M.W. and Harding A. Predicting Construction Cost Using Multiple Regression Techniques. *Journal of Construction Engineering and Management*, 132 (7): 750–758, 2006.
- [23] Matel E., Vahdatikhaki F., Hosseinyalamdary S., Evers T. and Voordijk H. An artificial neural network approach for cost estimation of engineering services. *International Journal of Construction Management*, 22 (7): 1274–1287, 2019.

- [24] Coduto D., Yeung M. and Kitch W. Geotechnical Engineering: Principles & Practices 2nd Edition. Pearson, London, United Kingdom, 2010.
- [25] Kumaraswamy M.M. and Chan D.W.M. Determinants of construction duration. *Construction Management and Economics*, 13 (3): 209–217, 1995.
- [26] Wu Z., Pan S., Chen F., Long G., Zhang C. and Yu P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks* and Learning Systems, 32 (1): 4–24, 2021.
- [27] Python Software Foundation. Python. On-line: https://www.python.org/, Accessed: 22/11/2022.
- [28] Linux Foundation and Meta AI PyTorch. On-line: https://pytorch.org/, Accessed: 28/12/2022.
- [29] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M. and Duchesnay É Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12 (85): 2825–2830, 2011.
- [30] Goodfellow I., Bengio Y. and Courville A. Deep Learning, Illustrated Edition. The MIT Press, Cambridge, Massachusetts, United States, 2016.
- [31] Kingma D.P. and Ba J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 2015.

Automatic Assessment of Leanness of Construction Contract Provisions: A Proof-of-Concept

Ramesh Balaji¹, Sivakumar K.S², Murali Jagannathan² and Venkata Santosh Kumar Delhi³

¹Tata Consultancy Services, India ²Indian Institute of Technology Madras, India ³Indian Institute of Technology Bombay, India <u>ramesh.balaji@tcs.com</u>, <u>ce22s018@smail.iitm.ac.in</u>, <u>muralij@civil.iitm.ac.in</u>, <u>venkatad@iitb.ac.in</u>

Abstract

Construction projects significantly contribute to a nation's economic development. However, the sector is synonymous with delays and disputes for various reasons, often due to non-productive work practices. Researchers and practitioners recommend applying lean construction principles to mitigate non-value addition activities and improve productivity and performance. However, existing contract forms may contain provisions that are counter-productive, thereby making lean implementation a challenge. Therefore, when planning to go lean, it becomes important to assess the extent to which a contract provision aids or hinders lean implementation, in other words, 'leanness' assessment. A manual analysis is possible but time-consuming and prone to subjective decision-making. Artificial Intelligence (AI)-backed Language Model (LM)-based tools can be potentially used to quickly and efficiently classify a contract clause based on lean implementationfriendliness. Therefore, a dataset containing 734 contract clauses is manually classified into 14 labels based on the literature review, and a part of this data to train Bidirectional used Encoder is Representations from Transformers (BERT)-LM. With an F1 score of 77%, the study shows that LMbased solutions can be potentially employed for construction contract leanness assessment. The study, which is an initial attempt towards developing a reliable leanness prediction model in the future, also noted that the Bert-base-cased LM performs better than its 'large' counterpart under both the 'cased' and 'uncased' conditions.

Keywords -

Contracts; Lean; Language Models; Finetuning; Artificial Intelligence

1 Introduction

Lean construction principles can potentially help the

construction industry enable timely project completion, thereby preventing expensive cost overruns and associated disputes. Lean construction principles also help make construction more sustainable by reducing process and material waste [1]. Having said that, construction projects are stakeholder-intensive, and without their buy-in, lean implementation may not yield results [2]. Stakeholder acceptance of such initiatives is not easy, especially for an industry less friendly to innovativeness than sectors like manufacturing and automobiles [3]. Nevertheless, there should be a start, and academicians, practitioners, and researchers are important in triggering the start.

Over the last few decades, there has been a sustained effort by the research community to experiment with lean implementation in construction projects and disseminate the findings to the world at large. Construction firms, at least the leading ones from most developed and developing nations, have either implemented or shown interest in implementing lean construction practices [4]. However, there are still a large number of firms that are yet to realize the importance of lean implementation [5]. This difference in reaction times of various firms within the construction sector has triggered a separate research sub-domain studying the motivation for lean implementation in construction [6]. At the outset, it appears that some firms implement lean through a topdown approach where the firm implements lean as a response to some diktat or directive that insists on lean implementation, and some others embrace lean through a bottom-up approach, leaving the implementation exercise a largely self-driven activity [7], [8]. While there is no single answer to the question of the "better" way to implement lean, it is clear that the role of a "trigger" is inevitable, and it is here that the role of stakeholder practices comes into the picture.

In the construction industry, the demand for construction services is created by the developers and investors in infrastructure development belonging to the public or private sectors. When demand comes with the requirement to implement lean construction practices

while designing and executing projects, the supply is more or less assured. The suppliers, being construction contracting and sub-contracting firms, automatically adapt to deliver lean construction practices when they are asked to [9]. However, few researchers have pointed out that when lean is implemented as a reaction to a contractual or a policy requirement, it turns out to be superficial, merely to satisfy bare minimum requirements [6]. Nevertheless, the authors believe that when employers demand implementing lean practices through their contracts, it can at least serve as an initial trigger for firms that may not be self-motivated to embrace change. With this premise, this study attempts to develop a proofof-concept tool to automatically assess whether a given contractual provision promotes lean implementation in construction projects or does not. While this assessment can be manually performed, given the verbose contractual documents often running into hundreds of pages, it becomes time-consuming to assess documentintensive construction contracts [10] for their leanfriendliness or 'leanness.' Here, using artificial intelligence (AI) and machine learning (ML) can potentially be of immense help to researchers and practitioners to analyze contract documents quickly and efficiently. Building a classification model using the Language Model (LM) - Bidirectional Encoder Representations from Transformers (BERT) [11], the study recorded an F1 score of 77%, showing that AIbased solutions can be potentially employed for construction contract leanness assessment.

2 Background and Literature Review

This study attempts to answer the research question, "Can LMs be used to assess the leanness of a contract clause?" Accordingly, the literature is reviewed for LM and lean construction studies.

2.1 Language Models (LM)

The LM (or its larger counterpart, Large LM or LLM), a text model pre-trained on a large corpus, typically from general domains, is at the core of text analytics. For example, the Generative Pre-trained Transformer or GPT is an LLM on which the "ChatGPT" application is created [12]. Essentially, ChatGPT is a "questionanswering" platform. However, LLMs can be useful in developing applications for tasks such as text summarization, topic modeling, and text classification, among other things [13]. As most of the base LMs and LLMs are pre-trained on a large corpus of publicly available data (like Wikipedia), they may have limitations when used in highly specialized or domainspecific application development. For instance, in the instant case, where the objective is to classify a given contract clause based on its lean-friendliness, using a

base LM for application development may not yield satisfactory results as the model's training data may not have a sufficient concentration of lean-related information. In such cases, finetuning is one way to improve the model's output accuracy.

2.2 Finetuning

Finetuning is a supervised training process wherein base LMs, or LLMs, are trained with a dataset containing sample question-answer pairs or classified or labeled paragraphs [14]. Through this process, the base LMs generally trained on generic datasets get trained to answer questions or classify as per user requirements, as the case may be, thereby improving the output quality by making it more specific to the domain requirements. During the process of finetuning, the model interprets the pattern in the input data such that the pre-trained machine learning model (in this case, Bert) adapts to a new specific task (in this case, leanness assessment). While it is possible to train models from scratch, full training is computationally intensive, and in such cases, fine-tuning proves helpful. While research has shown considerable improvement, the process of finetuning is as good as the quality and quantity of the training data [14].

2.3 Choice of Base Model

The output accuracy (the extent to which the model helps assess the lean-friendliness of a contract clause) depends on the model size (the training data size and the number of parameters) and the application being developed. While some models like GPT are not opensource beyond a limit, many are open-source models (BERT, FLANT5, etc.). Among the open-source models, depending on the transformer architecture, there are encoder-only models, encoder and decoder models, and decoder-only models. Given the constraints of the scope of the study, and without getting into the technicalities, it is observed by researchers that encoder-only models like BERT (Bidirectional Encoder Representations from Transformers), an open-source Language Model (LM) can perform well for applications aimed at text classification, especially in the context of construction contracts [10], [14]. Based on dataset size and parameters, BERT models have the "base" and "large" and "cased" and "uncased" models. In this study, the 'BERT-basecased" model is chosen for finetuning. The review undertaken to arrive at the finetuning data is explained next.

2.4 Lean construction principles

Worldwide, there is a push for embracing sustainable construction techniques with the United Nations, Sustainable Development Goal (SDGs) 9 (resilient infrastructure and fostering innovation), specifically target 9.4, pushing for increased resource use efficiency. Here, lean construction principles take center stage in bringing sustainable practices into construction by minimising waste and increasing productivity.

Fundamentally, lean implementation boils down to seven principles that advocate eliminating or reducing non-value-adding activities, in other words, 'waste.' Firstly, lean implementation involves stakeholders, leading to a collaborative decision-making process [15]. When decision-making is collaborative, it is more likely that the parties are committed to the decision. Secondly, lean implementation promotes open communication among stakeholders, thereby helping in the early resolution of conflicts [16]. Thirdly, in construction projects that are often known for an adversarial environment [17], a lean and a "no blame game" culture go hand-in-hand, improving the trust among the stakeholders [15], [18]. Fourthly, while long-term plans are definitely relevant, lean advocates focus on shortterm goals as they are within the reach of stakeholder control [16]. However, the proponents of lean construction also advocate that while retaining a greater focus on short-term goals, it was also important to get into details in the form of weekly plans and six-month or 8-month lookahead schedules [16]. The fifth and sixth lean principles refer to identifying constraints through "pull planning" rather than the traditional "push" approach [16]. "Pull planning" refers to the process where the frontline engineers, the process owners, or "last planners" are encouraged to commit to what is achievable, given the resource constraints, rather than being blindly pushed by the top management to achieve unrealistic targets. If the commitment by the last planners is not aligned with the project requirements, the top management must ease the constraints so that the last planners can commit more. Through this, the last planners take ownership, improving productivity [16] and eliminating waste. Finally, lean implementation is about continuous improvement, a process through which parties identify risks and evolve mitigation plans for future projects [16], [19]. To summarise, any contract clause that promotes collaboration, timely decisionmaking, dispute-prevention, and supports waste reduction can be considered 'lean.'

A recent article reviews the studies presented at the annual conferences conducted by the International Group for Lean Construction (IGLC) for the synergies between lean construction and AI and observes that the LLM-based question-answering application – "ChatGPT" – can potentially empower lean researchers and practitioners [20]. The study also further summarises IGLC articles that directly or indirectly discuss the potential benefits of AI in lean construction, and it is clear that a contract document's lean-friendliness assessment

is not explored [20]. Articles with keywords "Construction," "contract management," "classification," "AI," and "label" mostly dealt with classifying the provisions of a contract document based on risk management [14], requirements identification [21], and scope and obligations identification [22], [23]. However, AI-based contract content classification to assess the 'leanness' of the contents is not evidenced. To assess LM's potential for leanness assessment, the study's objective is to develop an automated classification model that classifies a contract clause into a "Lean" or "Not Lean." The rationale for the classification labels is explained in Table 1.

Table 1. Labels for contract clause classification (number in the parentheses indicates count under each

Clause/Provision	Ref	Label
Clauses on Liquidated	[24]	Laber -
Damages drafted with clarity	[27]	Prevents
and certainty		Delay (11)
Clauses that provide fair cost	[25]	Lean -
and time compensation in	[20]	Prevents
case of delays caused by the		Disputes
Employer		(61)
A clause that is drafted with	[26]	Lean -
clear timelines for the	[=0]	Prevents
fulfillment of certain		Waiting (58)
obligations and mentions the		(e c)
implication of non-		
compliance		
A clause that discourages a	[27]	Lean -
contractual party from		Prevents
reworking/material		Rework,
wastes by focusing on skills		Material
and training		Wastes (54)
Clauses that enable early	[28]	Lean -
completion bonus		Promotes
-		Early
		Completion
		(18)
Clauses intended to check and	[29]	Not Lean –
approve the Contractor's		Liability
submissions and activities but		undefined
fail to specify the		(24)
intent/liability of such		
approvals. Will the approver		
be liable, or is the check only		
limited to the review of		
conformance of the		
work/submission to		
specifications/contract		
requirements? Without such		
clarification, there will likely		
be a "blame game."		

Clauses that contain provisions or terms that may have multiple meanings or are in contradiction with other provisions of the contract. Such provisions result in conflicts and non- collaborative behavior, often leading to delays and disputes	[30]	Not Lean – Misuse (ambiguous) (33)
Clauses that protect a party from being held responsible for its shortcomings	[31]	Not Lean – Misuse (No Waiver) (74)
Clauses that provide one party with absolute powers can be potentially misused, leading	[32]	Not Lean – Misuse (one-sided)
to non-collaborative behavior Clauses, which are a kind of "disclaimers", in which one party expects the other to bear risks that may not be reasonable and/or foreseeable even after due diligence. In such cases, the affected party tends to approach arbitration or judiciary for relief, resulting in delays and the development of adversarial relationships	[33]	(122) Not Lean – Misuse (unfairness) (44)
Clauses that prevent the non- breaching party from seeking compensation for losses incurred due to the delays or breaches by the breaching party. While such clauses are common, they can often lead to disputes when the breaching party excessively relies on such "no damage" provisions to protect itself. This can lead to adversarial and non-collaborative behavior, preventing stakeholder involvement and delaying decision-making.	[34]	Not Lean – No Damages Provision (38)
delaying decision-making. Clauses with no clear timelines for action, and therefore there is a tendency for delayed decision-making, inducing "waiting "	[26]	Not lean – No Timelines (84)
Clauses that do not explain the implication of not complying with the instructions/orders/contractual promises. In such cases, there	[35]	Not Lean – Non- compliance Unaddressed (91)

is no certainty on how the		
other party will react to the		
inaction of the non-complying		
party. This can lead to		
disagreements, non-		
collaborative behavior, and		
breaking the stakeholder		
involvement (incompleteness)		
Clauses, when acted upon,	[36]	Not Lean –
can lead to unnecessary		Waiting (21)
"waiting" at project sites.		

3 Methodology

The study employs contract document content analysis using an LM-based classification technique. A three-step methodology is adopted to develop a proof-ofconcept model to assess the leanness of a given construction contract provision.

3.1 Step 1: Developing Training Dataset

The initial idea was to approach using a standard supervised classification method in which a training dataset is prepared to develop a classification model annotated with labels derived from the literature.

3.2 Data Pre-processing, Classification Training, and Testing using LM

In this step, as a prerequisite for data preprocessing, the developed file (in the .csv format) is ingested as a Pandas dataframe through the algorithm in the Python programming language. After setting up the environment with key libraries (Datasets, Transformer, Accelerate, sklearn metrics, Pandas, Torch), the data is pre-processed to enable its use in the LM, which, in this case, BERT.

To enable training the language model, the input data will be split in an 80:20 ratio, with 80% of the data used for training and the rest for testing. Since there are more than two labels (14, in this case), the stratification technique is used for sampling the data.

Stratification ensures dividing the labels into homogenous subgroups, called strata, and then applying simple random sampling within each subgroup. As a result, the test set is representative of the population since the percentage of each stratum is preserved. In this context, the stratification is performed on the "label" column with 14 groups. The key idea is to ensure that the train and test dataset has all 14 groups represented in complete. After formatting the current dictionary into a tokenizer-based dataset, the "Bert-base-cased" model is downloaded through the Auto model function and prepared for the training process. Regarding training arguments, 18 epochs are set up, meaning all the training records will run 18 training cycles. In each epoch, the size of the batch will be 16.

3.3 Step 3: Model Evaluation

Lastly, since the evaluation process is also run simultaneously, the evaluation batch of 64 records will be used. In this experiment, "Accuracy" and "F1 score" are the primary metrics that will be evaluated. Considering that the input data is imbalanced, the F1 score eliminates any anomalies in calculating model accuracy.



Figure 1. Label Histogram

4 Results

4.1 Step-1 – Developing a Training Dataset

As discussed earlier, 14 labels are considered to classify a contractual provision under lean/not lean. The initial list of clauses for training consisted of 307 clauses sourced from the General Conditions of Contract (GCC) of public sector contract documents from India. However, with the number of instances being low, it was decided to leverage the power of ChatGPT to generate multiple instances of 'Not Lean' provisions through paraphrasing. This helped create paraphrased clauses. After validating the meaning through a manual reading of the paraphrased text, such clauses were added to the original list. The final list consisted of 531 clauses labeled as 'Not Lean' and 202 'Lean' clauses, totaling 733 classified clauses with label count as indicated in the last column of Table 1 in parenthesis and in Figure 1.

4.2 Steps 2 and 3 – Classification and Testing Results

In the training process with 80 to 20 splits, it is observed that the accuracy and F1 score is a maximum of 77%. On increasing the training size to 90% and then to 95%, it is noted that the F1 score improved to 78% and 85%. The results of trying with the Bert-base-uncased, Bert-large-cased, and uncased models are shown in Table 2. In terms of the epoch standpoint, considering the mechanics of double descent, there was no improvement even with 50 epochs. In the above technique, all the weights of the original model are used. However, another technique called a Low-Rank Adoption [37] is used by which the original weights of the Bert model remain untouched, and a new set of weights is created based on the rank of the matrix, which is a user parameter. A snapshot of the finetuning results is shown in Figure 2.

Epoch	Training Loss	Validation Loss	Accuracy	F1 Score
13	No log	1.141195	0.768707	0.768707
14	0.623200	1.204195	0.775510	0.775510
15	0.623200	1.236950	0.775510	0.775510
16	0.623200	1.239883	0.775510	0.775510
17	0.623200	1.249316	0.775510	0.775510
18	0.623200	1.262342	0.775510	0.775510

Figure 2. A Snapshot of the Finetuning Result

Table 2. Results from different Bert Models

Model Name	F1 Score (in percentage)
Bert-base-cased	77
Bert-base-uncased	71
Bert-large-cased	68
Bert-large-uncased	65
Bert-base-LoRA	45

4.3 Discussion

In this study, two different categories of Bert Models are used, namely, the Bert-base model and the Bert-large model. Bert-base has a total of 12 attention heads and 110 million parameters. Meanwhile, Bert-large has 16 attention heads with 340 million parameters. Although from the size standpoint, Bert-large models are 3 times the size of the Bert-base models, the performance of the Bert-base model is superior, indicating that model size parameters do not always matter and that smaller models like Bert-base are less prone to overfitting and more capable of generalizing to new data, making them dependable and robust in real-world settings. Nevertheless, larger models still outperform smaller ones in specific use cases. However, this study indicates that smaller models are often more useful for retraining with recent data or fine-tuning for specific tasks.

Regarding the choice between the cased and the uncased models, the results indicate that the cased model's performance outweighs its uncased counterpart (Table 2) in both large and small Bert models. Cased models have separate vocabulary entries for differentlycased words. For instance, the words "the" and "The" have distinct uses in English. "Contractor" and "contractor" will typically have different meanings in construction contracts. While "Contractor" refers to the specific organization defined in the contract's definitions clause, "contractor" can refer to the word with the dictionary meaning. The cased models, sensitive to such differences, have performed better than the uncased models.

Overall, recalling the research question, "Can LMs be used to assess the leanness of a contract clause?" the results inform the potential of language models to understand and assess the 'leanness' of a given contract clause. However, only the publicly available standard form contracts are used for analysis. The model needs further improvement by including more illustrative contract provisions from various contract documents worldwide. Illustrative provisions from internationally used contract forms like the International Federation of Consulting Engineers (FIDIC) can help generalise the model usage. However, in most cases, contract documents are considered confidential and may not be available to the researchers for analysis. While confidentiality is important, in the absence of relevant data, the power of AI techniques is severely underutilized. Therefore, it will be helpful if the industry worldwide works on a common platform to share anonymized contract provisions so academicians and researchers can help develop AI-based contract risk assessment tools. It is here that the role of industry-academia bodies such as the International Group for Lean Construction (IGLC), Lean Construction Institute (LCI), and Institute for Lean Construction Excellence (ILCE) (Indian lean body) becomes crucial as a bridge to connect industry, academia, and researchers.

Adequately drafted contract documents can be crucial in bringing out lean adoption. However, to enable the catalyzing role of the contract, it is important that the contract documents are drafted in a manner that can promote lean principles in projects. It is here that this study, when fully ready, comes in handy to practitioners to assess the 'leanness' of their construction contracts. In the absence of such an assessment and if the contract provisions do not support lean implementation, attempts to bring a positive change in the project turn futile. In terms of its contribution to theory, the study exposes the power of AI to understand the underlying implicit features in contract provisions, a distinctive feature of AI that can be potentially used to assess many such underlying features in a contract document. Specifically, the differences in the performances of Bert-base and the large models, under both the cased and uncased conditions, are analyzed in the context of assessing a contract document.

5 Limitations and Future Scope

The inference window for this proof-of-concept available model is at the link: https://huggingface.co/RameshBal/LeanContractModel/ blob/main/README.md. However, a major limitation of this model lies in the limited training dataset size, which comprises clauses only from Indian public sector contracts, which diminishes model reliability. Therefore, notwithstanding the reasonable F1 score, the model is not industry-ready at this stage. Future studies can consider contract forms from different jurisdictions to enrich the training data, making the model application-friendly to test the leanness of various contract forms. The model development is in a preliminary stage, and with additional data and specific expert validation, this proofof-concept can be developed to an application scale. Nevertheless, the study helps understand the adaptability of models to real-world situations in construction management tasks. Secondly, only the BERT LM is evaluated in this study, whereas there are other LMs and LLMs whose robustness for similar studies has not been explored. Accordingly, future studies can focus on improving the dataset with additional clauses. Secondly, researchers can attempt to use various LMs and LLMs and evaluate their performance; and lastly, develop lean domain-specific language models for the exclusive use of AI-based lean studies.

6 Future Work

Considering limited data sources, the number of clause illustrations per label is improved by reducing the label count yet retaining the essence. Roberta-Large, with LoRA, is being explored to develop the classification model. Initial trends show a promising improvement in the F1 score. However, the study will be reported after a detailed analysis of results, the role of model architecture, and model validation by industry experts.

7 Conclusion

The study's objective was to develop a proof-ofconcept AI-based tool to assess the leanness of contractual provisions in construction. A supervised algorithm-based approach was adopted, and a BERTbase LLM was finetuned with contract provisions labeled as 'lean' or 'not lean.' The resulting model could predict the leanness of a given contract clause with an F1 score of 77%. However, at this stage, the model is just a proofof-concept to demonstrate the robustness of AI applications to understand the implicit meaning of contract provisions and can potentially be developed into an industry-ready assessment tool. Nevertheless, this study is a step closer to realizing the potential of data in making lean implementation a reality.

References

- Jamil, A. H. A. and Fathi, M. S. The Integration of Lean Construction and Sustainable Construction: A Stakeholder Perspective in Analyzing Sustainable Lean Construction Strategies in Malaysia. *Procedia Comput Sci*, 100:634–643, 2016.
- [2] Simon, S. M. and Varghese, K. Assessment of organizational culture in construction – A case study approach. In Proceedings of the 26th Annual Conference of the International Group for Lean Construction: Evolving Lean Construction Towards Mature Production Management Across Cultures and Frontiers, The International Group for Lean Construction, pages 348–357, Chennai, India, 2018.
- [3] Nesensohn, C., Bryde D., Ochieng E., Fearon D. and Hackett V. Assessing lean construction maturity. In Proceedings of the 22nd Annual Conference of the International Group for Lean Construction: Understanding and Improving Project Based Production, International Group for Lean Construction, pages 1157–1168, Oslo, Norway, 2014.
- [4] Roy, D. and Jagannathan, M. Exploring the Reach of Lean Philosophy in Indian Construction Industry. In Proceedings of the fourth biennial conference of the Indian Lean Community - Indian Lean Construction Conference, pages 203–212, Ahmedabad, India, 2021.
- [5] Ranadewa, K. A. T. O., Sandanayake, Y. G. and Siriwardena M. Enabling lean through human capacity building: an investigation of small and medium contractors. *Built Environment Project and Asset Management*, 11(4):594–610, 2021.
- [6] Bygballe, L. E., Endresen, M. and Fålun, S. The role of formal and informal mechanisms in implementing lean principles in construction projects. *Engineering, Construction and Architectural Management*, 25(10):1322–1338, 2018.
- [7] Devkar, G. Delhi, V.S.K. Enablers and inhibitors of lean construction in Indian construction projects. *Proceedings of the second biennial conference of the Indian Lean Community - Indian Lean Construction Conference*, pages. 179–185, Chennai, India, 2017.
- [8] Narayanan, S. R. Adopting lean, cycle time got reduced from 15 days to 7 days. *Construction World*, pages 48–48, 2021.
- [9] PMI&KPMG, Revamping Project Management. Project Management Institute, New Delhi, India, 2019.

- [10] Eken, G. Using natural language processing for automated construction contract review during risk assessment at the bidding stage. *Doctoral Thesis*, Middle East Technical University, 2022.
- [11] Devlin, J., Chang, M.-W., Lee, K. and Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv Preprint*, 2018.
- [12] Brockman, G. Sutskever, I. and OpenAI, Introducing OpenAI. On-line: https://openai.com/blog/introducing-openai, Accessed 10/12/2023.
- [13] Ghimire, P. Kim, K. and Acharya M. Generative AI in the Construction Industry: Opportunities & Challenges. arXiv Preprint, 2023.
- [14] S. Moon, S. Chi, and S. B. Im, "Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT)," Autom Constr, vol. 142, Oct. 2022, doi: 10.1016/j.autcon.2022.104465.
- [15] Lichtig, W. A. Sutter Health: Developing a Contracting Model to Support Lean Project Delivery. *Lean Construction Journal*, 2(1):105–112, 2005.
- [16] Forbes, L. H. and Ahmed S. M., *Modern Construction: Lean Delivery and Integrated Practices.* CRC Press, New York, 2011.
- [17] Tazelaar, F. and Snijders, C. Dispute resolution and litigation in the construction industry. Evidence on conflicts and conflict resolution in The Netherlands and Germany. *Journal of Purchasing and Supply Management*, 16(4):221–229, 2010
- [18] Aidibi, H. I. Studying the effect of BIM on Construction Conflicts and Disputes using agentbased modeling. *Master Thesis*, American University of Beirut, Lebanon, 2016.
- [19] LCI. Work Structuring. *LCI White Paper*, Lean Construction Institute, pages 1–15, 1999.
- [20] Hatoum, M. B. and Nassereddine, H. Unleashing the Power of ChatGPT for Lean Construction: An Early Outlook. In Proceedings of the 31st Annual Conference of the International Group for Lean Construction (IGLC31), International Group for Lean Construction, pages 208–219, Lille, France, 2023.
- [21] Hassan, F. U. and Le, T. Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 12(2):1-14, 2020.
- [22] Yang, J., Chen, Y., Yao, H. and Zhang, B. Machine Learning–Driven Model to Analyze Particular Conditions of Contracts: A Multifunctional and Risk Perspective. *Journal of Management in*

Engineering, 38(5):1-16, 2022.

- [23] Chalkidis, I. and Androutsopoulos, I. A Deep Learning Approach to Contract Element Extraction. *JURIX*, 2017:155–164, 2017.
- [24] Sweet, J. Liquidated Damages in California. California Law Review, 60(1):84–145, 1972.
- [25] Osama Jannadia, M., Assaf, S., Bubshait, A. A. and Naji A. Contractual methods for dispute avoidance and resolution (DAR). *International Journal of Project Management*, 18:41–49, 2000.
- [26] Jaeger, A.-V. and Hök, G.-S. FIDIC A Guide for Practitioners, 1st ed. Springer, Berlin, Heidelberg, 2010.
- [27] Mojahed, S. and Aghazadeh, F. Major factors influencing productivity of water and wastewater treatment plant construction: Evidence from the deep south USA. *International Journal of Project Management*, 26(2):195–202, 2008.
- [28] Bubshait, A. A. Incentive/disincentive contracts and its effects on industrial projects. *International Journal of Project Management*, 21:63–70, 2003.
- [29] Abu Dief, M. and Metwally, A. Construction Claims Mitigation in Design and Build Contracts. *PM World Journal*, IX:2330–4480, 2020.
- [30] Anish, P. R., Sainani, A., Ramrakhiyani, N., Pawar, S, Palshikar, G. K. and Ghaisas, S. Towards Disambiguating Contracts for their Successful Execution-A Case from Finance Domain. In Proceedings of the First Workshop on Financial Technology and Natural Language Processing, pages 8-13, Macao, China, 2019.
- [31] Padhy, J., Jagannathan, M. and Delhi, V. S. K. Application of Natural Language Processing to Automatically Identify Exculpatory Clauses in Construction Contracts. *Journal of Legal Affairs* and Dispute Resolution in Engineering and Construction, 13(4):1–9, 2021.
- [32] Agrawal, A. K., Jagannathan, M. and Delhi, V. S. K. Control Focus in Standard Forms: An Assessment through Text Mining and NLP. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 13(1):1–11, 2021.
- [33] Thomas, H. R. Differing Site Conditions Indications and Disclaimers. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 4(3):86–89, 2012.
- [34] Ramnath International Construction Ltd. v. Union of India, Appeal (civil) 3167-3168 of 2005, Court Judgement, Supreme Court of India, 2006.
- [35] Badenfelt, U. Fixing the contract after the contract is fixed: A study of incomplete contracts in IT and construction projects. *International Journal of Project Management*, 29(5):568–576, 2011.
- [36] Koskela, L., Bølviken, T. and Rooke, J. Which are the wastes of construction? *21st Annual Conference*

of the International Group for Lean Construction, pages 905–914, 2013.

[37] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. arXiv Preprint, 2021.

BIM and IPA – Excerpt of an automated assessment system for an autodidactic teaching concept

Christian Heins¹ and Gregor Grunwald²

1,2 Jade University of Applied Sciences, Department of Architecture, Germany

christian.heins@jade-hs.de, gregor.grunwald@jade-hs.de

Abstract -

This article introduces an innovative approach to assessing and enhancing Building Information Modeling (BIM) skills using Digital Process Automation (DPA) and Robotic Process Automation (RPA) based on [1] [2]. Aimed at university BIM courses, the study addresses the challenge of evaluating students' modelling proficiency. The autodidactic teaching concept utilizes a Bizagi-based work portal, offering modules categorized by BIM roles. Learners engage in application-specific tasks, assessed through automated evaluations facilitated by RPA bots. The integrated system successfully combines DPA and RPA components, highlighting challenges and prospects for an error-free portal.

Keywords -

Building Information Modeling (BIM); Intelligent Process Automation (IPA); Digital Process Automation (DPA); Robotic Process Automation (RPA); Information Management System (IMS); Work portal; fully automated BIMbased model checking (BMC); Education and training

1 Introduction

Robotic Process Automation (RPA) technology can automate manual, repetitive and error-prone Computer Aided Engineering (CAE) applications [1]. Particularly in university courses in the field of "Building Information Modeling" (BIM), lecturers are increasingly faced with the challenge of assessing students' modelling skills. For grading purposes, lecturers need to assess not only the geometric representation and causal relationships between the component objects but also whether the associated component information (so-called free characteristics, known as property sets) has been correctly added to the 3D model by the students in response to the client's information requirements. In the BIM Games [3] [4], a new teaching and training format developed and implemented by Jade University of Applied Sciences (Jade-HS), the building model check was previously carried out manually by downloading the Industry Foundation Classes (IFC) models [5] from the work portal, checking them in a model checking application and uploading the results back to the work portal. Due to the limited duration of the BIM games developed by

the Jade-HS (usually three days [4]), checking the building models represented the greatest effort for the teachers. To reduce this effort, RPA technology is used to activate software robots from the central work portal, that fully automatically verify digital building models, provide initial plausibility statements and check the model skills of the participants. In addition to presenting a work portal based on an information management system used as a Digital Process Automation (DPA), this article presents RPA technology, and it is porting to the work portal to present a fully automated assessment system based on a new autodidactic teaching concept that checks modelling skills using RPA bots and presents and document the learning progress in the work portal in real-time.

2 Autodidactic teaching concept

The learning concept is accessible via a work portal and is shown in principle in Figure 1. The module content is divided into different roles (see Figure 1 (a)). Each module consists of application-related tasks. For each task, a unique example solution has been developed, which is superimposed on the end-user for checking the results. When it comes to acquiring modelling skills, e.g., to obtain a certificate as a BIM author, the end-user logs in to the work portal using any browser on a terminal device (see Figure 1 (b)). After logging in, the user accesses his personalized area, where he has access to his completed, activated and running modules, his login credentials, and the module overview. If the user wants to edit an already running or a new module, he clicks on the corresponding module and is forwarded to the user interface, another front end of the work portal (see Figure 3).

The interaction between man and systems takes place in the user interface. Here, the end-user is provided with the corresponding learning objectives, tasks, and didactic aids as well as input fields for his solutions and value fields for reading the respective result check (cf. Figure 1 (b) bottom right). The solutions uploaded to the work portal are checked as soon as the end-user clicks the "Check Solution" button. The check is fully automated by the system, either by a tabular check using Microsoft Excel or by a check of a third-party CAE application performed by an RPA bot. The principle of the autodidactic learning concept thus offers the end-user the possibility to apply the most common CAE applications and their interoperable workflows in a standards-compliant manner based on the interest-specific selected BIM roles (BIM Author, BIM User, BIM Coordinator, BIM Manager) selected according to their interests. Here is an example from the first task of a BIM author in module 1.1 "Revit", which an end-user would have to process:

Scenario for Module 1.1 "Revit" - Task 1.1 [2]: "Model the foundation of the single-family house shown in Figure 5, using the general plan in Figure 6 and the project origin specified. Note the orientation of the project. You must model exactly within the predefined reference plan and use the correct materials. Once you have modelled the foundation, export the foundation as an IFC file (using the didactic tools) and upload the IFC file in the work portal under the appropriate input field (see Figure 8 (a)). Once your IFC file has been uploaded, click on the button "Check solution" to automatically check your result. If your result is correct, you will be notified that you have passed the task and the "Next" button will be enabled to proceed to the next task. If your result is incorrect, please read the result log carefully and adjust your model. Then repeat the above procedure until you have fully completed the task."

If the first task is solved correctly, the following task "Modeling the first floor" is unlocked. The tasks in this module are repeated until the single-family house has completely been modelled according to the standard. In module 1.2 "Allplan" the CAD program is changed, and the task is repeated. In total, the user must select three CAD programs to fully understand the differences between the most common CAD programs. Provided that the user has completed three modules, the person will receive a BIM Author certificate. For the acquisition of modelling skills, various CAD programs are provided (e.g. Revit, Archi-CAD, Allplan, Tekla, Vektorworks) to independently acquire BIM-compliant modelling skills using didactic tools for topics such as modelling guidelines, input, and organization of property sets, IFC export settings, text files for IFC mapping etc.

3 DPA-based work portal

For the teaching concept, the work portal of Bizagi, a process-oriented information management system (IMS) is used. For this purpose, Bizagi offers a low-code portal which can be adapted to the needs of an autodidactic teaching concept based on process and data models, business rules, read and write permissions, software distribution and an open integration layer [6, p. 779 ff.].





Figure 1. Principle of the autodidactic teaching concept: (a) Module concept; (b) Roles and rights of the work portal

As shown in Figure 2, the Bizagi suite consists of the Modeler, Studio and "Automation to Map" process flows (Modeller) to develop process instances (Studio) and to make the portal executable and persistent for end-users via a browser (Automation) [6, p. 36 ff.]. For the work portal presented here, Bizagi is used to provide the task packages, request the solutions, activate the RPA bots and display the achieved RPA results back to the end-user as well as to document the learning success through integrated reports. In terms of technical requirements, Bizagi requires a database and an application server to host the web server instance. Bizagi offers two ways to develop the integrated runtime environment (a predefined work portal) as a learning portal: Editing as a single user, for which a quick start setup is provided, and editing in multi-user mode, where further development as teamwork is supported [6, p. 324 ff.]. It should be noted that the Bizagi work portal is also accessible via various mobile devices running iOS (e.g. iPad, iPhone) or Android. The different user interfaces, shown as an example in Figure 3, present the end-user with the modules as cases in an overall overview (Figure 3 (a)) or the user interface in which the interactions described in Chapter 2 take place. Learning progress is displayed in the Reports ribbon (see Figure 3 (a) and (b) above).



Figure 2. Building blocks of the Bizagi Suite

3.1 Processes

The processes stored in the work portal are shown in Figure 4 (a) – (extracts for modules 1.1 to 1.3 only). Once the end-user has logged into the work portal and clicks on module 1.1 in the general overview, he is taken to an information page of the user interface. There he can inform himself about the selected module and decide whether he wants to edit the module or to look at another module. If the user does not wish to edit the module or view another module, the user interface is closed, and the user returns to the general overview. If the user wishes to edit Module 1.1 for example, he will be taken to the tasks to be completed, as shown in Figure 4 (b). Module 1.1 is completed as soon as the end-user has completed all tasks.

3.2 Data models

To match the solutions to the requirements of the different tasks, unique attributes must be stored in the process





Figure 3. Frontend of the work portal: (a) Stakeholder view; (b) User interface

instances of the user interface. The attributes are listed in different entities to store the information in data models and to structure the information flow. A master process entity combines all the individually extendable entities. This ensures, among other things, that the RPA bots integrated into the work portal deliver their results in the right place and the right order. For example, if the enduser has finished modelling the foundation in Module 1.1, he will upload the corresponding IFC model in the work portal by inserting the IFC model into the "Solution 1.1" input field (see Figure 8 (a)). This field stores the file attribute <uSolution11> belonging to the master entity <M11_Revit_ModelFoundation>. The RPA bot receives this information to execute the activities and to provide the results to the end-user again in the value field "Result 1.1" with the file attribute <uResult11>. Following this principle, the attributes of the work portal are migrated with the attributes of the RPA bots [6, p. 3216 ff.].

4 Automated evaluation system

The evaluation system integrated into the work portal consists of the third-party systems Microsoft Excel [7] and an RPA bot from the company UiPath [8]. Both systems are embedded in the integration layer of Bizagi (cf. [6, p. 2893 ff.], [6, p. 3216 ff.]), so that an automated evaluation can take place at either tabular, based on alphanumeric values (e.g. for the comparison of GUIDs to check the one-time submission), or by any external third-party sys-



Figure 4. Excerpt from the technical process model of the work portal: (a) process model for modules 1.1 to 1.3; (b) sub-process 1.1

tems. In the course of the autodidactic learning concept, the evaluation of content-related knowledge is done in tabular, whereas the evaluation of modelling skills is done by the RPA bot with the help of a model checker of the company Solibri [9]. For example, to check the modelling capabilities of an end-user fully automatically, the RPA bot takes the uploaded IFC file of the end-user, opens a Solibri project – in which the corresponding IFC example model already exists – and compares the two IFC models. The RPA bot then saves the corresponding Solibri result message and transfers the file to the work portal. The enduser can then view the results in the work portal and, if necessary, improve his IFC model until the task has been fully solved.

4.1 BIM-based model checking (BMC)

Before RPA automation could be implemented, it was first necessary to determine which rules should be used as the basis for comparing the end-user's IFC model with the IFC example model. It was also necessary to determine the behaviour of Solibri's BMC system in the presence of error-free and minor and major errors. For this purpose, different IFC models with different errors were generated and compared with the IFC example model. The following cases were considered:

Case-1: IFC test model too small

Case-2: IFC test model too large (cf. Figure 7)

Case-3: IFC test model on the coordinate origin 0.0.0

Case-4: IFC test model rotated by 1 degree

For testing purposes, an automated test process was carried out for the foundation of a single-family house (cf. Figure 5). The test process is therefore initially focused only on the geometric and alphanumeric modelling of the foundation. Solibri's BMC software is used as a test tool for the scenario described in Chapter 2.



Figure 5. Example project for acquiring modelling skills in the role of a BIM author [2]

In the case of Module 1.1, the strip footings in Revit must be modelled as beams or walls, as shown in Figure 6, because Revit does not provide objects for strip footings without a direct connection to the wall above. For example, since civil engineers focus primarily on shell construction, and it is common to model wall openings as individual walls connected by a lintel rather than walls with door openings, a beam object should be used instead of a strip footing object. This information is important for the IFC export as the strip footing to be exported will not be exported as IfcFooting as intended, but as IfcBeam (if no adjustments are made), which could cause other thirdparty systems to misinterpret e.g. the fabrication costs or the load-bearing function.

In addition to the geometric checks, the alphanumeric information to be added to the component objects must be checked. Table 1 shows the required alphanumeric information.

The results of the manual tests show that with modified rules the modelling skills of the end-user can be checked in an automated way, provided that the following requirements are met:



Figure 6. Reference plan for post-modelling of the foundation [2]



Figure 7. Example for the self-test error in Case-2

- During the automated verification, the BMC software Solibri must always remain open with the predefined classification.
- The IFC example model is never removed as an architectural model
- The IFC model of the end-user is added to Solibri as a static model (short name "B")
- The static model is removed after each check and added again before each check, as described above.

4.2 RPA-BIM-based model checking (RBMC)

For automated model checking, an attended bot from the company UiPath is used for testing purposes [10]. This robot (RPA bot) mimics human software activity on

Table 1. F	Required	property	set
------------	----------	----------	-----

Name	Mate- rial	IFC en- tity	PSet
Base plate	C20/25	IfcSlab	Loadbearing = yes
Foundation	C20/25	IfcFoot- ing	Loadbearing = yes
Insulation	EPS	IfcCov- ering	Loadbearing = only un- der foundation

the screen by reconstructing the corresponding actions in a sequence. Once programmed, the RPA bot can be activated at any time by pressing the "Check solution" button via the work portal, as shown in Figure 8 (a). For automated model checking, the sequence flow of the RPA bot is shown in excerpt form in Figure 10.

Performing an automated check with Solibri is relatively simple compared to report output, provided the above requirements are met. With a detailed report output, the RPA bot would now have to recognize the anomalies identified by Solibri, expand the corresponding categories, select the corresponding categories, select the corresponding messages, recognize the number of error messages in the results and create a presentation slide with predefined content for each error message. This type of reporting would be relatively complex for the RPA bot to program and would result in high throughput times for the RPA bot. Therefore, the initial focus is on the internal reporting of the underlying rules. To do this, the independent rules are combined into a rule set to create a summary report using RPA click applications (see Figure 9).

5 IT-Infrastructure

Bizagi's DPA can be hosted locally in an older version, see Figure 8 (b). Among other things, Bizagi offers a personalized space and can be accessed via an API or a web service on external websites or portals. However, it can be assumed that the field of view of the embedded user interface remains limited by the design of the home page during execution, so immersion in the learning process is distracted by other information that may interfere with the field of view. Furthermore, when embedding Bizagi into a third-party system, it is necessary to run Bizagi as a web service or an application integrated into the homepage. For this integration, it is necessary to coordinate the data models and clearly define the corresponding instructions on the homepage via program code. Figure 8 (b) shows the chosen IT infrastructure in which the learning portal is executed separately from the home page [11]. The learning portal is accessed via a link on the homepage. As soon as the end-users want to complete a certified course, a new, independent program is opened. There is therefore no need to transfer data between the home page and the IMS.





Figure 8. Digital Process Automation for an automated modelling skills assessment system: (a) User interface of the work portal (here is an excerpt for module 1.1) [2]; (b) IT infrastructure with linking via a landing page



(b)

Figure 9. Results report of the automated model check with Solibri: (a) Solibri report button; (b) Results as Excel file

6 Results

The results show that Intelligent Process Automation (IPA) with its individually composable components can provide an automated evaluation system for the assessment of modeling skills. In this work, the DPA and RPA components have been migrated together to provide endusers with tasks via a DPA-based work portal, to demand solutions at a specified time and to display the learning progress already achieved to the end-users in real-time. Furthermore, the DPA was migrated with RPA technology to provide fully automated model checks in addition to the classic Excel evaluations. Especially concerning an always-on portal that enables end-users to have a customized learning experience regardless of time zones and individual scheduling, RPA technology is necessary to evaluate the submitted digital building models at any time. However, it turns out that the RPA bots can only provide their services without any problems if they use a recurring screen layout. If the UI elements [12] (elements of the graphical user interface) change, for example, due to a smaller scaled toolbar, the RPA bot cannot find the UI element to be accessed and the routine of the RPA bot leads to a system abort.





Figure 10. Excerpt of the RPA bot automation as a sequence flow: (a) UiPath Studio [2]; (b) Section of the sequence flow

In addition, Solibri issues system-related discrepancies that do not clearly define whether a non-conformity exists. For this error message, the end-user must decide whether the reported comment represents a modelling error. To exclude this room for interpretation, it is necessary to further simplify and concretize the tasks and their example models. For example, an insulation envelope in the foundation should be omitted, as there are different ways of modelling the insulation (e.g. in the case of intersections of abutting components).

The preferred solution is unattended RPA bots that are called up via an orchestrator, instantiated on a local server and triggered via a work portal. This would have the advantage that the software to be automated and installed locally always uses the same screen layout. Furthermore, the tasks and example models are iteratively adapted until the results output by Solibri no longer leaves any room for interpretation. As soon as this process, consisting of DPA and RPA, has been tested several times without errors, the DPA-based work portal is to be expanded to include additional components such as artificial intelligence or machine learning to intelligently check the plausibility of the results achieved based on independent decisions in line with IPA.

Acknowledgement

The results of this work are based on the results of a cooperative doctorate between the University of Wuppertal and the Jade University of Applied Sciences [1] as well as the third-party project "AUFLADEN" funded by Stiftung Innovation in der Hochschullehre [13].



Stiftung Innovation in der Hochschullehre

References

- C. Heins. Entwicklung eines Statikportals für die partizipative Bearbeitung von Tragwerksberechnungen im Prozess des Vieraugenprinzips (Planen – Prüfen). In *Dissertation*, University of Wuppertal, Germany, Submitted: 21/06/2023; Disputation: 11/12/2023.
- [2] C. Heins, T. Luhmann (ed.), and T. Sieberth (ed.). Intelligente Prozessautomatisierung im Weiterbildungs- und Bauplanungsprozess. In Photogrammetrie-Laserscanning-Optische 3D-Messtechnik – Beiträge der 21. Oldenburger

3D-Tage, Wichmann Verlag, Berlin/Offenbach, 2024.

- [3] C. Pütz, C. Heins, M. Helmus, and A. Meins-Becker. Gamification and BIM: Teaching the BIM method through a gamified, collaborative approach. *IS-ARC. 37th International Symposium on Automation and Robotics in Construction*, 37:272–277, 2020. doi:10.22260/ISARC2020/0039.
- [4] C. Heins, G. Grunwald, and M. Helmus. Gamification and BIM - The didactic guidance of decentralised interactions of a real-life BIM business game for higher education. *ISARC. 38th International Symposium on Automation and Robotics in Construction*, 38:932–939, 2021. doi:10.22260/ISARC2021/0126.
- [5] ISO 16739-1:2018. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries. On-line: https:// www.iso.org/standard/51622.html, Accessed: 04/12/2023.
- [6] Bizagi. Bizagi Digital Business Platform. User Guide. On-line: https://download.bizagi. com/docs/suite/1124/en/SuiteHelpEng.pdf, Accessed: 04/12/2023.
- [7] Microsoft Corporation. Microsoft Excel. Online: https://www.microsoft.com/de-de/ microsoft-365/excel, Accessed: 04/12/2023.
- [8] UiPath. UiPath Business Automation Platform. Online: https://www.uipath.com/de, Accessed: 04/12/2023.
- [9] NEMETSCHEK SE. Solibri. On-line: https:// www.solibri.com/, Accessed: 04/12/2023.
- [10] UiPath. Robot user manual. Online: https://docs.uipath.com/de/ robot/standalone/2023.4/user-guide/ attended-vs-unattended-robots, Accessed: 04/12/2023.
- [11] AUFLADEN. Website. On-line: https://www. wissen-aufladen.de/, Accessed: 04/12/2023.
- [12] UiPath. About UI Elements. On-line: https: //docs.uipath.com/de/studio/standalone/ 2023.4/user-guide/about-ui-elements, Accessed: 04/12/2023.
- [13] Stiftung Innovation in der Hochschullehre. Aus- und Fortbildungslabor für digitales Entwerfen und Bauen. On-line: https://stiftung-hochschullehre.de/ projekt/AUFLADEN/, Accessed: 04/12/2023.

Flat and Level Analysis Tool (FLAT) for real-time automated segmentation and analysis of concrete slab point clouds

Nolan W. Hayes¹, Bryan P. Maldonado¹, Mengjia Tang¹, and Diana Hun¹

¹Buildings and Transportation Science Division, Oak Ridge National Laboratory, United States of America * havesnw@ornl.gov, maldonadopbp@ornl.gov, tangm@ornl.gov, hunde@ornl.gov

Abstract -

In the United States, the flatness and levelness of concrete floors during construction is traditionally specified by a maximum allowable gap under a 3 meter straightedge. However, the straightedge method is inexact and rarely representative of the entire floor since the technician is free to choose any location on the floor to perform the measurement. In cases requiring a higher degree of precision and repeatability, concrete floor flatness and levelness can be measured using the standard test method ASTM E1155. With the recent introduction of advanced surveying instruments such as robotic theodolites and terrestrial laser scanners (TLS), the means now exist to modernize and expedite the measurement of floor flatness and levelness. This paper details the development and demonstration of a digital tool, named the Flat and Level Analysis Tool (FLAT), to automate and expedite the segmentation and analysis of flatness and levelness from dense point cloud data of concrete floor slabs. Segmentation algorithms were developed using unsupervised machine learning to extract the set of points belonging to the concrete floor slab from a full 360° scan of a construction site. After segmentation, automated analysis algorithms report the results according to the standard method. The developed algorithms were demonstrated on a dense point cloud captured from a concrete slab-on-grade at a construction site. Results show that the digital tool can quickly provide estimates for floor flatness and levelness with minimal human involvement with comparable accuracy to manual methods.

Keywords -

concrete floor, flatness, levelness, automation, real-time

1 Introduction

In the United States, the flatness and levelness of concrete floors during construction is traditionally specified by a maximum allowable gap under a 3 meter straightedge. For example, when a 3 meter straightedge is placed on the floor, a technician may measure a maximum distance of 6 mm between the floor and straightedge at any point for the floor to be considered acceptable construction quality (a specified flatness of 6 mm in 3 meters). However, the straightedge method is inexact and rarely representative of the entire floor since the technician is free to choose any location on the floor to perform the measurement. Additionally, this practice is not typically controlled using a standard method and results may significantly vary based on technician, tools, and methods.

In cases requiring a higher degree of precision and repeatability, concrete floor flatness and levelness can be measured using the standard test method ASTM E1155 [1]. This standard method specifies the means to extract and analyze data from a constructed floor to determine floor flatness F_F and floor levelness F_L numbers. Additionally, for concrete floors, ACI 302.1 specifies acceptable ranges for F_F and F_L [2]. Different building applications require varying degrees of F_F and F_L . For example, a warehouse with high-stacking shelves requires a high degree of both F_F and F_L to ensure a small risk of storage shelves overturning.

Even in the modern age of construction, the methods and equipment used to measure F_F and F_L are largely the same since the 1970s. With the recent introduction of advanced surveying instruments such as robotic theodolites and terrestrial laser scanners (TLS), the means now exist to modernize and expedite such measurements. For example, modern commercial software eases the analysis of 3D point clouds for flatness and levelness [3]. However, the process does not perform automated, real-time analysis. Instead, an experienced user is required to collect, analyze, and report the data. Fully automated procedures to analyze floor flatness and levelness would allow realtime assessment and lower the user skill requirement for assessment.

This paper details the development and demonstration of the Flat and Level Analysis Tool (FLAT), a digital tool to automate and expedite the segmentation and analysis of flatness and levelness from dense point cloud data of concrete floor slabs. Segmentation algorithms were developed

^{*}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).

using unsupervised ML to extract the set of points belonging to the concrete floor slab from a full 360 ° scan of a construction site. After segmentation, automated analysis algorithms report F_F and F_L according to ASTM E1155.

1.1 Floor flatness and levelness control

Floor flatness (F_F) and levelness (F_L) numbers are quantitative measures of a floor's smoothness and levelness [4]. In general, floor flatness is a measure of the local smoothness of the floor. F_F is a function of the change in floor elevation between colinear points spaced at 60 cm increments. The flatness of the floor affects the installation of flooring, ride quality, safety, and drainage. On the other hand, the floor levelness number is a global measure of the levelness of the floor. F_L is a function of the change in floor elevation between colinear points spaced at 3 m increments. The floor's levelness can impact drainage as well as the placement and design of shelves. Many types of buildings require a high degree of flatness and levelness such as warehouses with stacking shelves, ice rinks, and movie studios.

In the United States, the traditional method of consistent measurement of concrete floor flatness and levelness is the ASTM E1155 standard. This standard procedure consists of subdividing the floor slab into test sections, marking sample measurement lines, collecting measurements every 30 cm, and calculating F_F and F_L along each line, each aggregate test section, and the entire floor slab. ACI 302.1 specifies acceptable ranges for F_F and F_L based on the building use case. Specified values for F_F and F_L range from 15 to 50 for most applications. For example, where flatness and levelness are noncritical such as mechanical rooms and nonpublic areas, local values of 15 F_F and 10 F_L and overall values of 20 F_F and 15 F_L are acceptable. However, for applications requiring a high degree of flatness and levelness such as movie studios, local values of 35 F_F/F_L and overall values of 50 F_F/F_L are required at the minimum.

The procedure to gather data to calculate F_F and F_L is manually time-consuming. While the standard does allow the use of manual instruments such as straightedges, laser levels, and taut level wires to measure change in elevation, the fastest measurement tools allowed by the standard are inclinometers or profilometers. These instruments are "walked" or rolled along the surface of the slab to measure the change in elevation between two points spaced apart by 30 cm. Depending on the size and complexity of the slab, it may be necessary to collect 200 or more data points using these instruments. After data collection, analysis must be conducted to determine F_F and F_L from the acquired data. The entire process can take anywhere from 2 - 12 hours depending on the size and complexity of the concrete slab. In most cases, the process is not fast enough to identify issues in real-time to correct mistakes. Additionally, the measurement process itself cannot be completed on concrete before setting because the operator must stand on the concrete. Modern surveying instruments have the potential to complete the same data collection procedures in a fraction of the time and enable real-time validation while the concrete is still workable such that errors in flatness can be corrected more easily.

When issues with flatness and levelness are detected by traditional methods, the concrete has already hardened. Additional testing must often be conducted to determine where exactly the flatness or levelness issues are located. After locating the issues, remediation efforts typically consist of grinding, planing, surface repair, re-topping, or removal and replacement [4]. These efforts are costly and time-consuming to the concrete contractor. Additionally, the concrete placement contract often stipulates that if flatness and levelness specifications are not met on the first attempt, a reduction in payment will occur. Thus, concrete contractors have a major interest in identifying flatness and levelness issues in real-time such that costly mistakes can be corrected while the concrete is still workable.

1.2 Laser scanning and point cloud segmentation

Recent developments in advanced surveying instruments have yielded simple, easy-to-use laser scanning devices that can measure more than 2 million points per second at large distances. These modern instruments have the potential to significantly expedite the calculation of F_F and F_L for concrete slabs. Modern instruments are accurate at long ranges with peak range accuracy of 2 mm + 2 ppm or better. With continued development, these instruments are also becoming more affordable and accessible with easy-to-use interfaces. Additionally, most modern laser scanners come prepackaged with software to automate and simplify the process of registering multiple scans into a single frame; a process which can require several man-hours of effort from an experienced surveyor. As a result, laser scanners have the potential to easily replace handheld measurement devices such as the inclinometers and profilometers that are used to measure F_F and F_L .

However, one aspect of laser scanning that currently limits adoption is segmentation. Because scanners operate by spinning a mirror or aperture and rotating, the fastest method of scanning a site/object using a TLS is usually performing a full 360° scan. During this process, the scanner measures every object in sight. After completion, the operator must import the data into software to separate the points belonging to the object of interest (the concrete slab) from the rest of the points. This process is called segmentation. One of the major topics of research in construction automation is the automated segmentation of points clouds for varying applications [5, 6]. In particular, automated point cloud segmentation techniques exist to perform tasks ranging from automated digital twin generation [7] to prefabricated tolerance compliance assessment [8]. By automating the segmentation process using artificial intelligence, the process can be turned from a several-hour manual process into a few second-long automated process, increasing the accessibility of laser scanning technologies to a wider array of applications and users.

2 Methodology

The primary research of this work details the development of (1) an automated segmentation algorithm for concrete slabs and (2) an automated analysis algorithm to perform and report on F_F and F_L following the ASTM E1155 standard. Other research efforts have focused on the development of methods to automate the process of flatness and levelness quality control for concrete slabs [9, 10, 11, 12]. However, no work exists to the authors' knowledge that combines both model-free automatic segmentation and automatic analysis of F_F and F_L according to ASTM E1155. The proposed methods were developed to automatically segment a concrete slab with no user interaction from a 360° dense point cloud and generate F_F and F_L according to ASTM E1155.

2.1 Segmentation

The automated segmentation of points belonging to the concrete slab is necessary to automate the analysis. The segmentation process must fully isolate the points belonging to the slab from the remainder of the point cloud data. Concrete slabs can be separated into two types: slabs ongrade and suspended slabs. Concrete slabs on-grade are simply constructed to bear directly on foundations and the soil beneath. Suspended slabs are constructed in multistory buildings at elevations above the ground. Suspended slabs are significantly more difficult to construct in conformance with stringent flatness and levelness requirements. For both types of floor slabs, the slab is usually constructed before the walls which presents a specific segmentation issue. For full 360° scans, the entire construction site is included in the point cloud. For suspended slabs, the scanner will capture the floor on which it is set and the structure above in multi-story constructions. As a result, the automated algorithm must be able to identify what points belong to the floor slab to analyze with potentially multiple near-planar surfaces existing in the point cloud data.

The proposed segmentation algorithm consists of several steps to isolate the points belonging to the concrete slab. Figure 1 illustrates the procedure. Each step in the procedure will be further discussed in more detail.

The first step in the analysis is to reduce the data in the

point cloud so that the subsequent steps can be completed more quickly. The goal is to reduce the data by at least 95%; however, the actual reduction will depend on several factors including the original point density and scan area. Data reduction is achieved by voxel-downsampling [13]. The process of voxel-downsampling generates a grid of cubes of specified dimension. Within each cube, all points are averaged, and the resulting voxel is reduced to a single point located at the mean. For this work, a voxel size of 5 cm was determined to be effective in reducing the point cloud while still accurately providing enough data for the remaining steps in the procedure.

Next, plane segmentation is performed using a random sample consensus (RANSAC) procedure [13] with some modifications. This modified RANSAC is performed to identify the near-level plane of the concrete slab. The procedure iteratively selects three points in the cloud to generate a plane. Valid points must be below the origin of the point cloud (which is the location of the scanner) to ensure that the identified plane is the intended plane of the floor. A distance threshold specifies the normal distance from the plane that points are selected as inliers. A modification to the procedure eliminates planes that are not within a specified threshold of level (i.e. for the plane described by z = ax + by + c where z is elevation, the parameters a and b must be within a specified threshold close to zero). After a fixed number of iterations, the plane with the largest number of inlier points is returned as the plane representing the concrete slab.

The points identified by the modified RANSAC frequently include scattered miscellaneous points that do not belong to the concrete slab's point cloud. Such outliers belong to parts of the surrounding area that fall within the plane of the slab. Because the slab is often slightly elevated from the adjacent soil, a gap most likely exists between the point cloud of the slab and the other planar outlier points. A density-based scan (DBSCAN) clustering method [13, 14] is performed to identify the largest cluster of points which belong to the concrete slab.

After the slab is isolated, a principal component analysis (PCA) [15] is conducted to align the edges of the slab to the major axes. Before conducting PCA, the points are projected to the xy-plane, removing the elevation, (z) axis. The PCA determines the first primary component of the points. The points are then rotated by the angle between the first primary component vector and the x-axis, aligning the slab with the major axes.

After alignment, edge optimization is performed to cleanly delineate the edge of the slab from miscellaneous features. Although the standard procedure requires that measurements must be sufficiently far from the slab edges, this process is necessary due to the possibility of miscellaneous features near the slab edges which could affect the



Figure 1. Workflow for automated segmentation of concrete slab.

automated analysis algorithm if not removed. The edge optimization is completed in the *xy*-plane. The median (*M*) and median absolute deviation (MAD) are calculated for all points projected onto the *xy*-plane. All points *p* are normalized according to n = (p - M)/MAD. Consider the probability density function (PDF) of the normalized points on each axis, then the inlier points satisfy $|PDF(n_x)| < \alpha$ and $|PDF(n_y)| < \beta$, where α and β are specified thresholds. A bounding box is generated for the inlier points. Points falling outside of the bounding box are not considered part of the automated analysis, as shown in Figure 2.



Figure 2. Edge optimization for trimming miscellaneous points.

The bounding box generated from edge optimization is used to crop the slab from the original point set before voxel downsampling, rotated by the same angle determined by the PCA. The points inside the bounding box are the resulting point cloud which includes only the isolated points belonging to the concrete slab. It is important to note that the bounding box is used to crop the original, raw point cloud before voxel-downsampling was performed which retains the original accuracy of the collected point cloud. Finally, the origin is set to the most negative point within the cloud of the isolated concrete slab. After this step, the resulting point cloud is ready for the automated analysis.

2.2 Analysis

After segmenting the point cloud to yield only the points belonging to the concrete slab, an analysis can be performed to automatically generate F_F and F_L according to ASTM E1155. As part of the measurement procedure of ASTM E1155, a series of lines are established from which to measure elevation changes every 12 inches (≈ 0.3 m).

There are several rules for conducting the setup of measurement lines according to the standard method. Some general requirements apply to the overall procedure. No portion of the entire test surface can be associated with more than one test section, and no test section boundary should cross any construction joint. Other standard requirements are described where appropriate during the algorithm descriptions.

Because the points belonging to the slab are segmented in the prior step, many of these criteria can be tested as part of the algorithm. First, the minimum and maximum bounds of the concrete slab are extracted from the point cloud by the minimal oriented bounding box. From these bounds, the slab dimensions are extracted, and the slab area can be calculated. The slab dimensions and area are checked against the approval criteria of ASTM E1155 (i.e., the test section cannot measure less than 8 ft (2.4 meters) on a side and the test section cannot have an area less than $320 \text{ ft}^2 (29.7 \text{ m}^2)$. If the slab is large enough, the algorithm is allowed to proceed.

The standard procedure also requires that no part of any sample measurement line fall within 2 ft (≈ 0.61 m) of any slab boundary, construction joint, isolation joint, block-out joint, penetration, or other similar discontinuity. However, two exceptions are provided. First, shrinkage crack control joints formed by either partial depth sawcuts or by partial depth inserts can be ignored. Second, if the area to be excluded from the measurement exceeds 25% of the test section area, then the 2 ft (≈ 0.61 m) boundary exclusion does not apply. The boundary exclusion area is calculated and tested to determine if it exceeds 25% of the test area. If the boundary exclusion area is not too large, then the allowable sample measurement area is inset within the slab perimeter; otherwise, the entire surface area of the slab is treated as the allowable sample measurement area.

The standard methods provide conditions for sample measurement lines within each test section. Each line must be arranged to blind the test results by using one of two methods: (1) orienting all lines at 45° to the longest construction joint abutting the test section, or (2) placing equal numbers of lines of equal aggregate length both parallel to and perpendicular to the longest test section boundary. However, when the short dimension of the slab being measured is less than 25 ft (7.62 m), all measurement lines must be 45° diagonals. Sample measurement line generation criteria allow lines to be placed at $\pm 45^{\circ}$ to the longest construction joint regardless of slab dimensions. For this reason, the automated analysis procedure will always generate measurement lines at 45° diagonal to the longest dimension of the slab. As a reminder, the most negative point on the slab was set as the origin for the local slab coordinate system before automated analysis. Because the largest slab dimension is aligned with the x-axis as part of the automated segmentation, each measurement line can be described by the slope-intercept formula, y = mx + b, where x, y are the coordinates of a point along the line, m is the slope which is either -1or 1, and b is the y-intercept which will vary in uniform spacing to create lines that span across the entirety of the allowable sample measurement area.

A series of lines are constructed to fill the area that is generated from doubling the slab width and height. Points are generated along each line, spaced at 12 inches (30.48 cm). Adjacent parallel lines are separated by a perpendicular distance of 4 ft (1.22 m) as required by the standard procedure. The construction of these lines and measurement points are illustrated in Figure 3. Lines are constructed at angles of -45° and 45° from the x-axis (which is aligned with the longest dimension of the slab). Lines are trimmed such that all points fall within the allowable sample measurement area. The length of each line is calculated by determining the Euclidean distance between the start and end points. The standard procedure requires that no sample measurement line measure less than 11 ft (3.35 m). If any line is shorter than the allowable 11 ft (3.35 m), then the line is removed from the sample set. Additionally, at the end of this step, the total number of sample measurement points is determined according to the standard calculations. The standard method details that a minimum number of readings is required per test section conforming to the following equations: $N_{\min} = 2\sqrt{A}$ for $320 \le A \le 1600$ or $N_{\min} = A/30$ for A > 1600 where A is the test section area in square feet. If N_{\min} is larger than the actual number of sampled points, the test results will be deemed invalid and not reported; otherwise, the algorithm is allowed to proceed.



Figure 3. Automated sample generation.

After sample line generation, the next step in the algorithm is to iterate over each point within each line to determine the closest point in the slab point cloud from which to extract the height measurement. First, a kd-tree is generated for the slab point cloud projected to the xyplane [16]. The kd-tree is queried with each point along a sample measurement line, also projected to the xy-plane, to determine the nearest neighbor within the slab point cloud ignoring the elevation (*z*-axis). The elevation, *z*component, of the nearest neighbor in the slab point cloud is recorded as the measured sample height of the queried point of the sample measurement line. This process is repeated iteratively for each point along each sample measurement line. The result is the recorded elevations of the slab every 12 inches (30.48 cm) along the sample measurement lines. After the collection of slab elevations along each measurement line, the analyses detailed by ASTM E1155 [1] are performed to determine F_F and F_L for each sample measurement line and the overall slab (hereby referred to as composite). In summary, floor flatness is a function of the standard deviation of curvature of difference in elevation between points separated by 24 inches (60.96 cm). Floor levelness is a function of the standard deviation of curvature of difference in elevation between points separated by 10 ft (3.048 m). Composite floor flatness and composite floor levelness numbers are calculated by iteratively combining floor flatness or levelness numbers for each sample measurement line weighted by the number of samples. The reader is referred to the ASTM E1155 standard for complete details on how F_F and F_L are calculated [1].

2.3 Limitations

As is, the automated segmentation and analysis algorithms include some limitations. First, both algorithms assume that the concrete slab is rectangular. For complex shapes beyond rectangles, the edge optimization process fails to cleanly extract the edge of the slab. Additionally, the automated segmentation algorithm assumes that the slab is subdivided into a single test area. If construction joints are present, according to ASTM E1155, the slab must be subdivided into multiple test areas. This subdivision process is currently not automated within the algorithm. Moreover, if the slab includes multiple elevations (there are steps in elevation), then the automated analysis will only identify a single elevation. In future work, these limitations are planned to be addressed to allow the automated segmentation to analyze multiple elevations and multiple combinations of rectangular slab areas.

3 Demonstration

The automated segmentation and analysis algorithms were tested on an actual concrete slab located on the Oak Ridge National Laboratory site. A concrete slab-on-grade with design dimensions of 80 ft (24.4 m) by 50 ft (15.2 m) was scanned within 7 days of concrete placement. The slab was designed to be noncritical in terms of flatness and levelness requirements; as a result, flatness and levelness numbers were expected to be on the low end. A Leica MS60 was placed near the center of the concrete slab and set to scan the entire surrounding area using a full-dome scan setting with a maximum point resolution of 3 mm at the furthest edge of the slab. All scanned points within the perimeter of the concrete slab had a point spacing of less than 3 mm. The laser scan included the full 360° view of the construction site. Figure 4 shows the raw .pcd file of

the scan that was loaded into the FLAT digital tool.



Figure 4. A portion of the full-dome scan of the construction site with concrete slab.

The automated segmentation and analysis algorithms of FLAT were written in Python. The automated procedures use several input parameters to perform various functions. Table 1 shows the parameters used in the demonstration. Voxel Size is the dimension of the voxel cube used for downsampling. The Plane Distance Threshold is the minimum normal distance between the RANSAC plane and points to be considered as inliers. Cluster Neighbor Radius is the maximum distance between points to be considered members of the same cluster. Cluster Minimum Points is the minimum number of points that a cluster must contain. Because only the largest cluster is taken, this parameter has minimal effect on the overall algorithm. Edge Optimization Bin Size controls the size of bins used to generate the PDF (n_x) and PDF (n_y) for edge optimization. The peak accuracy of slab dimensions is the same as this specified parameter. The Edge Optimization Threshold is the minimum threshold (α and β) for the probability density function for a bin to be considered an inlier (part of the slab).

Table 1. Parameters for the demo	<u>nstratio</u> n.
Parameter	Value
Voxel Size	5 cm
Plane Distance Threshold	5 cm
Cluster Neighbor Radius	30 cm
Cluster Minimum Points	50
Edge Optimization Bin Size	3 mm
Edge Optimization Threshold	0.10

One proposed feature of the FLAT digital tool is the real-time evaluation of floor flatness and levelness which enables measurement and correction of flatness and levelness issues while concrete is still workable. For the MS60 used in this study, the single-view dense point cloud with

Table 2. Run times. 100 iterations.						
Taale	Run Times (sec)					
Task	Min	Max	Mean	Std		
Downsample	1.03	1.60	1.17	0.11		
Segment Plane	0.07	47.31	9.95	10.48		
Cluster	0.92	1.71	1.17	0.11		
Align	0.20	0.27	0.22	0.02		
Optimize Edge	0.03	0.06	0.04	0.01		
Crop	2.11	2.91	2.35	0.17		
Generate	0.13	0.31	0.16	0.02		
Measure	5.47	8.22	6.08	0.44		
Calculate	0.001	0.02	0.002	0.002		
Entire Procedure	10.63	58.21	21.13	10.51		

a specified resolution was collected in approximately 30 minutes. However, many TLS boast faster scanning rates and can capture point cloud data with higher resolution in a fraction of the time. Initial running time targets for the automated segmentation and analysis of FLAT were proposed at 60 seconds per 1000 m² of slab area. Table 2 shows the running times for each task of the FLAT digital tool on a standard laptop computer. Tasks of the automated segmentation process include, in order, Downsample by voxel downsampling, Segment Plane by RANSAC level plane segmentation, Cluster by density-based scan (DB-SCAN) clustering, Align by principal component analysis (PCA), Optimize Edge by the proposed method, and Crop the original point cloud using the minimal oriented bounding box determined from the algorithm. Tasks of the automated analysis process include, in order, Generate sample measurement lines according to ASTM E1155, Measure the elevation of the nearest neighbor point in the cloud to sample measurements, and Calculate floor flatness and levelness according to ASTM E1155.

The mean total run time was approximately 21 seconds. Considering the surface area of the slab (370.88 m²), the resulting average run time is 56.6 seconds per 1000 m² of slab area. Most tasks within the algorithms have consistent run time except for RANSAC. This variability in execution time is expected due to the random nature of the procedure. One option to reduce the time of this procedure, which may be introduced in the future, is to change to plane fitting of three user-specified points. Although this change could reduce processing time variability and improve repeatability, the resulting algorithm would have vastly reduced autonomy.

The output of the algorithms was also assessed. Table 3 shows the output of algorithms for slab dimensions, composite flatness, and composite levelness for 100 iterations of the algorithm on the same dense point cloud. Composite flatness and composite levelness refer to the calculated value by joining all flatness and levelness numbers for all sample measurement lines. Of the 100 iterations, 5 yielded

Table 3. Reported	output.	100	iterations.
-------------------	---------	-----	-------------

Outrast	Value				
Output	Min	Max	Mean	Std	
Slab Length (m)	24.258	24.553	24.419	0.025	
Slab Width (m)	15.283	15.563	15.304	0.048	
Composite F_F	5.3	8.2	7.0	0.61	
Composite F_L	12.3	13.5	13.1	0.27	

segmentation results that incorrectly identified the plane of the concrete slab surface resulting in a failure rate of 5%. The results of failed runs were not included in the statistical analysis of the output results.

As shown among the 100 iterations, there is some variability in the output results. The processes within the algorithm that could induce the most variability include plane segmentation, clustering, and alignment because they use unsupervised machine learning methods which can produce differing results between iterations. Since these processes are stacked upon each other, the variability increases further. For example, for the slab dimensions, the range of extracted values is approximately 25 centimeters. Although this value is a small percentage of the overall dimensions, the variability in dimensional output is too large to accurately assess the slab's dimension with a single execution of the algorithm. Slab dimensional tolerances are commonly 0.75 in (19 mm) according to ACI 117-10 [17]. Ideally, the algorithm should produce results with variation less than this tolerance.

Additionally, there is variability in the output for floor flatness and floor levelness. Floor flatness numbers ranged from 5.3 to 8.2. This variability is hypothesized to occur within the plane segmentation, clustering, and edge optimization procedures. An intermediate step after plane segmentation where the user confirms that the slab has been correctly identified with near-perpendicular corners could help minimize variability between iterations. In future work, the source of variability will be addressed to reduce the range of output.

To determine the true values for composite floor flatness and floor levelness, ASTM E1155 was manually completed by extracting point elevations along sample measurement lines from the point cloud data gathered from the laser scan. Manually created sample measurement lines were created at similar locations as sample measurement lines automatically generated by the analysis algorithm. This manual analysis yielded a composite floor flatness of 7.6 and composite floor levelness of 14.1. In future work, the deviations between manual and automated method results will be investigated. The total run time of all 100 iterations was 35 minutes. Considering that 100 iterations of the analysis take 35 minutes, which is a significantly shorter amount of time compared to the traditional method, it is still reasonable for the digital tool to reduce the amount of time required to determine floor flatness and levelness with a high degree of accuracy. Future work will include a validation study to compare the digital tool output with conventional methods to perform ASTM E1155.

3.1 Discussion

A major finding of this study is that ASTM E1155 has several major shortcomings compared to modern surveying methods. Although the standard method has been clearly established for some time, it is not exhaustive even when optimized using automated procedures. The standard procedure limits the placement of sample measurement lines so that adjacent parallel lines are not closer than 4 ft (1.22 m). The purpose of this limitation in the standard is unclear to the authors and excludes a significant portion of the slab surface area from measurement, potentially hiding flatness defects. Modern surveying instruments can quickly capture sample elevations at high density (3 mm or more). A better alternative is to generate a topographic map of the slab topography to highlight the degree and locations of imperfections in flatness.

4 Conclusions and next steps

A Flat and Level Analysis Tool (FLAT) was developed to automate the process of determining floor flatness and levelness for concrete slabs according to ASTM E1155. An automated segmentation algorithm and automated analysis algorithm were developed and tested on a demonstration concrete slab on grade. Results of the demonstration of the algorithms show that the digital tool can quickly provide estimates for floor flatness and levelness with minimal human involvement.

In future work, FLAT will be optimized to minimize the variability of output results to ensure consistent output of slab dimensions and floor flatness/levelness numbers. Additionally, the automated segmentation limitations on slab shapes, dimensions, and complexity will be addressed to extend applicability to more types of concrete floors. The output results of FLAT must also be compared with standard output from conventional methods to perform ASTM E1155. Ultimately, additional point cloud data of different slabs must be collected to test the algorithms.

5 Acknowledgements

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-accessplan). This research was supported by the DOE Office of Energy Efficiency and Renewable Energy, Building Technologies Office, under the guidance of Sven Mumme, and used resources at the Building Technologies Research and Integration Center, a DOE-EERE User Facility at Oak Ridge National Laboratory.

References

- [1] ASTM E1155-20. Standard Test Method for Determining F_F Floor Flatness and F_L Floor Levelness Numbers. Standard, American Society for Testing and Materials, 2020.
- [2] ACI PRC-302.1-15. Guide to Concrete Floor and Slab Construction. Standard, American Concrete Institute, 2015.
- [3] Faro. Floor Flatness and Level ASTM E1155 with BuildIT Construction. Online: https://knowledge.faro.com/Software/ BuildIT/BuildIT_Construction/Floor_ Flatness_and_Level_ASTM_E1155_with_ BuildIT_Construction. Accessed: 11/03/2024.
- [4] Mark A. Cheek. The Floor Flatness Report. Technical report, American Concrete Institute, 2011.
- [5] Min-Koo Kim et al. Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3d point cloud data. *Automation in Construction*, 112:103077, 2020.
- [6] Mads A Kuhlmann-Jørgensen et al. Plastr: Planning for autonomous sampling-based trowelling. *IEEE Robotics and Automation Letters*, 8(8):5069–5076, 2023.
- [7] Bryan P. Maldonado et al. Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) using Machine Learning. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 48–55, Chennai, India, 2023.
- [8] Pingbo Tang et al. Automatic reconstruction of as-built building information models from laserscanned point clouds: A review of related techniques. *Automation in Construction*, 19(7):829–843, 2010.
- [9] Zhipeng Chen et al. Rapid inspection of large concrete floor flatness using wheeled robot with aidedins. *Remote Sensing*, 14(7):1528, 2022.
- [10] Dongsheng Li, Jiepeng Liu, et al. Terrestrial laser scanning assisted flatness quality assessment for two

different types of concrete surfaces. *Measurement*, 154:107436, 2020.

- [11] Frédéric Bosché and Emeline Guenet. Automating surface flatness control using terrestrial laser scanning and building information models. *Automation in construction*, 44:212–226, 2014.
- [12] Pingbo Tang et al. Characterization of three algorithms for detecting surface flatness defects from dense point clouds. In *Three-Dimensional Imaging Metrology*, volume 7239, pages 197–208. SPIE, 2009.
- [13] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018.
- [14] Martin Ester et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231, 1996.
- [15] F. Pedregosa, G. Varoquaux, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [16] Songrit Maneewongvatana and David M. Mount. Analysis of approximate nearest neighbor searching with clustered point sets. *CoRR*, cs.CG/9901013, 1999.
- [17] ACI 117-10. Specification for Tolerance for Concrete Construction and Materials. Standard, American Concrete Institute, 2010.

Evaluation of Mapping Computer Vision Segmentation from Reality Capture to Schedule Activities for Construction Monitoring in the Absence of Detailed BIM

Juan D. Núñez-Morales^{1,3,†} and Yoonhwa Jung^{1,3,†} and Mani Golparvar-Fard^{2,3}

¹Department of Civil and Environmental Eng., and Computer Science

²Department of Civil and Environmental Eng., Computer Science, and Technology Entrepreneurship

³University of Illinois at Urbana-Champaign, USA

[†] These authors contributed equally.

{jdnunez2, yoonhwa2, mgolpar}@illinois.edu

Abstract -

Over the past few years, research has focused on leveraging computer vision in construction progress monitoring, particularly in comparing construction photologs to Building Information Modeling (BIM), with or without schedule data. The practical application of these techniques and a large number of startups that have brought hyper AI and human-in-the-loop services around progress monitoring have revealed several gaps: 1) Current BIM-driven projects do not have model disciplines at the right level of maturity and Level of Development; 2) definitions of states of work-in-progress that are detectable from images are not formalized; 3) poor schedule quality and lack of frequent progress update challenges the incorporation of detailed 4D BIM for progress tracking. Such gaps are addressed in this work by exploring the requirements for mapping modern computer vision techniques for object segmentation with construction schedule activities to automate progress monitoring applications using computer vision without BIM as a baseline. The approach utilizes reality mapping practices to offer time machines for construction progress, organizing photologs over space and time. Additionally, this work shows how Large Language Models can structure schedule activity descriptions around <Uniformat Object Classification, Location>, focusing on how vision and language models can be trained separately with limited annotated data. ASTM Uniformat classification is utilized to map triangulated object segments from images to color-coded 3D point clouds aligned with schedule activities without the need for image and language feature alignments. Exemplary results on tied new transformer-based models with few-shot learning are shown, and the requirements for full-scale implementation are discussed.

Keywords -

Automated Progress Monitoring; Artificial Intelligence; Computer Vision; Natural Language Processing

1 Introduction

The field of construction progress monitoring has witnessed significant advancements in recent years, primarily driven by the integration of computer vision techniques and BIM. Specifically, leveraging devices such as LiDAR (Light Detection and Ranging), 360-degree cameras, and drones, a prevalent approach involves comparing generated as-built point clouds with the as-planned BIM model to assess construction progress by identifying deviations or discrepancies [1]. The integration between computer vision and BIM has enabled stakeholders to gain comprehensive insights into the construction process, facilitating improved decision-making, resource allocation, and offering real-time data analysis and project visualization [2].

However, despite these advancements, several critical gaps persist, posing challenges to the effective implementation of computer vision-based progress monitoring applications:

- Lack of BIM model disciplines at the right level of maturity and Level of Development (LoD): The current computer vision-based progress monitoring relies heavily on the quality and completeness of BIM models. Insufficient BIM LoD and maturity in model disciplines, particularly around work zones and ASTM Codes, create discrepancies when attempting to align as-built point clouds with the as-planned BIM model [3, 4].
- The absence of formalized definitions for states of work-in-progress detectable from images: The lack of formalized definitions for work-in-progress states detectable from images poses challenges, especially in visually complex indoor environments with occlusion complexities like wall layers. [5, 6]. Establishing clear and universally accepted definitions for construction states is crucial for enhancing the accuracy and comparability of progress assessments in diverse project environments.
- Poor schedule quality and lack of frequent progress

update: The effectiveness of 4D BIM, integrating the temporal dimension, relies heavily on schedule quality and BIM LoD. However, the suboptimal schedule quality, varying schedule granularity, and infrequent progress updates impede the alignment of real-time progress with the as-planned 4D BIM model [7].

This paper addresses these gaps by investigating the requirements for mapping modern computer vision techniques, specifically object segmentation, to construction schedule activities. The focus is on automating progress monitoring applications using computer vision as a standalone tool without relying on BIM as a baseline. To achieve this, this work proposes leveraging reality mapping practices that effectively organize photologs over space and time, providing time machines for construction progress.

Furthermore, this work explores how Large Language Models (LLMs) can play a pivotal role in structuring schedule activity descriptions, emphasizing key elements such as Action, Location with an emphasis on utilizing the ASTM Uniformat classification to map triangulated object segments from images to space. Exemplary results from a novel transformer model, coupled with few-shot learning, are presented, and the paper concludes by discussing the requirements for the full-scale implementation of these proposed methodologies. Identified gaps and specific challenges and solutions are covered in detail in the following sections.

2 Literature Review

This section reviews the current state-of-the-art in construction for the creation of semantically rich point clouds, and how both visual and text feature information have been automatically aligned in different fronts of computer vision. Additionally, this section covers the main strategies widely utilized for mapping progress against schedules, specifically focusing on the challenges presented when aligning progress information against construction schedule documents. Lastly, BIM-based progress extraction methods are explored and further discussed as a function of their advantages and challenges.

2.1 Point cloud generation and segmentation

Point clouds, obtained via 3D laser scanning or similar technologies, consist of data points in a three-dimensional coordinate system. LiDAR, using laser beams, accurately generates detailed 3D representations of structures and environments, capturing geometry and spatial information at construction sites. Photogrammetry, another method, extracts 3D details from 2D images taken by cameras or drones. The process involves estimating camera parame-

ters and correlating pixels between cameras to triangulate depth information [8, 9].

After creating a 3D point cloud, machine learning algorithms are used to classify individual points into categories such as building components and Mechanical-Electrical-Plumbing (MEP) systems[10]. However, existing detection systems often depend on complete object visualizations, making them sensitive to occlusions and capture completeness. In the context of photogrammetry-based point cloud reconstruction, classification to correlate 2D information with the 3D point cloud is achieved through numerous novel image-based classification and segmentation techniques, only using 2D features.

For example, [11] employs traditional vision-based algorithms using shape and color information to infer indoor partition wall progress. Others adopt deeplearning approaches like MaskRCNN [12], YOLO [13], and Transformer-based Swin [14] to directly detect construction elements from 2D images [15, 16]. Despite some photogrammetry approaches mapping camera pixels to real-world coordinates, the projection of semantic information from 2D inputs onto point clouds remains underexplored [16].

2.2 Image and Text Information Mapping

In the realm of Computer Science, significant progress has been made in mapping information between images and text as multimodal learning. Notable developments include pre-trained vision-language models like CLIP [17] and mPLUG [18], which demonstrate the capability to understand and associate textual descriptions with visual content. However, construction projects involve unique terminologies, specialized jargon, and context-specific information that may not be adequately addressed by generic image-text mapping models. Addressing these limitations is essential for creating effective tools that can assist professionals in the construction industry, providing them with streamlined access to information and enhancing communication between textual project descriptions and visual project elements.

2.3 Progress Against Schedules in Construction

Proactive construction workflows rely on critical schedule updates. Researchers have explored automated methods for optimizing schedules based on underlying reasons of sequencing logic [19], physical building component relationships [20, 21], aligning different levels of schedules [22], and ensuring consistency with BIM, schedule, and payment applications via ASTM Uniformat classifications [7]. These approaches often involve natural language processing and machine learning algorithms. However, the usefulness of the ASTM Uniformat II to properly report visual progress information against construction schedules is often ignored and poorly explored.

2.4 BIM-based Progress Monitoring Applications

Traditionally, progress monitoring has been based on comparing documented reality against plans, and through the standardization of the usage of BIM, numerous research have explored this data format for calculating construction progress. This is seen in BIM methods to drive Earned Value Analysis (EVA) [23] through geometry and time-based heuristics. Similarly, other works have relied on geometry information from BIM models to compute progress when compared against photogrammetry-based point clouds and proximity-based heuristics [3, 4], eventually incorporating 4D BIM [24] to determine the presence of built elements over time. More advanced implementations eventually considered detecting semantics from images to increase the correct detection of construction elements against BIM elements [25, 26].

While these lines of work pioneered one of the great avenues of progress monitoring using BIM, they were exposed to dependencies on the BIM LOD, leaving a wide gap to properly comparing reality against plan given the lack of 4D adoption that could connect progress against construction documents and schedules or the incorrect comparison between visible elements without semantics in a point cloud cluster and their intersection against BIM elements.

3 Method

The proposed methodology addresses one of the applications proposed in [27]. This method consists of a four-step approach model ensemble to create connections between vision and language features. At a high level, this is achieved by aligning orthographic projections of semantically rich photogrammetry point clouds against quantity-take-off (QTO) construction drawings to detect completed quantities of each classification of construction object classes based on the ASTM Uniformat II. Simultaneously, this method leverages corresponding logical sequences from a construction schedule to determine the completeness constraints of overlapping construction classes as part of the sequence logic, constrained to a specific location, generally denoted in a construction drawing. Such imposed logical constraint allows for the correction of observed quantities in segmented point clouds due to occlusion and point cloud completeness issues, which are present in most photogrammetry-based point cloud reconstructions. Figure 1 presents a graphical representation of the description presented above.

The first step leverages modern computer vision Structure-from-Motion (SfM) [8] and Multi-View Stereo (MVS) [9] algorithms to create unstructured threedimensional as-built representations of the construction environment from video capture frames or cameras C_i . At this stage, and specifically during the depth estimation for each pixel P_j of a registered and localized camera C_i , mappings between pixel and three-dimensional point coordinates are captured using camera matrix transformations M_i for each camera C_i . For a more in-depth understanding of the employed SfM with MVS approaches, readers are recommended to read the work shown in [28].

In parallel, using the image inputs from the first step, the second step leverages a few-shot trained Swin [14] transformer model, pretrained with synthetic data to create perpixel semantic classes or segmentations. The model training step considers a class structure based on the ASTM Uniformat classification of construction objects to detect partial construction of different construction elements. A semantic label class L_k is stored for each pixel P_i in camera C_i , and added to a general dictionary D containing tuples $D = [C_i, P_j, L_k, point_{x,y,z}]$. Such a dictionary is used to create a semantically rich segmented point cloud based on ASTM Uniformat II classes for each visible object. By leveraging camera vectors and orientations, an orthographic projection O is automatically created and overlayed against corresponding IFC drawings with QTOs using a three-point aligning process.

Concurrently, during the third step, corresponding schedules are parsed and classified Uniformat Level 2 & 3 instances using UniformatBridge [7], built on the pretrained BERT model. At this stage, for each activity line item Act_m , similar ASTM Uniformat classification labels L_k as assigned, together with location-based constraints Loc_n based on the schedule text usign PoAT [29]. At this stage, using the schedule activity relationships, sequential constraints are extracted for each detected Uniformat classification (i.e., [*Floor* > *Wall* > *Windows* > *Ceiling*]). These extracted sequences provide the logical constraints in which overlapping detected activity class orthographic projections from a segmented point cloud.

The fourth step attempts to create a polygon-based completion coverage to extract current object quantities. For each polygon Pol_u from a QTO drawing containing construction class label L_k , its completeness is evaluated using the overlapping segmented pixel class from the aligned orthographic point cloud projection as a function of the percent complete of an object and its corresponding activity, as shown in equations 1 and 2, respectively:

$$\label{eq:complete_Objm} \begin{split} & & \\ & & \\ & Drawing_{Pol_u} \cap PointCloud_{Pol_u} \qquad (1) \\ & & \\ &$$



Figure 1. The proposed methodology of this work for automating progress monitoring using Images, IFC Drawings with QTOs, and construction schedules.

$$\%Complete_{Act_m} = \%Complete_{Obj_m} \in Loc_n |Act_{m,Loc_n} = Drawing_{Pol_u,L_k}$$
(2)

Where $\[mathcal{Complete}_{Obj_m}\]$ is the estimated percentage completion of a construction object, $Drawing_{Pol_u}\]$ is the IFC Drawing polygon object, and $\[mathcal{Complete}_{Act_m}\]$ is the estimated percentage completion of an activity Act_m , upon constraining quantities to a location polygon Loc_n provided by the activity classification step, and extracting locations from the IFC Drawing.

As part of the logical constraint imposed by the extracted sequence in the second step of this method, quantities of initial objects in such sequence are corrected to reflect their completion imposed by constructability constraints of subsequent dependent objects (i.e., floor slabs quantities are corrected as a function of the detected ceiling quantities), as defined in the expression of equation 3, and the rules imposed by equation 4:

$$Seq = Obj_a, Obj_b, Obj_c$$
 (3)

$$\% Complete_{Obj_a} =$$

$$100\% |\% Complete_{Obj_b} > \% Complete_{Obj_a}$$
(4)

4 Data and Experiment Settings

The experiment of this work evaluates a real-world scenario in the construction of a high-rise hotel building. It considers the utilization of an IFC drawing with computed quantities, a construction schedule, and a set of images encompassing five different ASTM Uniformat II classifications corresponding to floors, partition walls, windows, ceilings, and pipe installation.

For the creation of dense reality point clouds, this work makes use of the OpenSfM library [30] and a set of 120 images corresponding to a room area of the construction to evaluate, as shown in Figure 2. The available IFC drawings are manually scaled and annotated based on the Uniformat II object classifications and stored as a rasterized document, as shown in Figure 3.

Moreover, the detection of construction elements makes use of a Swin model, pretrained with synthetically created construction scene images, following the training strategies from work presented in [31]. For visualization purposes, the projected RGB values of the point cloud semantics are set to differ from those from the IFC drawing.

Given the limited visibility of information when comparing the orthographic projection of a point cloud against the rasterized IFC drawing, different elevations are considered to detect and measure objects properly. Specifically, three elevation values – each at one-third (1/3) incremental of the total height – are selected to extract key orthographic projections from the point cloud to compare and estimate detected quantities. To account for elements that may be detected but outside of the designated point cloud orthographic projection height slice, a sampling threshold of one-sixth (1/6) in the vertical direction is utilized, as shown in Figure 4.

To extract schedule sequential information, a pretrained



Figure 2. Sample view of the point cloud reconstruction results. As made evident by using photogrammetry in construction environments, common challenges include point cloud completion due to homogeneous surfaces and object occlusions due to non-construction-related elements. The current reconstructed elements in this view correspond to wall aluminum framing, windows, gypsum board on ceilings, and concrete floors.



Figure 3. An a) Issued for Construction (IFC) drawing, along with b) polygon-based annotations for computing construction object quantities (QTYs). The scope of the experiment focuses on room areas marked with a hash pattern.

BERT model following the work proposed in [7] is used. A total of 1,700 construction activities are parsed and classified as a function of levels two and three of the ASTM Uniformat II. Specifically for this experiment, the scope of activities is focused on activities present in the room shown in Figure 3, focusing on the scope of structural and interior work. In addition to such classification, and to account for the mapping between a visually detected object and its correct activity line item, location information is extracted following the work presented in [29] and compared against the matched segmented point cloud segments, whose location information is extracted from the overlapping annotated IFC drawing. Additionally, Uniformat II sequences are created based on the extracted activity relationships present in the schedule and stored as separate recipes.

Lastly, the automatically detected progress quantities

are evaluated against ground-truth actual progress quantities, estimated from the project visual documentation and available daily construction reports. The accuracy of the detected progress quantities using the proposed method is evaluated using the Mean Average Percentage Error (MAPE) and reported against each schedule line item.

5 Result and Discussion

The following section focuses on the results pertaining to the computation of actual quantities based on the comparison between actual and planned orthographic projections, the mapping accuracy against schedule line items via using Uniformat and Location information, and the ability to correct detected quantities as a function of sequential information from a schedule's Uniformat classifications and relationships.

As evident in Figure 5, the sparsity of the mesh in a point cloud may create inaccurate estimates of progress values. Table 1 showcases how each detection is assigned to the corresponding activity Uniformat code and location, the results of comparing the estimated quantities based on comparing a segmented mesh against the IFC drawing, and the corrected completion estimates for each entry.

This case study shows the success of utilizing the ASTM Uniformat II as the bridge to align schedules against detected vision information. The contrast provided between the results in Figure 5 and Table 1 shows how elements with large pixel coverage (such as walls, floors, and ceilings) contain more pixel-level information that allows for dense point cloud reconstruction, improved feature segmentation, and better heuristically-controlled calculations due to corresponding to middle-sequence tasks, which can make use of a predecessor and successor thresholds for progress-based dependencies against other completed or to-be-completed elements.

In contrast, having significant occlusions denotes how the precision of the reality capture is less significant than the strategies utilized for correcting or assuming completed quantities. Such a case becomes especially challenging for slender objects such as MEP components, thus explaining the higher degrees of error in the experiment. This is mainly attributed to the lower number of pixels available for reconstructions and feature segmentation. With fewer pixels, a reconstruction engine may produce non-dense point clouds, decreasing the degree of label projection against the point cloud. Moreover, fewer pixels decrease the success rate of correctly classifying pixel features against predetermined classes. Lastly, due to MEP components being at the very last step of a sequence, the rule-based checking of the quantities becomes less impactful compared to the middle steps of a sequence.

Nevertheless, occlusions are a phenomenon that would similarly impact LiDAR-based scanning strategies. Still,



Figure 4. Pixel-based mapped against the orthographic representation of a point cloud at three different heights, where case a) shows the bottom projection, b) shows the middle height projection, and c) the top projection. These projections are scaled to the mesh representation of the resulting point cloud.



Figure 5. Initial orthographic overlay between the segmented point cloud and an IFC drawing with annotated quantities at three different heights. Different coverage is observed for each evaluated height, showcasing the need for heuristics based on Uniformat II sequences.

Table 1. Reported quantities for each construction activity. Classification results are based on the Uniformat Level 3, and Location information is automatically extracted from the Activity Name. Actual quantities (QTY) correspond to the actual quantity takeoff for completed elements from the IFC drawing (ground truth), while the Detected QTY comes from the area completeness ratio of the segmentation overlay against IFC drawings. The Corrected QTY is determined based on the extracted sequence from the classified activities and their relationships, and is compared against the ground truth to compute their mean average percentage error (MAPE).

Activity Name	Uniformat Code	Location	Actual QTY	Percentage Coverage (%)	Detected QTY	Corrected QTY	MAPE (%)
5th Floor - Wall Covering	C3010	Level 5	0 SQFT	0	0	0	0
5th Floor - Install Flooring	C3020	Level 5	182.6 ŠQFT	45.8	83.6 SQFT	179.6 SQFT	1.6
N-E 5th Floor - Hang Gyp at Ceiling - NE	C3030	Level 5 Zone NE	191.5 SQFT	93.8	179.6 SQFT	179.6 SQFT	6.2
5th Floor - Frame Walls	C1010	Level 5	579.4 SQFT	100	579.4 SQFT	579.4 SQFT	0
Exterior Skin - Window Installation Complete - 5th Floor	B2020	Level 5	1 EA	100	1 EA	1 EA	0
5th Floor - Plumbing Trim - Shower	D2010	Level 5	0 LF	0	0	0	0
5th Floor - Start Final MEP Wall/Ceiling Rough-In	D3050	Level 5	10.6 LF	4.6	0.5 LF	0.5 LF	95.3

using language models to extract sequential information encoded through the ASTM Uniformat II allows for the correction of certain construction elements that poor point cloud reconstructions may impact.

6 Conclusion

This work presented an application of automated progress monitoring, combining vision and language outputs only (i.e., without the use of BIM), and a practical way forward to creating automated progress monitoring applications, leveraging modern deep learning and computer vision algorithms. The proposed method utilizes novel transformer-based architectures such as Swin and BERT, each trained to classify detected construction objects based on the ASTM Uniformat II of Construction Objects from reality capture images and construction schedules. Additionally, the proposed method introduces an approach and heuristic to calculating the actual progress of construction. This approach is done by leveraging sequential and location-based information extracted from classified activity line items. Additionally, it imposes constraints on computed quantities from comparing per-pixel construction semantics projected on orthographic point cloud representations against IFC drawings with annotated QTOs.

The utilization of schedule-based-heuristics and sequential constraints addresses two types of limitations from the utilization of photogrammetry for the detection of progress: 1) the miscalculation of percent completed due to occluded construction objects in typical scenes, and 2) the incomplete computed progress from low-density reconstructions showcasing incomplete objects. This method is validated by a real-world case study, which computes progress quantities for a high-rise hotel as a function of provided reality captures, IFC drawing QTOs, and construction schedules. In this study, the detected pixel segmentations are projected against point cloud reconstructions and compared against drawing QTOs to compute actual progress. Such progress is corrected based on the automatically extracted schedule sequences and locations.

Given the current state of BIM standardization across the industry, the applicability of this method in real-world scenarios may fill the gap that low LOD in BIM models create by bringing an alternative to progress monitoring based on globalized practices when working with IFC drawings and QTOs. Future steps in this line of work will evaluate the validity of the established heuristics for other ASTM Construction Object categories.

References

- Varun Kumar Reja, Koshy Varghese, and Quang Phuc Ha. Computer vision-based construction progress monitoring. *Automation in Construction*, 138:104245, 2022. ISSN 0926-5805. doi:10.1016/j.autcon.2022.104245.
- [2] 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*, 108:103677, 2021. ISSN 0886-7798. doi:10.1016/j.tust.2020.103677.
- [3] Sebastian Tuttas, Alex Braun, Andre Borrmann, and Uwe Stilla. Comparision of photogrammetric point clouds with bim building elements for construction progress monitoring. volume 1, pages 341–345, 08 2014. doi:10.5194/isprsarchives-XL-3-341-2014.
- [4] Seungho Kim, Sangyong Kim, and dong-eun Lee. 3d point cloud and bim-based reconstruction for evaluation of project by as-planned and as-built. *Remote Sensing*, 12:1457, 05 2020. doi:10.3390/rs12091457.
- [5] Biyanka Ekanayake, Johnny Kwok-Wai Wong, Alireza Ahmadian Fard Fini, and Peter Smith. Computer vision-based interior construction progress monitoring: A literature review and future research directions. *Automation in Construction*, 127:103705, 2021. ISSN 0926-5805. doi:10.1016/j.autcon.2021.103705.

- [6] Hélène Macher, Tania Landes, and Pierre Grussenmeyer. From point clouds to building information models: 3d semi-automatic reconstruction of indoors of existing buildings. *Applied Sciences*, 7(10):1030, 2017. doi:10.3390/app7101030.
- [7] Yoonhwa Jung, Julia Hockenmaier, and Mani Golparvar-Fard. Transformer language model for mapping construction schedule activities to uniformat categories. *Automation in Construction*, 157:105183, 2024. ISSN 0926-5805. doi:10.1016/j.autcon.2023.105183.
- [8] Johannes L. Schönberger 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.
- [9] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Multi-view photometric stereo revisited. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3126–3135, 2023. doi:10.48550/arXiv.2210.07670.
- [10] Qian Wang, Yi Tan, and Zhongya Mei. Computational methods of acquisition and processing of 3d point cloud data for construction applications. *Archives of computational methods in engineering*, 27:479–499, 2020. doi:10.1007/s11831-019-09320-4.
- [11] Hesam Hamledari, Brenda McCabe, and Shakiba Davari. Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction*, 74:78–94, 2017. ISSN 0926-5805. doi:10.1016/j.autcon.2016.11.009.
- [12] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017. doi:10.48550/arXiv.1703.06870.
- [13] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. doi:10.48550/arXiv.1506.02640.
- [14] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In Pro-
ceedings of the IEEE/CVF international conference on computer vision, pages 10012–10022, 2021. doi:10.48550/arXiv.2103.14030.

- [15] Nipun D. Nath and Amir H. Behzadan. Deep convolutional networks for construction object detection under different visual conditions. *Frontiers in Built Environment*, 6, 2020. ISSN 2297-3362. doi:10.3389/fbuil.2020.00097.
- [16] Wei Wei, Yujie Lu, Yijun Lin, Ruihan Bai, Yichong Zhang, Haisong Wang, and Peixian Li. Augmenting progress monitoring in soil-foundation construction utilizing solov2-based instance segmentation and visual bim representation. *Automation in Construction*, 155:105048, 2023. ISSN 0926-5805. doi:10.1016/j.autcon.2023.105048.
- [17] Ron Mokady, Amir Hertz, and Amit H. Bermano. Clipcap: CLIP prefix for image captioning. *CoRR*, abs/2111.09734:pp. 1–10, 2021. doi:10.48550/arXiv.2111.09734.
- [18] Chenliang Li, Haiyang Xu, Junfeng Tian, Wei Wang, Ming Yan, Bin Bi, Jiabo Ye, Hehong Chen, Guohai Xu, Zheng Cao, et al. mPLUG:: Effective and efficient vision-language learning by cross-modal skip-connections. arXiv preprint arXiv:2205.12005, 2022. doi:10.48550/arXiv.2205.12005.
- [19] Bonsang Koo, Martin Fischer, and John Kunz. A formal identification and re-sequencing process for developing sequencing alternatives in cpm schedules. *Automation in Construction*, 17(1):75–89, 2007. ISSN 0926-5805. doi:10.1016/j.autcon.2007.03.005.
- [20] Yibrah Weldemihret Weldu. Automated Generation and Visualization of Initial Construction Schedules from Building Information Models. PhD thesis, Louisiana State University, 2016. URL https://digitalcommons.lsu. edu/gradschool_dissertations/175.
- [21] Zhiliang Ma, Songyang Li, Yong Wang, and Zhenqing Yang. Component-level construction schedule optimization for hybrid concrete structures. *Automation in Construction*, 125:103607, 2021. ISSN 0926-5805. doi:10.1016/j.autcon.2021.103607.
- [22] Fouad Amer, Yoonhwa Jung, and Mani Golparvar-Fard. Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Automation in Construction*, 132: 103929, 2021. doi:10.1016/j.autcon.2021.103929.

- [23] Yelda Turkan, Frédéric Bosché, Carl Haas, and Ralph Haas. Toward automated earned value tracking using 3d imaging tools. *Journal of Construction Engineering and Management*, 139:423–433, 01 2012. doi:10.1061/(ASCE)CO.1943-7862.0000629.
- [24] Sebastian Tuttas, Alex Braun, Andre 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, 85:3–15, 02 2017. doi:10.1007/s41064-016-0002-z.
- [25] Christopher Kropp, Christian Koch, and Markus König. Interior construction state recognition with 4d bim registered image sequences. *Automation in Construction*, 86:11–32, 02 2018. doi:10.1016/j.autcon.2017.10.027.
- [26] Alex Braun, Sebastian Tuttas, Andre Borrmann, and Uwe Stilla. Improving progress monitoring by fusing point clouds, semantic data and computer vision. *Automation in Construction*, 116, 08 2020. doi:10.1016/j.autcon.2020.103210.
- [27] Juan D. Nunez-Morales, Yoonhwa Jung, and Mani Golparvar-Fard. Bi-directional image-to-text mapping for nlp-based schedule generation and computer vision progress monitoring. In *Construction Research Congress 2024*, 2024. (In press).
- [28] Moch Kholil, I Ismanto, and M 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*, 1073:012066, 02 2021. doi:10.1088/1757-899X/1073/1/012066.
- [29] Fouad Amer and Mani Golparvar-Fard. Automatic understanding of construction schedules: Part-ofactivity tagging. In *EC3 Conference 2019*, volume 1, pages 190–197, 2019. doi:10.35490/EC3.2019.196.
- [30] Pablo Fernández Alcantarilla, Jesús Nuevo, and Adrien Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *British Machine Vision Conference*, 2013. URL https://api. semanticscholar.org/CorpusID:8488231.
- [31] Juan D. Núñez-Morales, Shun-Hsiang Hsu, and Mani Golparvar-Fard. Synthetic Image Generation for Training 2D Segmentation Models at Scale for Computer Vision Progress Monitoring in Construction, pages 273–281. doi:10.1061/9780784485224.034.

Automating Construction Safety Inspections using Robots and Unsupervised Deep Domain Adaptation by Backpropagation

Vimal Bharathi¹, Samuel A. Prieto², Borja Garcia de Soto², Jochen Teizer¹

¹ Department of Civil and Mechanical Engineering, Technical University of Denmark

² S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi

vvebh@dtu.dk, samuel.prieto@nyu.edu, garcia.de.soto@nyu.edu, teizerj@dtu.dk

Abstract

Due to the dynamic aspect of construction sites, constant implementation and removal of safety equipment is a required practice. This leads to frequent manual and time-consuming inspections to make sure the safety measures are in place. There is the potential to automate the inspection process using robots and Deep Learning. Such an approach can save time and cost while improving safety. Using images collected by an Autonomous Ground Vehicle, a Deep Learning model with Domain Adaptation techniques is trained to detect and segment safety guardrails. The results of the model indicate a promising method to assist in automating site safety inspection that can make construction sites safer. Further work is necessary to validate this effort under more realistic and harsh construction site conditions.

Keywords -

Construction safety measures, deep learning, domain adaptation, prevention through design and planning, site inspection, robot, YOLOv8.

1. Introduction

In the dynamic realm of construction safety, technological advancements are reshaping traditional practices. The construction industry stands out as one of the riskiest sectors due to the ever-changing work environment. Ensuring the safety of construction workers is a vital aspect of a construction business, leading to the development of comprehensive regulations and guidelines. These measures are designed to protect workers during construction activities. Given the demanding nature of crafting a secure construction plan, it remains a top priority. Consequently, substantial time and effort are dedicated to promoting the health and wellbeing of workers, preventing fatalities, severe and minor injuries, and close-call incidents, also known as prevention through design (PtD) [1].

An analysis of industries and their associated hazards in the UK [2] reveals that the private construction sector is the second leading sector in workplace injuries and fatalities. A recent report summarizing construction safety statistics for 2023 shows that falls make up 35% of fatalities in construction sites [3]. This underscores the significance of the focus on fall hazards and, more specifically, protective guardrails.

Another critical aspect of ensuring safety in construction operations involves inspecting and pinpointing missing or subpar safety equipment, such as protective guardrails. Inspecting collective safety equipment is also a labor-intensive task, given the dynamic nature of construction sites. Therefore, inspections must occur frequently. Additionally, the actual installation of safety measures often deviates from the intended quality outlined in digital models.

Hence, in this study we propose a preliminary stage of a more automated inspection that improves the current human-based solutions onsite. The data collection process is based on the work by Gopee et al. [4], which uses the existent BIM to generate waypoints of interest for the navigation of an Autonomous Ground Vehicle (AGV) to collect images (e.g., RGB information) as it traverses the designated areas. The collected data can be processed and used to train Deep Learning models designed for real-time object detection of the elements of interest, such as protective guardrails. In this study, a YOLOv8 network enhanced by a Domain Adaptation technique is used. The differences in construction sites due to factors like weather, sunlight, and location result in numerous scenarios for the deep learning model to analyze and predict outcomes. Each of these cases could be called a "Domain," as the images from these cases could have different features for the model to identify and learn. Due to the lack of available data and time, the technique of domain adaptation has been used to achieve the maximum possible results in predicting the guardrails with less unlabeled data from any target domain. The presented model provides a binary classification (i.e., safe or unsafe) as an output for the guardrails detected. Evaluation metrics such as Precision, Recall and Mean Average Precision (mAP) are used to quantify the improvements in detection capabilities and choose the right model configuration.

The rest of the paper is structured as follows. Section

2 provides a background on segmentation in the construction field. Section 3 presents the methodology used for this study. Section 4 uses a case study to illustrate the main elements of the methodology. Section 5 summarizes key results from the case study and, finally, Section 6 includes the conclusions and future work.

2. Background

There have been many previous works on the usage of different methods and models for the detection of guardrails using data in the form of construction images, synthetic images, and point cloud data (from laser scanners or photogrammetry).

Kolar et al. [5] used synthetic data for training a model with VGG16 as the feature extractor. The construction images with guardrails were used as the validation dataset, while augmented synthetic data was used for training the model. Their method had a 96.5% accuracy but failed in cases with low-light images of guardrails, which shows that the model would work poorly in domain change (i.e., changes in the physical conditions of the environment) situations.

The lack of data available for training an effective and efficient Deep Learning model is challenging. The construction sector is widely diverse, with different types of visual features around the world, including different seasons and weather conditions. Each difference is a different domain of data, which presents an extra challenge to train Deep Learning models to work in all these different conditions. To overcome this, there is a need for a technique that could help to train a generalized Deep Learning model, to work in multiple domains and with low amounts of data by applying Transfer Learning. Ganin et al. [6] showed that domain adaptation could be used to learn Domain Invariant features, which could predict and segment objects across different domains with few or no labeled target domain data. Using the unlabeled target domain data in the training process is called "Unsupervised Domain Adaptation".

Li et al. [7] employed a Domain Adaptation Technique in combination with the YOLOv5 architecture [8], applying it separately to the source and target domains. In this approach, they used CSPDarknet-53 [9] as the feature extractor, which forms the backbone, neck, and head components of the YOLOv5 structure. In their study, a significant amount of source labeled data was used, along with small amounts of target labeled data, to train the YOLOv5 pipeline. Features from 3 different resolutions from the backbone of the source and target pipeline were used to calculate the transfer loss (Maximum Mean Loss). The transfer loss shows the distance between the features extracted by both the source and target data, which is added to the overall loss function as a regularization term. The drawback is that the sample target data that has been used for training has to be labeled, which is a labor-intensive task given the number of different domains.

The performance of YOLOv8 on specific datasets [10] and its anchor-free detection technique that increases the model's ability to detect objects of various shapes and sizes without the constraints imposed by predefined anchors has made it an ideal choice for our task. This work shows how the method of Unsupervised Domain Adaptation with backpropagation [6] can be used with a YOLOv8 [11] model to improve the detection and segmentation of objects with labeled source domain and unlabeled target domain datasets.

3. Methodology

The overall process used in this study is shown in Figure 1. It can be divided into two sections: (1) data collection and (2) data processing/model development. For the data collection, given a BIM of the environment, it is possible to extract a set of waypoints for an AGV to stop and collect data. While the AGV autonomously moves towards the waypoints, it can collect data with multiple sensors (i.e., an RGBD camera and a 360 camera). This study focuses on data processing and model development. For more information about data collection, readers are referred to previous work by the authors (e.g., [12-14]).

3.1 Data Pre-processing and Labeling

To build a Deep Learning model, a set of training data needs to be used. Different labeling tools can be used for the collected data.

3.1.1 Data Augmentation

In the case when the amount of training data is not enough to ensure good results in the segmentation stage, data augmentation can be used to increase the size of the training dataset. Once the data has been labeled, various types of data augmentation are used during the training process, which greatly increases the chances for the model to explicitly learn the guardrail features.

3.2 Build Deep Learning Model

3.2.1 Deep Learning Network

Deep-feedforward architectures have brought significant advances to state-of-the-art models across a wide variety of machine-learning tasks and applications.

3.2.2 Deep Domain Adaptation

A Deep Learning architecture trained on one to work for another domain with a shift in the distribution needs features that are common to both domains. Learning a label classifier in the presence of a shift between source and target distribution is known as domain adaptation. There are several methods to perform Domain Adaptation (DA). Here, we use DA by backpropagation.



Figure 1: Main elements of the proposed methodology for data collection, processing, and model development.

4. Case Study

The proposed methodology has been tested on a condition with an abrupt change in elevation, representing conditions that could lead to fall hazards if proper protection was not in place during a construction project. The scenario used was a staircase on a university campus (Figure 2). A mock-up of a fall protection guardrail and an AGV equipped with different sensors were used to collect data.

The outdoor experiment consisted of different case studies, considering all the possible cases that could be present in a real construction site. These cases consider the proper installation of the guardrail (with all the elements installed correctly), missing elements of the guardrail (i.e., missing mid and toe board), and the presence of clutter that could potentially be a trip hazard.



Figure 2: (a) Overall view of the scenario used for the experimentation, and (b) view of the stairs representing the fall hazard.

4.1 Fall Protection Guardrail Mock-up

A wooden mock-up was built to resemble the most common safety guardrails typically used in construction sites as fall protection [15]. It was built according to the requirements specified by the Employer's Liability Insurance Association for the Construction Industry in Germany (BG BAU) [16]. The mock-up consisted of a modular system with three vertical poles (1m height) with top, mid and toe boards (in total 6 horizontal boards of 20cm height and 1.5m width). An overview of the mock-up is shown in Figure 3.



Figure 3: View of fall protection guardrail mockup developed and used for this study.

4.2 Autonomous Ground Vehicle (AGV)

The AGV used was a SUMMIT-XL platform by Robotnik Automation. The robot has holonomic locomotion (i.e., mecanum wheels). This allows the robot to move in all directions, providing a more accurate and reliable data collection in highly dynamic environments such as construction sites, where narrow passages are common. The AGV is shown in Figure 4, and a recording of the AGV collecting data can be watched in [17].

In terms of sensors for the data collection, the robot is equipped with a mid-range 3D scanner (BLK360) suitable for high-resolution dense point cloud acquisition, a long-range LiDAR (OUSTER OS1) suitable for lowresolution point cloud acquisition used for the navigation, an RGB-D camera (Orbbec Astra) used to collect both RGB and depth information of the robot front view, and a 360 camera (GoPro MAX 360) aimed to collect extra RGB data surrounding the robot as it moves through the environment. Key characteristics of the different equipment used are summarized in Table 1.



Figure 4: AGV and related payload used.

Table 1: Key specifications of the sensors used.

Sensor	Key specifications
Leica Geosystems	Accuracy: 6mm@10m
BLK360	Range: 60m
(Laser scanner)	Speed of capture: 360,000 pts/s
	FOV: 360°H x 300°V
OUSTER OS1	Accuracy: 2.5cm@10m
(LiDAR)	Range: 170m
	Speed of capture: 2,621,440 pts/s
	FOV: 360°H x 45°V
Orbbec Astra	Accuracy: 3mm@1m
(RGBD camera)	Range: 0.6m-8m
	Image resolution: 640x480
	@30fps
	FOV: 60°H x 49.5°V x 73°D
GoPro MAX	Image resolution: 4992 x 2496 x
(Action camera)	360°

4.3 Conditions Evaluated

4.3.1 Full Guardrail (Safe Condition)

For this condition, the ideal scenario is tested. This means that the guardrail contains all the horizontal boards, fulfilling all the safety standards (Figure 5).



Figure 5: (a) View of the mock-up placed on the stairs entrance, and (b) AGV collecting data.

4.3.2 Missing Boards (Unsafe Condition)

For this condition, multiple boards of the guardrail are missing. This reflects a potential fall hazard since the space between the top board and the floor is wide enough for a person to fall through (Figure 6). In addition, multiple objects (clutter) were added, presenting additional trip hazards that can potentially lead to a worker falling through the faulty installation. These elements also present occlusion for the Deep Learning detection algorithm, ensuring that the approach is tested under non-ideal conditions.



Figure 6: (a) View of the scenario where the mock-up misses the mid and toe boards and has clutter on the floor, and (b) same condition with only vertical poles installed.

4.4 Data Pre-processing and Labeling

To train the Deep Learning model, RGB data was collected manually in an indoor environment with controlled lighting conditions (source domain). The training data was labeled using "Label Studio", an opensource data labeling tool [14]. For the specifics of a segmentation model, the labeled data needs to be in the form of masks (i.e., vertices of a polygon surrounding the segmented object). To further enhance the robustness and performance of the Deep Learning model under the dynamic conditions of the construction site. Unsupervised Deep Domain Adaptation is also used with indoor training images as the source domain and the outdoor collected images as the target domain. The target domain does not require any data processing or augmentation.

4.4.1 Data Collection

The data collected can be split into two categories: indoor and outdoor. To train the YOLOv8 network, a set of pictures of the fall protection guardrail taken in an indoor controlled environment was used. The source domain corresponds to the labeled training indoor data, and the target domain corresponds to the sample unlabeled data that needs to be classified. The scenarios were classified into safe (Figure 7 (a)) and unsafe (Figure 7(b-d)).

4.4.2 Source Data Collection

A total of 56 images in the indoor setting were taken. The source images were split into training and validation images with a ratio of 3:1. The pictures were taken from different points of view and two sets of distances (Figure 8) and different cases (i.e., all horizontal boards, only the top board, only the vertical poles, and with no guardrail) (Figure 7) to collect as many features as possible.



Figure 7: Source images used for training with guardrails: (a) full board, (b) top board and vertical poles, (c) poles only, and (d) none present.



Figure 8: Setup for indoor data (images) collection.

4.4.3 Target Data Collection

The outdoor images, which correspond to the target domain, were collected with the AGV in different conditions. Examples of the collected data with the AGV are shown in Figure 9.

4.4.4 Data Augmentation

Since the amount of training data was not enough to train a robust object detection model, data augmentation was used to increase the training data tenfold. Continuous and random selection of various techniques of augmented data is used during the training of the Deep Learning segmentation model. The data augmentation techniques used in this study are summarized in Table 2.

Table 2: Augmentation techniques used on the original training dataset.

Type of augmentation	Value
Rotation	[0-30] (+/- deg)
Translation	0.1 (fraction)
Scaling	0.5(+/- gain)
Shear	0.5 (+/- deg)
Flip left/right	0.5 (probability)
Mosaic	[0.1-0.9] (probability)
Mix-up	[0.1–0.9] (probability)



Figure 9: Target images: (a) Full board guardrail, (b) top board with and without clutter, and (c) only vertical poles, with and without clutter.

4.5 Model Development (YOLOv8 Model)

The main goal of this model is to classify images into safe and unsafe categories, as indicated in Section 4.3, by segmenting the presence (or lack of) and state (i.e., if all the components are present) of protection guardrails. For this study, we used YOLOv8 as the base model for the training, coupling it with other techniques, such as data augmentation and Unsupervised Deep Domain Adaptation (DA), to enhance guardrail detection in varying environments (i.e., indoor, outdoor, sunny, dark, etc.). We focus on DA using backpropagation. We use a feed-forward network or feature extractor to extract the domain invariant features and use them to train both the label classifier and Domain Adaptation Network. The label predictor uses the features to detect the guardrails and if it is a safe or unsafe condition. Whereas the domain classifier is used to predict whether the input belongs to the source or target domain.

The YOLOv8 network can be built with different architecture configurations by modifying the amount and type of layers, and the different hyperparameters that affect the weights of the network. To properly choose the best architecture, several tests need to be performed to assess the performance of the network. The domain predictor uses the same features from the last layers of the network to predict whether the input belongs to the source or target domain. The domain predictor layer consists of the gradient reversal layer, which backpropagates the loss for the optimization of the weights of the network with a negative value. This makes the Domain Adaptation network and feature extractor (Deep Learning network backbone) oppose each other, making it possible to learn domain invariant features. The weights of the underlying deep feature extractor (i.e., YOLOv8 network backbone) are optimized to minimize the loss of the label classifier and to maximize the loss of the domain classifier. The latter encourages domain-invariant features to emerge while training or optimization. The three different networks (namely the feature extractor, label predictor and domain classifier) can be embedded into a single deep feed-forward network using standard layers and loss functions and can be trained using standard backpropagation algorithm [6]. The crucial layer of this architecture for the task of domain adaptation is the gradient reversal layer, which leaves the input unchanged during forward propagation and reverses the gradient by multiplying it by a negative scalar (-1 × lambda) during backpropagation.

To achieve the best possible results, different characteristics of the YOLOv8 model architecture, target data, and changes in hyperparameters, such as *lambda* in the Gradient Reversal layer, needed to be fine-tuned.

A summary of the different tested models and their corresponding results can be seen in Table 3 and Figure 10, respectively. The results include metrics: Precision, Recall, and Mean Average Precision (mAP) for the segmented guardrails in the images. These metrics show how well the model identifies the different configurations of guardrail elements in the images.

As seen in Figure 10, the results from training the YOLOv8 only with the source images with and without augmentation (Model#1 and Model #2) are not good on the validation Target data. The training "Model #3" with only 1 layer Domain Adaptation Layer (P5) performs poorer than the previous training as it was given only very few variations in Target data for training. The training of "Model #4" (same as "Model #3" but with more variation in target data for training) performs much better than "Model #3". This shows that even though the Target data are fed into training without the ground truth labels, the model can learn more Target domain features with the Domain Adaptation Network. The best model training was achieved with "Model 5," built with 3-Layers of DAN with features from P3, P4, P5 and more variation in the Target Training Data. "Model #6" is the same as "Model #5" except for the value of lambda for the Gradient reversal layer. "Model #5" was trained with lambda 1, and "Model #6" was trained with lambda 5, which has obtained slightly poor results. This shows that "Model #6" with lambda 5 punishes the rest of the model more to predict the correct domain class during backpropagation and yields worst results on label prediction. Hence, a better value of lambda for this training is 1, as it punishes the model the right amount during training.

Table 3: Training Model description and parameters.

Model	Details			
#1	Training only source images without Data			
π 1	Augmentation			
#2	Training only source images with Data			
112	Augmentation			
#3	Training with source images and fewer target			
	images and only the P5 DAN layer			
#4	Training with source images and more target			
	images and only the P5 DAN layer			
#5	I raining with source images and more target	[
#3	images and P3, P4, P5 DAN layers and			
	Training with source images and more target	f		
#6	images and P3 P4 P5 DAN layers and			
	lambda=5			
1				
	li i			
0.75	Model #	#1		
lie e	Model =	#2 #3		
0.5	Model	#4		
Met	Model /	#5		
0.25	Model :	#6		
0				
	Precision Recall mAP50 mAP50-95			

Figure 10: Results for different metrics of each model.

The selected YOLOv8-m architecture is configured with varying channel sizes for the P3, P4, and P5 layers, specifically 192, 384, and 576 channels, respectively. This configuration is dependent on the type of YOLOv8 variant used, which includes n, s, m, l, x models. In this architecture, the extracted features from the training phase are employed by the up-sampling process in the YOLOv8 head. This up-sampling is crucial for reconstructing the image back to its original size, and it is particularly instrumental in the segmentation tasks, where maintaining image resolution and detail is essential. The overall network structure, with the YOLOv8 classifier network and the Domain Adaptation, can be seen in Figure 11.



Figure 11: YOLOv8 with domain adaptation network for selected model (Model #5).

5. Results

5.1 Classification Results

The results show that Model #5, with a 3-layer Domain Adaptation Network, has achieved the highest accuracy on the validation target data. The values for the evaluation metrics for Model #5 are summarized in Table 4, and representative results to classify the target images are shown in Figure 12.

Precision	Recall	mAP50	mAP50-95
0.947	0.868	0.909	0.818
2cdoff 1a8-trame00			

Table 4. Evaluation metrics for Model #5.

Figure 12: Validation results of Model #5.

The Confusion matrix (Figure 13) shows that Model #5 performs well in identifying the 'Unsafe' class with a high number of true positives (14). This indicates that the model is effective at detecting 'Unsafe' scenarios. There are no instances where 'Safe' is confused with 'Unsafe' or vice versa, which is positive as it suggests that the model can distinguish between these two classes effectively.

Overall, it can be said that the model has a high precision for the 'Unsafe' class since there is only one false positive; however, its recall is affected by the five false negatives. The presence of false negatives for 'Unsafe' (5 instances misclassified as 'Background') suggests that there may be room for improvement in distinguishing between 'Unsafe' and 'Background'. This could be attributed to the debris occluding the view, dark shadows, or being far from the camera, which is acceptable as the model has not been trained enough for such harsh cases. This could be improved in the future by training with more diverse and harsher data so the model learns more features.



Figure 13: Confusion matrix of Model #5.

5.2 Failed cases

The guardrail detection by the best model is not always accurate. The failed cases, as shown in Figure 14, either have some debris lying in front of the guardrails, the guardrails are at a far distance, or the guardrails are too close and partially captured in the images. These are reasonable failed cases since the training data had no such cases and was not the objective of this work. The guardrails prediction with the AGV was designated to work with the guardrails being at a reasonable distance (i.e., from 2 meters to 4 meters) as can be seen in Figure 8, showing the setup for indoor data collection. Therefore, the failed cases in the target images with guardrails at distances less than 2 meters or more than 4 meters are understandable. In addition, the training data did not contain guardrails with debris lying in front of them. In future work, we will solve these shortcomings with more data, capturing all these situations and using depth information to overcome these.



Figure 14: Failed predictions by Model #5.

6. Conclusion and Future Work

From this work, we have achieved detection and segmentation of guardrails with YOLOv8 trained on source domain (i.e., indoor training data) and predicted on target domain (i.e., outdoor collected data) using Domain Adaptation with Backpropagation. This indicates that the model, needing fewer target domain images, can be used for guardrail detection across different domains without having to label the ground truth for each one of the images.

The model had some limitations that could be overcome by using images from failed cases in the target data during training, allowing the model to learn extra features. The significance of the unlabeled target training data is crucial, as it determines the performance of the model prediction on the target domain. The same method, along with transfer learning, can be effectively used to retrain the model repeatedly onto different domains with less unlabeled training data. This can save a significant amount of time while not sacrificing performance.

Some challenges faced by a vision-based approach can be overcome by adapting the same techniques to RGB-D images or by using point cloud data. The use of depth information might lead to better results. Another future aspect of this work lies in the real-time use of safety information to deploy predictive models in an online Digital Twin. Further work can also target alternative data collection and inspection methods, for example, unmanned aerial vehicles [18] for tailored human-assisted safety management software [19].

Acknowledgment

The research presented in this paper was in part funded by the European Union Horizon 2020 research and innovation program under grant agreement no. 101058548. This work also benefited from the NYUAD (CITIES) and NYUAD (SHORES) funded by Tamkeen under Awards CG013 and CG001.

References

- Teizer, J. Right-time vs real-time pro-active construction safety and health system architecture, *Constr. Innov.*, 16(3): 253–280, 2016, https://doi.org/10.1108/ci-10-2015-0049
- [2] HSE. Statistics Industries. 30.11.2023, https://www.hse.gov.uk/statistics/industry
- [3] Finity, J. 36 Construction Safety Statistics for

2023," Procore. Accessed: 30.11.2023. https://www. procore.com/library/constructionsafety-statistics

- [4] Gopee M.A., Prieto S.A., García de Soto B. Improving autonomous robotic navigation using IFC files. *Constr. Robot.*, 2023, https://doi.org/10.1007/s41693-023-00112-8.
- [5] Kolar Z., Chen H., Luo X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. *Autom. Constr.*, 89:8–70, 2018.
- [6] Ganin Y., Lempitsky V. Unsupervised Domain Adaptation by Backpropagation. arXiv, 2015.
- [7] Li C., et al. A domain adaptation YOLOv5 model for industrial defect inspection. *Measurement*, 213, 112725, 2023.
- [8] Jocher G., et al. ultralytics/yolov5: v7.0 -YOLOv5 SOTA Realtime Instance Segmentation. Zenodo, 2022. doi: https://doi.org/10.5281/zenodo.7347926.
- [9] Wang C.-Y. et. al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN, 2019.
- [10] Solawetz J. What is YOLOv8? The Ultimate Guide. https://blog.roboflow.com/whats-new-inyolov8/
- [11] Ultralytics. YOLOv8. https://docs.ultralytics.com/
- [12] Prieto S., García de Soto B., Adan, A. A Methodology to Monitor Construction Progress Using Autonomous Robots. 37th ISARC https://doi.org/10.22260/ISARC2020/0210.
- [13] Prieto S.A., Giakoumidis N., García de Soto, B. AutoCIS: An Automated Construction Inspection System for Quality Inspection of Buildings, 38th ISARC, 2021.
- [14] Prieto S.A., Giakoumidis, N., García de Soto B. Multi-agent robotic systems and exploration algorithms: Applications for data collection in construction sites. arXiv, 02, 2023.
- [15] Johansen, K.W., de Figueiredo R.P., Golovina O., Teizer J. Autonomous Safety Barrier Inspection in Construction: An Approach Using Unmanned Aerial Vehicles and Safe BIM. 38th ISARC, 2021. https://doi.org/10.22260/ISARC2021/0085.
- BG Bau. B 100, Absturzsicherungen auf Baustellen, 2023. https://www.bgbau-medien.de/
- [17] *Guardrail video data*, 2023. https://drive.google. com/drive/search?q= SMART%40NYUAD
- [18] Siebert S., Teizer J. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Autom. Constr.*, 41:1-14, 2014, https://doi.org/10.1016/j.autcon.2014.01.004
- [19] Golovina O., Teizer J., Johansen K.W., König, M. Towards autonomous cloud-based close call data management for construction equipment safety. *Autom. Constr.*, 132, 103962, 2021, https://doi.org/10.1016/j.autcon.2021.103962

From Unstructured Data to Knowledge Graphs: An Application for Compliance Checking Problem

Ankan Karmakar¹, Chintan Patel¹, and Venkata Santosh Kumar Delhi¹

¹Department of Civil Engineering, Indian Institute of Technology Bombay, India ankank@iitb.ac.in, 23m0619@iitb.ac.in, venkatad@iitb.ac.in

Abstract -

The rule requirements of a building code are frequently violated to create financially viable designs. These deviations are subjected to condonation by the municipal commissioner if recognizable hardships are faced. The historical concession applications for similar cases are stored in an unstructured manner, creating a barrier to knowledge transfer. The subjective statements given by applicants are composed of logical structure, language, and embedded knowledge that requires years of experience from the domain expert to decipher. A knowledge graph (KG) representation of the problem can capture concepts and represent them visually, which is easy for novice stakeholders to understand. A Large Language Model (LLM)-based method is used in this study for ontology extraction in the form of concepts and relationships. Also, unstructured input preprocessing and entity disambiguation were performed to evaluate the applicability of KG in this domain. The performance of the proposed method was checked qualitatively in a case study from reallife project examples. The limitations and scopes for improvements were also highlighted. The outcome of this study indicates KG as a potential candidate for knowledge generation from the unstructured archival data of compliance checking. The target audience for this application can be the new architects, reviewers, and programmers working on developing the end-toend automated compliance checking systems. Finally, applying these Artificial Intelligence (AI)-based knowledge transfer mechanisms can ignite future research on automated concession applications and approvals, laying a path to the digital transformation of the industry.

Keywords -

Code Compliance Checking; Knowledge Graphs; Artificial Intelligence

1 Introduction

The construction industry behaves in a fragmented manner where information is not transferred across a building lifecycle, projects, and stakeholders. The siloed data stored prohibits knowledge transfer even among stakeholders from similar domains. Post-design coordination code compliance checking emerges as an essential step for design information transfer. Communication happens between the liaison architects the developers with government-authorized and personnel at the local urban bodies (ULB). According to the literature, automated code compliance systems comprise four steps, i.e., rule interpretation, input model preparation, rule execution, and checking [1]. With the advent of Building Information Modeling (BIM), several commercial and academic research studies have been performed to verify design information at the compliance-checking stage [2]. However, the rule-based compliance checking process is not the end of the complete permit checking process.

After rule checking, the architects apply to the municipal commissioner (MC) to condone several rules that were violated during the design stage. The hardships faced are stated in the form of subjective statements, cross-referencing intertwined concepts. In the Indian context, these problems are detailed in an unstructured manner through natural human languages. Furthermore, the logical reasoning given for achieving a concession in one of the violations can vary across projects. The new architects applying for new concessions or new reviewers looking at applications are deprived of such enriched data captured in a longitudinal timeframe. Even with the availability of the documents, it seems humanly impossible to summarize the tacit knowledge behind these submissions from thousands of documents. Thus, even with the focus on translating code clauses to machine-readable formats, the other half of the compliance checking process lacks the knowledge transfer aspect. Also, during knowledge transfer, an individual actor's own understanding can influence the successor's method of analyzing a present situation, thus

reinforcing the requirement of an artificially developed knowledge base.

The system has to capture the unwritten implicit assumptions and complex causal relationships to capture the knowledge from documents prepared by the domain experts [3]. A graph-based framework was proposed in this study to facilitate the knowledge transfer process for compliance checking. The paper applies a Machine Learning (ML)-based graph generation technique to evaluate the applicability of knowledge graphs. Reports were gathered from real-life projects from the municipal corporation of Mumbai, India. Further, the knowledge graph's limitations and potential application areas in the automated code compliance checking domain are also discussed.

2 Literature Review

The literature review is divided into two sub-sections where the knowledge graph (KG) development and its working principle are initially discussed. The following section highlights KG's state-of-the-art applications in the construction industry.

2.1 Knowledge Graphs

Knowledge extraction methods were first developed to address the data available across the web in the forms of text or HTML. These human-created data are not in the form of machine-readable language, leading to the requirement of a system that can extract information from text. This action of extracting insights and inference from the information is knowledge extraction [4]. The steps for the method involve named entity and relationship recognition, entity disambiguation, and relation linking. These methods of knowledge extraction form syntactical relations but lack contextual relations within the knowledge graphs [5]. Knowledge graphs are used to overcome the limitation of contextual relationships among the entities. KG are representations of unstructured data in the form of graphs. It consists of nodes that store the entities, which are interconnected by embedded links, i.e., relationships. Google was the first to introduce its knowledge cloud in 2012, which converted literals to knowledge [6]. Subsequently, several knowledge graphs were developed, such as Wordnet for Natural Language Processing (NLP), YAGO, and DBnet trained over data from Wikipedia. Dbnet model stored data in the form of a Resource Description Framework (RDF) containing subject, predicate, and object triplets [7].

KGs created from unstructured data were indicated to have three components, i.e., ontology extraction, entity extraction, and relation extraction. Further, it was highlighted that relations are not created initially, as manual tagging is required for further training on the extractor network, essentially making it a supervised approach [8]. As KGs have different data formats, several methods are proposed for embeddings, such as - rulebased [5], semantic-based cross-lingual [9], and more. This idea can also be extended to different input formats like text, images, and videos. Further, text and image KGs were combined for visual queries, and text-based embeddings on image data were illustrated, enhancing the visual understanding of the objects by describing them [10], [11].

A few limitations of knowledge graphs include their inability to find the semantic relations with less accuracy and the incapability to explain the relationships. Hence, several applications, such as financial investments, cannot be run based on it. Another limitation is the data's sparsity, especially when working with domain-specific knowledge graphs. As KG are static in nature, but the links or relationships may change over time, the temporality of data also adds to its inadequacy list [12].

2.2 KGs in the Construction Industry

In construction, data is maintained in a siloed manner. Due to this, there is a lack of integration among the data, and engineers lack insights into the decision-making process. KG provides a knowledge management framework that holistically stores the information [13]. KG has various applications in the construction industry. In the case of pavement engineering, maintaining the data related to pavement materials, quality tests during the pavement construction, monitoring data, and updating it in the maintenance phase were identified to be difficult [14]. KG provides the platform to holistically store data in graph formats, with semantic enrichment, which helps project managers make data-driven decisions. A similar use case of the KG graph is explored in Bridge maintenance, where an ontology-based framework for knowledge creation and KG for knowledge storage is used [15]. It also helps in the project management for interoperability in the project teams and document management. A KG model was used with semantic web technologies to highlight the impact of changing design decisions [16].

KGs can also be used to build domain-specific knowledge graphs to analyze construction safety reports [17]. KGs were also used to automate the process of checking the fire safety drawings in combination with BIM. KGs were able to extract the specific information from the clause and match it with the information from the BIM document for the review process [18]. In the digitization of built assets, i.e., digital twins, various information comes in from BIM, IoT sensors, and legal documents [19]. In this scenario, KGs emerge as a valuable tool for project information management. It is based on the ifcOWL-based ontology, providing more openness to the project data [20]. On the other hand, Modular construction, a widely applied technique in the construction of high-rise buildings, requires a high number of customizations, leading to difficulties in managing the work packages. Researchers use KG to manage the data for the extraction of work packages. It maps products to tasks based on granularity and tasks to work packages; subsequently, relationships are learned using the ML models [21].

Knowledge extraction from the video has also been targeted through KG. Computer Vision extracts the entities from the video input, subsequently updating the KG [22]. Similarly, the problem of extracting knowledge from images is also attempted with the ontology of humans, actions, and objects, and the relationships are extracted using ML and deep learning (DL) techniques [23]. Object detection techniques were also used at the construction site to capture the construction progress and derive insights based on the KG created from highly structured data [24].

KG has been applied to the Industry Foundation Class (IFC) graph structure to extract data from BIM models and transfer the knowledge across projects [25]. The BIM-based variants from KGs were retrieved through case-based reasoning and pattern matching for early-stage designs to aid the architects [26], [27]. The BIM KGs were able to suggest alternative design decisions depending on local code restrictions, the similarity of design requirements, and cost-effectiveness [28], [29]. The interoperability among the different BIM authoring tools was also addressed through semantically enriched BIM knowledge graphs produced over the backbone of the IFC graph structure [30].

KGs were recommended for knowledge transfer from subject matter experts to programmers in the compliance checking domain. Researchers also proposed KG quality assessment parameters to generalize the KG application beyond the variations of building codes [31]. Conversely, in the case of unstructured archival data, the variation exists due to the different ways in which the user speaks. According to research, PDF documents hold the lowest rank in the 5-star linked data matrix [32]. Despite the drawbacks of a significant amount of archival data being PDF documents, it cannot be denied that knowledge graphs can be widely applied. The KGs can solve the complex problems of siloed data due to their adaptability in updating the knowledge and cognitive ability to get the semantic links.

3 Application of KGs over Unstructured Data Used for Compliance Checking

The Architecture, Engineering, and Construction (AEC) industry has a plethora of data stored in different formats across projects. The complexity of information increases as the project traverses through the planning to

execution phases. A necessary process playing a pivotal role in this information transfer process is the code compliance check. In the Indian ULBs, the checking process is done through AUTODCR [33], which is an automated rule-checking engine for 2D CAD drawings. The information exchanged and reports generated are stored in PDF format, which loses applicability and automatability, thus restricting the transfer of knowledge generated across projects. Applying knowledge graphs to these data helps understand the hidden concepts in these documents and identifies the crucial concepts that need to be stored as a structured database for future use. Where the ULBs hold a large amount of unstructured data, KGs can be a starting point for discovering the paths to move forward.

3.1 Subjective Inputs for Concession Reports

According to the local code "Development Control and Promotional Regulations" (DCPR 2034) for the Greater Mumbai municipal region of India, architects can apply for a concession to the Municipal Commissioner (MC), even after failing to meet the code requirements. The allowable concessions document lists 35 special cases where the MC can provide leeway to the design deviations from the DCPR 2034 code. However, according to clause 6[b] of DCPR - "In specific cases where a clearly demonstrable hardship is caused, the Commissioner may, for reasons to be recorded in writing, by special permission permit any of the dimensions prescribed by these Regulations to be modified, ..., provided that the relaxation will not affect the health, safety, fire safety, structural safety, and public safety of the inhabitants of the building and the neighborhood."

The above requirements demanded by the regulation force the architects to demonstrate the hardship faced through writing. The logical reasonings provided under health, fire, structural, public, and neighborhood safety are thus expressed in natural languages through paragraphs explaining the necessity of concessions to make the project feasible. Depending on the arguments provided in the report, a group of concessions can be approved or rejected for that specific project. Rejection leads to a wastage of time for both designing firms and the ULBs in the form of reworks. Further, the knowledge gathered across these projects is stored as unstructured documents, making it impossible for designers to refer to the archival records for future reference.

3.2 Methodology

The unstructured PDF documents obtained from the ULBs of historical projects are used as input documents. These documents are assumed to consist of concepts or ideas that drive the acceptance or rejection of a concession by the MC. The methodology used aims to

extract these concepts and their intertwining relationships. This combination of complex concepts might have been challenging to inculcate for the users across multiple projects.

The first step in the process is to preprocess the archival data. The concession pleadings are sometimes presented in the form of paragraphs or nested tables. The headers of these tables also vary across projects, making the automation process challenging to generalize. Once the tables and nested tables are converted into paragraphs, the document is distributed in chunks for each concession type to maintain its inherent contextual proximity. Contextual proximity is the indicator of concepts originating from the same concession type.

The generated chuck length is ensured to be within the capability limits of the Large Language Model (LLM) to be used. In the LLM model, it is prompted to generate concept and relationship pairs for each chunk. The prompt is engineered to instruct the LLM to extract ontology between key concepts. These key terms may include object, location, roles of person, entity, acronyms, documents, or conditions. However, the names of persons, units of measurement, and acronyms are to be excluded.

Further, it was clarified that the concepts can also have one-to-many relationships. The final output was a JSON object with 'node1', 'node2', and 'edge' parameters. The concept nodes generated are further refined for entity disambiguation. In this study, the concepts were clustered manually. However, this step can also use NLPs, a second pass over LLMs, or graph-based models. In the next step, the contextual proximities and communities are created depending on the relationships' degree of centrality and edge weights through the Girvan-Newman method. Finally, the knowledge graph is generated from the acquired data. The methodology is also depicted in 'Figure 1'.



Figure 1. Methodology for developing knowledge graphs from unstructured data

4 Case Study

On submission of designs for pre-construction permits to the ULBs, first rule-based checks on the designs are performed. However, in congested metro cities, many deviations are allowed depending on the planning constraints faced. These concessions are approved through the discretionary power of the MC by reading and understanding the hardships encountered by the architects. PDF reports containing arguments and decisions for these subjecting judgments are stored as archival data in the municipality.

The input data for knowledge graph generation were collected from real-life projects from the Municipal Corporation of Greater Mumbai (MCGM), a ULB from the state of Maharashtra, India. The inputs were concession reports submitted by architects to the Assistant Engineer (AE) of MCGM. Hence, the critical comments from AE were also considered in the analysis. The archival data consisted of submissions that were approved for a set of concessions. Therefore, the entities generated from these reports can be considered as concepts leading to successful condonation.

The LLM used for this study was a quantized opensource model that can be run on a local machine. An open-source model was preferred over a proprietary 'GPT-4' model for trial and error runs. These large models can be used once the basic nodes and relationships are saturated and prompt engineering is finalized. The LLMs used are Mistral-instruct-7b and Zephyr-7b models, where Zephyr generated more profound meaningful entities and relationship triplets than Mistral for this study. Hence, the results discussed in this paper will be output from the Zephyr model.

'Figure 2' depicts KGs generated from the full concession report for two projects with contextual proximities. Meanwhile, 'Figure 3' depicts KGs generated for a single concession case, 'Open Space Deficiency' (OSD). As this graph is based on the conceptual arguments for the condonation of a single concession, contextual proximities are not generated. The final KG' Figure 4' is an amalgamation of concepts generated across ten projects for the OSD concession. The outputs from this graph are further analyzed in the next section.

5 Discussion

The KGs generated from complete reports carry a higher number of relational connections due to contextual proximities. Contextual proximity helps identify the clusters of concepts employed for a specific type of concession. This graph also highlights the potential connections between different clusters. Concession reports are legal documents, so it is a widespread practice to use cross-references to DCPR clauses and previous concession numbers. A central node's high density of sub-nodes also emphasizes the concession type's criticality. Different node clusters and connection patterns can be visualized in 'Figure 2', which points out the variability of the concession reports as an input document for automated code compliance. This finding indicates the complexity of new users /architects going across multiple project types and gathering the logical reasoning required to receive a condonation from the municipal commissioner for any violation in design.



Figure 2. Knowledge Graphs with contextual proximities from two different concession reports



Figure 3. KGs without contextual proximities for OSD Concession for two different projects



Figure 4. KG for OSD across ten different projects

A specific concession of OSD was considered to identify and analyze the relationship between conceptual entities in detail. In a metropolitan city like Mumbai, with a population density of 73,000 people per square mile, the OSD application was found to be one of the most common concessions across projects. MCGM follows strict guidelines to prohibit the congested development of high-rises. However, after adding allowable fungible floor space index (FSI) and incentive FSI (according to clauses 31(3) and 33(7)) over the basic FSI, the architects are not able to reach the maximum permissible FSI of the building considering the required setback distances. Therefore, the margin requirements are frequently not fulfilled by the design. The OSD application consists of explanations from the architects stating that the margin violation in the required open space will not violate the fire, health, and safety requirements of the inhabitants and the neighborhood. This explanation can vary between projects with a few critical common concepts. The KGs depicted in 'Figure 3' tries to tie these concepts together.

The concepts extracted in part (b) of 'Figure 3' highlight rehabilitation of non-resident tenements, chief fire officer (CFO) requirements, odd plot shapes, placement of refuge area, and consumption of full permissible FSI as significant challenges. The liaison architects also ensured the use of Indian standards for the seismic resistivity of the building alongside the supervision of a registered structural engineer and site supervisor. In the case of project (a), similar hardship concepts were found in a different relationship pattern. The graph generated in project (b) seemed to enlist more concepts than in project (a). However, project (b) seemed to lack connections between the concept nuclei. Due to the variation in the way the architects express, the LLMs failed to create a generalized concept-relationship ontology. OSD concession arguments across ten projects were concatenated to generate a master concept graph, as shown in 'Figure 4', to tackle this challenge.

The KG generated for ten projects confronted challenges regarding concept name ambiguity. For example, different projects used different acronyms to mention the requirement of a no objection certificate (NOC) from the CFO. Hence, entity resolution became crucial for generating the merged concept graph. The likes of entities such as 'cfo', 'c.f.o', 'c.f.o.', 'cfo NOC', 'NOC from cfo', and 'chief fire officer' were clustered under a single entity, 'CFO'. Post entity resolution, the KG developed a complex relationship among different concepts imitating the tacit knowledge, similar to liaison architects. A combination of concepts in the graph can generate an argument to receive condonation of the required OSD successfully. This graph can help architects decide where their design stands in the form of possibilities for concession application.



Figure 5. Interaction between concepts leading to hardship in achieving code requirements

The yellow box from 'Figure 4' is zoomed into 'Figure 5' for better visibility. In part (a) of the diagram, an impact of the redevelopment project, whose financial viability is driven by the consumption of fungible compensatory FSI and transferrable development rights (TDR), is found. Furthermore, the interaction between

redevelopment projects and public safety was also found. The architects also emphasized planning constraints due to the requirement of clear open spaces and the requirement of existing rehabilitation tenants. Concepts like aesthetic beauty for the environment were an exciting finding, as such hidden concepts might be missed by the human mind while reading reports in the form of PDF documents. In the (b) part, the hardship driven by the plot can be found. On further analysis, the clearance required for existing roads, odd plot shapes, and narrow plot sizes can be identified as the driving factors.



Figure 6. Impact of roles of government personnel in concession acceptance

From the blue box in Figure 4, we derive Figure 6, which shows the role of the CFO in concession approval. With the CFO's NOC, the deficit open space can be condoned, which violates the DCPR 2034 requirements.

Further, it highlights the CFO's decision to influence the completion and building of occupation certifications. Fire safety was also linked with Regulation 47, which indicates the number of fire protection clauses. Thus, the KG performs as a perfect amalgamation for interaction between concepts, government personnel roles, and code requirements.

A similar identification is found in the green box, which is enhanced in 'Figure 7', where the discretionary power of the municipal commissioner and chief engineer is detected. The KG also highlights historical data, such as the maximum joint deficiency approved without charging a premium. The joint deficiency applies to designs with more than one building/wing on a plot, which is correctly connected to neighborhood deficiency, where buildings already exist on the plot.



Figure 7. Edge feature and identification of hidden information

A few limitations of the LLM-based entity generation can be identified in 'Figure 8' and 'Figure 9'. In Figure 8, the height restrictions by the civil aviation department near the airport area are correctly identified. However, another entity identified as connected with the height node only summarizes building height, which was also influenced by the Airport Authority of India (AAI) regulations. Even though both these height restrictions indicate the same relationship, the non-standardized way of representing data influenced erroneous concept generation. On the other hand, the entities generated in 'Figure 9' are disconnected from central KG. However, as the concepts identified are highly case-specific, they can get subdued when a large number of project entities are merged, as found in the red box in 'Figure 4'.

6 Conclusion

Deviations and subjective judgments to design hardships in a densely populated metro city are integral to the compliance verification process. The project characteristics leading to successful concession of deviations are often stored as unstructured reports in the ULBs. A method of knowledge transfer for these postrule-check concession applications is discussed in this paper. Facilitated by the concept graphs, meaningful tacit knowledge was expressed by combining entities and their relationships. The final graph represented complex relationships among concepts, all of which might not be manually extractable without years of domain knowledge. Thus, it can be concluded that, with directed prompt engineering, quantized LLMs can perform noticeably well in generating the node and edge triplets from unstructured data. However, the output suffers from entity disambiguation due to a lack of standardization among report formats and a natural way of expressing arguments. Language models with a significantly higher number of parameters, such as GPT-4, can be used to resolve erroneous results.



Figure 8. Erroneous entity generation by LLM



Figure 9. Highly contextualized concepts in KG

The concept graphs generated with the concatenation of knowledge gathered across projects develop a master database. These KGs can be used by new architects, reviewers, or programmers who are likely to develop automated concession application systems in the near future. These systems can gather data for concept nodes identified in the KG from enriched BIM models and predict the acceptability of concession, given the design specifications through ML models. On the other hand, future research can also explore the possibility of generating enriched XML files from KG that can address complex relationships through multiple paths to improve machine readability. Further, the complex reasoning capabilities of KGs can be used across the construction project lifecycle, such as construction progress monitoring and constraint checking. Thus, applying AIbased knowledge transfer mechanisms will only solidify the path for structured data acquisition protocols, leading to the industry's highly required automated compliance checking system.

Acknowledgment

The authors acknowledge and thank for the support received from Softech Engineers Limited through sponsored research project number RD/0120-SOFTT18-001 of IIT Bombay.

References

- C. Eastman, J. min Lee, Y. suk Jeong, and J. kook Lee, "Automatic rule-based checking of building designs," *Automation in Construction*, vol. 18, no. 8. Elsevier, pp. 1011–1033, Dec. 01, 2009. doi: 10.1016/j.autcon.2009.07.002.
- [2] A. Karmakar and V. S. K. Delhi, "Requirements of Machine Learning and Semantic Enrichment for BIM-Based Automated Code Compliance Checking: A Focus Group Study," in *International Conference on Computing in Civil & Building Engineering*, no. 19, Cape Town, 2023, pp. 65–74. doi: 10.1007/978-3-031-35399-4_6.
- W. Solihin and C. Eastman, "Classification of rules for automated BIM rule checking development," *Autom. Constr.*, vol. 53, pp. 69– 82, May 2015, doi: 10.1016/j.autcon.2015.03.003.
- [4] A. McCallum, "Information Extraction," *Queue*, vol. 3, no. 9, pp. 48–57, Nov. 2005, doi: 10.1145/1105664.1105679.
- [5] A. S k or, "Knowledge Extr c tion from Unstructured D t ," Gottfried Wilhelm Lei niz Universität, 2023. doi: https://doi.org/10.15488/13721.
- [6] A. Singhal, "Introducing the Knowledge Graph: things, not strings," *Google Blog*, 2012. https://www.blog.google/products/search/introd ucing-knowledge-graph-things-not/ (accessed Dec. 19, 2023).
- S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A Nucleus for a Web of Open Data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), vol. 4825 LNCS, Springer, Berlin, Heidelberg, 2007, pp. 722–735. doi: 10.1007/978-3-540-76298-0_52.
- [8] R. Reinanda, E. Meij, and M. de Rijke, "Knowledge Graphs: An Information Retrieval Perspective," *Found. Trends*® *Inf. Retr.*, vol. 14, no. 4, pp. 289–444, 2020, doi: 10.1561/1500000063.

- [9] M. Chen, Y. Tian, M. Yang, and C. Zaniolo, "Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment," in *IJCAI International Joint Conference on Artificial Intelligence*, Nov. 2016, vol. 0, pp. 1511–1517. doi: https://doi.org/10.48550/arXiv.1611.03954.
- [10] Y. Zhu, C. Zhang, C. Ré, and L. Fei-Fei, "Building a Large-scale Multimodal Knowledge Base System for Answering Visual Queries," Jul. 2015, doi: https://arxiv.org/abs/1507.05670v2.
- [11] R. Krishna *et al.*, "Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations," *Int. J. Comput. Vis.*, vol. 123, no. 1, pp. 32–73, Feb. 2016, doi: 10.1007/s11263-016-0981-7.
- [12] R. Reinanda, E. Meij, and M. de Rijke, "Knowledge Graphs: An Information Retrieval Perspective," *Found. Trends*® *Inf. Retr.*, vol. 14, no. 4, pp. 289–444, 2020, doi: 10.1561/150000063.
- [13] D. Fensel *et al.*, "Introduction: What Is a Knowledge Graph?," in *Knowledge Graphs*, Cham: Springer International Publishing, 2020, pp. 1–10. doi: 10.1007/978-3-030-37439-6_1.
- [14] Z. Yang, Y. Bi, L. Wang, D. Cao, R. Li, and Q. Li, "Development and application of a field knowledge graph and search engine for pavement engineering," *Sci. Rep.*, vol. 12, no. 1, p. 7796, May 2022, doi: 10.1038/s41598-022-11604-y.
- [15] Y. Zhang, J. Liu, and K. Hou, "Building a Knowledge Base of Bridge Maintenance Using Knowledge Graph," *Adv. Civ. Eng.*, vol. 2023, pp. 1–16, Apr. 2023, doi: 10.1155/2023/6047489.
- [16] M. H. Rasmussen, M. Lefrançois, P. Pauwels, C. A. Hviid, and J. Karlshøj, "Managing interrelated project information in AEC Knowledge Graphs," *Autom. Constr.*, vol. 108, p. 102956, Dec. 2019, doi: 10.1016/j.autcon.2019.102956.
- [17] W. Wu, C. Wen, Q. Yuan, Q. Chen, and Y. Cao, "Construction and application of knowledge graph for construction accidents based on deep learning," *Eng. Constr. Archit. Manag.*, no. ahead-of-print, Sep. 2023, doi: 10.1108/ECAM-03-2023-0255.
- [18] J. Wang, L. Mu, J. Zhang, X. Zhou, and J. Li, "On Intelligent Fire Drawings Review Based on Building Information Modeling and Knowledge Graph," Constr. Res. Congr. 2020 Comput. Appl. - Sel. Pap. from Constr. Res. Congr. 2020, pp. 812–820, 2020, doi:

10.1061/9780784482865.086.

- [19] A. Karmakar and V. S. K. Delhi, "Construction 4.0: what we know and where we are headed?," *J. Inf. Technol. Constr.*, vol. 26, no. July, pp. 526–545, 2021, doi: 10.36680/j.itcon.2021.028.
- [20] C. Ramonell, R. Chacón, and H. Posada, "Knowledge graph-based data integration system for digital twins of built assets," *Autom. Constr.*, vol. 156, p. 105109, Dec. 2023, doi: 10.1016/J.AUTCON.2023.105109.
- [21] X. Li, C. Wu, Z. Yang, Y. Guo, and R. Jiang, "Knowledge graph-enabled adaptive work packaging approach in modular construction," *Knowledge-Based Syst.*, vol. 260, p. 110115, Jan. 2023, doi: 10.1016/J.KNOSYS.2022.110115.
- [22] Z. Pan, C. Su, Y. Deng, and J. Cheng, "Video2Entities: A computer vision-based entity extraction framework for updating the architecture, engineering and construction industry knowledge graphs," *Autom. Constr.*, vol. 125, p. 103617, May 2021, doi: 10.1016/J.AUTCON.2021.103617.
- [23] Z. Pan, C. Su, Y. Deng, and J. Cheng, "Image2Triplets: A computer vision-based explicit relationship extraction framework for updating construction activity knowledge graphs," *Comput. Ind.*, vol. 137, p. 103610, May 2022, doi: 10.1016/J.COMPIND.2022.103610.
- [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: Data, Sensing, and Analytics*, 2023, pp. 186–193. doi: 10.1061/9780784485224.023.
- [25] F. Heger and D. Napps, "Floorplan2IFC: Transformation of building floor plans into hierarchical IFC-graphs," in 34. ForumBauinformatik, no. September, 2023, pp. 19–26. doi: 10.13154/294-10084.
- [26] D. Napps, D. Pawlowski, and M. König, "BIMbased variant retrieval of building designs using case-based reasoning and pattern matching," in *Proceedings of the International Symposium on Automation and Robotics in Construction*, Nov. 2021, vol. 2021-Novem, no. December, pp. 435– 442. doi: 10.22260/ISARC2021/0060.
- [27] D. Napps, A. Zahedi, M. König, and F. Petzold, "Visualisation and graph-based storage of customised changes in early design phases," in *Proceedings of the International Symposium on Automation and Robotics in Construction*, 2022,

vol. 2022-July, no. July, pp. 191-198. doi: 10.22260/isarc2022/0028.

- [28] D. Napps, J. Maibaum, and M. König, "Identifying cost-efficient alternatives to building designs based on building regulations," in Proc. of the 30th Int. Conference on Intelligent Computing in Engineering (EG-ICE), 2023, pp. 1–10.
- [29] D. Napps and M. König, "Consideration of detailing in the graph-based retrieval of design variants," eWork Ebus. Archit. Eng. Constr. -Proc. 14th Eur. Conf. Prod. Process Model. ECPPM 2022, pp. 389–396, Mar. 2023, doi: 10.1201/9781003354222-50/CONSIDERATION-DETAILING-GRAPH-BASED-RETRIEVAL-DESIGN-VARIANTS-NAPPS-K.
- [30] Z. Wang, H. Ying, R. Sacks, and A. Borrmann, "CBIM: A Graph-based Approach to Enhance Interoperability Using Semantic Enrichment," Apr. 2023, Accessed: Feb. 18, 2024. [Online]. Available: https://arxiv.org/abs/2304.11672v1
- [31] Y. Zhou, W. Solihin, and J. K. W. Yeoh, "Facilitating Knowledge Transfer during Code Compliance Checking Using Conceptual Graphs," J. Comput. Civ. Eng., vol. 37, no. 5, 2023, doi: 10.1061/jccee5.cpeng-4884.
- P. Pauwels, A. Costin, and M. H. Rasmussen, "Knowledge Graphs and Linked Data for the Built Environment," *Struct. Integr.*, vol. 20, pp. 157–183, 2022, doi: 10.1007/978-3-030-82430-3_7/COVER.
- [33] SoftTech Engineers Limited, "AutoDCR." https://softtech-engr.com/autodcr/ (accessed Dec. 15, 2023).

Towards Automated Physics-Based Modeling: Fusion of Construction Equipment Data for Efficient Simulation

Liqun Xu¹, Dharmaraj Veeramani², and

Zhenhua Zhu³

^{1,3}Department of Civil and Environmental Engineering, University of Wisconsin-Madison, USA ²Department of Industrial and Systems Engineering, University of Wisconsin-Madison, USA <u>liqun.xu@wisc.edu</u>, raj.veeramani@wisc.edu, zzhu286@wisc.edu

Abstract

Physics-based simulations play a crucial role in design and development of autonomous the construction equipment. Currently, a key challenge in these simulations is the time-intensive task of preparing construction equipment models that accurately represent both the equipment's geometry and physics. Manual model creation for simulations becomes particularly laborious due to the integration of diverse mechanical data such as materials, joints, and drives with the geometric data. Extant methods for automatic physics-based modeling of standard modular robots are inadequate for addressing the complexities of construction equipment. Therefore, this paper investigates the feasibility of automating and streamlining the physics-based modeling process by fusing the construction equipment's mechanical data into its 3D computer-aided design (CAD) model. The proposed method involves converting the construction equipment 3D CAD model into a universal scene description (USD) model for efficient data fusion. Subsequently, the method automatically configures material parameters, collision meshes, establishes component relations, and incorporates joints and drives for the USD model. To validate the efficacy of this approach, the proposed method is applied to create a physics-based model of a Caterpillar 390F LME excavator, and simulated in a scalable robotic simulator (NVIDIA Isaac Sim). The findings demonstrate that the proposed method significantly reduces the time required for physicsbased modeling compared to traditional manual methods.

Keywords -

Construction equipment modeling, Data fusion, Physics-based simulation

1 Introduction

Physics-based simulations play a pivotal role in facilitating the design and advancement of autonomous construction equipment such as automated excavators and trucks [1]. These simulations have provided an accelerated and safe approach to train, validate, and test the control algorithms and prototype designs of autonomous construction equipment before real-world implementation [1,2]. Moreover, in the quest to leverage deep learning for developing AI-enabled autonomous construction equipment, physics-based simulations can generate a wealth of annotated training data in a short amount of time [3]. This is particularly valuable in situations where data is difficult to obtain in the real world. In addition, physics-based simulations are increasingly employed by researchers to apply and refine reinforcement learning algorithms, thereby enhancing the operational intelligence of autonomous construction equipment [4,5].

Despite the advantages, a key challenge in utilizing physics-based simulations is the preparation and generation of construction equipment models that can accurately represent equipment geometry and physics. Inaccurate models can result in a sim-to-real gap, where algorithms and designs proven in simulations fail when applied in the real world [6]. Although simulation platforms such as Unity, Gazebo, and Isaac Sim offer environments to build these models, the physics-based modeling process is still time-consuming and requires modeling expertise [1,7]. This is because complex data required for the simulation, such as construction equipment materials, joints, drives, etc. need to be integrated during the modeling process, which requires a lot of manual work [2].

Some studies have proposed the use of automatic physics-based modeling methods to reduce the manual effort in the context of modular robots. Modular robots are systems composed of standardized modules, which can be combined in various ways to adapt to different tasks or environments [8]. Jace et. al [9] presented an automatic approach to model the robot kinematics and dynamics for modular robots, given only the module data and their arrangements. Maddalena et. al [7] proposed an algorithm that takes as input the Unified Robotics Description Format (URDF) files of the single modules with their desired arrangement and provides the final URDF of the assembled robot as a result. However, these methods still require manual configuration of modeling data such as materials and joints for each module. This is not a burden for modeling modular robots because only a few modules need to be configured and they can be reused for robot modeling. However, these methods are not effective in reducing the workload of physics-based modeling of construction equipment as they are far more complex and varied than standard modular robots.

This paper introduces an automated physics-based modeling method through data fusion to streamline the process of creating equipment models for physics-based simulation. Inputs to this method include a construction equipment 3D CAD model along with required data for physics-based modeling, including materials, collision meshes, component relationships, joints, and drives. Initially, the 3D CAD model of construction equipment is converted into a universal scene description (USD) using Isaac Sim to facilitate data fusion. Then, our method automatically fuses the required data for simulation into the USD model. To demonstrate the effectiveness of our method, we created models of an excavator (Caterpillar 390F LME) using both the proposed method and the manual modeling method. Then we compared the time required by these two methods to complete the modeling. The results show that our method can greatly improve the modeling efficiency, and thereby can help promote the application of physics-based simulation in the development of autonomous construction equipment.

2 Literature Review

2.1 Physics-based Modeling and Simulation Platforms

The evolution of simulation platforms such as Unity, Unreal Engine, Gazebo, Isaac Sim, and Webots has significantly impacted the field of robotics. These platforms offer diverse functionalities and environments for robot modeling, each with unique characteristics that distinguish them from one another [10].

Unity, primarily known for its widespread use in game development, has emerged as a versatile platform for robot simulation. Its user-friendly interface and robust physics engine make it an attractive choice for simulating complex robotic systems [11]. Unity's real-time 3D development capabilities enable the creation of detailed and dynamic environments, which are essential for testing the interaction of robots with their surroundings. The platform supports a wide range of robot models, from simple wheeled robots to more complex humanoid robots, allowing for extensive experimentation and research in robotics [10].

Unreal Engine stands out for its high-fidelity graphics and realistic simulation environments [12]. This platform is particularly favored for applications requiring photorealistic rendering, such as autonomous vehicle testing [13]. Unreal Engine's advanced lighting and shading capabilities contribute to creating highly immersive simulation scenarios. It is adept at simulating sophisticated robot models, including drones and autonomous vehicles, providing a realistic platform for testing sensors and navigation algorithms [14].

Gazebo, an open-source simulation platform, is renowned for its strong community support and extensive library of robot models and environments [15]. Its ability to simulate both indoor and outdoor environments with various physics engines makes it a versatile tool for robotics research. Gazebo is particularly popular for simulating multi-robot systems, such as swarm robots, and has been instrumental in numerous robotics competitions and research projects [16].

Webots is a user-friendly, cross-platform simulation software widely used in education and research. Its ease of use and comprehensive documentation make it accessible to both beginners and experienced users [17]. Webots support a broad range of robot models, from simple mobile robots to more advanced humanoid robots, making it versatile tool for various robotic applications.

Isaac Sim, developed by NVIDIA, is tailored for robotics applications involving artificial intelligence (AI) and machine learning (ML). Its integration with NVIDIA's GPU technology enables high-performance simulations, crucial for training and testing AI algorithms [18]. Isaac Sim is adept at simulating complex robotic systems, such as robotic arms and mobile robots, and is particularly beneficial for scenarios involving ML and sensor processing.

2.2 Physics-based Modeling of Construction Equipment

Physics-based modeling has been widely used in autonomous construction equipment training and testing. To demonstrate control of large robots to perform construction tasks, Lei et al. [19] created a construction robot hand model in Isaac Sim, and trained it via reinforcement and imitation learning to conduct operations with 6 types of construction tools, such as power drill, flat screwdriver, adjustable wrench, etc. Similarly, Sungjin et al. [20] employed Gazebo for dynamic modeling of spraying robots, evaluating their performance in construction tasks like indoor wall painting. Jaco et al., [21] built a wheeled robot model using Gazebo and then trained a map corner-based navigation model in a virtual world. Lofgren et al. [22] advanced this field by simulating an underground loader in Unity, training a deep reinforcement learning controller that autonomously adapts to varying terrains and soil conditions. Azulay and Shapiro [23] also used Gazebo for wheel loader modeling, achieving a controller adept at complex earthmoving tasks, and showcasing the potential for automation in construction.

Furthermore, physics-based simulation is used for generating synthetic data. Wilfredo et al. [24] used Unity to simulate excavator postures, creating a dataset that bypasses the need for time-intensive manual annotation. Jia et al. [25] established a drone model in Unity for capturing simulated dam images, facilitating the training of dam defect detection algorithms.

2.3 Automatic Physics-based Modeling

Despite these available physics-based modeling and simulation platforms, manually building models in

simulation platforms is still time-consuming and requires modeling expertise [7]. Some studies have proposed the use of automatic modeling methods to reduce manual modeling effort, and have investigated the automatic modeling process in the context of modular robots. For example, [9] introduced an automated approach for modeling robot kinematics and dynamics, requiring only module data such as joints, drives, etc. and their arrangements. Maddalena et. al [7] proposed an algorithm that takes as input the URDF files of the single modules with their desired arrangement and provides the final URDF of the assembled robot as a result.

However, these methods have shortcomings that limit their use for modeling complex construction equipment. They require manual configuration of joints and additional data within the modeling software, obligating users to acquire proficiency in the software itself. Furthermore, for construction equipment lacking a substantial array of universal modules, these methods offer no advantage over direct manual modeling, thereby confining their applicability primarily to modular robotics.

3 Methodology



Figure 1. Overview of the automatic physics-based modeling method

The proposed method is designed to automate the physics-based modeling process of construction equipment. The inputs for the method include a 3D CAD model of construction equipment and other data required for physics-based modeling including materials, collision meshes, relations of components, joints, and drives. As shown in Figure 1, the 3D CAD model is initially converted into a universal scene description (USD) model using Isaac Sim to facilitate data fusion. Then, the various data are fused into the USD model. The overall fusion process comprises four steps: Firstly, the materials of components are set based on the materials data, which enables the components to have material information such as density and friction coefficient. Secondly, collision meshes are set to their corresponding components. This allows the components to emulate collision behaviors. Thirdly, motion dependencies between the components are established based on their relationships, which allows components to move with their dependent components accordingly. For example, bucket bolts will move with the bucket. Lastly, joints and drives are created based on their mechanical properties, which allows the components connected with the joints to move accordingly. At this point, the USD model integrated with the data is ready for simulation.

3.1 Model Conversion

To facilitate the fusion of data requisite for the simulation, the initial step involves converting the 3D CAD model of construction equipment into a USD model. USD is an open-source 3D scene description file format developed by Pixar. It can be used for 3D content creation and interchange among different tools [26,27].

There are two reasons for choosing the USD format. The first is its dual support for both intricate construction equipment modeling and complex environmental constructs [28], thus facilitating the import of construction equipment models into its operating environment. Meanwhile, USD has a Python Application Programming Interface (API), and the USD model can be easily customized through Python script [29]. This conversion lays the groundwork for subsequent data fusion. Moreover, USD supports a variety of simulation platforms including Unity [30], Unreal [30], Isaac Sim [31], etc.

3.2 Data Fusion

Subsequent to the model conversion, data fusion can start. The first step of data fusion focuses on setting material properties to each component of the USD model. This involves a process where both the mechanical and aesthetic properties of materials are defined. Mechanical properties include aspects such as density, elasticity, and friction coefficients, which are crucial for accurate physical interactions in the simulation. These mechanical properties can be retrieved from technical specifications provided by construction equipment manufacturers. Aesthetic properties, on the other hand, involve visual characteristics like color, texture, and reflectivity, enhancing the visual realism of the model. Color codes and textures for construction equipment can be obtained through Internet search. Reflectivity can be determined by selecting the corresponding material from the rendering software, such as stainless steel, paint, etc. To organize and store these properties, a JavaScript Object Notation (JSON) document, designated as JSON-1, is compiled. This document serves as a comprehensive repository for the material attributes. Following this, an automated process retrieves the components and their material parameters from the document. The USD model is then systematically scanned, and the documented material parameters are applied to each respective component. This mapping ensures that each component of the USD model is a true-to-life representation of its physical counterpart, mirroring it in both functionality and appearance.

The method then employs an automated process to parse JSON-1 and integrate the material properties into the USD model, ensuring each component reflects its real-world counterpart both in function and form.

Following the materials setting, the second step focuses on setting collision meshes to components of the USD model. Collision meshes are simplified representations of the physical shape of each component. It provides an efficient way to detect and respond to collisions between robot components and/or their surrounding environment. The setting of collision meshes enables the construction equipment model to have realistic interaction within the simulation environment. Our method provides a range of collision mesh estimation methods for users to choose from, including triangle mesh, convex decomposition, convex hull, etc. In this step, JSON-2 document is created to encapsulate the collision mesh estimation method for each component in the USD model. The proposed method then systematically parses this document, extracting component names and their corresponding collision mesh estimation methods, and cataloging them into a Python dictionary structure where the component name is the key, and the collision mesh method is the value. Following this, all components in the USD model are traversed and collision meshes are set for the components listed in the dictionary accordingly.

The third step involves setting up relationships between components. The mutual relationship amongst components determines which components can move together as a group. For example, the excavator bucket body and bolts on it move when the bucket moves, as the bolt and body are connected to the bucket, as shown in Figure 1. In order to encapsulate the relationships among components, this method creates a relationship document JSON-3. This document employs a dictionary structure to chronicle the relationships among the various components. Keys are utilized to denote the names of dependent components, while values enumerate the associated followers. For example, the above relation is encoded as {"bucket": ["bucket body", "bucket bolt1", "bucket bolt2" ... "bucket bolt6"]}. Within the USD model, these intercomponent relationships are depicted through parent-child hierarchies. All components that are children of a parent component are expected to move together when the parent component moves. Our technique proceeds to parse these relationships from the dictionary and then traverse all the components in the USD model hierarchically to check whether the current child-father relationship of each component is consistent with that in the dictionary. If any discrepancies are found, the parent and child designations are realigned accordingly. This iterative process continues until every component relationship in the USD model has been validated, guaranteeing an accurate representation of movement in the simulation environment.

The last step is to add the joints and drives. In simulations, a joint refers to a functional connection between rigid bodies that facilitates a specific range of relative motion between them. This motion is typically enabled by a drive mechanism. For instance, the rotational movement of car wheels around an axle is attributed to revolute joints. If a wheel is designated as powered, a corresponding drive will be added to actuate it. In our approach, details pertaining to the joints and drives are stored in a document called JSON-4. This document includes the designation of the joint, the components it connects to, the associated drive mechanisms, and the physical parameters of the joint, such as damping coefficient, and stiffness, among others. These parameters can be obtained by from construction equipment manufacturers or by theoretical calculations. After obtaining this data, this method will add these joints and drives accordingly to the USD model.

4 Implementation and Results

To verify the effectiveness of the proposed method, we established the same excavator model on the Isaac Sim platform using both the manual modeling method and the proposed method. Then we compared the modeling time required by the two methods.

4.1 Experimental Environment

The experimental environment used in this study includes a server with an AMD Ryzen 9 5950X CPU running Ubuntu 20.04 system, NVIDIA GeForce RTX 3090Ti GPU with 24G memory of a single graphics card, Nvidia Isaac Sim 23.01.

The construction equipment selected for this experiment is a Caterpillar 390F LME excavator. Its 3D CAD model is downloaded from GRABCAD, as shown in Figure 2. This particular model is engineered with four hydraulic cylinders responsible for actuating the movement of its boom, arm, and bucket. Additionally, it features a swing joint that facilitates the rotation of its upper structure and two actuated sprockets that empower the excavator to advance, retreat, and turn. A breakdown of all the joints and drives incorporated in this model is systematically cataloged in Table 1. This excavator is composed of 142 joints, with a majority of 138 revolute joints that allow for rotational movements and 4 prismatic joints that facilitate linear actions. In conjunction with this, the model incorporates 7 distinct drives, which are instrumental in actuating the various movements of this excavator.



Figure 2. The 3D CAD model of Caterpillar 390F LME

Table 1 Joints and drives statics

Joint/drive type	Number
Revolute joint	138
Prismatic joint	4
Angular drive	3
Linear drive	4

4.2 Results

Upon integrating the CAD model of the excavator with the JSON documents comprising the requisite data for physics-based modeling, we successfully generated an excavator model ready for simulation. Subsequently, in our evaluation conducted on the simulation platform, we tested an array of motion functions pertinent to the excavator. The outcomes of these tests demonstrated that our model is adept at replicating all the essential motion functions of the excavator. This is depicted in Figure 3, which showcases a sequence of video frames extracted from the simulation outcomes. These frames distinctly highlight the model's capability to accurately simulate the dynamic movements of critical components, including the boom, arm, bucket, track, and the excavator's upper body. To determine the efficiency of our proposed method, we compared the modeling time required by the proposed method and the manual method. Before starting this modeling experiment, we had no experience in robot modeling. During the experiment, we recorded the time spent learning to model using Isaac Sim, which took a total of 28 hours. The duration of excavator modeling from start to completion was then recorded, which lasted 72 hours. Finally, the time required for excavator modeling through the method proposed in this study was recorded, which was 8 hours. Almost all of these 8 hours were used to prepare the data required for modeling. The running time of the automatic modeling program was almost negligible, lasting less than 2 seconds. This comparison, shown in Table 2, reveals a stark contrast in time investment. The manual modeling process required approximately 100 hours. In contrast, our method needed only 8 hours to prepare documentation for materials, collision meshes, component relations, and joints and drives needed for physics-based modeling. Upon completion of this preparatory phase, the true efficacy of our approach becomes evident. It automatically fuses all the data from these documents into the final model, accomplishing this complex integration in an instant. This significant reduction in time, without compromising accuracy or detail, underscores the potential of our method to revolutionize the efficiency of physics-based modeling processes.



Figure 3. Video frames of excavator simulation in Isaac Sim platform

Modeling method	Time (hour)
Manual modeling	100
Our method	8

5 Conclusion

Physics-based simulations offer a rapid and secure platform for training, validating, and testing control algorithms, as well as prototyping designs for autonomous construction equipment. Our paper introduces an innovative method to automate the physicsbased modeling process using data fusion. This method transforms a 3D CAD model of construction equipment into a USD model, seamlessly integrating the necessary data for physics-based modeling through data fusion. When users apply the proposed method to build other construction equipment models for physics-based simulation, they only need to prepare the 3D CAD model of the construction equipment and the corresponding physical data. Our method can then be used to automatically integrate these data into the CAD model, enabling physics-based modeling. The study benchmarks the efficiency of this automated method against traditional manual modeling within an identical experimental setup, revealing marked enhancements in modeling efficiency.

However, it is imperative to acknowledge certain limitations of our method. Firstly, it requires users to manually prepare the initial data required for physicsbased modeling. Additionally, as of now, the method does not possess the capability to incorporate sensors into the models. In the future, we will integrate a large language model such as ChatGPT, into the physics-based modeling pipeline. It will be used to extract and prepare initial data from construction equipment technical specifications, eliminating the need for users to manually prepare this data. In addition, we will develop a function for adding sensors. Using this functions, users would only need to input the sensor type, location, and other parameters, and the sensor would be automatically integrated into the model.

References

- O. Wong Chong, J. Zhang, R.M. Voyles, B.C. Min, BIM-based simulation of construction robotics in the assembly process of wood frames, Autom Constr 137 (2022) 104194. https://doi.org/10.1016/J.AUTCON.2022.10419 4.
- C.K. Liu, D. Negrut, The Role of Physics-Based Simulators in Robotics, Annu Rev Control Robot Auton Syst 4 (2021) 35–58. https://doi.org/10.1146/annurev-control-072220-093055.
- [3] S. Mukhopadhyay, Y. Chen, S. Morais, N. Cennamo, J. Lee, J. Boone, C. Goodin, L. Dabbiru, C. Hudson, L. Cagle, D. Carruth, Training Artificial Intelligence Algorithms with Automatically Labelled UAV Data from Physics-Based Simulation Software, Applied Sciences 2023, Vol. 13, Page 131 13 (2022) 131. https://doi.org/10.3390/APP13010131.
- [4] A.A. Apolinarska, M. Pacher, H. Li, N. Cote, R. Pastrana, F. Gramazio, M. Kohler, Robotic assembly of timber joints using reinforcement learning, Autom Constr 125 (2021) 103569. https://doi.org/10.1016/J.AUTCON.2021.10356 9.
- [5] O. Azulay, A. Shapiro, Wheel Loader Scooping Controller Using Deep Reinforcement Learning, IEEE Access 9 (2021) 24145–24154. https://doi.org/10.1109/ACCESS.2021.3056625.
- Y. Chebotar, A. Handa, V. Makoviychuk, M. MacKlin, J. Issac, N. Ratliff, Di. Fox, Closing the sim-to-real Loop: Adapting simulation randomization with real world experience, Proc IEEE Int Conf Robot Autom 2019-May (2019) 8973–8979. https://doi.org/10.1109/ICRA.2019.8793789.
- [7] M. Feder, A. Giusti, R. Vidoni, An approach for automatic generation of the URDF file of modular robots from modules designed using SolidWorks, Procedia Comput Sci 200 (2022) 858-864.

https://doi.org/10.1016/J.PROCS.2022.01.283.
[8] V. Nezhadali, O.K. Kayani, H. Razzaq, M. Tarkian, EVALUATION OF AN AUTOMATED DESIGN AND

OPTIMIZATION FRAMEWORK FOR MODULAR ROBOTS USING A PHYSICAL PROTOTYPE, DS 68-4: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 4: Product and Systems Design, Lyngby/Copenhagen, Denmark, 15.-19.08.2011 (2011)195-204. https://www.designsociety.org/publication/3054 5/EVALUATION+OF+AN+AUTOMATED+D ESIGN+AND+OPTIMIZATION+FRAMEWO RK+FOR+MODULAR+ROBOTS+USING+A +PHYSICAL+PROTOTYPE (accessed December 8, 2023).

- [9] C. Nainer, M. Feder, A. Giusti, Automatic Generation of Kinematics and Dynamics Model Descriptions for Modular Reconfigurable Robot Manipulators, IEEE International Conference on Automation Science and Engineering 2021-August (2021) 45–52. https://doi.org/10.1109/CASE49439.2021.9551 680.
- [10] J. Collins, S. Chand, A. Vanderkop, D. Howard, A review of physics simulators for robotic applications, IEEE Access 9 (2021) 51416– 51431.

https://doi.org/10.1109/ACCESS.2021.3068769.

- [11] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, D. Lange, Unity: A General Platform for Intelligent Agents, (2018). https://arxiv.org/abs/1809.02627v2 (accessed December 13, 2023).
- [12] C. Symeonidis, N. Nikolaidis, Simulation environments, Deep Learning for Robot Perception and Cognition (2022) 461–490. https://doi.org/10.1016/B978-0-32-385787-1.00023-3.
- Tom. Shannon, Unreal Engine 4 for Design Visualization Developing Stunning Interactive Visualizations, Animations, and Renderings., (2017). https://books.google.com/books/about/Unreal_E ngine_4_for_Design_Visualization.html?id=Hf YtDwAAQBAJ (accessed December 13, 2023).
- [14] B. Alvey, D.T. Anderson, A. Buck, M. Deardorff, G. Scott, J.M. Keller, Simulated Photorealistic Deep Learning Framework and Workflows To Accelerate Computer Vision and Unmanned Aerial Vehicle Research, (2021) 3889–3898. https://github.com/ (accessed December 13, 2023).
- [15] N. Koenig, A. Howard, Design and use paradigms for Gazebo, an open-source multirobot simulator, 2004 IEEE/RSJ International

Conference on Intelligent Robots and Systems(IROS)3(2004)2149–2154.https://doi.org/10.1109/IROS.2004.1389727.

- J. Harbin, S. Gerasimou, N. Matragkas, A. Zolotas, R. Calinescu, Model-Driven Simulation-Based Analysis for Multi-Robot Systems, Proceedings 24th International Conference on Model-Driven Engineering Languages and Systems, MODELS 2021 (2021) 331–341. https://doi.org/10.1109/MODELS50736.2021.0 0040.
- [17] O. Michel, WebotsTM: Professional Mobile Robot Simulation, (2004). https://arxiv.org/abs/cs/0412052v1 (accessed December 13, 2023).
- V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, G. State, Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning, (2021). https://arxiv.org/abs/2108.10470v2 (accessed December 13, 2023).
- [19] L. Huang, W. Cai, Z. Zhu, Z. Zou, Dexterous manipulation of construction tools using anthropomorphic robotic hand, Autom Constr 156 (2023) 105133. https://doi.org/10.1016/J.AUTCON.2023.10513 3.
- [20] S. Kim, M. Peavy, P.C. Huang, K. Kim, Development of BIM-integrated construction robot task planning and simulation system, Autom Constr 127 (2021) 103720. https://doi.org/10.1016/J.AUTCON.2021.10372 0.
- [21] J.C. Virgolino Soares, G.F. Abati, G.H. Duarte Meggiolaro, Lima, M.A. Autonomous Navigation System for a Wall-painting Robot based on Map Corners, 2020 Latin American Symposium, Robotics 2020 Brazilian Symposium on Robotics and 2020 Workshop on Robotics in Education, LARS-SBR-WRE 2020 (2020).https://doi.org/10.1109/LARS/SBR/WRE51543.

https://doi.org/10.1109/LARS/SBR/WRE51543. 2020.9306998.

- S. Backman, D. Lindmark, K. Bodin, M. Servin, J. Mörk, H. Löfgren, Continuous Control of an Underground Loader Using Deep Reinforcement Learning, Machines 2021, Vol. 9, Page 216 9 (2021) 216. https://doi.org/10.3390/MACHINES9100216.
- [23] O. Azulay, A. Shapiro, Wheel Loader Scooping Controller Using Deep Reinforcement Learning, IEEE Access 9 (2021) 24145–24154. https://doi.org/10.1109/ACCESS.2021.3056625.

- W. Torres Calderon, D. Roberts, M. Golparvar-Fard, Synthesizing Pose Sequences from 3D Assets for Vision-Based Activity Analysis, Journal of Computing in Civil Engineering 35 (2021) 04020052. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000937/ASSET/4F7E1E4F-2244-4815-B1E8-4398B6104731/ASSETS/IMAGES/LARGE/FI GURE14.JPG.
- [25] J. Xu, C. Yuan, J. Gu, J. Liu, J. An, Q. Kong, Innovative synthetic data augmentation for dam crack detection, segmentation, and quantification, Struct Health Monit 22 (2023) 2402–2426. https://doi.org/10.1177/14759217221122318/A SSET/IMAGES/LARGE/10.1177_1475921722 1122318-FIG19.JPEG.
- [26] M.A. Bolstad, Large-Scale Cinematic Visualization Using Universal Scene Description, 2019 IEEE 9th Symposium on Large Data Analysis and Visualization, LDAV 2019 (2019) 85–86.

https://doi.org/10.1109/LDAV48142.2019.8944 362.

- [27] Pixar Animation Studios, (n.d.). https://www.pixar.com/usd (accessed March 9, 2024).
- [28] Introduction to USD Universal Scene Description 24.03 documentation, (n.d.). https://openusd.org/release/intro.html#usd-canrepresent (accessed March 9, 2024).
- [29] Working with USD Python Libraries | NVIDIA Developer, (n.d.). https://developer.nvidia.com/usd/tutorials (accessed March 9, 2024).
- [30] Universal Scene Description in Unreal Engine | Unreal Engine 4.27 Documentation, (n.d.). https://docs.unrealengine.com/4.27/en-US/WorkingWithContent/USD/USDinUE4/ (accessed March 9, 2024).
- [31] Universal Scene Description (USD) 3D Framework | NVIDIA, (n.d.). https://www.nvidia.com/en-us/omniverse/usd/ (accessed March 9, 2024).

Reinforcement Learning-Enhanced Path Planning for Mobile Cranes in Dynamic Construction Environments: A Virtual Reality Simulation Approach

Rafik Lemouchi¹, Mohamed Assaf¹, Ahmed Bouferguene² and Mohamed Al-hussein¹

¹Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada ²Campus Saint-Jean, University of Alberta, Edmonton, Canada <u>lemouchi@ualberta.ca</u>, <u>massaf2@ualberta.ca</u>, <u>hemouchi@ualberta.ca</u>, <u>malhussein@ualberta.ca</u>

Abstract

This paper introduces a novel approach to crane path planning on construction sites through the utilization of Reinforcement Learning (RL) and Virtual Reality (VR) simulations. The strategy includes a comprehensive simulation model that incorporates an agent, actions, states, environment, and a reward system. After undergoing extensive training across millions of episodes, the crane agent has acquired optimal path-planning techniques that enhance lifting time, manage energy consumption, and improve collision detection. The results highlight the agent's impressive growth from initial exploration to peak efficiency, represented by cumulative rewards and evolving simulation times. The findings also demonstrate the effectiveness of RL-based path planning in maneuvering dynamic construction environments and optimizing crane operations.

Keywords –

Path Planning, Simulation, Reinforcement Learning, Virtual Reality.

1 Introduction

The construction industry constantly evolves, aiming to maximize efficiency and minimize costs. Over the past few years, off-site construction has gained traction due to its ability to save time and money. This approach involves transporting prefabricated modules to the construction site for installation, making cranes an essential component.

Despite numerous attempts to improve crane operations, outdated tools and planning methodologies still need to be used. According to [1,2], current practices involve lift engineers generating CAD-based 2D and 3D simulations of various lift scenarios in a static and timeconsuming manner. The planning process is often trial and error, with better alternatives not being discovered until later stages.

To address the path planning problem, many studies have explored using automated planning tools and information technology to enhance current practices. The initial works focused on using deterministic algorithms for path planning, as [3] outlined. To automate the path planning task, they employed two heuristic search methods, hill climbing and A*. Hill climbing involves an iterative approach of adjusting a solution to minimize the distance or any desired cost function. In comparison, A* uses a nodes approach, where it tries to find the shortest path between the start and end nodes using a cost function similar to hill climbing. However, these methods proved time-consuming and often stuck in local optima instead of finding the optimal solution.

Several studies have explored using metaheuristic algorithms to improve crane path planning. [4] used an ant colony to achieve collision-free path planning for mobile cranes, while [5] employed genetic algorithms to plan lifts in complex environments. However, metaheuristic algorithms are only sometimes the optimal solution and can be heavily influenced by initial conditions. At the same time, other works, such as [6], attempted to use hybrid approaches to tackle the issue, with similar results as previous works.

Recently, many researchers have turned to 3D/4D simulations and Building Information Modelling (BIM) to simulate and generate feasible solutions for lift planning. [1] integrated 4D crane simulation and BIM to manage operations on a construction site, while [7] presented a data-driven crane management system for industrial projects. Although BIM-based simulations offer detailed visualization, they may struggle to handle dynamic scenarios, making them less adaptable. [8] developed a methodology that enables automatic replanning of lifting paths for robotized tower cranes in dynamic BIM environments to address this issue. They used a GPU-based parallelization approach for discrete and continuous collision detection. However, the

methodology was built using a genetic algorithm, which may generate premature solutions.

According to [9], Reinforcement Learning (RL) is a type of machine learning that relies on learning through experience without needing previous data. The learner is tasked with discovering which actions yield the highest rewards by experimenting with different approaches, as explained by [10]. RL has been applied successfully to challenging tasks, such as game-solving. AlphaGo was developed to solve the game of Go by Silver et al.2017 [11], notably self-driving and robotics. As for the implementation of RL in construction projects, there have been some works using it to enhance the overall planning process. [12]explored RL applications during the design phase for decision-making purposes. As for crane operations, [9] used a hybrid greedy and RL approach for crane mat layout optimization. The same techniques used to tackle these complex issues can also be used to address the crane path planning issue.

This study identifies several areas of improvement where further work is needed, including the following:

1. Better tools must be provided for crane operators to develop practical and achievable lifting paths where less redundant and efficient lifts are desired.

2. Current path planning methodologies often fail to account for the unique complexities of mobile crane operations and their planning procedures. These complexities relate to the congested nature of construction sites and the requirements needed to perform a successful lift effectively.

3. A fully automated path planning process that explores all possible lifting approaches while considering the changing nature of construction sites needs to be developed.

4. There is a need for better-optimized procedures that can learn from the built environment and adjust to changes that occur during construction. Therefore, a framework with an adaptive learning approach is required.

This methodology combines VR simulations with RL to address previously mentioned research gaps. This work aims to address a need for developing a more comprehensive understanding of optimal solutions in complex and dynamic environments. By integrating 3D environmental elements, the methodology aims to identify precise solutions that consider aspects such as time, complex movement, and realistic scenarios that previous methods have overlooked. Furthermore, using RL, the methodology seeks to enhance exploration within construction sites, particularly in the automated path planning of mobile cranes, which can be complex and challenging to adapt to dynamic site conditions. This integration of VR simulations and RL allows a thorough evaluation of diverse alternatives, ultimately selecting

the most optimal solutions. The methodology fills the gap left by the absence of fully automated path-planning methods tailored specifically for mobile cranes and dynamic construction settings. Ultimately, this methodology aims to enhance the understanding of adapting to changes in the built environment during the construction phase, leading to better solutions for complex and dynamic building projects.

2 Methodology

In this section, a brief description of the methodology followed in developing the RL-based path planning framework. Figure 1 displays the main steps followed; where after identifying the lack of a realistic solution based on a 3D environment is needed to optimize the lift planning resources, data was collected to develop a simulation model that would assess various lift alternatives, and based on the main KPI the best solution is suggested for the user.



Figure 1. The overall research methodology.

2.1. Data Collection

In order to realistically replicate the lifting planning process, the authors determined that three main types of data are necessary: Crane data, Lifting Module data, and Building data.

2.1.1. Crane Model

The crane model used in the simulation is a highly accurate replica of a crawler crane widely used in industry. The model boasts a boom length of 36.5 78meters, for which the 36 meters configuration was selected, a track length of 10.0 meters and a width of 7.85 meters, and a maximum capacity of 300 tons, making it a highly versatile and reliable piece of machinery for a wide range of construction projects. The 3D rendition of the crane model is presented in Figure 2, showcasing the equipment's intricate details and precise specifications.



Figure 2. The crawler crane 3D model.

2.1.2. Modules and Buildings

Regarding the modules, a diverse array of payloads was brought into a 3D format, each possessing unique dimensions, weight, and physical attributes. Figure 3 contains a sample of a payload. A commonly used type of construction trailer is used, with dimensions of $6.01 \times 2.34 \times 2.69$ cubic meters and a total weight of 2850kg. The trailer is transported to provide a suitable workshop area for welders and is needed in the vicinity of the construction site.





A) Construction trailer 3D model.

B) 3D representation of the lifting task.

Figure 3. The 3D model of payload and construction site.

As for the buildings were imported via a BIM format, with the models' measurements and characteristics preserved, albeit with simplified component properties and reduced detail to optimize simulation performance. A traditional construction project is selected to display the efficiency of the developed tool. The crane agent is assigned a task of lifting a payload from its pick location, highlighted in red, and delivering said payload to its set location, highlighted in green, effectively performing a single lift.

2.2. Objectives Definition

The main focus of this framework is to provide a lift plan that enhances the following components:

2.2.1. Lifting Time

Crane operators are responsible for ensuring the safe and efficient transport of payloads. To achieve this, they rely on lifting times as a crucial factor. The longer the payload spends in the air, the higher the risk of accidents, which is why keeping the lifting time to a minimum is essential. Additionally, the operator must follow the shortest and safest paths during transport to reduce the risk of damage to the payload or any surrounding structures. Maintaining the payload's elevation level throughout the lift is also preferred, as sudden changes in height can cause instability and increase the risk of accidents. According to a recent study by [13], heavy lifts are often planned to remain low until they're close to their destination, ensuring maximum safety and efficiency during transportation.

2.2.2. Energy consumption

There exist three distinct alternatives to move a load in a more sustainable manner. The first one is walking, which involves transporting the object from point A to point B while the mobile crane moves. This type of movement is deemed suboptimal, primarily because it significantly increases planning costs. The need for crane mats for crane movement is a major contributor to the high costs. To mitigate mat costs, one of the most common measures is to plan the lifts in a project such that the crane does not need to walk while performing the lifting procedure. This is according to [13].

The duration of the lifting process is critical in ensuring a safe and efficient transport of the payload when operating a crane. The primary objective of the crane operator is to move the load from one location to another using the shortest and safest route possible while minimizing lifting time. To further enhance safety measures, it is recommended to keep the payload at a low elevation until it reaches its destination. A recent study by [13] revealed that heavy lifts are often planned to maintain a low elevation during transport.

The equipment's Hook Movement feature is equipped with a primary hook that offers a singular degree of freedom in the vertical direction. This exceptional feature allows the hook to effortlessly lift or lower attached objects, making it an incredibly versatile and efficient tool suitable for a wide range of applications.

2.2.3. Collision Detection

Typically, construction sites implement numerous safety measures to prevent collisions. However, for this particular project, it was determined that five types of collisions would be of particular concern. Two of these occur before lifting the load, including collisions between the crane and humans and between the crane and the building. An additional three types of collisions are considered once the load has been lifted, including collisions between the payload and the crane, collisions between the payload and humans, and collisions between the payload and the building.

3 Simulation Model Development

In this section a brief description of the methodology used to develop the simulation model is presented.it consists of five main components which are the agent, the actions, the states, the environment, and the rewards. The interaction between the different components is represented in Figure 6.



Figure 4. The reinforcement learning model.

3.1. The Agent

Initially, a crane agent is needed to take different actions to transport the payload from its original loading point to a predetermined set point. To achieve this task, a proper locomotion system is needed.

A mobile crane's locomotion system is more complex than that of a tower crane. Where the crane is made up of many parts that the crane operator needs to coordinate in order to perform a lift using the crane's full capacity. For the purposes of this work, three components' movements were modeled and are used in the training process, which is seen in Figure 5.

The main boom, where the boom is capable of rotation around the central axis of the crane. Both movements were taken into consideration, and the impact the angle has on the loading capacity was considered as well.



A) Crane Tracks



B) Main hook



C) Crane Boom Figure 5. The main Crane Components

The second comes the crane's tracks, which enable the crane to move forwards, backwards and rotate around the it's center.

Finally, the crane's hook's movement was considered. This component has a relatively simple movement, where it moves either up or down depending on the lift phase.

3.2. The Actions

Since the agent is comprised of three main components, each component was given a degree of freedom (DOF).



Figure 6. The degrees of freedom of the main crane components.

The tracks are given two DOFs, one rotation around the Y axis, and one translation along the x-axis, as seen in

Figure 6A. For the agent; this constitutes four different actions that can be taken.

The main hook was given one degree of freedom along the Y axis as seen in figure 6B, which allows the agent to take two different additional actions.

Finally, the main boom was given two DOFs, two rotations, one rotation around the Y axis and one around the Z axis, as seen in Figure 6C. This constitutes four additional actions.

3.3 The environment

Once, the crane agent is instantiated in the 3D environment it starts interacting with the different environment components to explore it. In order to improve the agent's interaction, a set of sensors are attached to it. The sensors use collisions in with the different 3D models and collects data about the geometrical properties of 3D component found in the simulation environment. In the following training episodes, the agent uses the data collected through sensors to avoid colliding with the different components of the environment. for instance, when a sensor collides with a building, it can identify physical dimensions, velocity, and its tag. Where there are four main tags Human, Obstacle, and Objective, Crane.

3.4. The states

The simulation is run for a predefined number of iterations/episodes. In each episode, the location of each major element is stored alongside the element's velocity local rotation.

In the initial runs of the simulation, and prior to adding a penalty to the collision of the crane model with the payload, the agent could move the payload by pushing it towards the set point to gain the final reward. To penalize the suboptimal behaviour, an additional variable was added for the payload, which is a Boolean variable. The Boolean value represents whether the object is being lifted; once the object is connected to the hook, the Boolean value is turned to true.

Finally, the environment is reset to its initial configuration in three cases:

- 1. The episode will end if the maximum number of steps has been attained and the payload is yet to be delivered.
- 2. If the agent leaves the environment, the episode will end.
- 3. If the payload is delivered to the set location, the final reward is given, and the episode is terminated.

3.5. Reward signal selection

While the agent is training, it is essential to select the appropriate reward; this task is perhaps the most critical. Moreover, selecting a reward that balances penalties and rewards is necessary. For instance, if the agent is being over-penalized, the behavior resulting from the training would be suboptimal. For this work, various rewarding strategies were explored, and the final selected strategy was selected based on its merits and final training results.

The reward pseudo code.	
Input:	
Crane initial location (Cl), Payload initial location	
(PL), Crane capacity (Cc), Payload weight (Pw)	
for steps $= 1$ to Number of iterations	
for steps $= 1$ to max steps	
if PL is disconnected	
Compute D	
if D > 0.5m	
Act from possible actions	
Apply Move penalty \times D.	
check for collision.	
if a collision is true	
Apply collision penalty.	
else if $D < 0.5m$	
Apply connection.	
Add Lift reward.	
if PL is connected	
Compute Dps.	
If $Dps > 0.5m$:	
Act from possible actions	
Apply Move penalty × Dps	
check for collision.	
If a collision is true:	
Apply collision penalty.	
Else if Dps < 0.5m:	
Set payload.	
Add Final reward.	
Reset environment.	

Where:

D: distance between from crane to the object. DPS: Distance between payload and set point.

In order to attain the previously determined objectives, which are the optimization of the lifting time, energy consumption, and collision detection, three different types of rewards were built into the training process. The first reward is related to the lifting time, which is highly affected by the number of actions taken and the overall time needed to deliver the payload to its set location. Furthermore, each action has its parameters, such as movement speed, damping ratio, and interaction with the previous movement. For instance, when the crane is moving, no other action can be taken until the crane is at a complete stop. All the previously mentioned parameters are combined to calculate the time needed to perform the lift, and the subsequent reward signal associated with the movement is developed to penalize the agent for each time spent lifting and transporting the object. This penalty ensures that the resulting lift uses the shortest lifting time. The penalty amount was selected after multiple iterations, where the initial penalties were severe enough to disable the agent from moving.

The second reward signal used is related to collisions or leaving the training area. The objective of the training is to discover all of the possible alternatives that can be used to perform a lift. It was decided that the agent would leave to explore the whole construction site. However, in some instances, the agent would leave the training area; thus, in those instances, a significant penalty is used to prevent the agent from leaving the said area. Furthermore, the episode is instantly terminated since, based on the author's experience with the model, the agent does not find its way back to the payload once it has explored the extremities of the training area.

Additionally, using the same approach, the agent sometimes collides with the surrounding obstacles and buildings despite the built-in sensors. An additional penalty is added based on the collision tag in those instances. The episode is terminated in some instances where the agent either collides with a human or a building. The other collisions are only penalized to allow the agent to train for an extended period.

The third type of reward is related to the type of action taken, where specific actions are preferred. For instance, since pick-up and walk operations are less favorable due to their increased cost, crane movements are more penalized than boom and hook movements. Next, main boom movements are expected to consume more energy than hook movements. Thus, they penalized more than hook movements. Finally, hook movements are given the lowest penalty.

Finally, the main reward signals for the crane are those related to lifting and delivering the payload. The first portion of the reward is connecting the payload to the hook; once the agent lifts the payload, it receives an enormous reward. The second portion relates to the agent setting the payload in the set place where the final reward is given and the episode is terminated. If the agent lifts the payload and fails to deliver it, the episode is terminated, and the environment is reset.

4 Results

The following section sheds light on the initial findings of our research and emphasizes the efficacy of our methodology. To quantify the impact of our methodology, three main indicators are used: cumulative rewards, episode length, and curiosity value indicators. The cumulative rewards graph displays the agent's improved training, which starts with negative outcomes (no solutions) and improves in value until it reaches a 298-point solution. To understand the cumulative rewards graph efficiently, the episode length indicator is used, which displays the number of steps needed by the agent to achieve the task; a lower number of steps indicates a more optimized solution. In order for the agent to decrease the number of episodes to achieve the lifting task, it must have a significant focus on exploring the actions, space, and the environment; this exploration task is achieved through curiosity, where the agent's focus on exploration in later episodes coincides with the decreased episode length and as a result of the latter higher cumulative rewards. To get a better understanding, you can refer to Figure 7, which displays the cumulative rewards achieved per episode and reveals some intriguing insights. In the first 500,000 episodes, the agent encountered a formidable challenge due to the task's complexity, leading to a high number of exploratory and negative rewards resulting in penalties for every action. However, the agent discovered a promising solution around the 550,000th episode, which significantly improved the outcomes. Nonetheless, further refinements were necessary to optimize the training process. In the subsequent episodes, the agent consistently improved its approach, gradually climbing towards the optimal policy. Eventually, it peaked at 298 total rewards before stabilizing at approximately 4 million episodes.



Figure 7. The cumulative reward value per episode.

The study conducted by the researchers involved a comprehensive examination of various factors that influence the efficiency of an agent in executing tasks proficiently. Amongst these factors, the duration of each episode was given special consideration, as it is a crucial parameter used to evaluate an agent's overall performance. The researchers analyzed the simulation time of the agent during its initial million episodes and

found that it ranged from 1900 to 800 seconds, which is significantly longer than the desired duration. This indicates that the agent was in the exploratory phase, trying to identify the best possible path to achieve the assigned task. The researchers observed a significant decrease in the simulation time as the simulation progressed, which continued until the 3.5 millionth episode. This trend indicates that the agent had identified an optimal path and slightly improved the discovered solution. This discovery led the agent to find the most efficient path, which could be completed within 17.8 seconds in the simulation environment. The agent's continuous improvement and progress ultimately revealed the most efficient and effective way to complete the task. The episode length results are highlighted in Figure 8.



Figure 8. The episode length is in seconds per episode.

To assess an agent's learning progress, it's vital to analyze its level of curiosity and how it interacts with the environment. Curiosity value estimates provide valuable insights into the agent's learning patterns, reflecting its interest in exploring information. Higher values indicate strong engagement and a deep desire to learn, while lower values suggest a lack of enthusiasm for exploration and learning. By measuring curiosity-driven behavior, we can gain valuable insights into the agent's interest levels and learning patterns, which help us understand its progress. Figure 9 presents a graphical representation of the curiosity value estimates for behavior analysis. The results indicate a clear pattern of curiosity value progression that aligns with expected trends. At the beginning of the learning process, the agent exhibits minimal curiosity, but its curiosity level gradually increases to larger values in later episodes.

This indicates a notable surge in engagement or interest as the agent continues to learn. Statistical mean and standard deviation measures support this progression, demonstrating a distinct shift from initial disinterest to intense curiosity. Evaluating the agent's curiosity-driven behavior is crucial for understanding its learning patterns and progress. This leads to more effective learning by ensuring the agent actively engages with the environment and explores new information. After the lift was developed, the authors assessed its validation using two approaches, which are face validation and VR lift assessment.

In the first approach, the authors used the computer model in the virtual environment and analyzed the different steps used to lift and deliver the payload. However, the computer version, although a 3D approach was deemed insufficient due to the presence of certain blind spots that the users could miss, to that extent a VR, the authors explored the lift using a gamified approach, where they moved along the environment during the lift's execution and assessed the validity of the lift in the virtual environment.



Figure 9. The curiosity value estimate per episode.

5 Conclusion and Future Directions

This work was used to develop a new tool for the construction industry that helps with crane path planning. The tool is designed to create lifting paths that are more comprehensive and take into account the changes that occur on a construction site. This leads to a smoother lifting process, which reduces planning time and produces more accurate results. The tool helps to improve collision detection, decrease lift complexity, and lower energy consumption. In addition, we have incorporated RL and VR simulations to make the planning process more resilient to unforeseen changes on

the construction site. Using a virtual environment enhances the planning process, allowing for updates to be made to the crane agent training and lifting capabilities based on changes in the construction site, resulting in more reliable and efficient lifting schemes.

The authors developed a comprehensive simulation model to achieve our goal. This model included an agent, actions, states, environment, and a reward system. By training the crane agent through numerous episodes, the agent could detect optimal path-planning strategies. The agent's learning journey was tracked through cumulative rewards, which progressed from exploratory phases to achieving peak efficiency after approximately 4 million episodes. Analysis of simulation times highlighted the agent's evolving learning patterns, ultimately identifying the most efficient path within the environment. Face validation and VR-based assessment were performed to validate the developed approach's results. The results demonstrated the agent's ability to navigate complex environments, optimize lifting processes, and minimize simulation times. The insights gained through the agent's learning patterns and performance metrics validated the effectiveness of RL-based path planning in dynamic construction scenarios.

Future works will involve conducting a comparative analysis with real-life path planning scenarios by testing the developed model in live construction settings. By aligning simulated results with real-world scenarios, we aim to refine and validate our methodology for seamless integration into actual crane operations. Ultimately, our approach will contribute to safer, more efficient, and adaptable construction practices.

References

- A.N. Tak, H. Taghaddos, A. Mousaei, A. Bolourani, U. Hermann, BIM-based 4D mobile crane simulation and onsite operation management, Autom Constr 128 (2021) 103766.
- [2] S. Hu, Y. Fang, Y. Bai, Automation and optimization in crane lift planning: A critical review, Advanced Engineering Informatics 49 (2021) 101346.
- [3] P.L. Sivakumar, K. Varghese, N.R. Babu, Automated Path Planning of Cooperative Crane Lifts Using Heuristic Search, Journal of Computing in Civil Engineering 17 (2003) 197– 207.
- [4] X. Wang, Y.Y. Zhang, D. Wu, S. De Gao, Collision-Free Path Planning for Mobile Cranes Based on Ant Colony Algorithm, in: Materials, Mechatronics and Automation, Trans Tech Publications Ltd, 2011: pp. 1108–1115.
- [5] P. Cai, I. Chandrasekaran, J. Zheng, Y. Cai, Automatic Path Planning for Dual-Crane Lifting

in Complex Environments Using a Prioritized Multiobjective PGA, IEEE Trans Industr Inform `(2018) 829–845.

- [6] K. Boutouhami, A. Bouferguene, R. Lemouchi, M. Assaf, M. AL-Hussein, J. Kosa, Hybrid Approaches For Handling Mobile Crane Location Problems In Construction Sites, in: 2023 Winter Simulation Conference (WSC), 2023: pp. 2722–2733.
- [7] S. Han, Z. Lei, U. (Rick) Hermann, A. Bouferguene, M. Al-Hussein, 4D-based automation of heavy lift planning in industrial construction projects, Canadian Journal of Civil Engineering 48 (2021) 1115–1129.
- [8] S. Dutta, Y. Cai, L. Huang, J. Zheng, Automatic re-planning of lifting paths for robotized tower cranes in dynamic BIM environments, Autom Constr 110 (2020) 102998.
- [9] G.M. Ali, A. Bouferguene, M. Al-Hussein, Crane Mat Layout Optimization Based on Agent-Based Greedy and Reinforcement-Learning Approach, J Constr Eng Manag 149 (2023) 4023067.
- [10] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, A Bradford Book, Cambridge, MA, USA, 2018.
- [11] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis, Mastering the game of Go without human knowledge, Nature 550 (2017) 354–359.
- [12] S. BuHamdan, A. Alwisy, A. Bouferguene, Explore the application of reinforced learning to support decision making during the design phase in the construction industry, in: Procedia Manuf, Elsevier B.V., 2020: pp. 181–187.
- [13] N. Kayhani, H. Taghaddos, A. Mousaei, S. Behzadipour, U. Hermann, Heavy mobile crane lift path planning in congested modular industrial plants using a robotics approach, Autom Constr 122 (2021) 103508.

Automated Decision Support System Based on Quantification of Defective Tubular Steel Temporary Materials for Quality Circles.

James Mugo Njoroge¹, Seojoon Lee², Kyuhyup Lee³, Junsung Seol¹, Younghee Chang² and Soonwook Kwon⁴

¹Department of Civil, Architectural and Environmental System Engineering, SungKyunKwan University, Republic of Korea.

²Department of Global Smart City, SungKyunKwan University, Republic of Korea.

³Convergence Engineering for Future City, SungKyunKwan University, Republic of Korea.

⁴School of Civil, Architectural Engineering and Landscape Architecture, SungKyunKwan University, Republic of

Korea.

<u>mugojames254@gmail.com</u>, <u>sjlee8490@naver.com</u>, <u>leekyuhyup@naver.com</u>, <u>seol987@naver.com</u>, <u>yhyhchang@g.skku.edu</u>, <u>swkwon@skku.edu</u>

Abstract

Temporary materials are stored in construction warehouses after and before they are used in construction sites. Most of these materials are made of steel tubular sections such as square, circular, and rectangular sections. However, due to frequent reuse and their storage conditions they become subject to defects such as rust and bend which affect their quality for use in construction projects. Since these materials are stored in stacks or batches, checking the materials individually for defects can be time consuming making end point surface defect detection more efficient for construction workers. Deep learning techniques have proven to be more efficient than manual inspection. However, quantification of materials with defects for decision making on reusing, repairing, and disposing actions and documentation is still a challenge for construction workers. Hence, this paper quantifies temporary tubular steel materials; square hollow section, circular hollow section and rectangular hollow section with common cross section area defects using deep learning technique connected to a web platform for decision making by construction quality circle workers. The proposed system achieved an average precision of 84.9 percent with 105.2 GFLOPS and a speed of 20 seconds per inference.

Keywords -

Deep Learning, Decision Support System, Temporary Materials, Defects, Quality Circles.

1 Introduction

Construction projects use temporary materials to provide support and safe working platforms for the workers. The dominant materials used for these structures are tubular steel materials such as square hollow sections, rectangular hollow sections, and steel pipes. However, with frequent reuse and storage conditions of these materials, they become subject to surface defects such as rust and bend which has the potential of causing accidents. These materials are stacked in batches as shown in figure 1 which makes end point surface defect detection faster and productive method of judging their quality.

Monitoring these surface imperfections is done by quality circle workers. Quality control circles or simply quality circles encourages construction workers to actively participate in continuous improvement of material quality in construction [1]. However, these workers rely on manual inspection checklist and reporting which not only reduces productivity but also, is prone to human error. Computer vision techniques such as object detection are being employed to replace manual inspection and identification of defects in areas such as concrete cracks and anomalies in dimensions and misalignment, however, there exist a gap on quantification of temporary tubular materials with surface defects for decision making by construction workers in quality circles for quality control.



Figure 1. End point surface of stacked steel sections.

Hence, this paper proposes an automated decision support system based on quantified number of materials with surface defects; rust, bend and, rust and bend using deep learning object detection yoloV7 model. The model is deployed on flask web framework for visualization, with a rectification action and reporting of the results which can be used on mobile devices by quality circle construction workers. This is also visualized on local computer system by the management in real time.

2 Background

2.1 Tubular Steel Temporary Materials

Temporary materials are very important for the erection of temporary structures such as scaffolding, falsework and formwork systems. Metal temporary materials such as steel and aluminium are frequently used in the construction industry for these temporary structures due to their guaranteed strength, rigidity, and the ability to erect and dismantle quickly. Additionally, they are easy to reuse which makes them much cheaper compared to other alternatives such as timber based temporary materials. However, due to frequent reuse and exposure to varied environmental conditions, these metal-based materials become subject to rust which undermine their structural strength creating unsafe working condition on construction sites. Additionally, hollow steel sections such as square hollow sections experience global and local deformation under excessive loading especially when stored in batches creating a warp or bend on the cross-section area. This undermines the hollow steel section's aesthetic nature and their structural strength [2].

2.2 Deep Learning-Based Inspection

The quality inspection of these materials is mainly done manually, where workers or inspectors check these surface defects for quality control. However, use of deep learning techniques such convolutional neural networks (CNN) on identifying these surface defects have gained attention in the past decade. They involve extracting useful information from images and videos to reduce the involvement of human inspectors and construction workers. For instance, Hussein et al. [3] employed VGG-16, a convolutional neural network on images to detect and localize defects such as mold and stains in buildings. Sergio et al. [4] used convolutional neural network to detect defects such as cracks on structural members of bridges through images. In terms of steel sections, Zhaoguo et al. [5] introduced new modules on the deep learning neural network such as multiscale feature extraction to improve extraction process of features on images and efficient feature fusion to improve the fusion mechanism on the neck layer for accurate detection and localization of surface defects such as crazing and patches. However, though there exists extensive research on detection of surface imperfections using computer vision techniques, quantification of materials with these defects is still lacking.

2.3 Quality Control of Materials

Checking the quality standard of individual materials in a batch can be expensive and time consuming, especially where the lot size is very large. This has led to the introduction of sampling plans to decide whether to accept or reject a lot based on a defined sample size which is a representative of the lot. The whole lot is either rejected or accepted based on whether a specific threshold of the number of defective items in the sample has been observed or not. The two main sampling techniques used to determine the acceptance of materials in a batch are sampling by attributes and sampling by variables. Sampling by attributes is based on whether there is presence or absence of a particular characteristic of a material, for example, whether a material contains defective patch or not. It is computed on the probability of accepting or rejecting a lot using the defective aspect of the batch. For instance, assuming we have a lot of N size, where the sample size is n, and the actual number of defective materials is M, the probability of getting x number of defective items in the sample is expressed as shown in equation (1)

$$p(X = x) = \frac{\binom{M}{x}\binom{N}{n} - \frac{M}{x}}{\binom{N}{n}}$$
(1)

To compute the probability of accepting or rejecting the lot, we assume that the fraction of non-defective items in the lot to be q, and the fraction of defective items to be p. This is expressed in a binomial distribution equation as shown in equation (2).

$$g(p) = \sum_{x=0}^{r} p(x=x) = \sum_{x=0}^{r} \frac{\binom{N_p}{x}\binom{N_q}{n-x}}{\binom{N}{n}}$$
(2)
Sampling by variables on the other hand is more quantitative in nature and is focused on measurement of actual values such as length, height, and weight. It is a continuous process used to determine the overall acceptability of the batch. The main goal of the two sampling techniques is to control quality of materials in the most economical and efficient way [6].

However, this manual quality control system is prone to human errors and the documentation process is tedious for the workers especially when we have large number of materials.

2.4 Normal Distribution

Normal distribution or gaussian distribution is a probability distribution where an average distribution of random samples tends to converge towards the normal distribution creating a symmetric bell-shaped curve. This works under the concept of the central limit theorem. There are two broad techniques for monitoring quality in construction management: management techniques and statistical techniques. Management techniques involve quality control, quality assurance and total quality management concepts. Statistical techniques include gaussian distribution and hyperbolic distribution which use the 6-sigma concept on quality control. Gaussian distribution has been applied in many fields such as construction management, manufacturing, and civil engineering through estimation of statistical properties such as number of defects and defects per million opportunities [7]. For instance, Vivian et al. [7] compared the use of gaussian distribution and hyperbolic distribution on defect detection to improve the construction quality and the yield percentage. Diego and Peter [8] proposed gaussian process for predicting product quality based on defective products per unit (fault density). Silva et al. [9] employed the gaussian distribution to detect harmful conditions in aquatic life. The method used 2D image visualization on production of fish under three classifications: dangerous, warning, and normal conditions. Zhou et al. [10] solved the problem of clutters on 3D feature descriptors using histograms of gaussian normal distribution. This enabled capturing of conspicuous features creating a homogenous scene with the 3D model.

2.5 Web Framework

The era of information technology has seen a huge increase of data in the construction industry. However, the visualization of this data to help stakeholders make informed decision has always been a challenge. This is due to variations on the type of data produced which comes in different formats such as images, text and videos causing data driven decision making in the construction industry difficult. Additionally, the transfer of these data between project stakeholders has been slow affecting the project productivity. Hence, a cross integration platform such as a web framework system which can visualize and share construction information data in different formats rapidly has been of necessity.

In order to improve access and sharing of integrated construction data, Chassiakos and Sakellaropoulos [11] proposed a web framework connected to a relational database for managing construction information. Gurmu et al. [12] developed a dashboard for visualization of building defects from inspection reports through data mining using python libraries and natural language toolkit. Changyoon et al. [13] made a construction management system for real time site monitoring and construction information sharing with the aim of visualizing on mobile devices. Do-Yeop et al. [14] used web system framework to link defective data with BIM environment for visualization. However, there is still a gap on a web framework which can visualize quantified defective materials from a deep learning algorithm through mobile devices and local computer and, report the output for documentation.

3 Methodology



3.1 Overview of the Proposed Methodology.

Figure 2. 3-Stage overview of the methodology.

The proposed methodology was divided into 3 main stages: computer vision, decision support and, reporting and visualization as shown in figure 2.

3.2 Stage I: Computer Vision.

3.2.1 Data Collection and Preprocessing

The first stage was collection of tubular steel temporary materials image dataset from two storage sites. The data collected consisted of circular hollow sections hereby referred to as steel pipes, rectangular hollow sections, and square hollow sections. The images taken focused on the temporary materials in batches as shown in figure 3. In the process a total of 500 images were collected for training the deep learning model.



Figure 3. a). Square Hollow Sections, b). Steel Pipes, c). Rectangular Hollow Sections dataset collected in batches.

The image dataset obtained had varied size in terms of the height and width pixel. Hence, before training the deep learning model, the data was preprocessed using a fixed height and width of 640 x 640 pixels. A custom labelling process using LabelMe software was adopted as shown in figure 4 employing a 2-tier naming nomenclature where the first part represented the material type: steel pipe, square hollow section, and rectangular hollow section, and the second part separated by the hyphen represented the state of the material: rust, bend, rust and bend, and non-defective state.



Figure 3. Labelling nomenclature using LabelMe software.

In the labelling process, occluded images such as those with concrete splatter on the cross-section area of hollow sections were included to make the model robust to occlusion. Additionally, before training, mosaic augmentation, mix-up augmentation and perspective transformation hyper parameters were adjusted to enable the model to generalize across various object configurations for occluded images.

During inference of the deep learning model, an image output of the detected object and text output are generated. The text output contains the class label, xcenter, y-center, height and width of each detected box and confidence scores as shown in figure 4. A special python code was developed which converted the detected text results to python list. The listed results were separated using white spaces to extract the material type and the special hyphen character to extract the surface defect of the materials.

Rectangular Hollow	Section-R	0.798437	0.101562	0.165625	0.140625	0.90625
Rectangular Hollow	Section-RE	0.624219	0.101562	0.157813	0.1375	0.907227
Rectangular Hollow	Section-B	0.297656	0.500781	0.160938	0.123438	0.908203
Rectangular Hollow	Section-B	0.790625	0.747656	0.15 0.11	4063 0.9	19922
Rectangular Hollow	Section-N	0.778125	0.373437	0.15625 0	.121875	0.92041
Rectangular Hollow	Section-RE	0.792188	0.85625	0.140625	0.10625	0.929199
Rectangular Hollow	Section-N	0.778906	0.239844	0.164062	0.129688	0.930664
Material Type - State		x-center	y-center	height	width	conf.

Figure 4. Detection text results file.

3.2.2 YoloV7 Deep Learning Model

The dataset was then divided to a train set and validation set ratio of 4:1 and trained on the YoloV7 deep learning model. The model is typically made of three parts: backbone, neck and head as shown in figure 5. The backbone layer contains convolutional layers for feature extraction from an input image. Each convolutional layer has a kernel size, number of stride operation and the number of channels. The neck layer is used to fuse or connect the extracted features from the backbone layer while the head is used to make prediction using bounding boxes. The head is subdivided into 3 parts depending on the level of feature extraction on obtaining detection results for large, medium, or small objects [15].

Feature extraction process from the backbone layer is followed by a spatial pyramid pooling (SPP-net) layer in the neck layer. The SPP-net allows the use of input images with variable sizes which help maintain feature information from images without warping them. Additionally, the network contributes to improved accuracy on training [16]. There are additional plug-andplay modules which can be used in place of the SPP-net such as the Ghost-net layer and the Bottleneck layer. The Ghost-net layer is aimed at reducing the computational cost of the deep learning model enabling a lighter model. This is achieved by stacking the ghost modules [17]. Bottleneck layer uses pointwise convolutions to make bottlenecks for reducing parameters and increasing the depth, enhancing efficiency and computational performance [18].

However, these layers still have the problem of heavy inference computation especially when deploying them on mobile devices and web platforms. To mitigate this problem the SPP-net, Ghost-net and Bottleneck layer are concatenated with Cross-Stage Partial network (CSP-net). This improves speed of inference and accuracy when deployed on various platforms. For instance, when using the Dense network, the base layer is concatenated with the subsequent layers directly to map the output assuming a network with k-layers of convolutional neural networks, with **F** as the mapping function using an input x_0 to the target **y**, an equation can be derived as follows:

$$y = F(x_0) = X_{k.} \tag{3}$$



Figure 5. Internal architecture of the YoloV7 model.

However, the cross stage partial network relies on optimization by dividing the F function for the base layer (x_{θ}) into two parts:

the dense network to another [19].

$$y = M([x_0, T(F(x_0))])$$
 (5)





Figure 6. Cross stage partial network in the DenseNet.

Hence, to map the target y, x0' is connected to the end stage while the x0'' undergoes the dense network operation under M transition function for combining the two separate parts of the network that is, F the mapping function and T the transition function from one layer in

Hence, to obtain a model which can perform inference fast with high accuracy, three networks with cross-stage networks SPPCSP-net, partial were evaluated: GhostCSP-net and BottleneckCSP-net. Their performances were assessed using Giga Floating Point per second, mean average precision and recall in order to determine the best network that could be integrated into the web framework. The training parameters were a batch size of 16, 150 epochs with a 0.01 initial learning rate. The model was trained on the pytorch framework, CuDA 11.8, on the Tesla V100, 16GB graphical processing unit (GPU).

3.3 Stage II: Decision Support System.

To determine the threshold for different classification of batch materials based on whether to reuse, dispose or repair, a statistical distribution of the surface defects of different tubular steel materials in lots was performed on two storage sites. The equation for calculating distribution is normally based on two key parameters; mean (μ) and standard deviation (σ) with the normalization factor and natural logarithm base (*e*).



Figure 7. Normal distribution curves of different defective states of materials.

Hence, the mean and standard deviation of different batches was computed for statistical distribution.

$$f(\mathbf{x}\mathbf{I}\boldsymbol{\mu}\boldsymbol{\sigma}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mathbf{x}-\mathbf{u})^2}{2\sigma^2}} \tag{6}$$

The normal distribution curve of different condition of materials was plotted as shown in figure 7 and an assumption was made where percentage values greater than the curve threshold line indicated that the cost of repair is more than the cost of disposing the defective materials hence the dispose decision should be employed on the defective materials in the batch. On the other hand, when the percentage value falls below the threshold line but is more than zero, the cost of repairing the defective material in the lot is cheaper compared to the cost of disposing. Hence, the repair decision is employed.

Using the statistical distribution results, a degree of severity table was formulated to create a classification criterion as shown on table 1. The mean and standard deviation for the statistical distribution was based on the defective material count to total material count ratio expressed as a percentage.

Table 1. Surface Imperfection Degree of Severity Criterion

Defect	Critical	Minor	Perfect
	Imperfection	Imperfection	(Reuse)
	(Dispose)	(Repair)	
Rust	>22%	0% <r≤22%< td=""><td>0%</td></r≤22%<>	0%
Bend	>28%	0% <b≤28%< td=""><td>0%</td></b≤28%<>	0%
Rust and Bend	>44%	0% <rb≤48%< td=""><td>0%</td></rb≤48%<>	0%

R - Rust, B - Bend, RB - Rust and Bend

3.4 Stage III: Web Framework.

The final phase connected the computer vision model with the classification criteria from the decision support stage using flask web framework. The web framework was used to visualize the detected image, the total count of materials for each individual lot, the count for detected materials with defects, the decision to reuse, dispose or repair the materials in the batch and reporting the information for documentation. The quality circle leader or the user is required to take an image of the temporary materials in a batch using a mobile device, the image is sent to the cloud server which runs inference based on the trained deep learning model. From the detected results a text file is generated as shown in figure 5 containing information on the type of material and their corresponding defective state. Python code embedded into the flask framework computes the total number of materials in a batch and the total number of materials with specific defective condition expressed as a percentage. The percentage obtained is then compared to the degree of severity criterion table 1 where a decision on whether to reuse, dispose or repair the materials in the lot is obtained. The results are rendered on an interactive webpage using hypertext mark- up language. The user interface has an input field for the quality circle leader to enable reporting and documentation. The web framework plays a central role in connecting the front end which is the user interface and the back end which is the cloud server for hosting the deep learning model for inference and visualization on the actions to be applied on the defective materials. It ties almost all the nodes in the system which are illustrated in figure 8, creating a unified automated decision-making system for the management and the quality circle workers.

		-		-	-		
Network	Epochs	GFLOPS	Precision	mAP@0.5	Recall	F1	_
						Score	
SPPCSP-net	150	105.2	82.4	84.9	74.8	78	_
GhostCSP-net	150	102.2	78.5	84.8	82.1	80	
BottleneckCSP- net	150	100.3	82.4	82	76.3	78	

Table 2. Performance comparison of the selected deep learning neural network modules



Figure 8. Web framework Concept.



Figure 9. Visualization of the web framework user interface

4 Results and Analysis

Evaluation of the system was done based on average precision, precision, inference speed, Giga Floating Point per Second (GFLOPS) and the count accuracy of the materials. The spatial pyramid network with the cross stage partial network achieved the highest GFLOPS of 105.2, mean precision of 82.9, precision of 84.9 with a speed of 20 seconds per inference.

In terms of count accuracy, the material batches were divided into 3 classifications:

- 1. Small batches batches with materials in between 1 to 100.
- 2. Normal batch batches with materials between 101 to 250.
- 3. Large batch batches with materials over 250

Different construction storage sites had specific number of materials to be stacked on an individual batch based on space availability. Hence, the classification was based on analysis of the stacking of materials at different temporary material warehouses and sites. Absolute error (Measured value – True Value) was calculated for each subgroup expressed as a percentage as shown in table 3.

 Table 3. Measure of absolute error of average count of different batch size.

Batch	Detected	True	Absolute
Size	Count	Count	Error
Small	35	35	0%
Normal	105	105	0%
Large	300	500	40%

5 Conclusion

This research proposed an automated decision support system which can be used to determine whether to dispose, reuse or repair of tubular steel materials based on their end point surface defects. Additionally, the study contributed to quantification of temporary steel tubular materials using deep learning and integration of the deep learning model to the web framework for management of materials. This can be further used to calculate the cost of repair and disposal of steel tubular temporary materials.

However, through the study, a very high absolute error was observed on large batch material sizes. This is because the target objects appear smaller during the detection process. In order to improve the accuracy on large batches, it is recommended to customize anchor boxes responsible for detecting small objects in addition to adding dataset with large batches of materials. Hence, a robust deep learning model which is capable of accurately quantifying large batch size material is suggested as an area of further study.

Acknowledgement

This work is financially supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as \lceil Innovative Talent Education Program for Smart City \rfloor .

This research was conducted with the support of the "National R&D Project for Smart Construction Technology (No. RS-2020-KA156488)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

References

- Bruno G., Ali T and Toshiyuki A. Quality control circles in construction. ASCE Journal of Construction Engineering and Management, 113(3):432, 1987.
- [2] Cosgun S.I. Experimental and computational study on local buckling of standard and improved coldformed square hollow sections under static and dynamic loading. *Structures*, 54:291, 2022.
- [3] Hussien P., Joseph T. and Amir M. Deep Learning for detecting building defects using convolutional neural networks. *Sensor*, 19(16):3556, 2019.
- [4] Sergio R., Angelo C. and Andrea N. Using machine learning approaches to perform defect detection of existing bridges. *Procedia Structural Integrity*, 44:2028, 2023.
- [5] Zhaoguo L., Xiumei W., and Hassaballah M. A deep learning model for steel surface defect detection. *Complex Intell. Syst.* 2023.
- [6] Hendrickson C. Project management for construction: fundamentals concept for owners engineers, architects, and builders, 2nd ed., Version 2.2. Carnegie Mellon University, Department of Civil and Environmental Engineering, Pittsburgh, PA, 2008.
- [7] Vivian T., Khoa L., and Hoang L. Using gaussian and hyperbolic distributions for quality improvement in construction: Case study approach. *Journal of construction engineering and management*, 134:7, 2008.
- [8] Diego R. and Peter G. Predicting product quality in continuous manufacturing processes using a scalable robust Gaussian Process approach. *Engineering applications of artificial intelligence*,

127, 2024.

- [9] Silva L., Lopes B., Blanquet I. and Marques C. Gaussian distribution model for detecting dangerous operating conditions in industrial fish farming. *Applied Sciences*, 11(13):5875, 2021.
- [10] Zhou W., Ma C., Yao T. et al. The histograms of gaussian normal distribution for 3D feature matching in cluttered scenes. *Vis. Comput*, 35, 489-505, 2019.
- [11] Chassiakos A. and Sakellaropolous S. A web-based system for managing construction information. *Advances in engineering software*, 39(11), 2008.
- [12] Gurmu A., Hosseini M.R., Arashpour M. and Lioeng W. Development of building defects dashboards and stochastic models for multi-storey buildings in victoria, Australia. *Construction Innovation*. 2023.
- [13] Chanyoon K., Taeil P., Hyunsu L. and Hyoungkwan K. On-Site construction management using mobile computing technology. *Automation in Construction*. 35:415-423, 2013.
- [14] Do-Yeop L., Hung L.C., Jun W., Xiangyu W. and Chan S.K. A linked data system framework for sharing construction defect information using ontologies and BIM environments. *Automation in Construction*, 68:102-113, 2016.
- [15] Chien-Yao W., Alexey B. and Hong-Yuan M.L. YOLOv7: Trainable bag-of-freebies sets new stateof-the-art for real-time object detectors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pages 7464 - 7475, Vancouver, BC, Canada, 2023.
- [16] Kaiming H., Xiangyu Z., Shaoqing R. and Jian S. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *Proceedings of the Computer Vision – ECCV*, Pages 346 - 361, 2014.
- [17] Kai H., Yunhe W., Qi T., Jianyuan G., Chungjin X., and Chang X. GhostNet: More features from cheap operations. In *Proceedings of the IEEE Conference* on Computer Vision and Pattern Recognition (CVPR), Pages 1557-1586, Seattle, WA, USA, 2020.
- [18] He K., Zhang X., Ren S and Sun J. Deep residual learning for image recognition. In *Proceedings of* the IEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770 -778, Las Vegas, NV, USA, 2016.
- [19] Wang C., Liao H., Wu Y., Chen P., Hsieh J. and Yeh I. CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1571 - 1580, Seattle, WA, USA, 2020

Towards autonomous shotcrete construction: semantic 3D reconstruction for concrete deposition using stereo vision and deep learning

Patrick Schmidt¹, Dimitrios Katsatos², Dimitrios Alexiou², Ioannis Kostavelis², Dimitrios Giakoumis², Dimitrios Tzovaras², Lazaros Nalpantidis¹

¹Technical University of Denmark, Kongens Lyngby, Denmark ²Centre for Research and Technology Hellas, Thermi, Greece

pasch@dtu.dk, {dkatsatos, dalexiou, gkostave, dgiakoum, dimitrios.tzovaras}@iti.gr, lanalpa@dtu.dk

Abstract -

The adoption of autonomous systems is a foreseeable necessity in the construction sector due to work hazards and labor shortages. This paper presents a semantic 3D understanding module that creates 3D models of construction sites with highlighted regions of interest for shotcrete application. The approach uses YOLOv8m-seg and SiamMask for robust semantic segmentation together with RTAB-Map and InfiniTAM for visual odometry and 3D reconstruction. Our method is the first step towards a novel, autonomous robot for shotcrete spraying and finishing. The effectiveness of our approach is shown on a mock-up construction site and provides evidence for the applicability of robotic construction.

Keywords -

Construction Robotics, 3D Reconstruction, Semantic Segmentation, Shotcrete Automation

1 Introduction

The construction sector is among the industries that have not undergone a major digital transformation. Within the European Union (EU), it stands out as one of the least digitalized industries [1]. Furthermore, construction work is skilled labor, posing demands on the labor market that are not met today: A European Labor Authority report shows that 13 out of 30 regions are reporting shortages of Concrete Placers and Finishers [2]. Construction work is also known to be highly hazardous with Eurostat reporting the highest incidence of non-fatal work-related accidents in the EU, reaching almost 3000 accidents per 100,000 employed persons [3].

Aiming to alleviate these challenges and enhance overall efficiency, the construction industry is gradually embracing digital advancements [4]. In particular, the application of semantic 3D reconstruction, through computer vision systems and deep learning algorithms, plays a critical role in enabling precise 3D digital models of the scene. This advancement allows for the creation of rich Building and Construction Information Models (BIM/CIM) and Mechanical, Electrical, and Plumbing (MEP) systems [5], targeting the development of digital models of the construction site to ensure time-sensitive decision-making and streamline project progress monitoring.

Additionally, real-time 3D reconstruction featuring semantic annotations can be used to measure the thickness level of ground support walls. This is especially valuable in shotcrete tasks, where dry concrete is sprayed onto the wall surface. Shotcrete processes subject workers to health hazards due to concrete rebound and rely heavily on the expertise of nozzlepeople to determine the amount of concrete necessary for effective filling [6]. Consequently, this method often leads to a substantial waste of concrete. Enabling advanced imaging systems to monitor and control the procedure can significantly improve accuracy and worker safety and reduce excess use of material [7].

This paper presents a robust real-time method for semantic 3D reconstruction tailored for robotic construction applications. It introduces a novel computer vision system for the detection and reconstruction of shotcrete construction sites, using YOLOv8m-seg for semantic segmentation and InfiniTAM for 3D reconstruction.

The main contributions of this paper are summarized as follows:

- Compilation of a new semantic segmentation dataset and training of a robust segmentation model.
- Introducing an integration strategy between 3D reconstruction and semantic segmentation, adapted for robotic applications within construction sites.

The paper is organized as follows: In Section 2, the latest advancements in semantic segmentation, 3D reconstruction, and computer vision datasets tailored to construction sites are outlined. Section 3 provides a comprehensive analysis of each proposed pipeline component and the interaction between them. Section 4 demonstrates the experimental assessment, and Section 5 concludes this paper.

2 Related work

2.1 Semantic segmentation

Semantic segmentation is one of the key components necessary for a robotic system to analyze its surroundings. With the advent of deep learning, a boost in performance gave the computer vision community momentum to research new, deep-learning-based, methods for semantic segmentation. These methods allow for parameter-less inference methods and do not rely on extensive domain knowledge. U-Net [8] is a popular, one of the first deeplearning-based semantic segmentation models [9] with an application in biomedicine. DeepLabV3 [10] is another popular model for semantic segmentation. It made its debut in 2017 and is the result of incremental developments based on the first [11] and second [12] versions of DeepLab. The use of atrous convolutions and atrous spatial pyramid pooling makes DeepLab have a larger receptive field and thus higher-resolution feature maps, retaining more information in deeper layers. This enables the integration of both local and global contexts when extracting features. HRNet [13] was released in 2020 and also addresses the issue of retrieving and maintaining highresolution representations from the input image. The main contributions of HRNet are the so-called parallel multiresolution convolutions and repeated multi-resolution fusions.

YOLO (You Only Look Once) [14] is a popular series of Convlutional Neural Network (CNN)-based object detection models, first released in 2016 and consequently developed up to this date in various versions [15, 16, 17, 18]. Its latest version, YOLOv8, provides a unified framework for image classification, object detection, instance segmentation, object tracking, and pose estimation. Architectural details concerning this model can be found in [19]. It provides an easy-to-use interface for training the various tasks and integrates well with experiment tracking tools, ensuring quality control.

Other recent advancements like Segment Anything [20] or BEiT-3 [21] are departing from CNN-based architectures and use Transformer-based models trained on largescale datasets, providing foundation models capable of tackling multiple downstream tasks.

In our work, we combine YOLOv8m-seg with a mask tracking model to stabilize predictions obtained from a low-data model.

2.2 3D reconstruction

Recently there has been a notable effort to integrate deep learning techniques into 3D reconstruction approaches. Specifically, this effort focuses on Neural Radiance Fields (NeRF) [22], and its variants such as Instant-NGP NeRF [23]. These methods are pivotal elements that boosted

3D reconstruction. Typically, NeRF-based methods use a regression technique for opacity and color together with numerical integration, predicting the real step of the volumetric rendering function based on images with known camera poses. Robotic applications of 3D reconstruction in the construction industry include Inspection-NeRF [24] as a method for capturing surface defects in the form of RGB-D images, collected by a wall-climbing robot to create a 3D model and its bounding box, aligning it with the NeRF implicit boundary. In the work of [25], semantic segmentation is applied to a 3D model generated by Instant-NGP to construct BIM models from a sequence of construction site images. Additionally, conventional methods such as Clustering Multi-View Stereo (CMVS) and Patch-based Multi-View Stereo (PMVS) [26], are deployed for dense 3D reconstruction in construction sites.

Notwithstanding the advancements in NeRFs, these techniques require absolute scale recovery and entail significant computational time requirements, especially when handling large-scale scenes such as construction sites [27]. More specifically, 3D reconstruction pipelines that integrate camera pose estimation methods and dense meshing frameworks have demonstrated their superior feasibility for 3D reconstruction in time-sensitive construction projects [28]. KinectFusion [29] paved the way for approaches using real-time volumetric Truncated Signed Distance Fields (TSDF), resulting in InfiniTAM [30]. This method uses RGB-D input to perform real-time reconstruction. To accomplish that, it enables a module for estimating camera poses with a keyframe-based relocalization system and provides globally consistent reconstruction, using either TSDF or surfel methods. Additionally, most recent RGB-D or stereo approaches are using real-time TSDFs from Euclidean Signed Distance Fields (ESDFs) to formulate implicit surfaces [31].

In this study, InfiniTAM is investigated as a realtime modular method and its integration with the wellestablished visual SLAM algorithm, namely RTAB-MAP [32], is explored to achieve more accurate results.

2.3 Computer Vision datasets in construction

Examples of datasets focusing on computer vision in construction include the Alberta Construction Image Dataset [33], datasets for safety helmet detection [34, 35], and the SODA dataset [36], designed for general-purpose object detection in construction environments. For specific construction activities, datasets such as CODEBRIM [37] address concrete inspection, while others like [38] focus on window installation. Regardless, publicly available datasets captured in construction environments, specifically tailored for robotic application in construction areas, remain a scarce resource. Consequently, we compiled custom datasets for our application.



Figure 1. System overview of the proposed and integrated methods

3 Methodology

The structure of our method is illustrated in Figure 1. It is composed of several modules with their respective submodules which are covered in detail in the following.

3.1 Semantic understanding

The proposed end-to-end deep-learning-based pipeline consisting of YOLOv8 and SiamMask needs to be trained to yield sensible results. YOLOv8 and SiamMask were selected due to their state-of-the-art performance. Furthermore, YOLOv8 has a well-known architecture, ensuring reliable and interpretable results. We use an unmodified, pre-trained SiamMask model. Thus, the training efforts concern YOLOv8m-seg. The following section describes experiments carried out to obtain a well-performing semantic segmentation module.

3.1.1 Training setup

The training goal is to teach the model how to segment areas where shotcrete needs to be applied. We stipulate that such areas are easily identified by exposed reinforcement bars. The model is therefore trained to detect exposed reinforcement bars. We compiled a small dataset for this task: detecting exposed reinforcement bars in wooden structures, without a significant presence of other objects. The dataset restricts the appearance of areas in which shotcrete is applied, to rectangular areas with a rather uniform, wooden background and good lighting conditions. In real-life conditions, these areas are less regular in shape and have a bigger variety of backgrounds and lighting conditions. The dataset consists of three splits:

- Training: 515 frames, 580 instances
- Validation: 191 frames, 167 instances
- Testing: 210 frames, 210 instances

We train the YOLOv8m-seg model for 100 epochs with default hyperparameters recommended by [19].

3.1.2 Pre-training procedure

We perform a custom pre-training schedule as follows:

- Train the YOLOv8m bounding box detection model from COCO pre-trained weights provided by [19] on the COCO dataset with augmented, synthetic reinforcement bars [39], for 100 epochs. All COCO classes are used, plus an "ExposedBars" category.
- 2. Considering the epochs from step 1., we use the weights that achieved the best bounding box mean Average Precision (50-95) (mAP50-95) as a starting point and train on the CODEBRIM dataset [37] with CODEBRIM classes, as well as on an augmented version of the dataset with synthetic reinforcement bars.
- 3. Use weights from the last epoch of step 2. as the custom pre-trained weights to start training the segmentation model on the dataset described in 3.1.1.

Figure 2 shows the values of the segmentation loss (both evaluated on the training and validation dataset split) and the development of mask precision and recall (evaluated on the validation dataset split) over training epochs. The losses and metrics show no sign of overfitting, i.e., the validation loss increasing while the training loss is decreasing.



Figure 2. Training and validation losses per epoch (left) and Mean Mask Precision and Mean Mask Recall per epoch (right)

3.1.3 Robustness enhancement

Perturbations of the input image sequences, e.g., rotations and distortions, which are often encountered from cameras mounted on mobile robots can lead to cases where the segmentation model cannot re-detect previously seen areas with exposed reinforcement bars. To counteract this, an object-tracking model is deployed. The semantic segmentation model in combination with the tracking model constitutes the complete semantic understanding module. The tracking model used for this task is SiamMask [40], using the default pre-trained model on the DAVIS dataset [41] as well as the default configuration provided by the developers.

When the segmentation model detects an object, the tracker is initialized with the respective image and bounding boxes. They are stored in a first in, first out (FIFO) queue of images and bounding boxes. When the segmentation model fails, the tracker pops a bounding box and the corresponding image from the queue to find the content of the bounding box in the current image.

We evaluate the effect of the tracking model on the semantic understanding module. The module is tested on the validation and testing sequence of the dataset, both with disabled and enabled tracking. We use a FIFO queue length of 1 and default confidence thresholds for both the segmentation model and the tracker model. Figure 3 shows plots of the mean Intersection over Union (mIOU) calculated on all available IOUs up to that frame. Both plots indicate the superior performance of the semantic understanding module when the tracking model is used as an additional layer to recover false negatives. The benefit is mainly pronounced on the test dataset split, as the segmentation model fails and thus activates the tracking submodule more often. A video showcasing the benefit of the module can be found at [42]. In this video, frames with red masks show the output of a tracked mask when the segmentation module failed to produce an output.

In conclusion, the SiamMask tracking model can increase robustness in the case of the segmentation model failing on objects/areas it has detected once before.

3.1.4 Hyperparameter tuning

This section evaluates the influence of tracker and segmentation model hyperparameters on the overall performance measured in mean precision, mean recall, and mIOU. We perform a grid search on these parameters, to maximize mIOU on the test set:

- 1. Tracker FIFO queue length (QL), search range: $\{1, 2, ..., 11\}$
- 2. Tracker Confidence Threshold (TCT), search range: $\{0.1k | k \in \{1, 2, ..., 9\}\}$
- Segmentation Confidence Threshold (SCT), search range: {0.1k|k ∈ {1, 2, ..., 9}}

Contrary to common practice, the sweep is conducted on the test split. This is due to the minimal effect of the tracker on the validation split.

We found that the 10 best-performing runs exclusively use a QL of 2 together with medium to high TCTs (0.6 - 0.8) and medium to low SCTs (0.1 - 0.4). The 10 worstperforming runs tend to use medium to high QLs (6-10), medium to high TCTs (0.6-0.9), and a high SCT (0.9). This suggests the usage of a QL of 2 together with a low SCT and a high TCT. We have chosen QL=2, SCT=0.4, TCT=0.9, achieving a mIOU of 0.85 together with mean Precision (mP)=0.96 and mean Recall (mR)=0.88 on the test split.



Figure 3. Temporal development of detections and mIOU, on the validation (left) and test (right) splits. Green dashed line shows the current mIOU with tracking enabled, whereas the orange line indicates the scenario with tracking disabled.

3.2 3D Reconstruction

The 3D reconstruction process uses RGB-D input, obtained through a stereo-vision camera to infer a real-time 3D mesh. The 3D reconstruction module comprises the subsequent submodules:

3.2.1 3D mesh creation

Aiming to develop an end-to-end modular pipeline, we strategically applied the InfiniTAM algorithm. Its key advantage over recent approaches is the capability to synthesize accurate 3D surfaces in real-time, addressing absolute scale recovery and long computational time issues common in other methods, as reported in [27], which is crucial for dynamic and expansive scenes such as construction sites. Furthermore, it seamlessly integrates with stereo sensors, typically used for robotic vision applications.

In particular, InfiniTAM infers TSDF-based volumetric reconstruction, using hash tables. It relies on RGB-D input to segment the scene into rigid subscenes and refines their relative poses to build a coherent overall map. During the camera pose prediction, it adjusts the current camera position to track the sensor frame, aligning surface measurements with the model estimation [30].

It is worth noting that our approach strongly depends on the performance of the stereo camera, as conventional RGB-D cameras may face limitations in composing depth images in semi-indoor scenes with changing illumination conditions. To fulfill these objectives, the Roboception RC-Visard 160 stereo camera was deployed.

3.2.2 Odometry estimation

During our experiments, we observed poor relocalization performance of vanilla InfiniTAM, when dealing with partial and noisy surface measurements, captured under conditions of reduced overlap. To address this challenge, we conducted a thorough investigation of the camera pose estimation. Leveraging InfiniTAM's modularity, we looked into the integration of a more robust camera pose estimation module. For this, we deployed the well-established visual odometry method RTAB-Map [32], which is a flexible Graph-Based SLAM technique, to create dense 3D reconstructions. Consequently, the integration of an advanced external camera pose estimation played a significant role in developing a robust real-time 3D reconstruction method, demonstrating adaptability in handling complex environments such as construction sites.

3.3 Integrated system

The proposed system consists of two interconnected major submodules: semantic understanding and 3D reconstruction. The semantic understanding module deploys a YOLOv8m-seg [19] segmentation model, stabilized with the SiamMask [40] segmentation mask tracker, adept at robustly identifying exposed reinforcement bars. The 3D reconstruction module incorporates an external odometry estimation component and employs TSDF volumetric reconstruction, leveraging RTAB-Map [32] and InfiniTAM [30], respectively.

As illustrated in Figure 1, the integrated method follows these steps: Given the sensor data (RGB image, depth image) as an input, the semantic understanding module infers a mask, marking areas with exposed reinforcement bars in each image. Simultaneously, the sensor data is used to obtain a visual odometry estimation. Combined with the segmentation mask and sensor depth measurements, the 3D reconstruction module constructs a semantic 3D model of the environment. This model can be used to acquire a digital representation of the construction site in real-time, streamlining the planning and execution of robotic shotcrete operations.

4 Experimental evaluation

4.1 Experimental process

We tested our pipeline at a semi-indoor construction site featuring wooden panels, both with and without exposed reinforcement bars. We tested before shotcrete application, as depicted in the upper part of Figure 4. The 3D reconstruction module processed data captured by a Roboception RC-Visard 160 stereo camera. The semantic understanding module inferred masks for areas with exposed rebars.

4.2 Results

Qualitative results are presented in the lower part of Figure 4. The semantic understanding module precisely identified regions of interest, denoted by cyan masks.

We measured the effectiveness of the integrated framework in terms of its geometric precision by applying a manual procedure to measure the point-to-point distances of the wooden frames within the 3D model, corresponding to the segmented area. The vertical and horizontal dimensions of the real wooden frame were measured to be 2 meters long and 1 meter wide. We selected several 3D point pairs in the point cloud (X, Y, Z) to measure their distances and assess the accuracy of the scene reconstruction. Specifically, the experiment involved 12 point-to-point measurements of the frame's dimensions, as illustrated in Figure 5. Subsequently, we compute the root mean square error (RMSE) by comparing the distances of the real wooden frame with the distances of the reconstructed 3D model. The RMSE amounts to 0.564 centimeters.

5 Conclusion

We proposed a real-time method for semantic 3D reconstruction for robotics-based construction applications. Our method uses a robust semantic understanding module using a custom YOLOv8m-seg segmentation model and the SiamMask mask tracker, together with RTAB-Map for odometry estimation and InfiniTAM for 3D reconstruction. The resulting semantic 3D mesh model is an important step towards introducing robotic systems in shotcrete construction to improve worker safety and alleviate labor shortages. Our experimental evaluation concluded that the developed method is applicable for semantic understanding and reconstruction of semi-indoor construction scenes, highlighting regions of interest before performing



Figure 4. Upper part: samples of two images of the testing area within the construction site. Lower part: semantically annotated 3D mesh model, where regions of interest are highlighted in cyan.



Figure 5. Samples of point-to-point measurements of the 3D reconstruction accuracy

shotcreting. The novel approach was tested under realistic construction site conditions, showcasing good performance in terms of RMSE, mIOU, mean precision and mean recall. We believe that the method can be extended to accommodate other common shotcrete application scenarios, as well as to spark general interest in introducing advanced computer vision approaches to challenging construction tasks.

6 Acknowledgements

This work has been funded and supported by the EU Horizon Europe project "RobetArme" under the Grant Agreement 101058731. We extend our gratitude to Christiansen & Essenbaek A/S (CEAS) for providing access to their premises and for the mock-up construction site.

References

- European Construction Sector Observatory. Digitalisation in the construction sector. Apr. 2021. URL: https://ec.europa.eu/ docsroom/documents/45547/attachments/ 1/translations/en/renditions/pdf.
- [2] European Labour Authority. Report on labour shortages and surpluses: November 2021. en. LU: Publications Office, 2021. URL: https://data. europa.eu/doi/10.2883/746322.
- [3] URL: https://ec.europa.eu/eurostat/ statistics-explained/index.php?title= Accidents_at_work_-_statistics_by_ economic_activity.
- [4] Zhiliang Ma and Shilong Liu. "A review of 3D reconstruction techniques in civil engineering and their applications". In: *Advanced Engineering Informatics* 37 (2018), pp. 163–174.
- [5] Pingbo Tang et al. "Automatic reconstruction of as-built building information models from laserscanned point clouds: A review of related techniques". In: *Automation in construction* 19.7 (2010), pp. 829–843.
- [6] Ioannis Kostavelis et al. "RoBétArmé Project: Human-robot Collaborative Construction System for Shotcrete Digitization and Automation through Advanced Perception, Cognition, Mobility and Additive Manufacturing Skills". In: Open Research Europe 4 (Jan. 3, 2024), p. 4. ISSN: 2732-5121. DOI: 10.12688/openreseurope.16601.1.
- [7] A. Sawhney, M. Riley, and J. Irizarry. Construction 4.0: An Innovation Platform for the Built Environment (1st ed.) Routledge, 2020.
- [8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: arXiv:1505.04597 (May 2015). arXiv:1505.04597 [cs]. URL: http: //arxiv.org/abs/1505.04597.
- Shervin Minaee et al. "Image Segmentation Using Deep Learning: A Survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021), pp. 1–1. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2021.3059968.
- [10] Liang-Chieh Chen et al. "Rethinking Atrous Convolution for Semantic Image Segmentation". In: arXiv:1706.05587 (Dec. 2017). arXiv:1706.05587
 [cs]. URL: http://arxiv.org/abs/1706.05587.

- [11] Liang-Chieh Chen et al. "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs". In: arXiv:1412.7062 (June 2016). arXiv:1412.7062 [cs]. URL: http://arxiv.org/ abs/1412.7062.
- [12] Liang-Chieh Chen et al. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: arXiv:1606.00915 (May 2017). arXiv:1606.00915 [cs]. URL: http://arxiv.org/ abs/1606.00915.
- [13] Jingdong Wang et al. "Deep High-Resolution Representation Learning for Visual Recognition". In: (2019). DOI: 10.48550/ARXIV.1908.07919.
- [14] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: arXiv:1506.02640 (May 2016). arXiv:1506.02640
 [cs]. URL: http://arxiv.org/abs/1506. 02640.
- [15] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection". In: arXiv:2004.10934 [cs, eess] (Apr. 2020). arXiv: 2004.10934. URL: http://arxiv.org/abs/ 2004.10934.
- [16] Zheng Ge et al. "YOLOX: Exceeding YOLO Series in 2021". In: arXiv:2107.08430 [cs] (Aug. 2021). arXiv: 2107.08430. URL: http://arxiv.org/ abs/2107.08430.
- [17] Joseph Redmon and Ali Farhadi. "YOLOv3: An Incremental Improvement". In: arXiv:1804.02767
 [cs] (Apr. 2018). arXiv: 1804.02767. URL: http: //arxiv.org/abs/1804.02767.
- [18] Chien-Yao Wang, I.-Hau Yeh, and Hong-Yuan Mark Liao. "You Only Learn One Representation: Unified Network for Multiple Tasks". In: arXiv:2105.04206 [cs] (May 2021). arXiv: 2105.04206. URL: http: //arxiv.org/abs/2105.04206.
- [19] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8. 2023. URL: https://github. com/ultralytics/ultralytics.
- [20] Alexander Kirillov et al. "Segment Anything". In: (2023). DOI: 10.48550/ARXIV.2304.02643.
- [21] Wenhui Wang et al. "Image as a Foreign Language: BEiT Pretraining for All Vision and Vision-Language Tasks". In: arXiv:2208.10442 (Aug. 2022). arXiv:2208.10442 [cs]. uRL: http:// arxiv.org/abs/2208.10442.
- [22] Ben Mildenhall et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *CoRR* abs/2003.08934 (2020).

- [23] Thomas Müller et al. "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding". In: CoRR abs/2201.05989 (2022). arXiv: 2201.05989. URL: https://arxiv.org/abs/ 2201.05989.
- [24] Kunlong Hong, Hongguang Wang, and Bingbing Yuan. "Inspection-Nerf: Rendering Multi-Type Local Images for Dam Surface Inspection Task Using Climbing Robot and Neural Radiance Field". In: *Buildings* 13.1 (2023). ISSN: 2075-5309.
- [25] Martin Fisher Shun Hachisuka Alberto Tono. "Harbingers of NeRF-to-BIM: a case study of semantic segmentation on building structure with neural radiance fields," in: 2023.
- [26] Johannes L. Schönberger et al. "Pixelwise View Selection for Unstructured Multi-View Stereo". In: *Computer Vision – ECCV 2016*. Ed. by Bastian Leibe et al. Cham: Springer International Publishing, 2016, pp. 501–518. ISBN: 978-3-319-46487-9.
- [27] Dimitrios Katsatos et al. "Comparative Study of Surface 3D Reconstruction Methods Applied in Construction Sites". In: 2023 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE. 2023, pp. 1–6.
- [28] Zhexiong Shang and Zhigang Shen. "Real-Time 3D Reconstruction on Construction Site Using Visual SLAM and UAV". In: 2018.
- [29] Richard A. Newcombe et al. "KinectFusion: Realtime dense surface mapping and tracking". In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality. 2011, pp. 127–136. DOI: 10. 1109/ISMAR.2011.6092378.
- [30] Victor Adrian Prisacariu et al. "InfiniTAM v3: A Framework for Large-Scale 3D Reconstruction with Loop Closure". In: *CoRR* abs/1708.00783 (2017). arXiv: 1708.00783.
- [31] Helen Oleynikova et al. "Voxblox: Building 3D Signed Distance Fields for Planning". In: CoRR abs/1611.03631 (2016). arXiv: 1611.03631. URL: http://arxiv.org/abs/1611.03631.
- [32] Mathieu Labbé and François Michaud. "RTAB-Map as an open-source lidar and visual SLAM library for large-scale and long-term online operation". In: *Journal of Field Robotics* 36.2 (2019), pp. 416–446. DOI: https://doi.org/10.1002/ rob.21831.

- [33] Bo Xiao and Shih-Chung Kang. "Development of an Image Data Set of Construction Machines for Deep Learning Object Detection". en. In: *Journal of Computing in Civil Engineering* 35.2 (Mar. 2021), p. 05020005. ISSN: 0887-3801, 1943-5487. DOI: 10. 1061/(ASCE)CP.1943-5487.0000945.
- [34] URL: https://www.kaggle.com/datasets/ andrewmvd/hard-hat-detection.
- [35] Liangbin Xie. *Hardhat*. 2019. DOI: 10.7910/DVN/ 7CBGOS.
- [36] Rui Duan et al. "SODA: A large-scale open site object detection dataset for deep learning in construction". In: Automation in Construction 142 (2022). Citation Key: DUAN2022104499, p. 104499. ISSN: 0926-5805. DOI: https://doi.org/10.1016/j.autcon.2022.104499.
- [37] Martin Mundt et al. "Meta-Learning Convolutional Neural Architectures for Multi-Target Concrete Defect Classification With the COncrete Defect BRidge IMage Dataset". In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, June 2019, pp. 11188–11197. ISBN: 978-1-72813-293-8. DOI: 10.1109/CVPR.2019.01145.
- [38] Jiesheng Yang et al. "Computer Vision for Construction Progress Monitoring: A Real-Time Object Detection Approach". In: arXiv:2305.15097 (May 2023). arXiv:2305.15097 [cs]. uRL: http: //arxiv.org/abs/2305.15097.
- [39] Haoyu Wang et al. "Synthetic Datasets for Rebar Instance Segmentation Using Mask R-CNN". en. In: *Buildings* 13.3 (Feb. 2023), p. 585. ISSN: 2075-5309. DOI: 10.3390/buildings13030585.
- [40] Qiang Wang et al. "Fast Online Object Tracking and Segmentation: A Unifying Approach". In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, June 2019, pp. 1328–1338. ISBN: 978-1-72813-293-8. DOI: 10.1109/CVPR.2019.00142.
- [41] F. Perazzi et al. "A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, June 2016, pp. 724–732. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.85.
- [42] Patrick Schmidt. RobetArme Stabilized SiamMask output. Youtube. 2024. URL: https://www. youtube.com/watch?v=In6euNHw_Sk.

Effect of Missing Data on Machine-Learning Algorithms for Real-time Safety Monitoring in Scaffolds

Laura Alvarez¹, Mahendra Ghimire² and JeeWoong Park³

^{1,2,3}Department of Civil and Environmental Engineering and Construction, University of Nevada Las Vegas, USA alvarezq@unlv.nevada.edu, ghimim2@unlv.nevada.edu, jee.park@unlv.edu

Abstract –

In the context of real-time data acquisition and processing, dealing with missing data (MD) is a common challenge that can compromise the quality and effectiveness of machine learning (ML) algorithms. Previous research focuses on creating a real-time safety monitoring system that predicts safety conditions in scaffolds by analyzing strain measurements from sensors placed in the structure's columns. However, it does not address the effect of sensor failures and the resulting MD. This paper explores how the presence of MD, caused by faulty sensors, affects the performance of eight ML algorithms in a safety monitoring scaffolding system: gaussian naive Bayes (GNB), random forest (RF), multi-layer perceptron (MLP), support vector machine (SVM), decision tree (DT), XGBoost (XGB), logistic regression (LR), and linear support vector classification (LSVC). This study identifies how these algorithms perform when processing datasets with missing values. As the amount of MD in the datasets increases, there is a consistent negative influence on the performance of each algorithm, resulting in reduced predictive accuracy. Among all the tested ML algorithms, RF and DT have shown to be the most sensitive to MD.

Keywords -

Scaffolds; classification; missing data; machine learning.

1 Introduction

Real-time data acquisition and processing often face the challenge of missing data (MD), impacting data quality and machine learning (ML) algorithm performance [1], impacting pattern identification [2]. Strategies to handle MD are crucial [3] because its presence introduces a risk of process failures, failing to accurately represent the true reality of the system [4]. The collective findings from several studies [5–8] underscore the significant influence of MD on the ML algorithms' performances, and stress the necessity of handling MD effectively to ensure accurate and dependable ML results.

Previous research focuses on real-time monitoring of intricate scaffolds using ML techniques to forecast safety conditions [9]. This study delves into a technique for categorizing instances of scaffold failure and accurately predicting safety conditions, using data from strains installed on the scaffold columns. The authors successfully improved the accuracy of ML classification through a self-multiplication technique [10]; nonetheless, prior research did not account for the influence of sensor failures and the subsequent absence of data on the system's acquisition. In this specific case, MD can cause an incorrect prediction in the scaffold structure's safety conditions. For instance, if the scaffold structure is about to overturn due to unbalanced loads, the MD in a sensor measurement can result in a safe classification instead of an overturning one.

These real-time sensing systems should demonstrate high accuracy in safety monitoring to promptly detect temporary structures' potential structural failures. The primary purpose of research toward reliable safety predictions is to safeguard the lives of workers [10]. Accurate predictions help to identify potential hazards or structural failures in advance, allowing for timely interventions or preventive measures to protect workers from injuries. This could involve reinforcing a structure, evacuating an area, or adjusting working conditions to prevent accidents. In addition, an accurate safety prediction can save costs associated with medical expenses, structural or property damage, legal liabilities, and potential project delays.

In response, this project studies the problem of MD in a scaffold dataset. It investigates MD effects on the performance of eight ML classification algorithms: gaussian naive Bayes (GNB), random forest (RF), multilayer perceptron (MLP), support vector machine (SVM), decision tree (DT), XGBoost (XGB), logistic regression (LR), and linear support vector classification (LSVC).

2 Literature Review

To study the effects of MD on data mining processes, it is crucial to understand MD mechanisms [11]. The methods used to handle missing values are often based on assumptions tied to the underlying mechanisms causing the MD [12]. Four mechanisms of the occurrence of MD are defined [13]: when the entry is not supposed to have a value in a given field, it is said that it is structurally missing (SMD) [14]. Missing completely at random (MCAR) is when the cause of data absence is independent from the observed and unobserved entries. Missing at random (MAR) is when the cause of data absence is related to the other observed entries. Finally, missing not at random (MNAR) is when the cause of data absence is related to the missing entry and the other observed entries. According to the previous definitions, this study deals with MCAR data, because the missing values originate when a sensor failure occurs due to a communication error or device malfunction. This event is not related to any other measured variable in the system.

The influence of MCAR on ML algorithms for safety monitoring cannot be overstated. Studies [5,8] highlight how the presence of MCAR significantly distorts the integrity of datasets, leading to biased predictions and compromised algorithmic performance. This bias stems from the randomness of MD, affecting the statistical properties of the dataset and, consequently, the MS models' training and generalization capabilities [7].

In a comprehensive review [15] of 152 ML-based clinical prediction model studies, the reporting quality regarding MD was generally poor, aligning it with similar reviews. Excluding participants with MD and insufficient details was found to be a common practice of MD handling methods. Other studies [16] utilized techniques like mean imputation or complete-case analysis for healthcare data. The review highlighted the pressing need for improved reporting guidelines, adherence, and understanding the repercussions of improper MD handling in ML-based prediction studies in healthcare.

The gaps and challenges related to MD are evident in ML-based heart disease prediction models [17]. The limited exploration of MD effects reveals a lack of comprehensive understanding regarding model performance. The study demonstrates by varying accuracy percentages in predicting heart disease, yet fail to explicitly address how to handle MD or its influence on these outcomes. This highlights the need for a more nuanced comprehension of data preprocessing methods. Another study [6] centers on how MD affects ML algorithms used in hydrologic predictions and proposes a method to fill in the gaps. A comprehensive overview [18] of MD in ML emphasizes the importance of appropriately addressing and evaluating various imputation techniques. Additionally, the study notes the use of smaller, domain-specific datasets, accenting the need for exploring MD handling in larger, real-world datasets with diverse features.

Applications in construction [19] focus on critically evaluating concrete strength predictions for enhanced sustainability. However, challenges arise due to MD, noise, and model interpretability. The authors used a dataset of manufactured sand concrete and various ML algorithms to demonstrate predictive performance. They found it necessary to bridge MD concerns and enhance interpretability for reliable concrete strength predictions in construction applications.

In safety applications, researchers addressed MD issues on high-plateau flights [20] by proposing an improved method based on least squares support vector machines (LS-SVM). This method approaches the challenges placed by missing or abnormal quick access recorder data due to harsh environmental conditions. Through advanced ML techniques, this method enhances the reliability and accuracy of flight data processing and analysis, contributing to improved aviation safety.

Estimations of increased risk of crashes on freeways need to be accurate and reliable, but the utilization of real-time traffic data in proactive safety management systems is lacking due to MD. To fill these gaps, past research [21] has proposed a framework for real-time risk assessment on freeways by integrating data from multiple detection systems, real-time weather, and roadway geometry. The development of the framework mitigates the effects of MD, contributing to the system with high estimation accuracy, robustness, and reliability.

To detect high-severity accidents in the construction industry, accurate prediction models are needed. While occupational accidents are common in construction, the challenge lies in determining the combination of preprocessing techniques that yield the most accurate severity prediction, considering issues such as MD, outliers, feature scaling, and imbalanced class distribution. Specifically, in dealing with MD, the researchers [22] have experimented with different scenarios of preprocessing techniques to determine the best combination. One of the scenarios involved not removing MD, indicating that missing values were retained in the dataset rather than being imputed or deleted. This approach allowed the model to learn from the available data without discarding potentially valuable information.

Prior studies developed a real-time safety monitoring system [9] based on strain measurements from sensors installed in the columns of the structure but did not consider the implications of MD introduced by sensor malfunctions. The previous articles contribute to understanding the influence of MD and how to improve model accuracy, comparability, and reliability in different areas of study. These contributions are used as guidelines, as the nature of the described problems present similar challenges to the ones found when a faulty sensor generates MD in a scaffold safety monitoring system's dataset. Leveraging existing knowledge from various domains helps in developing strategies to handle MD effectively in the context of scaffold safety monitoring systems. Understanding how MD affects the performance of various ML algorithms contributes to the development of strategies to address specific types of MD and the mitigation of their impacts on algorithmic performance.

3 Objective and Scope

This research aims to assess the influence of MD, stemming from sensor failures, on the ML algorithms' predictive accuracy and performance in a real-time safety monitoring system for scaffolds. This paper evaluates the performance of eight ML algorithms: GNB, RF, MLP, SVM, DT, XGB, LR, and LSVC when exposed to datasets that feature MD caused by the simulation of faulty sensors. This paper expands upon a previous study [10] through additional investigation, evaluation, and discussions of MD across various ML models. By employing a diverse set of classification algorithms, the study explores different modeling approaches to handle the same type of data and capture the patterns. While reviewing different safety-related studies conducted with ML application, it was observed that GNB [23-25], RF [26–28], MLP [26,28,29], SVM [26,28,30], DT [28,31,32], XGB [33-35], LR [32,36,37] and LSVC [26,28,38] are mostly employed by researchers. Therefore, these eight ML models are used in this investigation.

4 Approach

The proposed approach comprised four fundamental steps, as illustrated in Figure 1. General approach of the studyFirst, it conducts an in-depth analysis of the initial dataset, establishing a benchmark for understanding the complete data's characteristics. Subsequently, simulated MD is introduced to replicate sensor faults, enabling the evaluation of ML algorithms' performance under these conditions. Then, the algorithms are assessed for their predictive capabilities using both the original complete dataset and the data affected by missing values.

A crucial aspect involves progressively evaluating algorithm performance as the number of faulty sensors or MD increases, highlighting the repercussions of sensor failures on algorithm accuracy. This comprehensive approach provides insights into the ML classification algorithms' behavior in scenarios involving faulty sensors, aiding in understanding their robustness.

The results may contribute to making decisions about algorithm selection and implementation in real-world applications. The previous steps are explained and justified as follows:



Figure 1. General approach of the study.

Step 1: Data analysis. This step involves thoroughly examining the characteristics of the original complete dataset. It is necessary to establish a baseline understanding of the nature of the system, including data distribution, variability of measurements, and any underlying patterns. This analysis provides essential context for subsequent steps and helps to identify potential issues or anomalies in the dataset.

Step 2: Data amputation. Simulated MD is introduced to replicate sensor faults, replicating real-world scenarios where data may be incomplete due to sensor malfunctions. This step is crucial to evaluate ML algorithms' performance in the presence of MD. By simulating sensor faults, it is possible to estimate how well the algorithms may handle these conditions and whether they can effectively make predictions despite incomplete information.

Step 3: ML algorithm evaluation. In this step, ML algorithms are trained and tested using the original complete datasets. This allows the evaluation of the algorithms ability to make accurate predictions, while operating under ideal conditions before introducing MD. This step is a starting point assessment to compare algorithm performance under different conditions and determine how MD affects predictive accuracy. It provides a quantifiable perception of the algorithms' robustness and their reliability in real-world applications.

Step 4: Progressive evaluation with increasing MD. As the number of faulty sensors or MD increases, algorithm performance is progressively evaluated. This step is essential to understand how algorithm accuracy changes as data quality deteriorates due to sensor failures. By systematically increasing the severity of MD scenarios, it is possible to identify thresholds where algorithm performance significantly degrades. This evidence helps to understand the limitations of ML algorithms in handling MD and informs decision-making regarding algorithm selection and deployment.

4.1 Real-time safety monitoring system

This study incorporates a real-time safety monitoring system designed for scaffold structures, based on a previous investigation [10]. The system relies on strain sensors embedded within the columns to gather crucial data indicative of potential scaffold failures. Figure 2 (Figure 5 in [10]), illustrates the scaffold's configuration, comprising 10 columns distributed across 3 stories with 20 sensors strategically positioned at various locations. The sensor measurements are used to predict the scaffold safety condition as overload, uneven, sideways, and safe by processing the data with ML classification algorithms in Python (GNB, RF, MLP, SVM, DT, XGB, LR, and LSVC). The algorithms are a built-in function from sklearn package, except for XGB, which has its own package called xgboost. Training was carried out with complete datasets without including MD.



Figure 2. Scaffolding Structure with sensor measurement (Figure 5 in [10]).

4.2 Dataset

The original complete dataset was obtained from a

previous study [9]. The dataset for strain measurement was created based on structural conditions observed in scaffold usage and distinguished between safe and unsafe conditions. The unsafe category was divided into global and local failures. Global failures involved overturning in both lateral X and Y directions. Local failures included uneven settlement and overloading issues. The scaffold model comprised 10 vertical members, each equipped with 20 evenly distributed strain-measuring sensors placed on them, as shown in Figure 2 (Figure 5 in [10]). The output is divided into 23 classes representing different safety conditions. The dataset contains 1,000 samples for each of the 23 classes.

4.3 MD generation and follow-up analysis

The initial dataset was generated based on the automated monitoring system's optimal operation conditions without accounting for MD [10]. Therefore, data amputation is needed to simulate the MCAR scenario.

This paper considers the progressive occurrence of failure. In this progression, the number of failing sensors starts from 1 to 20, equal to the entire sensors. While this may not make sense in a practical manner, it is important that this research investigates the effect of incremental sensor failures on the classification accuracy from a theoretical perspective. To simulate faulty sensors, or MD, null values were introduced, which are considered as $0.000\mu\epsilon$ as an indication of absence of measurement.

The conducted analysis involved observing how different ML algorithms responded to the introduced failure cases in the dataset and thus overall prediction performance. By systematically introducing these cases across all sensors progressively, how each algorithm adapted to and handled the simulated errors was evaluated. algorithm presents a summary of the analysis output by showing the prediction accuracy in relation to the number of faulty sensors the entire dataset across all the tested ML algorithms.

The evaluation metric used for ML classification algorithms is accuracy instead of other metrics because is straightforward to understand. It represents the proportion of correctly predicted classes out of the total classes in the dataset and intuitively captures how well the ML model performs overall. In this specific case of study, accuracy is an appropriate metric due to the balanced dataset, i.e., the dataset contains 1,000 samples for each of the 23 classes.

Accuracy is evaluated by dividing the strain dataset into two portions: a training set and a testing set. This splitting is done to assess how well the ML algorithms perform in making predictions on new, unseen data. Specifically, the testing set comprises 20% of the entire dataset. The purpose of this separation is to use the larger portion (80%) as the training data to teach the model how to make predictions based on patterns and information within that data. The remaining 20% is set aside as the testing data, which is kept separate and not used during the training phase.

After training with the training dataset, the obtained model is then evaluated using the testing dataset. The predictive accuracy is determined based on how well it predicts or classifies the outcomes within this separate testing data.

The following sections further discuss these results.

5 Results and Discussion

Prior to discussing the results, Table 1. ML algorithm performance in optimal operation conditions shows the accuracy achieved by each studied ML classification algorithm on a complete strain dataset. As this is based on a complete dataset, high accuracies are anticipated.

Table 1. ML algorithm performance in optimal operation conditions.

ML Algorithm	Accuracy
RF	0.9998
XGB	0.9998
DT	0.9991
SVM	0.9989
LR	0.9974
MLP	0.9941
GNB	0.9937
LSVC	0.9937

The high accuracy percentages reported in the study for various ML algorithms indicate their effectiveness in correctly classifying a complete dataset without MD presence. The reported accuracies in Table 1 are achieved under optimal operating conditions without any faulty sensor performing. In such conditions, these algorithms perform up to 99%, achieving near-perfect accuracy (100%). Note that more complex algorithms with a larger number of hyperparameters are prone to overfitting and have difficulty handling MD, especially if not appropriately tuned or validated.

Figure 3. Effect of MD generated from faulty sensors on the performance of ML classification algorithms while processing a strain dataset of a scaffold structure.illustrates how the accuracy of the ML classification algorithms is negatively affected by the inclusion of MD in the dataset. The accuracy of GNB drops gradually as the number of faulty sensors increases. With all sensors functioning, it achieves an accuracy of 99.37%. However, as the number of faulty sensors increases, the accuracy decreases progressively, reaching 4% when all sensors are faulty. It still presents at least 50% accuracy with about 8 faulty sensors out of 20 sensors.

RF shows a similar trend, but after the second sensor

fails, the accuracy is more negatively affected than that of GNB.



Figure 3. Effect of MD generated from faulty sensors on the performance of ML classification algorithms while processing a strain dataset of a scaffold structure.

DT shows a pattern where accuracy significantly drops from the third faulty sensor. The accuracy drops from analyzing with 2 sensors in failure to 3 sensors in failure, exhibiting the most detrimental rate at about 50%.

In general, all ML algorithms experience a decrease in accuracy as the number of faulty sensors increases. Except for DT's case, about 50% accuracy was still achieved when 5-7 failing sensors were included in the analysis. Overall, GNB and LSVC appear to be relatively more robust against MD compared to the other algorithms listed here. DT and RF exhibit higher sensitivity to MD, showing a significant decrease in accuracy as the number of faulty sensors increases.

Several reasons could contribute to RF and DT being more sensitive to MD compared to other algorithms, and they can be more susceptible to noisy or inconsistent data. DT create biased nodes when encountering MD, affecting subsequent decision-making and accuracy. MD, which can be considered a form of noise, might be challenging for these algorithms to handle effectively. These algorithms might lack the robustness to handle MD compared to other algorithms like GNB or LSVC, which can handle missing values more effectively due to their underlying mechanisms.

The study also evidences the decline in accuracy as the number of faulty sensors or amount of MD increases. This decrease in accuracy indicates that the algorithms are sensitive to MD, which may be due to their inability to effectively handle such inconsistencies.

It is important to note that the behavior of these ML algorithms concerning MD can depend on various factors, including the specific dataset used in training. Tuning hyperparameters or using specific techniques for MD handling might help to mitigate these algorithms' sensitivity to missing values.

6 Conclusions

This study investigates the effect of MD caused by simulated sensor failures on the performance of ML classification algorithms used in a real-time safety monitoring system for scaffolds. The research focuses on assessing the predictive accuracy of eight ML algorithms when confronted with a dataset containing different amounts of MD.

It was evident that MD has a significant negative influence on ML algorithms' performance, and there is a need to effectively handle it to ensure accurate results. Previous studies in various domains have stated the challenges posed by MD and the need for improved reporting guidelines and understanding the repercussions of improper MD handling in ML-based prediction studies. In the context of scaffold safety monitoring systems, developing strategies to handle MD effectively is required, considering the influence MD has on algorithmic performance.

Results indicate that most ML algorithms achieve over 99% accuracy on the complete dataset, and RF, DT, and XGB exhibit the highest accuracy. However, when introducing progressively incremental MD, all ML algorithms experience a decrease in accuracy. Notably, GNB and LSVC appear relatively robust to MD, while DT and RF exhibit higher sensitivities to MD, showing a significant decrease in accuracy as the number of faulty sensors increases.

To conclude, MD significantly affects ML algorithms' performance, particularly DT and RF, which show higher sensitivity to MD. Possible reasons for this sensitivity include susceptibility to noisy data and a lack of robustness in handling MD compared to that of other algorithms. Dataset characteristics and proper handling techniques must be considered to mitigate the algorithms' sensitivities to MD.

7 Limitations

The study primarily focuses on theoretical simulations of sensor failures, introducing null values to simulate MD. This approach allows for controlled experimentation and may not fully capture the complexities of real-world sensor malfunctions. Although the assumption of sensor failing one by one does not reflect a realistic scenario, this study conducts a progressive analysis to systematically assess MD's effects. By simulating failures in a progressive manner, the study can observe the incremental degradation in algorithm performance with each additional failing sensor. This helps to understand MD's cumulative effect on the algorithms' reliability for safety monitoring applications.

As the study's primary objective is to investigate MD's influence on ML algorithms' performance, the hyperparameter tuning is not considered. The performance comparison of the ML algorithms is made with default hyperparameter settings; this approach provides a baseline for comparison and allows for assessing the algorithms' robustness without additional tuning. To extend and generalize the results, validation with real-time acquisition systems and actual sensor failures are required.

8 Future Work

As future work, MD handling techniques like imputation could be applied to fill in the MD. The realtime safety monitoring system's accuracy and precision could be improved by optimizing these methods. To do this, tuning hyperparameters or data preprocessing could be used. Once the imputation stage is complete, it would be possible to evaluate the imputed values' effects on the predicted safety conditions in the real-time safety monitoring system.

References

- [1] Schauer, J. M., Diaz, K., Pigott, T. D., and Lee, J., Exploratory Analyses for Missing Data in Meta-Analyses and Meta-Regression: A Tutorial. *Alcohol* and Alcoholism, Vol. 57, No. 1, 2022, pp. 35–46. https://doi.org/10.1093/ALCALC/AGAA144
- [2] Alvarez Quiñones, L. I., Lozano-Moncada, C. A., and Bravo Montenegro, D. A., Machine Learning for Predictive Maintenance Scheduling of Distribution Transformers. *Journal of Quality in Maintenance Engineering*, Vol. 29, No. 1, 2023. https://doi.org/10.1108/JQME-06-2021-0052
- Sharma, S., Chmaj, G., and Selvaraj, H., Sensor Data Restoration in Internet of Things Systems Using Machine Learning Approach. Vol. 611 LNNS, 2023, pp. 21–30. https://doi.org/10.1007/978-3-031-27470-1 3
- [4] Rioux, C., and Little, T. D., Missing Data Treatments in Intervention Studies: What Was, What Is, and What Should Be. *International Journal of Behavioral Development*, Vol. 45, No. 1, 2021, pp. 51–58. https://doi.org/10.1177/0165025419880609
- [5] Blomberg, L. C., and Ruiz, D. D. A., Evaluating the

Influence of Missing Data on Classification Algorithms in Data Mining Applications. *Brazilian Symposium on Information Systems*, 2013, pp. 734– 743. https://doi.org/10.5753/SBSI.2013.5736

- [6] Gill, M. K., Asefa, T., Kaheil, Y., and McKee, M., Effect of Missing Data on Performance of Learning Algorithms for Hydrologic Predictions: Implications to an Imputation Technique. *Water Resources Research*, Vol. 43, No. 7, 2007. https://doi.org/10.1029/2006WR005298
- [7] Marlin, B. M., Missing Data Problems in Machine Learning. 2008. Retrieved 5 November 2023
- [8] Radišić, B., Seljan, S., and Dunđer, I., Impact of Missing Values on the Performance of Machine Learning Algorithms. 2023. Retrieved 5 November 2023
- [9] Cho, C., Park, J., Kim, K., and Sakhakarmi, S., Machine Learning for Assessing Real-Time Safety Conditions of Scaffolds. 2018. https://doi.org/10.22260/isarc2018/0008
- [10] Sakhakarmi, S., Park, J., and Cho, C., Enhanced Machine Learning Classification Accuracy for Scaffolding Safety Using Increased Features. *Journal of Construction Engineering and Management*, Vol. 145, No. 2, 2019, p. 04018133. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001601/ASSET/534952B9-0A71-450E-982D-8B7E10F06AD2/ASSETS/IMAGES/LARGE/FIG

URE8.JPG[11] Enders, C. K., and London, N. Y., Applied Missing Data Analysis the Guilford Press. 2010.

- [12] Kambach, S., Bruelheide, H., Gerstner, K., Gurevitch, J., Beckmann, M., and Seppelt, R., Consequences of Multiple Imputation of Missing Standard Deviations and Sample Sizes in Meta-Analysis. *Ecology and Evolution*, Vol. 10, No. 20, 2020, pp. 11699–11712. https://doi.org/10.1002/ECE3.6806
- [13] Bo, N., Little, R. J. A., and Rubin, D. B., Statistical Analysis with Missing Data. *Population (French Edition)*, Vol. 43, No. 6, 1988, p. 1174. https://doi.org/10.2307/1533221
- [14] Petrazzini, B. O., Naya, H., Lopez-Bello, F., Vazquez, G., and Spangenberg, L., Evaluation of Different Approaches for Missing Data Imputation on Features Associated to Genomic Data. *BioData Mining*, Vol. 14, No. 1, 2021. https://doi.org/10.1186/s13040-021-00274-7
- [15] Nijman, S. W. J., Leeuwenberg, A. M., Beekers, I., Verkouter, I., Jacobs, J. J. L., Bots, M. L., Asselbergs, F. W., Moons, K. G. M., and Debray, T. P. A., Missing Data Is Poorly Handled and Reported in Prediction Model Studies Using Machine Learning: A Literature Review. *Journal of Clinical*

Epidemiology, Vol. 142, 2022, pp. 218–229. https://doi.org/10.1016/J.JCLINEPI.2021.11.023

- [16] Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R., Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. *BMJ*, Vol. 338, No. 7713, 2009, pp. 157–160. https://doi.org/10.1136/BMJ.B2393
- [17] Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., and Siddique, Z., Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies*, Vol. 9, No. 3, 2021. https://doi.org/10.3390/TECHNOLOGIES903005 2
- [18] Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., and Tabona, O., A Survey on Missing Data in Machine Learning. *Journal of Big Data*, Vol. 8, No. 1, 2021. https://doi.org/10.1186/S40537-021-00516-9
- [19] Lyngdoh, G. A., Zaki, M., Krishnan, N. M. A., and Das, S., Prediction of Concrete Strengths Enabled by Missing Data Imputation and Interpretable Machine Learning. *Cement and Concrete Composites*, Vol. 128, 2022, p. 104414. https://doi.org/10.1016/J.CEMCONCOMP.2022.1 04414
- [20] Chen, N., Sun, Y., Wang, Z., and Peng, C., Improved LS-SVM Method for Flight Data Fitting of Civil Aircraft Flying at High Plateau. *Electronics* (*Switzerland*), Vol. 11, No. 10, 2022. https://doi.org/10.3390/electronics11101558
- [21] Ahmed, M., and Abdel-Aty, M., A Data Fusion Framework for Real-Time Risk Assessment on Freeways. *Transportation Research Part C: Emerging Technologies*, Vol. 26, 2013. https://doi.org/10.1016/j.trc.2012.09.002
- [22] Koc, K., and Gurgun, A. P., Scenario-Based Automated Data Preprocessing to Predict Severity of Construction Accidents. *Automation in Construction*, Vol. 140, 2022. https://doi.org/10.1016/j.autcon.2022.104351
- [23] Munian, Y., Martinez-Molina, A., Miserlis, D., Hernandez, H., and Alamaniotis, M., Intelligent System Utilizing HOG and CNN for Thermal Image-Based Detection of Wild Animals in Nocturnal Periods for Vehicle Safety. *Applied Artificial Intelligence*, Vol. 36, No. 1, 2022. https://doi.org/10.1080/08839514.2022.2031825
- [24] Sibarani, J. N., Sirait, D. R., and Ramadhanti, S. S., Intrusion Detection Systems Pada Bot-IoT Dataset Menggunakan Algoritma Machine Learning. *Journal Masyarakat Informatika*, Vol. 14, No. 1, 2023. https://doi.org/10.14710/jmasif.14.1.49721

- [25] Xue, L., Jiang, H., Zhao, Y., Wang, J., Wang, G., and Xiao, M., Fault Diagnosis of Wet Clutch Control System of Tractor Hydrostatic Power Split Continuously Variable Transmission. *Computers* and Electronics in Agriculture, Vol. 194, 2022. https://doi.org/10.1016/j.compag.2022.106778
- [26] Qi, C., Fourie, A., Ma, G., and Tang, X., A Hybrid Method for Improved Stability Prediction in Construction Projects: A Case Study of Stope Hangingwall Stability. *Applied Soft Computing Journal*, Vol. 71, 2018. https://doi.org/10.1016/j.asoc.2018.07.035
- [27] Liu, Z., and Li, S., A Sound Monitoring System for Prevention of Underground Pipeline Damage Caused by Construction. Automation in Construction, Vol. 113, 2020. https://doi.org/10.1016/j.autcon.2020.103125
- [28] Antwi-Afari, M. F., Li, H., Seo, J. O., and Wong, A., Y. L. Automated Detection and Classification of Construction Workers' Loss of Balance Events Using Wearable Insole Pressure Sensors. *Automation in Construction*, Vol. 96, 2018. https://doi.org/10.1016/j.autcon.2018.09.010
- [29] Hu, J., Huang, M. C., and Yu, X., Efficient Mapping of Crash Risk at Intersections with Connected Vehicle Data and Deep Learning Models. Accident Analysis and Prevention, Vol. 144, 2020. https://doi.org/10.1016/j.aap.2020.105665
- [30] Liu, P., Xie, M., Bian, J., Li, H., and Song, L., A Hybrid Pso–Svm Model Based on Safety Risk Prediction for the Design Process in Metro Station Construction. *International Journal of Environmental Research and Public Health*, Vol. 17, No. 5, 2020. https://doi.org/10.3390/ijerph17051714
- [31] Abbasianjahromi, H., and Aghakarimi, M., Safety Performance Prediction and Modification Strategies for Construction Projects via Machine Learning Techniques. *Engineering, Construction* and Architectural Management, Vol. 30, No. 3, 2023. https://doi.org/10.1108/ECAM-04-2021-0303
- [32] Zhu, R., Hu, X., Hou, J., and Li, X., Application of Machine Learning Techniques for Predicting the Consequences of Construction Accidents in China. *Process Safety and Environmental Protection*, Vol. 145, 2021. https://doi.org/10.1016/j.psep.2020.08.006
- [33] Alkaissy, M., Arashpour, M., Golafshani, E. M., Hosseini, M. R., Khanmohammadi, S., Bai, Y., and Feng, H., Enhancing Construction Safety: Machine Learning-Based Classification of Injury Types. *Safety Science*, Vol. 162, 2023. https://doi.org/10.1016/j.ssci.2023.106102
- [34] Koc, K., Ekmekcioğlu, Ö., and Gurgun, A. P.,

Integrating Feature Engineering, Genetic Algorithm and Tree-Based Machine Learning Methods to Predict the Post-Accident Disability Status of Construction Workers. *Automation in Construction*, Vol. 131, 2021. https://doi.org/10.1016/j.autcon.2021.103896

- [35] Geng, X., Wu, S., Zhang, Y., Sun, J., Cheng, H., Zhang, Z., and Pu, S., Developing Hybrid XGBoost Model Integrated with Entropy Weight and Bayesian Optimization for Predicting Tunnel Squeezing Intensity. *Natural Hazards*, Vol. 119, No. 1, 2023. https://doi.org/10.1007/s11069-023-06137-0
- [36] Halabi, Y., Xu, H., Long, D., Chen, Y., Yu, Z., Alhaek, F., and Alhaddad, W., Causal Factors and Risk Assessment of Fall Accidents in the U.S. Construction Industry: A Comprehensive Data Analysis (2000–2020). Safety Science, Vol. 146, 2022. https://doi.org/10.1016/j.ssci.2021.105537
- [37] Makki, A. A., and Mosly, I., Predicting the Safety Climate in Construction Sites of Saudi Arabia: A Bootstrapped Multiple Ordinal Logistic Regression Modeling Approach. *Applied Sciences* (*Switzerland*), Vol. 11, No. 4, 2021. https://doi.org/10.3390/app11041474
- [38] Baker, H., Hallowell, M. R., and Tixier, A. J. P., AI-Based Prediction of Independent Construction Safety Outcomes from Universal Attributes. *Automation in Construction*, Vol. 118, 2020. https://doi.org/10.1016/j.autcon.2020.103146

Digital Twin in Prefabricated Construction – Approaches, Challenges and Requirements

Veerakumar Rangasamy¹ and Jyh-Bin Yang¹

¹Graduate Institute Program of Construction Engineering and Management, Department of Civil Engineering, National Central University, Taiwan <u>sakthirgmv3@gmail.com, jyhbin@ncu.edu.tw</u>

Abstract -

Digital Twin (DT) is a virtual replica that mirrors physical objects, systems, or entities. In the construction sector, DTs play a crucial role in building management, optimizing energy usage, predicting maintenance, and enhancing building lifecycle management. Although effectively integrated into traditional construction practices, its application in the prefabricated construction (PC) field remains relatively limited. Despite its potential to transform the PC, a significant research gap exists in comprehensive literature addressing the approaches, challenges, and requirements specific to this field.

This study aims to bridge this gap through a mixed-method research approach that includes bibliometric analysis and systematic review. Publications related to PC and DT are collected from the Scopus database, followed by bibliometric analysis to offer a quantitative insight into the current research trends, prolific authors, and geographical distribution. Furthermore, this study conducts a systematic review that qualitatively analyzes the objectives of DT applications in PC, explores its methodological approaches, identifies challenges and recommendations for practitioners and researchers, and suggests future needs. The findings highlight the pressing need for standardization, robust data management, and scalable solutions to navigate the complexities. Conclusively, this paper presents a comprehensive conceptual framework to facilitate the application of DT in various phases of PC. The critical limitation is the reliance on the publications on the Scopus database and its potential biases. Excluding the need to expand on more emergent publications, discussing empirical DT studies in PC to highlight interoperability and standardization needs is also required for future research.

Keywords -

Prefabricated Construction; Digital Twins; Cyber-Physical System; Building Information Modeling; Internet of Things

1 Introduction

Traditional construction (TC) methods have faced criticism for low productivity, delayed project completion, unskilled workforce, subpar performance, and inefficient resource management. In response, prefabricated construction (PC) has emerged as a contemporary construction method, gaining popularity for its efficiency and benefits [1]. Though PC's origins trace back to the 1960s, its recent surge is primarily attributed to advancements in computing and technology like Building Information Modeling (BIM), Artificial Intelligence (AI), and the Internet of Things (IoT) [2]. BIM, in particular, is widely adopted in PC to address issues such as the need for early design decisions, logistical and supply chain complexities, and inadequate collaboration among stakeholders. As a digital representation, BIM encompasses a facility's physical and functional characteristics, serving as a shared and comprehensive knowledge base for information [3]. This information provides a dependable foundation for decision-making throughout the building's life cycle, from the initial concept stage to eventual demolition [4].

Using the cyber-physical system (CPS), Digital Twin (DT) is gaining significant traction across various industries, involving real-time monitoring, performance optimization, predictive analysis, simulations, and testing to aid critical decision-making processes. DT is emerging as a pivotal technology in the Industry 4.0 era, finding integration across diverse sectors. It essentially creates a digital replica or model of a physical entity (known as the physical twin), with both entities being interlinked through real-time data exchange [5]. This concept allows the Digital Twin to accurately reflect the state of its physical counterpart in real-time, and vice versa. DT's applications are varied, encompassing realtime monitoring, design and planning, optimization, maintenance, and remote operation [6]. Initially developed in the aerospace sector, DT is now poised to transform other industries, including manufacturing, construction, and healthcare [5]. By bridging the physical

and virtual worlds in real-time, DT offers an accurate approach to modeling complex, unpredictable scenarios. This ability, especially when combined with BIM, significantly improves and addresses PC challenges. Given BIM's interaction with CPS, DT is seen as a mature evolution of BIM [7].

Although DT was initiated in product lifecycle management by Michael Grieves in 2002 [5], its significant existence in construction began in the 2010s and the early 2020s in PC, as the technology matured in managing complex construction projects became more evident. However, those are related to the application rather than the overview of DT in PC, and a critical gap exists. To address this gap, this study aims to bring out the following objectives: (1) scientometrically analyze the status of DT application in PC; (2) understand the level of DT integration and approaches in PC; and (3) identify the challenges and their requirements. To achieve these objectives, articles searched from the Scopus database are analyzed bibliometrically and systematically.

2 Digital Twin in PC

2.1 DT in Construction

With the rise of IoT, which is the application of unique identifiers to physical objects that enable them to be connected to a network, allowing the transfer of data to and from those objects, the application of DT in the construction field is increasing. The evolution of various AI and machine learning (ML) technologies has made it even more possible in recent years [8]. From the articles published in the Scopus database, Figure 1 depicts a sharp rise in DT across industries from 2009 to 2023. A modest start with a few yearly publications until 2016 precedes a significant surge from 2017 onward, indicating increasing adoption of DT technology. The numbers more than double annually, from 38 in 2017 to a projected 2866 in 2023, highlighting the rapidly growing integration of DTs in the era of Industry 4.0.

In the construction sector, the application of DT has been increasing in the past few years. This can be depicted in Figure 2, which presents the distribution of articles between the engineering and construction sectors from 2014 to 2023. While engineering publications have grown substantially, reaching 1944 in 2023, DT in PC has also seen a notable increase, constituting approximately 15% of the engineering total. This percentage underlines the rising significance of DT applications in construction.

Initially, DT usage in the construction sector predominantly focused on post-construction operations and maintenance [9]. However, advancements in AI and IoT technologies have expanded DT applications to include logistics, supply chain management, design simulation, and energy optimization through sensors and machine learning. In PC, the synergy of DT with BIM is increasingly recognized for maximizing efficiency and promoting sustainability [10]. Some predominant applications include streamlining design, manufacturing, and assembly processes and improving lifecycle management through real-time data analysis and simulation. While many articles have addressed DT applications across PC phases, a thorough overview of DT applications is limited. For instance, Xie and Pan (2017) explored practices and challenges, proposing a basic framework for DT's integration with PC [9]. Yevu et al. (2023) reviewed DT in PC, focusing only on supply chain and carbon emissions [11], while Zhuo et al. (2023) discussed all digital technologies in PC [12]. This indicates the need for a more holistic review of DT applications in PC.



Figure 1. Application of DT over the years





2.2 Fundamental Components of DT in PC

The fundamental components of a DT in PC include:

- **Physical assets:** The PC components, such as panels, modules, or entire building sections, are the core objects that the DT will replicate.
- **Data collection system:** IoT sensors and devices embedded in the components gather real-time data on performance, conditions, and usage [13].
- **Connectivity system:** A network infrastructure enabling continuous data transmission from the physical to the digital, often utilizing cloud computing.

- Digital replica: A virtual model, usually created using BIM software, that reflects the physical characteristics and behavior of the PC components.
- Data processing and analytics: Software systems and algorithms that analyze the data collected incorporating AI and ML for advanced insights [11].
- Integration layer: Middleware that facilitates the integration of the DT with other systems, such as BIM, ensuring cohesive data flow and functionality.
- User interface: A dashboard to visualize data, interact with the DT, and make informed decisions based on the insights provided [9].
- **Simulation and modeling tools:** Applications to simulate various scenarios on data collected, predict outcomes, and plan for future actions in the virtual world.
- Feedback loops: Mechanisms that allow for the information and insights gathered from the DT to be fed back into the design, manufacturing, or construction process to improve future PC projects.

The integration of DT differs significantly between TC and PC. As mentioned earlier, DT in TC primarily focuses on managing on-site activities and adapting to changing project conditions, thus enhancing project management and resource allocation [14]. Conversely, in PC, DT is applied to optimize factory-based manufacturing, logistics, and assembly processes, benefiting from a more controlled and systematic environment, thus leading to improved precision and efficiency. Building on this foundation, Sections 3 and 4 delve into DT's detailed application in PC, highlighting empirical studies and integration strategies for overcoming identified challenges.

3 Research Approach

This study adopted a mixed review approach [4] that integrates a bibliometric analysis and a systematic review to mitigate biased conclusions. Bibliometric analysis is a quantitative research method to explore the patterns, trends, and impact within a body of literature. It involves statistical analysis of articles' metadata to understand the development and dissemination of knowledge in a specific field. On the other hand, a systematic review is a methodical and comprehensive literature review focused on a particular research question. It aims to identify, evaluate, and synthesize all available research evidence relevant to that question. Figure 3 illustrates the threestep methodology sequentially adopted in this research. As a first step, the Scopus database, known for its extensive coverage and stringent peer review process [4, 9], was selected as the primary source for literature retrieval. This search was confined to English-language articles, specifically targeting journals and conference proceedings in the Engineering subject field. The

research focused on a set of keywords intertwining DT and PC, encompassing terms such as "digital twins," "virtual replica," "predictive twin," and various related phrases in prefabrication such as "prefabricated construction," "prefabricated structures," "modular construction," "precast construction," "offsite "modular construction." integrated construction." "ppvc," "industrialized construction," using the Boolean operator "AND." The initial search yielded 79 articles, providing extensive bibliometric data. However, not all are aligned with the study's focus. After thorough abstract and full-text reviews, 47 articles are included in this study.

Subsequently, VOSviewer, which facilitated the construction and visualization of a comprehensive bibliometric network, was employed for a bibliometric exploration, concentrating on the co-occurrence of keywords and the geographic distribution of publications. Finally, the publications were thoroughly examined to the objective-oriented analyze application, methodological approaches, challenges, and requirements, and a conceptual framework was devised to facilitate the application of DT in PC.



Figure 3. Research steps and approaches

4 Analysis and Discussion

This section delves into the bibliometric content analysis and discussions based on the final dataset derived from step 1, as outlined in Figure 3.

4.1 Bibliometric Analysis

The bibliometric analysis concentrates on three aspects: discerning main research themes via keyword co-occurrence, mapping global research through country co-authorship, and identifying leading contributors through author co-authorship networks. This approach provides insights into the development, geographic distribution, and critical DT and PC research contributors.

4.1.1 Research Themes

The "co-occurrences of keywords" analysis is vital to identifying dominant themes and topics in DT and PC

literature [4]. Figure 4 illustrates a network map that visualizes the interconnectedness of keywords based on their co-occurrence in the literature. Central to the network is "digital twin," with 25 occurrences and 11 links, demonstrating a total link strength of 22. This highlights its pivotal role and frequent association with other key topics in the field. Adjacent to 'digital twin,' the term 'modular construction' features prominently, with 13 occurrences and a total link strength of 12, indicative of its significant overlap with DT concepts. The "construction industry," with ten occurrences and a total link strength of 9, bridges multiple topics, suggesting its encompassing impact across various research themes.

Further analysis in Table 1 identifies "architectural design" and "prefabricated construction" as influential nodes within the network, having 8 and 7 occurrences, respectively. This reflects their integral relationship with DT in streamlining design/construction methodologies. BIM emerges as a crucial facilitator within the DT framework, evidenced by its six occurrences and a link strength of 6, pointing to its role in enhancing the digital representation of construction. Other notable keywords like "offsite construction," "construction projects," "decision making," "modular integrated construction," and "robotics" are featured, each with five occurrences, signifying their relevance. The link strengths and occurrences of these terms, as reported in Table 1, underscore the multi-faceted nature of DT and PC, revealing a rich tapestry of interconnected research areas.





4.1.2 Contributor's Geographical Distribution

Analyzing the geographical distribution of research offers insights into DT and PC's leading research countries and the degree of international collaboration. Figure 5 visualizes the international collaboration network with a minimum of 3 articles and 20 citations, and their network details are presented in Table 2. China emerging as the most prolific contributor, with 18 publications and a link strength of 3, indicating focused research efforts within the country as of 2022. Hong Kong follows, with 9 publications and higher connectivity, reflected by 4 links and a total strength of 6, suggesting its central role in research collaborations.

Kauword	Occurrences	Links	Total link
Keywolu	Occurrences	LIIIKS	strength
digital twin	25	11	22
modular construction	13	10	12
construction industry	10	11	9
architectural design	8	9	8
prefabricated construction	7	7	7
BIM	6	6	6
modulars	6	7	6
offsite construction	6	8	5
construction projects	5	8	4
decision making	5	4	4
modular integrated construction	5	7	5
robotics	5	8	5
unite fizies our unite figeom	işçîla	hon	dhina

Table 1. List of keywords in the network

• •	Figure 5	Network	of co-aut	horship	of cou	ntries
-----	----------	---------	-----------	---------	--------	--------

T 11 A			
Table 2.	lon	contributing	countries
	- ~ P	• • • • • • • • • • • • • • • • • • •	

Country	Publication	Links	Total link strength	Avg. vear
			strongtin	year
China	18	1	3	2022
Hong Kong	9	4	6	2022
United States	7	5	6	2022
United Kingdom	5	5	6	2023
Australia	4	3	3	2022
Canada	3	2	2	2023
Italy	3	2	2	2021

The US is highlighted as a key player with 7 publications and 5 links, demonstrating its collaborative ties and a balanced link strength of 6, matching its research output in 2022. With 5 publications and 5 links, the UK continues its active research engagement into 2023, indicating an ongoing commitment to advancing the field. Although smaller, with 4 publications, Australia's contribution indicates a strong within-country focus, as shown by a link strength of 3 in 2022. Canada and Italy, each with 3 publications and a link strength of 2, demonstrate their involvement, with Canada's activity projecting into 2023 and Italy's contributions noted

earlier in 2021. These underscores DT and PC research's dynamic and interconnected nature, with varying degrees of international collaboration and research focus.

4.1.3 Top Contributors

Identifying principal authors and contributors highlights influential researchers and groups in DT and PC. This analysis underscores leading voices, offering insights into the field's key drivers. It aids in pinpointing potential collaborators and grasping the scholarly communication network. Out of 163 authors from the data collected, Figure 6 and Table 3 depict the coauthorship network among scholars in this domain with a minimum of 2 articles and ten citations. The network map reveals a robust collaborative cluster centered around Huang, George Q., Jiang, Yishuo, and Zhong, Ray Y., each with four documents to their credit, sharing six links and a total link strength of 17 in the year 2022. This trio forms a core collaborative group, underlining their significant contribution to the field.



Figure 6 Network of co-authorship of authors

Table 3. Top contributing authors

Country	Documents	Links	Total link strength	Avg. Year
Huang, George Q	4	6	17	2022
Jiang, Yishuo	4	6	17	2022
Zhong, Ray Y.	4	6	17	2022
Dong, Miaosi	3	4	11	2021
Li, Ming	3	6	13	2022
Liu, Xinlai	3	6	14	2022
Liu, Zhansheng	3	1	2	2022
Wang, Zhichen	3	4	11	2021
Yang, Bin	3	4	11	2021
Zhang, Binghan	3	4	11	2021

Adjacent nodes represent authors such as Dong, Miaosi, Wang, Zhichen, and Zhang, Binghan, each with three documents and four links, indicating active participation and a link strength of 11 in 2021. Li, Ming, and Liu, Xinlai, also with three documents but possessing stronger collaborative ties, evidenced by higher link strengths of 13 and 14, respectively, in 2022, are key figures within the network. Liu, Zhansheng, with 3 documents but a solitary link and a minimal link strength of 2 in 2022, appears as an emerging contributor within the network. The visual data from Figure 6, combined with the quantitative metrics from Table 3, highlight the influential authors and the intensity of their collaborations, offering insight into the community driving DT and PC research forward.

4.2 Content Analysis

This section delves in-depth into the objectiveoriented application of DT and PC, exploring their methodological connections, proposing a conceptual framework, and dissecting the challenges and requirements to leverage DT's full potential.

4.2.1 Objective-Oriented Application

The integration of DT in PC represents a pivotal shift towards leveraging advanced digital technologies to address various challenges and objectives, such as enhancing efficiency, innovation, and sustainability. Figure 7 illustrates the spectrum of objectives in the integration of DT in PC. The paramount objective is efficiency optimization, highlighted in 17 articles, with significant contributions such as ChainPM, a blockchain paradigm enhancing 3.0 construction project management by Zhao et al. [15], resulting in a 99.8% reduction in information synchronization latency. Jiang et al. introduced a DT-enabled smart PC system for optimizing on-site assembly processes [16]. Other notable integrations include supply chain coordination [17] and immersed tunnel works [18]. These examples show a coordinated industry effort to use DT to streamline PC processes and reduce downtime, highlighting a broader trend toward increased efficiency.



Figure 7. Distribution of objectives of DT in PC

Innovation and R&D, encapsulated in 8 articles, are represented by advances such as exploring multi-scale and multi-scenario digital twins by Jia et al. [19], underscoring the industry's endeavor to tackle complex digitalization challenges. Lifecycle Management and

Risk Mitigation are also well-represented; for instance, Nguyen et al. developed a digital engineering model combining the DfMA with object-oriented parametric modeling techniques [20]. Osadcha et al. systematically review DT geometry updating, emphasizing the need for accurate, real-time data to manage and mitigate risks throughout a building's lifespan [21]. Quality Enhancement, highlighted in works like Shi et al.'s method for geometry checking of precast culvert segments [22], 3D scanning-based geometric DT, and assembly [23], showcases the precision that DT can bring to construction quality control.

Although Safety and Compliance, and Collaboration and Communication appear less frequently, they are no less critical. For instance, Yang et al. [24] explored computer vision for fall hazard detection on construction sites. Sun et al.'s virtual reality training system based on IoT technology demonstrates DT's potential in improving safety and fostering collaboration [25]. Meanwhile, the objectives of sustainability improvement, illustrated by works such as DT-enabled prefabrication supply chain for carbon emissions evaluation by Yevu et al. [11], align with global environmental sustainability goals, showing the promise of DT in driving green construction practices. This reflects a multi-dimensional approach where DT serves as a cornerstone for innovation, addressing efficiency, safety, quality, and sustainability. As such, Figure 7 illustrates the current state of DT application in PC and points to potential areas for future research and development within the industry.

4.2.2 Methodological Connections

Integrating various technologies with DT in PC exemplifies the industry's movement toward digitalization. Exploring their methodological connection is essential to understanding how different technologies interplay with DT, enhancing PC efficiency, informed decision-making, and fostering innovation for a competitive, sustainable, and safe construction industry. Figure 8 represents the frequency of other technologies integrated with DT and PC. Among the technologies, BIM is the most integrated with DT in PC [15, 17, 20, 26], showcased by its highest occurrence. The pivotal role of BIM reflects its comprehensive digital representation capabilities essential for the life cycle management of a building [9]. Following BIM, IoT, and sensors are instrumental in real-time data acquisition, essential for the dynamic updating of DT to reflect the current state of PC elements [26-29].

AI and ML stand out as the third most integrated technology, often with BIM and IoT, highlighting their role in processing and analyzing data for predictive insights and decision-making [20, 26-29]. Conversely, augmented reality/virtual reality (AR/VR) [25], blockchain [30], cloud computing [31], finite element

method (FEM) $[\underline{32}]$, robotics $[\underline{26}]$, and geographic information systems (GIS) [17] show a nascent integration. These technologies, despite their lower integration frequency, bring distinct advantages such as enhanced visualization (AR/VR),secure data transactions (blockchain), scalable data storage (cloud computing), precise structural analysis (FEM). automation (robotics), and spatial data analysis (GIS).

In conclusion, while BIM and IoT form the backbone of DT integration in PC, the full spectrum of digital technologies contributes to a more intelligent, efficient, and interconnected construction ecosystem. Each technology, irrespective of its current integration frequency, has the potential to revolutionize aspects of DT in PC, indicating a trend towards an increasingly digital future in the construction sector.



Figure 8. Frequency of other technologies in DT-PC integration

4.2.3 Conceptual Application Framework

Integrating DT into PC is a multi-stage process involving careful planning, continuous data collection and analysis, and advanced technologies like AI, IoT, and BIM. Figure 9 encapsulates a comprehensive framework for integrating DT in PC, encompassing the entire lifecycle of a structure. From initial BIM-based design to IoT-enhanced manufacturing and logistics, the framework ensures real-time data flow and optimized construction processes. The subsequent stages of on-site assembly, quality control, and operation benefit from continuous DT interaction, facilitating predictive maintenance and efficient building management. Finally, the DT supports sustainable decommissioning and material recycling, aligning with circular economy principles. This integrated approach signifies a transformative step towards intelligent, sustainable construction practices.



Figure 9. Stage-wise application framework

4.2.4 Challenges and Requirements

The meticulous management of extensive data is a formidable challenge in DT-PC integration, necessitating unwavering accuracy, uniformity, and instantaneous updates [8]. Thus, ensuring data integrity becomes paramount [9]. Interoperability poses its trials, with disparate software and platforms often at odds, hindering seamless integration [33]. Scalability demands attention, calling for solutions adapting to varying project magnitudes within PC operations. The technological intricacies of DT require specialized acumen, forming a barrier to entry for broader adaptation [33]. The initial financial outlay for DT implementation is substantial, which may deter smaller enterprises from embracing the Furthermore, the absence of technology [33]. standardization within the construction domain results in poor integration procedures and quality outcomes [11].

This research highlights certain gaps in the existing methodologies and practices concerning the assimilation of DT within the PC sector. Therefore, to foster a cohesive integration of DT in PC, a concerted effort towards standardization of data protocols is imperative to guarantee inter-system compatibility. Α robust infrastructure for data management must be established, capable of safeguarding data integrity and facilitating real-time processing [9, 12]. Using sophisticated analytics and AI is essential to distill actionable insights from the collected data [12], enhancing predictive maintenance and decision-making. Moreover, highspeed and reliable connectivity is a prerequisite for effective communication across the DT ecosystem [9]. User interfaces must be intuitively designed to enable stakeholder engagement across varying levels of technical expertise [9]. Hence, educational initiatives are critical to equip the workforce with the necessary skills navigate these advanced technologies. The to convergence of DT with BIM is paramount, enriching the visualization and analytical facets of construction projects. DT solutions must be inherently scalable to accommodate a spectrum of project requirements. Additionally, compliance with regulatory standards and privacy legislation is essential. Consequently, investment in R&D is vital to drive innovation and refine DT applications tailored to the PC industry.

5 Conclusions and Limitations

This study presents the emergent route of DT integration within the PC sector, unraveling the methodological synergies and potential for industry-wide transformation. The research delineates the escalating adoption of DT, propelled by advancements in BIM, AI, and IoT, signifying a paradigm shift from traditional construction methods towards a digitized, efficient, and sustainable approach. Notably, the bibliometric analysis explains the focal research themes, geographical research

proliferation, and the influential scholarly contributions shaping the DT-PC narrative. The content analysis accentuates its multifaceted objectives, from efficiency optimization to sustainability enhancement.

The article offers a conceptual framework outlining the integration of DT across PC stages, advocating for a seamless digital continuum from design to decommissioning. The findings highlight the pressing need for standardization, robust data management, and scalable solutions to navigate the complexities inherent in DT-PC integration. Notably, the study acknowledges the limitation of relying solely on the Scopus database. Future research can expand more emergent publications and should focus on empirical validation of DT models in PC, exploring advanced technologies for enhanced DT functionality, and developing interoperability standards. Investigating scalable solutions, data security, and environmental impacts, alongside economic analyses, will address critical gaps and advance the field significantly, paving the way for practical, scalable, and sustainable DT applications in PC.

Acknowledgement

The authors express their gratitude to the Ministry of Education and the National Science and Technology Council of Taiwan, ROC, for funding this research under Contract No. NSTC-111-2221-E-008-025-MY3.

References

- Iacovidou E. Purnell P. Tsavdaridis K D and Poologanathan K. Digitally Enabled Modular Construction for Promoting Modular Components Reuse: A UK View. *Journal of Building Engineering*, 42: 102820, 2021.
- [2] Zhong R Y. Peng Y. Xue F. Fang J. Zou W. Luo H. Thomas Ng S. Lu W. Shen G Q P and Huang G Q. Prefabricated Construction Enabled by the Internetof-Things. *Automation in Construction*, 76: 59-70, 2017.
- [3] What is a BIM? On-line: https://www.nationalbimstandard.org/faqs, Accessed: 11/12/2023.
- [4] Yin X. Liu H. Chen Y and Al-Hussein M. Building Information Modelling for Off-site Construction: Review and Future Directions. *Automation in Construction*, 101: 72-91, 2019.
- [5] Singh M. Fuenmayor E. Hinchy E P. Qiao Y. Murray N and Devine D. Digital Twin: Origin to Future. *Applied System Innovation*, 4(2): 36, 2021.
- [6] Kritzinger W. Karner M. Traar G. Henjes J and Sihn W. Digital Twin in Manufacturing: A Categorical Literature Review and Classification. *IFAC-PapersOnLine*, 51(11): 1016-1022, 2018.
- [7] Andersson T. Biörck J. Sjödin E and J M. The Next Normal in Construction Material Distribution. On-

line: <u>https://www.mckinsey.com/</u>, Accessed: 10/12/2023.

- [8] Jacoby M and Usländer T. Digital Twin and Internet of Things—Current Standards Landscape. *Applied Sciences*, 10(18): 6519, 2020.
- [9] Xie M and Pan W. Opportunities and Challenges of Digital Twin Applications in Modular Integrated Construction. In Proceedings of the Proceedings of the International Symposium on Automation and Robotics in Construction, pages 278-284, 2020.
- [10] Ghansah F A and Lu W. Cyber-Physical Systems and Digital Twins for "Cognitive Building" in the Construction Industry. *Construction Innovation*, Early Access, 2023.
- [11] Yevu S K. Owusu E K. Chan A P C. Sepasgozar S M E and Kamat V R. Digital Twin-enabled Prefabrication Supply Chain for Smart Construction and Carbon Emissions Evaluation in Building Projects. *Journal of Building Engineering*, 78, 2023.
- [12] Cheng Z. Tang S. Liu H and Lei Z. Digital Technologies in Offsite and Prefabricated Construction: Theories and Applications. *Buildings*, 13(1), 2023.
- [13] Boje C. Guerriero A. Kubicki S and Rezgui Y. Towards a Semantic Construction Digital Twin: Directions for Future Research. *Automation in Construction*, 114: 103179, 2020.
- [14] Münker S. Wildemann P R. Göppert A. Brell-Cokcan S and Schmitt R H. Online Capabilitybased Resource Allocation for On-site Construction Operations Utilizing Digital Twin Models. *Construction Robotics*, 5(3): 211-226, 2021.
- [15] Zhao R. Chen Z and Xue F. A Blockchain 3.0 Paradigm for Digital Twins in Construction Project Management. *Automation in Construction*, 145, 2023.
- [16] Jiang Y. Li M. Guo D. Wu W. Zhong R Y and Huang G Q. Digital Twin-enabled Smart Modular Integrated Construction System for On-site Assembly. *Computers in Industry*, 136, 2022.
- [17] Lee D and Lee S. Digital Twin for Supply Chain Coordination in Modular Construction. *Applied Sciences*, 11(13), 2021.
- [18] Wen C and Ying Z. Research and Application of Digital Twinning for the Immersed Tunnel Prefabrication. In *Proceedings of the Chinese Control Conference, CCC*, pages 6959-6963, 2023.
- [19] Jia W. Wang W and Zhang Z. From Simple Digital Twin to Complex Digital Twin Part I: A Novel Modeling Method for Multi-scale and Multiscenario Digital twin. Advanced Engineering Informatics, 53, 2022.
- [20] Nguyen D C. Park S J and Shim C S. Digital Engineering Models for Prefabricated Bridge Piers. Smart Structures and Systems, 30(1): 35-47, 2022.

- [21] Osadcha I. Jurelionis A and Fokaides P. Geometric Parameter Updating in Digital Twin of Built Assets: A Systematic Literature Review. *Journal of Building Engineering*, 73, 2023.
- [22] Shi X. Xu Z and Zhang X. A Geometry Checking Method of Precast Culvert Segments Based on 3D Photography Measurement. In *Proceedings of the FIB Symposium*, pages 1724-1731, 2020.
- [23] Rausch C. Lu R. Talebi S and Haas C. Deploying 3D Scanning Based Geometric Digital Twins During Fabrication and Assembly in Offsite Manufacturing. *International Journal of Construction Management*, 23(3): 565-578, 2023.
- [24] Yang B. Zhang B. Zhang Q. Wang Z. Dong M and Fang T. Automatic Detection of Falling Hazard from Surveillance Videos based on Computer Vision and Building Information Modeling. *Structure and Infrastructure Engineering*, 2022.
- [25] Sun K. Li S. Zhang P and Deng X. Virtual Reality Interactive Training System of Prefabricated Building Based on Internet of Things Technology. In *Proceedings of the AICIT*, 2023.
- [26] He R. Li M. Gan V J L and Ma J. BIM-enabled Computerized Design and Digital Fabrication of Industrialized Buildings: A Case Study. *Journal of Cleaner Production*, 278, 2021.
- [27] Zhao Y. Cao C and Liu Z. A Framework for Prefabricated Component Hoisting Management Systems Based on Digital Twin Technology. *Buildings*, 12(3), 2022.
- [28] Liu Z. Li A. Sun Z. Shi G and Meng X. Digital Twin - Based Risk Control during Prefabricated Building Hoisting Operations. *Sensors*, 22(7), 2022.
- [29] Liu Z. Meng X. Xing Z and Jiang A. Digital Twinbased Safety Risk Coupling of Prefabricated Building Hoisting. *Sensors*, 21(11), 2021.
- [30] Jiang Y. Liu X. Kang K. Wang Z. Zhong R Y and Huang G Q. Blockchain-enabled Cyber-Physical Smart Modular Integrated Construction. *Computers in Industry*, 133, 2021.
- [31] Jiang Y. Li M. Li M. Liu X. Zhong R Y. Pan W and Huang G Q. Digital Twin-enabled Real-time Synchronization for Planning, Scheduling, and Execution in Precast On-site Assembly. *Automation* in Construction, 141, 2022.
- [32] Thomas L. Kernicky T. Whelan M. Park Y and Cox R. Application of a Structural Digital Twin on a Laboratory Model for Performance Monitoring of Aging and Degradation. In Proceedings of the Conference Proceedings of the Society for Experimental Mechanics Series, pages 5-13, 2024.
- [33] Lei B Y. Janssen P. Stoter J and Biljecki F. Challenges of Urban Digital Twins: A Systematic Review and A Delphi Expert Survey. *Automation in Construction*, 147: 15, 2023.

Exploring Digital Twin platforms across industries: A multiaspect SNA

Amin Khoshkenar¹ and Hala Nassereddine, Ph.D.²

¹Department of Civil Engineering, University of Kentucky, USA ²Department of Civil Engineering, University of Kentucky, USA <u>amin.khoshkenar@uky.edu</u>, <u>hala.nassereddine@uky.edu</u>

Abstract

Digital Twins have emerged as a transformative solution enabling organizations across sectors to digitally replicate physical assets and processes to extract operational insights. Implementing Digital Twin systems involves diverse stakeholders, ranging from providers to end-user developers and adopters. At the crux of Digital Twin implementation lies the need of Digital Twin platform - the foundational infrastructure on which solutions are built, integrations are executed, and data flows are While substantial research managed. targets advancing Digital Twin platforms' capabilities, investigations analyzing real-world implementations spanning industries remain scarce. This research profiles 19 platforms harnessing data aggregated from provider websites, white papers, press releases and user documentation to compile understanding on platform purpose, inbuilt security and interaction mechanisms, integration architectures, predominant use cases, real users' locations, and supported solutions. Social Network Analysis (SNA) conducted in Pajek detected valuable adoption patterns in the Digital Twin platforms market while community identification analysis linked predominant platformcapability combinations to industry and locational preferences, arming stakeholders to strategize road mapping. Results showed that Azure Cloud, IBM Cloud, and MindSphere were ranked highest in centrality among the platforms mapped. In parallel, to determine platform capability dimensions and their acceptance across geographies and use contexts, normalized centrality metrics were performed for other data types. Also, 58 solutions provided by platforms were classified into five categorical purpose groupings. The findings expand visibility into the dynamics of Digital Twin platforms and can be evolved by expanding sample diversity and blending functional, operational, and economic perspectives in future studies supporting stakeholders in implementation processes.

Keywords -

Digital Twin systems, Digital Twin Platforms, Social Network Analysis

1 Introduction

In the rapidly evolving landscape of digital transformation, organizations worldwide are increasingly turning to innovative solutions to enhance efficiency, streamline operations, and gain a competitive edge [1]. Among these, Digital Twins have emerged as a transformative force, revolutionizing the way industries approach data, simulation, and real-world processes. As the significance of digitalization continues to grow, the adoption of Digital Twins has become not only a strategic imperative but also a key driver in reshaping industries and fostering unparalleled advancements in technology [2]. Digital Twin provides the means to depict, emulate, forecast, optimize, and control physical space through connectivity, mapping, analysis, real-time and interaction with a specific fidelity and frequency [3]. This communication between the physical and virtual systems is enabled by explicitly defining the data produced by the system, augmenting it with information about system entities, and realizing "value adding services" on top of this data driven definition [4]. Core components of Digital Twin technology include high fidelity data collection via Internet of Things (IoT), sensors, predictive data analytics and process simulations to create the model, as well as visualization dashboards to provide services for users [5]. With the advantages of model, data, and service, Digital Twin offers superior solutions for enhancing quality, increasing efficiency, cutting costs, mitigating losses, ensuring safety, and conserving energy [6]. Therefore, Digital Twin enables maximum optimization of processes and value chains of the physical system. Moreover, the application of Digital Twin in diverse industries can effectively expedite their digitalization, networking, and intelligence development processes.

As the adoption of Digital Twin increases across sectors such as manufacturing, healthcare, automotive, construction, and urban development, dedicated software platforms have emerged to facilitate the Digital Twin development and scalability of Digital Twin [7]. A Digital Twin platform provides the core infrastructure to enable the creation, management, and utilization of Digital Twins across an organization's ecosystem while establishing the digital thread that connects the physical and virtual worlds [8]. According to [9], Digital twin platform ecosystem refers to collaborative environment involving Digital Twin platform providers, application developers, technology providers, system integrators, consultants, and user organizations for enabling digital transformation across industry verticals. The key capabilities of a Digital Twin platform include data ingestion and integration, simulation engine, visualization dashboards, analytics, reporting and notifications, collaboration tools, Application Programming Interfaces (APIs) and integration capabilities, security features, and scalability [10], [11]. Digital Twin platforms empower organizations to unlock key benefits, including predictive maintenance, improved asset performance, and accelerated innovation [12]. Prominent examples of Digital Twin platform include Microsoft Azure Digital 1Twins, GE's Predix, Siemens MindSphere, PTC's ThingWorx, Dassault Systemes' 3DEXPERIENCE, and Ansys Twin Builder. While these platforms are applicable across various industries, companies often opt for customized solutions based on Digital Twin platforms, avoiding a one-sizefits-all approach. Despite numerous investigations into developing Digital Twin platforms in recent years, there is a gap in gathering, categorizing, and analyzing the existing platforms provided by developer companies. An analysis of the implementation of existing Digital Twin platforms in different industries, along with their definitions, purposes, solutions, and other aspects is needed.

This paper aims to assist the evolving landscape of Digital Twin platforms and ecosystem partners dedicated to delivering asset-class or industry-specific Digital Twin solutions. This objective is accomplished by applying Social Network Analysis (SNA) techniques to model and interpret the relationships between various Digital Twin ecosystem participants on a global scale. Additionally, existing Digital Twin platforms are analyzed with respect to their functional and geographic concentration of interconnected providers and user communities.

2 Background

Digital Twins have gained traction in recent years across various industries, including manufacturing, aviation, healthcare, construction, and smart cities. Recent surveys on industry adoption levels indicate that manufacturing has taken an early lead in implementing Digital Twin prototypes and applications [13]. Although adoption levels vary across industries and applications, the common thread involves leveraging connected IoT devices, predictive analytics, and simulations through Digital Twin platforms. This is accomplished by utilizing core features such as physics-based equipment modeling, real-time data integration via IoT APIs, Artificial Intelligence (AI)-powered analytics, monitoring dashboards, and simulation tools [14]. Several studies have been conducted to develop Digital Twin platforms to address specific problems in different industries. In the manufacturing sector, [15] developed a Digital Twin for steel pipe weld quality control. A Digital Twin platform based on a microservices architecture and offering solutions for continuous deployment, data infrastructure and I4.0 business services was developed by [16]. [17] developed an AI-based injection molding machine Digital Twin able to prevent failures by recognizing machine deterioration patterns. [18] established a Digital Twin platform for the medical device assembly machine to diagnose the anomalies' root causes and predict the quality of the products with more confidence, higher speed, and less invasive methods.

While manufacturing leads in piloting Digital Twin platforms, momentum also exists in other sectors such as construction and smart cities. [19] developed a cyberphysical interconnection method for computational design and robotic construction in a wooden architectural realm. [20] employed Digital Twin within a human-robot collaborative system to assist in assembling complexshaped architectures and tested it through a real system. [21] created a Digital Twin-enabled anomaly detection system for asset monitoring and with a data integration method based on extended Industry Foundation Classes (IFC) in daily Operation and Maintenance (O&M) management, which was successfully tested on a real case. [22] proposed a Digital Twin-enabled real-time synchronization system (DT-SYNC) aiming to facilitate Planning, Scheduling, and Execution (PSE) using realtime resource status and construction progress information obtained from high-fidelity Digital Twins. [23] introduced a digital-twin based multi-information intelligent early warning and safety management platform to address high safety risks during tunnel construction. [24] developed a blockchain-enabled Digital Twin collaboration platform for Modular Integrated Construction (MiC) fit-out operations for modular construction. [25] proposed a geospatial platform based on the universal game engine Unity3D, to manage large-scale individual mobility data for an Urban Digital Twin (UDT) platform. [26] proposed a Digital Twin platform to address challenges in incorporating Photovoltaic (PV) systems and wind energy sources into smart city power grids. [27] presented the design, implementation, and use cases of the Chattanooga Digital Twin (CTwin) toward the vision for next-generation smart city applications for urban mobility management.. [28] presented an overview of different Digital Twin platforms that can be used in Electric Vehicle (EV) applications in smart cities.

Healthcare is also poised for Digital Twin adoption through "Digital Hospital" initiatives utilizing AI and simulation to optimize patient flow, resource allocation, and medical equipment maintenance [29]. [30] built a Digital Twin body through dynamic equations and pressure control mechanisms based on pressure reflexes. [31] built a Digital Twin coupled with blood flow and head vibration to develop diagnostic tools. Subramanian (2020) built a Digital Twin that integrated scientific information and clinical source information. [33] constructed a Digital Twin of lumbar spine based on AR, data analytics, motion capture system, Inverse Kinematic (IK) method and Finite Element Method (FEM). [34] developed a Digital Twin system for the vaccination process and tested it in a clinic. [35] built an emergency department Digital Twin simulation able to quantify the downstream impact of delayed or erroneous triage on patient outcomes. [36] developed a patient centric mathematical data model to formally define the semantic and scope of our proposed Healthcare Digital Twin (HDT) system based on Blockchain.On the other hand, there have been studies attempting to define the core infrastructure, tools, and capabilities of platforms to develop Digital Twins. As defined by[37], Digital Twin platforms aim to provide the technical foundation for virtual modeling, data orchestration, and digital thread management required in Digital Twin initiatives. Multiple conceptual reference models exist, detailing

potential components of Digital Twin platforms covering aspects such as physical counterparts, virtual models, connectivity, intelligence, and visualization, among others. [5], [38].

While logical representations are instructive, surveys of commercial platforms reveal differing priorities and configurations of key elements such as shadow maintenance, analytics digital engines, simulation services and front-end apps [11]. In addition, studies have identified several common Digital Twin platform capabilities. [39] discussed key services required in Digital Twin enabled smart manufacturing, including sensing, data analysis, modeling, simulation, and visualization. [5] reviewed core concepts and technologies behind Digital Twins, including system integration, simulation, machine learning, visualization and deep learning. [40] summarized developmental components of Digital Twins into four parts: Digital Twin modeling and simulation, data fusion, interaction, and service. Required features for Digital Twins such as interconnection, simulation models, data collection, and visualization were discussed by [41]. [38] and [42] reviewed Digital Twin capabilities which are defined and summarized in Figure1. While substantial research activity exists around conceptual Digital Twin platforms, architectures, and capability frameworks, a gap persists in empirical documentation and evaluation of specific platforms developed and deployed across industries. For all the prototyping and technical specification efforts, a systematic investigation into platforms supporting operational Digital Twin initiatives remains lacking. This limitation not only hampers the benchmarking of the



Figure 1 . Digital Twin platform capabilities

expansion of Digital Twin platforms adoption and deployment but also impedes the identification of obstacles and milestones crucial for guiding investment decisions. Asset owners, lacking evidence of platform viability and modernization pathways, face challenges in making informed trade-off decisions.

3 Data Overview

The Digital Twin platform dataset underpinning this research was compiled by drawing from secondary platform sources, press announcements, and providers' and users' websites. Structured information was captured for 18 unique Digital Twin platforms located in six different countries, covering details on the platform's launch year, core capabilities such as built-in security mechanisms and twin interaction methods, target use case, and roster of 58 solutions leveraging these platforms. Additionally, user community details were gathered related to leading industries and functional applications where available. Table 1 summarizes the findings and their definitions- the full data is available and will be provided upon request.

The dataset provides reasonable coverage across industry segments investing in Digital Twin, with over 40 industry segments identified including manufacturing, energy, healthcare, automotive, and construction. These major segments are located in 19 different countries. It is important to note that the depth of details on solutions varies considerably based on how much detail each provider or user provides for the solution they are using.

Data Attributes	Definition		
Platform provider	Owner name and headquarter		
i lationii providei	location		
Launch year	Year when platform first		
Launen year	released		
Platform purpose	Primary functions and digital		
i iationii puipose	twin focus areas		
Platform integration	Integration classification		
method	integration elassification		
Twin interaction	IoT connectivity and data		
methods	ingestion protocols		
Security	Data and processes security		
mechanisms	protocols and mechanisms		
Platform users	Industries that the platform		
r lationii users	can be deployed		
Current users'	Countries that are using each		
location	platform		
Digital Twin	Names and descriptions of		
platform's Solutions	developed solutions		
Solution focus	Solutions' purposes		

Table 1. Dataset summary

4 Analysis and Discussion

Social Network Analysis (SNA) was employed to analyze the dataset using Pajek software. In this analysis, Digital Twin platforms, users' industry, security mechanism, twin interaction method, and users' location are modeled as nodes with edges defined based on stated collaborations, integrations, and deployments. Both visual and quantitative SNA techniques are applied to identify key patterns.

SNA yielded several insightful observations regarding the current state and trajectory of the Digital Twin ecosystem. As depicted in Figure 2 and indicated by Degree of Centrality (DoC) rankings, a select few platforms have emerged as most influential in shaping today's deployments. Azure Cloud (DoC: 18), IBM Cloud (DoC: 18), MindSphere (DoC: 17), and Vuforia (DoC: 17) are identified as the top platforms, experiencing significant most adoption across various industry verticals. The dominance of these mainstream platforms aligns with the finding that cloud-based (DoC: 8) and hybrid (DoC: 6) integration methods now prevail in how Digital Twin solutions leverage platforms.

Additionally, the SNA results identified key security and twin interaction mechanisms that are central to Digital Twin platforms. Authorization (DoC: 18), authentication (DoC: 14), and encryption (DoC: 10) emerged as the most common security mechanisms. APIs (DoC: while service 15), bidirectional synchronization (DoC: 13), and interface simulation 13) lead in enabling integration and (DoC: communication with physical assets. The widespread use of these platform-enabled capabilities emphasizes their significance in ensuring the usability and security of Digital Twin platforms. However, the network topology also revealed gaps in the adoption of these leading practices, particularly among Asian user organizations. Providing more implementation guidance and sharing best practice could accelerate the assimilation of platform services in these regions.

Moreover, examining the industry and geographic distribution of Digital Twin platform adoption adds value by identifying demand patterns. Manufacturing (DoC: 10), automotive (DoC: 8), energy (DoC: 7), and construction (DoC: 5) industries emerged as leading segments actively leveraging Digital Twin platforms. Furthermore, and as depicted in figure 3, the community detection analysis provides particularly useful insights into industry and location combinations that tend to utilize specific platform and mechanism combinations. For instance, MindSphere platform with Authorization and secured protocols, paired with Hybrid integration method, see adoption in construction and agriculture use cases across USA and Germany. These patterns suggest that specific platform configurations meet the needs of these industries and locations.



Figure 3. Community identification analysis

Finally, the temporal social network analysis indicated that the development and implementation of Digital Twin platforms have entered a phase of rapid growth and acceleration over the past three years. The increasing rate of participation across both supply and demand sides of the market signals the crossing of a tipping point. Digital Twin providers are developing various solutions with specific purposes, moving beyond isolated proofs of concept into scaling adoption across industries. The analysis of the range of solutions associated with major Digital Twin platforms revealed that the purpose of these solutions largely falls into five key categories: 1) detailed virtual modeling, 2) simulation and predictive analysis, 3) connectivity and

data integration, 4) immersive interaction, and 5) lifecycle and change management. For example, solutions like CATIA, Creo, NX, OpenBuildings, and Unified Engineering provide sophisticated CAD and engineering environments to model highly accurate Digital Twin representations. Complementary simulation tools like SIMULIA, Discovery Live, Plant Simulation, and Twin Builder enable predicting real-world performance through digital prototypes. Solutions including IoT Hub, Watson IoT, and C3 IoT focus specifically on ingesting and managing streams of data from physical assets. Augmented and virtual reality solutions provide new ways of immersive interacting with Digital Twins on-site or remotely. Furthermore, solutions like Windchill, Teamcenter, and iModelhub help manage Digital Twin data and relationships throughout ongoing change.

While most platforms provide an integrated set of solutions spanning the five main areas, the analysis reveals some differentiation and specialization as well. For example, the ANSYS Digital Twin platform emphasizes simulation-centric solutions for digital mission engineering across complex systems. Autodesk brings generative design paired with construction site data flows. Azure offers cloud-native capabilities for scalable Digital Twin data and lifecycle management. Also, augmented reality leaders like Vuforia integrate immersive Digital Twin experiences as a key solution area. Observing these patterns provides perspective on the expanding functionality Digital Twin platforms now offer as enablers, spanning detailed modeling to operational connectivity to simulation-driven insights and beyond. Additionally, while Digital Twin platforms host an expanding roster of twin-enabled solutions, the depth of solution details wildly varies. Identifying use case patterns to determine which solutions best match the purpose will accelerate the large-scale adoption and implementation.

5 Conclusion

Numerous Digital Twin platforms, each designed for specific purposes, have been developed by providers globally. The comprehensive information gathered on these platforms and analyzed through SNA offers valuable insights into the current status, intentions, capabilities, and associated solutions of these platforms summarized as below:

- **Platforms:** SNA's analysis revealed that among these robust platforms, Azure Cloud, IBM Cloud, and MindSphere exhibit the highest centrality.
- **Platform users' industry:** When examining user industries across all platforms, Manufacturing emerged as the most prominent, followed by Automotive, Aerospace, and Marine, with Logistics

and Agriculture being the least common industries.

- **Platform users' location:** In terms of geographic distribution of users, the USA takes the lead, followed by European countries, while Asian countries having the least widespread adoption.
- Security mechanisms: According to the SNA, Authentication, Authorization and Encryption represented the most widely implemented security mechanisms, whereas Data Masking, Secured Protocol, and Trusted Hardware ranked as less common.
- Twin interaction methods: As a crucial capability of Digital Twin platforms, Service APIs were the predominant method of twin interaction, whereas Twin-to-Twin approaches ranked as the least used.
- Integration methods: Regarding integration methods, Cloud-Based platforms were deemed the most probable, with combinations of On-Premise with Hybrid, as well as On-Premise with Cloud-Based, ranking as the least probable.

Additionally, within such analyses, the identification of communities in the established relationship network can offer valuable insights. Community detection revealed groupings that include at least one node of each type: platforms, security mechanisms, interaction techniques, integration approaches, probable industry adopters and locational distribution. Mapping these associations highlights the most relevant combinations of platform capabilities for specific industry applications and geographies. This community perspective informs strategic decisions in selecting appropriate platforms capabilities for given use case requirements. Moreover, individual platforms offer differentiated solutions that serve distinct purposes, grouped into five categories: detailed virtual modeling, simulation and predictive analytics, physical-to-digital connectivity and data integration, immersive interaction modes, and lifecycle and change management. This research also confirms accelerated growth and maturation within the Digital Twin platform landscape over the past three years.

Despite limitations such as potential geographical bias toward platform providers and constraints related to documentation availability, the current analysis expands visibility into the dynamics of the Digital Twin platform ecosystem. Moving forward, several facets warrant deeper investigation in future research. For instance, expanding sample diversity and analyzing fringe use cases in terms of industry vertical and regional adoption to spotlight capability gaps restricting implementation. Longitudinal monitoring of advancing functionality across simulation fidelity, automation sophistication and cloud orchestration flexibility would reveal comparative platform suitability rates. Finally, incorporating financial partnership ties and usage data can provide tangible indicators of real-world sustainability alongside technical
capability. By blending functional, operational, and economic perspectives, evolving social network modeling and simulation techniques will empower stakeholders to chart technology maturation, predict partnership risks, and plan implementation timelines as industrial Digital Twins progress toward mainstream viability.

References

- X. Zhu, S. Ge, and N. Wang, "Digital transformation: A systematic literature review," *Comput. Ind. Eng.*, vol. 162, p. 107774, Dec. 2021, doi: 10.1016/j.cie.2021.107774.
- [2] T. Erol, A. F. Mendi, and D. Dogan, "Digital Transformation Revolution with Digital Twin Technology," in 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Istanbul, Turkey: IEEE, Oct. 2020, pp. 1–7. doi: 10.1109/ISMSIT50672.2020.9254288.
- [3] E. VanDerHorn and S. Mahadevan, "Digital Twin: Generalization, characterization and implementation," *Decis. Support Syst.*, vol. 145, p. 113524, Jun. 2021, doi: 10.1016/j.dss.2021.113524.
- [4] S. Boschert and R. Rosen, "Digital Twin—The Simulation Aspect," in *Mechatronic Futures*, P. Hehenberger and D. Bradley, Eds., Cham: Springer International Publishing, 2016, pp. 59–74. doi: 10.1007/978-3-319-32156-1_5.
- [5] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital Twin: Enabling Technologies, Challenges and Open Research," *IEEE Access*, vol. 8, pp. 108952– 108971, 2020, doi: 10.1109/ACCESS.2020.2998358.
- [6] S. Aheleroff, X. Xu, R. Y. Zhong, and Y. Lu, "Digital Twin as a Service (DTaaS) in Industry 4.0: An Architecture Reference Model," *Adv. Eng. Inform.*, vol. 47, p. 101225, Jan. 2021, doi: 10.1016/j.aei.2020.101225.
- [7] C. Cimino, E. Negri, and L. Fumagalli, "Review of digital twin applications in manufacturing," *Comput. Ind.*, vol. 113, p. 103130, Dec. 2019, doi: 10.1016/j.compind.2019.103130.
- [8] K. M. Alam and A. El Saddik, "C2PS: A Digital Twin Architecture Reference Model for the Cloud-Based Cyber-Physical Systems," *IEEE Access*, vol. 5, pp. 2050–2062, 2017, doi: 10.1109/ACCESS.2017.2657006.
- [9] R. Liyanage, N. Tripathi, T. Päivärinta, and Y. Xu, "Digital Twin Ecosystems: Potential Stakeholders and Their Requirements," in *Software Business*, vol. 463, N. Carroll, A. Nguyen-Duc, X. Wang, and V. Stray, Eds., in Lecture Notes in Business

Information Processing, vol. 463., Cham: Springer International Publishing, 2022, pp. 19–34. doi: 10.1007/978-3-031-20706-8_2.

- [10] W. Reim, E. Andersson, and K. Eckerwall, "Enabling collaboration on digital platforms: a study of digital twins," *Int. J. Prod. Res.*, vol. 61, no. 12, pp. 3926–3942, Jun. 2023, doi: 10.1080/00207543.2022.2116499.
- [11] F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji, "Digital twin modeling," *J. Manuf. Syst.*, vol. 64, pp. 372– 389, Jul. 2022, doi: 10.1016/j.jmsy.2022.06.015.
- P. Raj and C. Surianarayanan, "Digital twin: The industry use cases," in *Advances in Computers*, vol. 117, Elsevier, 2020, pp. 285–320. doi: 10.1016/bs.adcom.2019.09.006.
- [13] M. Singh *et al.*, "Applications of Digital Twin across Industries: A Review," *Appl. Sci.*, vol. 12, no. 11, p. 5727, Jun. 2022, doi: 10.3390/app12115727.
- [14] Q. Qi *et al.*, "Enabling technologies and tools for digital twin," *J. Manuf. Syst.*, vol. 58, pp. 3–21, Jan. 2021, doi: 10.1016/j.jmsy.2019.10.001.
- [15] Q. Wang, W. Jiao, P. Wang, and Y. Zhang, "Digital Twin for Human-Robot Interactive Welding and Welder Behavior Analysis," *IEEECAA J. Autom. Sin.*, vol. 8, no. 2, pp. 334– 343, Feb. 2021, doi: 10.1109/JAS.2020.1003518.
- M. Redeker, J. N. Weskamp, B. Rossl, and F. Pethig, "Towards a Digital Twin Platform for Industrie 4.0," in 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Victoria, BC, Canada: IEEE, May 2021, pp. 39–46. doi: 10.1109/ICPS49255.2021.9468204.

[17] Z. Wang, W. Feng, J. Ye, J. Yang, and C. Liu, "A Study on Intelligent Manufacturing Industrial Internet for Injection Molding Industry Based on Digital Twin," *Complexity*, vol. 2021, pp. 1–16, Jan. 2021, doi: 10.1155/2021/8838914.

- [18] F. Kakavandi *et al.*, "Towards Developing a Digital Twin for a Manufacturing Pilot Line: An Industrial Case Study," in *Digital Twin Driven Intelligent Systems and Emerging Metaverse*, E. Karaarslan, Ö. Aydin, Ü. Cali, and M. Challenger, Eds., Singapore: Springer Nature Singapore, 2023, pp. 39–64. doi: 10.1007/978-981-99-0252-1 2.
- [19] Y. Zhang, A. Meina, X. Lin, K. Zhang, and Z. Xu, "Digital Twin in Computational Design and Robotic Construction of Wooden Architecture," *Adv. Civ. Eng.*, vol. 2021, pp. 1–14, Apr. 2021, doi: 10.1155/2021/8898997.
- [20] Z. Ye, L. Jingyu, and Y. Hongwei, "A digital twinbased human-robot collaborative system for the assembly of complex-shaped architectures," *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.*, p.

095440542211109, Aug. 2022, doi: 10.1177/09544054221110960.

- [21] Q. Lu, X. Xie, A. K. Parlikad, and J. M. Schooling, "Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance," *Autom. Constr.*, vol. 118, p. 103277, Oct. 2020, doi: 10.1016/j.autcon.2020.103277.
- [22] Y. Jiang *et al.*, "Digital twin-enabled real-time synchronization for planning, scheduling, and execution in precast on-site assembly," *Autom. Constr.*, vol. 141, p. 104397, Sep. 2022, doi: 10.1016/j.autcon.2022.104397.
- [23] Z. Ye *et al.*, "A digital twin approach for tunnel construction safety early warning and management," *Comput. Ind.*, vol. 144, p. 103783, Jan. 2023, doi: 10.1016/j.compind.2022.103783.
- [24] Y. Jiang, X. Liu, Z. Wang, M. Li, R. Y. Zhong, and G. Q. Huang, "Blockchain-enabled digital twin collaboration platform for fit-out operations in modular integrated construction," *Autom. Constr.*, vol. 148, p. 104747, Apr. 2023, doi: 10.1016/j.autcon.2023.104747.
- [25] A. Lee, K.-W. Lee, K.-H. Kim, and S.-W. Shin, "A Geospatial Platform to Manage Large-Scale Individual Mobility for an Urban Digital Twin Platform," *Remote Sens.*, vol. 14, no. 3, p. 723, Feb. 2022, doi: 10.3390/rs14030723.
- [26] B. Li and W. Tan, "A novel framework for integrating solar renewable source into smart cities through digital twin simulations," *Sol. Energy*, vol. 262, p. 111869, Sep. 2023, doi: 10.1016/j.solener.2023.111869.
- [27] H. Xu et al., "Smart Mobility in the Cloud: Enabling Real-Time Situational Awareness and Cyber-Physical Control Through a Digital Twin for Traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 3145–3156, Mar. 2023, doi: 10.1109/TITS.2022.3226746.
- [28] M. Ibrahim, V. Rjabtšikov, and R. Gilbert, "Overview of Digital Twin Platforms for EV Applications," *Sensors*, vol. 23, no. 3, p. 1414, Jan. 2023, doi: 10.3390/s23031414.
- [29] Deloitte, "The hospital of the future," 2017.
- [30] O. Mazumder, D. Roy, S. Bhattacharya, A. Sinha, and A. Pal, "Synthetic PPG generation from model haemodynamic with baroreflex autoregulation: a Digital twin of cardiovascular system," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany: 5024-5029. IEEE, Jul. 2019, pp. doi: 10.1109/EMBC.2019.8856691.
- [31] H. Ahmed and L. Devoto, "The Potential of a Digital Twin in Surgery," Surg. Innov., vol. 28, no.

4, pp. 509–510, Aug. 2021, doi: 10.1177/1553350620975896.

- [32] K. Subramanian, "Digital Twin for Drug Discovery and Development—The Virtual Liver," *J. Indian Inst. Sci.*, vol. 100, no. 4, pp. 653–662, Oct. 2020, doi: 10.1007/s41745-020-00185-2.
- [33] L. Devoto, S. Muscroft, and M. Chand, "Highly Accurate, Patient-Specific, 3-Dimensional Mixed-Reality Model Creation for Surgical Training and Decision-making," *JAMA Surg.*, vol. 154, no. 10, p. 968, Oct. 2019, doi: 10.1001/jamasurg.2019.2546.
- [34] A. Croatti, M. Gabellini, S. Montagna, and A. Ricci, "On the Integration of Agents and Digital Twins in Healthcare," *J. Med. Syst.*, vol. 44, no. 9, p. 161, Sep. 2020, doi: 10.1007/s10916-020-01623-5.
- [35] T. Moyaux, Y. Liu, G. Bouleux, and V. Cheutet, "An Agent-Based Architecture of the Digital Twin for an Emergency Department," *Sustainability*, vol. 15, no. 4, p. 3412, Feb. 2023, doi: 10.3390/su15043412.
- [36] S. S. Akash and M. S. Ferdous, "A Blockchain Based System for Healthcare Digital Twin," *IEEE Access*, vol. 10, pp. 50523–50547, 2022, doi: 10.1109/ACCESS.2022.3173617.
- [37] M. Grieves, "Intelligent digital twins and the development and management of complex systems," *Digit. Twin*, vol. 2, p. 8, May 2022, doi: 10.12688/digitaltwin.17574.1.
- [38] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the Digital Twin: A systematic literature review," *CIRP J. Manuf. Sci. Technol.*, vol. 29, pp. 36–52, May 2020, doi: 10.1016/j.cirpj.2020.02.002.
- [39] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital Twin in Industry: State-of-the-Art," *IEEE Trans. Ind. Inform.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019, doi: 10.1109/TII.2018.2873186.
- [40] D. Yang, H. R. Karimi, O. Kaynak, and S. Yin, "Developments of digital twin technologies in industrial, smart city and healthcare sectors: a survey," *Complex Eng. Syst.*, 2021, doi: 10.20517/ces.2021.06.
- [41] E. Negri, L. Fumagalli, and M. Macchi, "A Review of the Roles of Digital Twin in CPS-based Production Systems," *Procedia Manuf.*, vol. 11, pp. 939–948, 2017, doi: 10.1016/j.promfg.2017.07.198.
- [42] F. Tao and M. Zhang, "Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing," *IEEE Access*, vol. 5, pp. 20418– 20427, 2017, doi: 10.1109/ACCESS.2017.2756069.

Exploring Self-Supervised GPR Representation Learning for Building Rooftop Diagnostics

Kevin Lee¹, Wei-Heng Lin¹, Bilal Sher^{2,1,*}, Talha Javed^{2,1,*}, Sruti Madhusudhan^{2,1,*}, Chen Feng^{1,2}

> ¹Tandon School of Engineering, New York University, USA ²Building Diagnostic Robotics, Inc., USA *Indicates equal contributions

k.lee@nyu.edu, whl318@nyu.edu, bas9876@nyu.edu, tj2186@nyu.edu, sm10183@nyu.edu, cfeng@nyu.edu

Abstract -

Ground Penetrating Radar (GPR), known for its applications in diverse domains, demonstrates potential for nondestructive diagnostic assessments on building rooftops. This study delves into the unique characteristics and data structure of GPR, investigating the novel approach of processing GPR as a "contextual neighborhood" of A-scans within their respective B-scans as opposed to the typical pixel-based approach. Given the challenge of obtaining a large corpus of annotated rooftop GPR data, we employ self-supervised deep learning methods for GPR representation learning. Experiments include training a vanilla Autoencoder, Variational Autoencoder, and a Transformer-based Autoencoder on GPR A-scans. Additionally, we extend our analysis by fine-tuning a pre-trained Masked Autoencoder on image based GPR B-Scans to investigate the differences between the conventional pixel-based approach and our proposed Ascan-based approach. Through a meticulous analysis of the learned latent spaces across these methods, we assess the viability of self-supervised deep learning in encoding meaningful GPR representations for downstream tasks. This research contributes to the exploration of GPR's applicability in building rooftop diagnostics and underscores the potential of self-supervised deep learning for efficient representation learning in the absence of annotated data.

Keywords -

GPR; Rooftop; Self-supervised; Transformer; Autoencoder

1 Introduction

Ensuring the structural integrity of building rooftops necessitates reliable assessment methods. Ground Penetrating Radar (GPR) emerges as a promising, cost-effective alternative to traditional techniques like infrared thermography and nuclear moisture gauges [1]. Utilizing electromagnetic waves, GPR captures changes in electromagnetic properties within the scanned medium of the propagating waves.



Neighborhood shares sequential and contextual information



Figure 1. Contextual neighborhood of A-scans within the same B-scan. A-scans from the same B-scan share contextual information that may aid in the identification of features of interest. The blue scans indicate typical GPR readings that may not be pertinent for analysis, and the red scans indicate scans that contain features of interest.

GPR data is typically presented in the form of A-scans and B-scans. An A-scan, a single-point, 1D scan collected at each sampling interval, records amplitude variations of reflected waves over travel time. Despite its 1D nature, an A-scan encapsulates information beyond its precise location due to the non-linear propagation of electromagnetic waves. This distinction can be observed in the appearance of point objects like pipes as hyperbolas in GPR B-scans.

In contrast, B-scans are a series of sequentially mea-

sured A-scans often treated as images, offering a humaninterpretable view of subsurface features. In rooftop diagnostics, many applicable downstream tasks appreciate higher levels of granularity in analyses. As B-scans may span many meters in length compared to the centimeterscale offered by A-scans, we investigate approaches that study GPR at the A-scan level. This allows us the opportunity to explore the potential advantages in studying GPR scans through leveraging their inherent data structure - as opposed to the conventional method of processing GPR B-scans as images. Figure 1 illustrates the concept of contextual neighborhood membership of A-scans within the same B-scan. Deep learning models may be able to extract richer semantics in GPR representations through the exploit of intrinsic contextual information shared between A-scans of the same B-scan, especially considering the capability of A-scans to capture information adjacent to the immediate position of the GPR sensor.

Despite GPR's capabilities in producing high-resolution radargrams, interpreting scans remains challenging even for skilled technicians [2]. This challenge underscores the motivation to integrate deep learning for automated interpretation. While prior applications of deep learning in GPR interpretation exist, such as for land mine detection and utility identification [3, 4, 5], they primarily relied on supervised training with annotated labels. Given the scarcity of such labeled data for building rooftops, our investigation centers on self-supervised methods.

Our contributions can be summarized as follows:

- 1. Exploration of Self-Supervised Deep Learning Models: We assess the efficacy of self-supervised deep learning models in extracting meaningful features within the latent space of GPR data.
- 2. Analysis of Inherent Data Structure: We investigate the advantages of addressing GPR data through its native data structure (A-scan, B-scan) as opposed to treating it as a conventional image.

These contributions collectively aim to advance the understanding of GPR applications in building diagnostics, particularly in the context of self-supervised learning.

2 Related Works

2.1 GPR for Subsurface Analysis

GPR has historically found great utility in imaging and analyzing subsurface features in various outdoor environments. For example, the sensor has found use in monitoring the structural integrity of river embankments and levees [6]. The non-destructive nature of GPR enabled the visualization and detection of underground animal burrows without disrupting the site. This advantage has also been leveraged in the inspection of concrete bridges [7]. Similarly to [6], GPR was used to detect and identify the different layers present in the composition of bridges, as well as the thickness of each layer. We consider the success in the use of GPR in these varied environments as an indication of the potential viability in the use of GPR in the building rooftop setting.

2.2 Use of GPR for Building Diagnostics

Though the use of GPR has predominantly existed outside of building diagnostics, investigations on the use of GPR in various applications on buildings and similar structures exist. The technology has been used to study and identify cracks and signs of moisture damage in historical and heritage buildings [8, 9], and has even been used to detect moisture within building walls in conjunction with a thermal camera in [10]. However, many of these methods rely on manual interpretation or an ensemble of processing methods. These limitations affect the scalability and generalizability of the methods, which are key factors to consider when addressing the sheer variability in the types of buildings, not to mention the number of buildings that may need to be serviced.

2.3 Learning-Based GPR Analysis

Several works [1, 3, 4, 5, 11] have delved into learningbased analysis of GPR scan data for feature detection. For instance, [11] utilizes the mask R-CNN [12] architecture to automatically detect and segment cracks in asphalt pavement at the pixel-level. However, these approaches predominantly rely on supervised learning methods, necessitating manual annotation of GPR scans. This reliance on labeled data poses a challenge in terms of scalability due to the labor-intensive nature of annotation. Additionally, these works predominantly study GPR feature detection from the perspective of GPR B-scans as images. To address these limitations, our study explores the potential of self-supervised learning, aiming to extract meaningful features from GPR scans without the need for extensive manual annotations. We also hope to differentiate our contributions through a comprehensive analysis addressing the utility of A-scans in context of B-scan neighborhoods over the typical image-based approach.

2.4 Autoencoders for Self-Supervised Learning

The Autoencoder (AE) architecture is a popular approach to self-supervised learning. This architecture typically involves an Encoder block that learns to compress the input into the latent space, and a Decoder block that learns to extract the original input from the compressed representation [13]. Over the years, many derivatives have emerged from the original Autoencoder idea including the Variational Autoencoder (VAE), which aims to encode the



Figure 2. A simplified diagram describing the end-to-end training and testing procedure for each proposed method.

input as a distribution over the latent space [14], and the Masked Autoencoder (MAE), which motivates effective self-supervised learning through the nontrivial task of uncorrupting masked inputs [15]. Transformers [16] have also been largely successful in challenging natural language and computer vision tasks. Recalling the unique property of A-scans sharing contextual information between their neighbors within the same B-scan, we consider a Transformer-based Autoencoder (TAE) that takes a window of sequential A-scans as input. We aim to evaluate these approaches to determine their capabilities in learning meaningful features from GPR data without the need for signals from manually annotated labels.

2.5 Challenges in the Use of Synthetic GPR Data

gprMax [17] is a GPR simulation software that has been leveraged in works such as [11, 18] to generate a large corpus of synthetic scans for training deep learning models. While effective in certain applications, particularly those requiring controlled environments, the use of synthetic scans presents challenges in the context of building rooftop diagnostics. Capturing the inherent variability in rooftop composition and the diverse range of surface features, such as pipes, vents, and walls, becomes complex. The fidelity of synthetic scans may struggle to emulate the nuanced characteristics of real-world GPR data from building rooftops. Due to the present obstacles, we decide against the use of synthetic GPR data in our analysis and instead collect data from real-world commercial and retail building rooftops.

3 Methodology

Our objective is to investigate various methods for learning representations of Ground Penetrating Radar data, specifically focusing on:

- 1. Vanilla Autoencoder
- 2. Variational Autoencoder
- 3. Transformer-based Autoencoder
- 4. Image-level Masked Autoencoder

These methods are compared against a baseline approach, involving the analysis of preprocessed data without a deep learning model. The Autoencoder, Variational Autoencoder, and Transformer-based Autoencoder are trained from scratch on GPR A-scans, while the ImageNet pretrained Masked Autoencoder will be finetuned on GPR B-scans treated as images. This approach aims to highlight any distinctions between leveraging the inherent data structure of GPR and treating GPR scans as holistic images. Evaluation of each method includes the presentation of reconstruction samples, visualizations of latent spaces using T-SNE for qualitative analysis, and quantitative metrics utilizing KMeans clustering. Figure 2 provides a simplified illustration of the described procedure.

3.1 Dimensionality Reduction

T-SNE, a common dimensionality reduction technique, is employed to visualize high-dimensional latent spaces. For the deep learning models, T-SNE is applied to feature vectors generated from the encoder block's forward pass. As is typically recommended to reduce the computation of T-SNE, models that have encoders that output to relatively high dimensions use PCA to reduce the feature dimension to 50 prior to the application of T-SNE. This procedure applies to the TAE and MAE, as well as the baseline. Feature vectors for AE, VAE, and TAE represent encoded A-scans, while those for MAE represent the 16x16 patches created from resized B-scans. Despite differences in representation, patches are reasonably analogous to A-scans as they capture local features from the entire B-scan. Indicators of meaningful encoding include visible clustering of feature vectors and separation between clusters.

3.2 Clustering

KMeans clustering is utilized to quantitatively evaluate the quality of the learned latent spaces. The algorithm is similarly applied to the generated feature vectors for each method. Two metrics, the Silhouette Score [19] and the average distance between cluster centroids, are employed for comparison. The Silhouette Score assesses cluster well-definition, while the average distance measures cluster separation. These metrics are calculated over a range of k = [2, 3, ..., 15] clusters, as the appropriate number of clusters needed to capture GPR representations is unknown. The value of the Silhouette Score can range between -1 and 1, inclusive, with a value closer to 1 indicating high-quality clusters. A score greater than 0.5 is generally interpreted as reasonable.

3.3 Data Collection

The dataset comprises diverse scans collected from realworld building rooftops using the Proceq GP8800 SFCW handheld GPR sensor. SFCW, or Stepped Frequency Continuous Wave, indicates that this sensor is capable of broadcasting a wide range of frequencies to scan at both shallow and deep depths. While capable of a broader range of sampling rates, the sensor was set to record one scan per centimeter traveled, focusing on an approximately 9 x 9 cm area. The GPR is mounted on a user-controlled differential drive robot, seen in Figure 3 approximately 3cm from the ground surface. The GPR employed a servo motor to facilitate scanning by matching the GPR scanning rate with the robot's speed. During the data collection process, the robot navigated the roof in a series of straight segments, limiting the GPR data to those linear paths. The scanning process was intentionally paused during turns. Multiple scans were performed per rooftop. The straight segments were positioned approximately 0.9m apart from one another, ensuring comprehensive coverage of the entire roof surface. Multiple paths across each rooftop were scanned, resulting in a dataset of over 1400 B-scans or over 1.3 million A-scans. Each B-scan contains a variable number of A-scans - from a few tens of scans to a few thousand - and each A-scan records 655ns in two-way travel

time. The dataset covers a diverse range of roofing surfaces, including built-up roofing with gravel overburden, Styrene-Butadiene-Styrene (SBS) modified asphalt roofing, Ethylene Propylene Diene Monomer (EPDM) roofing, Thermoplastic Polyolefin (TPO) roofing, and Polyvinyl Chloride (PVC) roofing. In addition, scans were acquired on both dry and moisture-saturated surfaces. The total area scanned exceeds 50k square meters.

3.4 Preprocessing

The dataset is partitioned into train, validation, and test sets (containing 1181, 144, and 144 B-scans, or over 1.1m, 136k, and 130k A-scans, respectively). While standard preprocessing techniques including signal gain and denoising are employed for better manual interpretation of GPR data, we have found that normalization by mean and standard deviation across training A-scans accomplishes the same as well as enhancing training performance.



Figure 3. Robot used for GPR Data Collection

3.5 Training

Each model is trained using varied regiments that are best suited for the model. The AE and VAE is trained for 100 epochs using a starting learning rate of 1.5e-4 and a batch size of 128. The batches are comprised of A-scans that are randomly sampled from the training dataset. The TAE is trained for 600 epochs using a starting learning rate of 1.5e-4 and a batch size of 16. The batches for the TAE are comprised of a random window of 64 sequential A-scans from a randomly sampled B-scan. If the B-scan is shorter than 64 in length, the scan is padded with the mean A-scan to make up the difference. The MAE is finetuned over 50 epochs using a starting learning rate of 1e-3 and a batch size of 16, with the masking ratio set to the default 0.75 as described in the original paper. During inference, the masking ratio is set to 0 to ensure each token is encoded, as the original architecture discards all masked tokens immediately after masking. Each batch is comprised of randomly sampled B-scans that are resized to (224 x 224). The resized scans are additionally normalized using the mean and standard deviation calculated from ImageNet. All models are trained using the AdamW optimizer with decay = 0.05 and betas = (0.9, 0.95) and a cosine decay learning rate schedule. Training is conducted on a machine with an Nvidia RTX 8000 GPU and the code is written using PyTorch. reading of some sort of point object, possibly a pipe, and on the right there is a response that is typical of "ponding" or moisture on the surface. All models are successful in reconstructing these features as well as the surrounding responses without excessive artifacts or noise. This is to demonstrate that the models are effective in encoding and reconstructing GPR signals, which allows us to further our investigation in determining whether anything meaningful is being encoded in the latent spaces of each model.



Figure 4. **Sample reconstruction outputs** from all trained deep learning methods. This scan is specifically chosen to study the diverse set of features that may be present in GPR scan data.

4 Results

4.1 Reconstruction

As shown in Figure 4, every deep learning model trained is successful in reconstructing GPR scans. To highlight the various features that may be present in GPR scans, a specific scan was chosen from the test dataset for visualization. On the bottom left corner of the scan, there is a



Figure 5. **Illustration of the learned latent spaces** generated from each method, with the inclusion of the baseline for reference. We note the similarities present in the baseline, AE, VAE, and TAE. This figure is best viewed in color.

4.2 Qualitative Evaluation

T-SNE visualizations are created with perplexity set to 50 to account for the large number of A-scans. Two visualizations are provided, one displaying the latent space using the entire test dataset and the other showing the encoded features of a single B-scan. This is done to examine how the models are organizing the latent space, as well as to provide a relevant visualization for interpretation.

Through analyzing the latent space, we may be able to understand whether the models are encoding certain A-scans similarly or if there is some latent structure to the data. For instance, A-scans that indicate the presence of objects like pipes would ideally be encoded closer together in a cluster and separate from A-scans that indicate moisture. Color is provided to further illustrate the location of points in 3D space. In Figure 4 we observe similar behavior to the baseline for the AE, VAE, and TAE models. While not well defined, weak clustering is apparent. Notably, the latent space appears to be organized similarly across the mentioned methods, particularly in the large cluster visible on the left. Despite the visible formation of weak feature clusters, it is not immediately apparent whether these groups are truly distinct or semantically meaningful.

This similarity is continued in Figure 6 where we observe the same pattern of a large cluster towards the left and smaller, tighter clusters on the right in the latent space visualizations. Displaying the latent spaces alongside the corresponding encoded scan provides additional visual cues that aid in understanding how the scans are encoded. We observe the possibility of a grouping of like A-scans which is particularly evident in the TAE. The A-scans belonging to the point object reading are encoded near each other, with the same being observed for the A-scans pertaining to the moisture-saturated surface.

The MAE stands out among the methods tested. The latent space visualization shown in Figure 5 for the MAE describes reasonably well-defined and separated clusters. However, the MAE results in Figure 6 are not as intuitive to interpret. The latent space appears sparse and there is no visible pattern in the encoding visualization. Therefore, it is uncertain whether the clustering exhibited in Figure 5 is representative of meaningful encodings. This result also suggests the validity of studying GPR scans as a collection of A-scans in a neighborhood of their respective B-scans.

4.3 Quantitative Evaluation

The following metrics are calculated over a range of k = [2, 3, ..., 15] clusters, inclusive. This is done to ensure a holistic understanding of method performance despite not knowing intuitively how many clusters are appropriate for GPR data. The number of clusters that adequately describe the GPR signal behavior is unknown. As mentioned prior, a positive Silhouette Score that is close to 1 and a large Average Centroid Distance indicates clusters that are well defined and distinct.

While the results in Figure 7 demonstrate the AE generating higher quality clusters among the methods tested, the scores observed indicate fairly weak performance. Similarly, the results shown in Figure 8 seem to suggest a substantial separation of clusters from the AE. However, the AE performs similarly to the MAE in this metric, which



Figure 6. Encoding visual of a specific scan through all methods. The colors seen in the encoding visual correspond to the colors of their respective points in the 2D latent space. For the baseline, AE, VAE, and TAE, the columns represent the encoded A-scans from the sample scan. The MAE visual represents the encoded 16 x 16 patches. This figure is best viewed in color.

measured a comparatively low Silhouette Score. Thus, the quantitative results also remain inconclusive.



Figure 7. Silhouette Score for 2-15 clusters. This figure is best viewed in color.



Figure 8. Average Centroid Distance for 2-15 clusters. This figure is best viewed in color.

5 Discussion and Conclusions

Self-supervision enables the possibility of leveraging the vast capabilities of deep learning in GPR processing at

a massive scale. We believe that advancements in this area can further refine the performance and generalizability of works such as [1] and [4] through training on far larger and more comprehensive unlabeled GPR datasets. Additionally, enabling self-supervised training on massive realworld datasets reduces the reliance on simulated synthetic data, whose characteristics may not translate accurately to the real world. As such, we recognize the significance of extending the utility of GPR in rooftop diagnostics through self-supervised GPR representation learning. Our study, however, reveals that this task is not without its complexities.

The challenges in learning meaningful representations likely stem from the inherent variability in GPR scan data. Rooftop scans encompass diverse materials, compositions, and features distributed somewhat unpredictably. This intricacy poses a substantial obstacle for models aiming to autonomously derive semantically significant encodings.

Despite the inconclusive results in representation learning, we remain optimistic in the approach of studying GPR as a sequence of A-scans within a shared B-scan context based on the results demonstrated in Figure 6.

Moving forward, addressing these challenges may involve exploring additional pre-processing techniques or alternative model architectures. Future research should investigate strategies to enhance the robustness of selfsupervised GPR representation learning, ensuring its adaptability across diverse scenarios in building diagnostics.

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

Disclosure: Bilal Sher, Sruti Madhusudhan, Talha Javed, and Chen Feng are founders of Building Diagnostic Robotics, Inc., a startup company that uses AI and robotics for building inspections.

References

- [1] Sher B. and Feng C. Deepgpr: Learning to identify moisture defects in building envelope assemblies from ground penetrating radar. In Borja García de Soto, Vicente Gonzalez-Moret, and Ioannis Brilakis, editors, *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 561–568, Chennai, India, July 2023. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-0-4. doi:10.22260/ISARC2023/0075.
- [2] Utsi E. Ground Penetrating Radar: Theory and Practice. Elsevier Science, 2017. ISBN 9780081022177.
- [3] Lameri S., Lombardi F., Bestagini P., Lualdi M., and Tubaro S. Landmine detection

from gpr data using convolutional neural networks. In 2017 25th European Signal Processing Conference (EUSIPCO), pages 508–512, 2017. doi:10.23919/EUSIPCO.2017.8081259.

- [4] Hou F., Lei W., Li S., and Xi J. Deep learningbased subsurface target detection from gpr scans. *IEEE Sensors Journal*, 21(6):8161–8171, 2021. doi:10.1109/JSEN.2021.3050262.
- [5] Jaufer R., Ihamouten A., Guilbert D., Todkar S., Yaram T., and Derobert X. Deep learning based automatic hyperbola detection on gpr data for buried utility pipes mapping. In 2021 11th International Workshop on Advanced Ground Penetrating Radar (IWAGPR), pages 1–6, 2021. doi:10.1109/IWAGPR50767.2021.9843151.
- [6] Di Prinzio M., Bittelli M., Castellarin A., and Pisa P. Application of gpr to the monitoring of river embankments. *Journal of Applied Geophysics*, 71(2):53–61, 2010. ISSN 0926-9851. doi:10.1016/j.jappgeo.2010.04.002.
- [7] Hugenschmidt J. and Mastrangelo R. Gpr inspection of concrete bridges. *Cement and Concrete Composites*, 28(4):384–392, 2006. ISSN 0958-9465. doi:10.1016/j.cemconcomp.2006.02.016. Non-Destructive Testing.
- [8] Orlando L. and Slob E. Using multicomponent gpr to monitor cracks in a historical building. *Journal* of Applied Geophysics, 67(4):327–334, 2009. ISSN 0926-9851. doi:10.1016/j.jappgeo.2008.09.003. Advanced Applications, Systems and Modelling for GPR.
- [9] Perez-Gracia V., Solla M., and Fontul S. Analysis of the gpr signal for moisture detection: application to heritage buildings. *International Journal of Architectural Heritage*, 18(2):230–253, 2024. doi:10.1080/15583058.2022.2139652.
- [10] Garrido I., Solla M., Lagüela S., and Fernández N. Irt and gpr techniques for moisture detection and characterisation in buildings. *Sensors*, 20(22), 2020. ISSN 1424-8220. doi:10.3390/s20226421.
- [11] Liu Z., Yeoh J., Gu X., Dong Q., Chen Y., Wu W., Wang L., and Wang D. Automatic pixel-level detection of vertical cracks in asphalt pavement based on gpr investigation and improved mask r-cnn. *Automation in Construction*, 146:104689, 2023. ISSN 0926-5805. doi:10.1016/j.autcon.2022.104689.
- [12] He K., Gkioxari G., Dollár P., and Girshick R. Mask r-cnn. In 2017 IEEE International Conference on

Computer Vision (ICCV), pages 2980–2988, 2017. doi:10.1109/ICCV.2017.322.

- [13] Ballard D. Modular learning in neural networks. In Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1, AAAI'87, page 279–284. AAAI Press, 1987. ISBN 0934613427.
- [14] Kingma D. and Welling M. Auto-encoding variational bayes, 2022.
- [15] He K., Chen X., Xie S., Li Y., Dollár P., and Girshick R. Masked autoencoders are scalable vision learners. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15979– 15988, 2022. doi:10.1109/CVPR52688.2022.01553.
- [16] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., and Polosukhin I. Attention is all you need. In Guyon I., Von Luxburg U., Bengio S., Wallach H., Fergus R., Vishwanathan S., and Garnett R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [17] Warren C., Giannopoulos A., and Giannakis I. gprmax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Computer Physics Communications*, 209:163–170, 2016. ISSN 0010-4655. doi:10.1016/j.cpc.2016.08.020.
- [18] Dai Q., Lee Y., Sun H., Ow G., Yusof M., and Yucel A. Dmrf-unet: A two-stage deep learning scheme for gpr data inversion under heterogeneous soil conditions. *IEEE Transactions on Antennas and Propagation*, 70(8):6313–6328, August 2022. ISSN 1558-2221. doi:10.1109/tap.2022.3176386.
- [19] Rousseeuw P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65, 1987. ISSN 0377-0427. doi:10.1016/0377-0427(87)90125-7.

Underground Utility Network Completion based on Spatial Contextual Information of Ground Facilities and Utility Anchor Points using Graph Neural Networks

Yuxi Zhang¹ and Hubo Cai¹

¹Lyles School of Civil Engineering, Purdue University, United States zhan2889@purdue.edu, hubocai@purdue.edu

Abstract -

Every year, accidental damage during excavation leads to numerous disruptions in utility services. These incidents cause not only financial losses but also injuries and fatalities. A major contributing factor to these incidents is the lack of accurate location data for utilities. The current practice involves a timeconsuming coordination process of obtaining utility maps from owners and field surveys, which is often hindered by delays and incomplete records. In response to these challenges, this paper proposes a novel method to predict underground utility lines in situations where records are unavailable or delayed. Our approach leverages visible utility anchor points, such as manholes, and the spatial context provided by nearby ground features like roads. The methodology involves three primary steps: constructing a relational data model of the utility network, transforming this data into graphs, and employing a graph neural network for prediction. This innovative approach demonstrates good performance, achieving a ROC AUC score of 95.24% in predicting sewer line connections between manholes. This method automates the inference of utility lines, providing utility owners and excavation contractors a solution for identifying unknown connections and reducing risks from inaccurate information.

Keywords -

Underground Utility Network Completion; Spatial Contextual Information; Graph Neural Networks

1 Introduction

The ongoing issue of inaccurate and incomplete information of buried utilities poses a significant challenge across the United States. Annually, numerous utility disruptions are caused by accidental excavation damage. These incidents impact communities and businesses, leading to injuries and tragically resulting in loss of life. According to Common Ground Alliance Damage Information Reporting Tool [1], 87.84% of these incidents occur due to missing or inaccurate location information. Current practice to mitigate these risks involves coordinating with utility owners to access utility maps and employ utility surveyors. The utility map serves as a crucial starting point, providing approximate line locations for further utility surveys. However, obtaining utility records faces prolonged delays in the coordination process, and some records may be entirely absent. Therefore, there's an urgent need to propose a method for inferring utility line locations when utility records are delayed or unavailable.

When records are inaccessible, inferring some utility lines is possible by examining visible utility anchor points like manholes and nearby ground facilities such as roads and buildings. These visible features imply the presence and general locations of utilities. Acquiring information about these visible features is feasible through field surveys or high-resolution satellite imagery. However, this inference relies on scarce professional judgment and expertise, which can be time-consuming, error-prone, and may further complicate the process.

This paper introduces a novel approach for automatically completing underground pipeline networks. It focuses on predicting utility line segments by using visible utility anchor points and ground facilities as spatial contextual cues. The objective is to aid users in inferring the existence and approximate locations of utility lines when utility records are not accessible.

2 Literature Review

2.1 Utility Parameters, Spatial Contexts, and Design Practices for Predicting Utilities

Existing studies [2-6] address the design and completion of utility networks by predicting the presence of pipelines based on their endpoints, such as manholes, and assessing the extent to which the network conforms

to design criteria and practices. For instance, Afshar [2] suggested minimizing the cost function related to pipe diameters and excavation depths while adhering to certain constraints to reflect design criteria. He compiled a list of sewer design practices as constraints, such as the minimum flow velocity required to prevent sediment buildup and the minimum pipe slope necessary to avoid adverse slopes due to inaccurate construction. Similarly, Izquierdo [3] formulated the problem for hydraulic systems, akin to Afshar's approach, but also incorporated the continuity and energy equations of hydraulics into the model. These studies concentrate on using the parameters of the pipeline and pipe endpoints to aid in the design of the pipeline network.

Furthermore, some research extends beyond the parameters of pipes and their endpoints. It also examines the relationship of these endpoints to visible ground elements and the surrounding environment, such as catchments, roads, and buildings near the pipes. For example, Bailly et al. [4] predicted the presence of pipelines based on the cumulative length of pipelines in relation to the catchment extent and network connectivity. Chahinian et al. [7] used manhole locations and elevations to predict the presence of pipelines and minimize instances of lines intersecting buildings and roads. Their top result achieved a precision and recall of 0.92 each, alongside a critical success index of 0.85. These studies underscore the importance of spatial contexts in enhancing pipeline network predictions.

2.2 Challenges and Limitations in Existing Studies

Existing studies carefully consider the information crucial for completing or designing pipeline networks. However, they face challenges in both mathematically modeling and solving the problem as follows:

- 1. One primary difficulty is the unknown correlation among pipe endpoint parameters, their connection parameters, spatial contexts, and pipeline presence. Existing studies simplify the problem by assumptions, leading to a lack of justification.
- 2. Another challenge is capturing the interdependency of variables within a network solely through human knowledge. This limits current methods to focusing only on parameters directly connected to the pipelines or nearby ground facilities, overlooking broader interdependencies.
- 3. Additionally, even when correlations and relationships are simplified and mathematically formulated, solving the model becomes computational expensive. These problems are often approached as combinatorial optimization, aiming to minimize costs while considering various constraints. The complexity of these problems is

compounded by non-linear functions and constraints, resulting in a solution space filled with numerous local minima and discontinuities. Consequently, studies have resort to computational expensive optimization methods such as heuristic algorithms, particle swarm, ant colony optimization, and others, in pursuit of the global optimal solution.

In summary, current research mainly utilizes rulebased approaches to predict pipeline connections between two endpoints, considering both their parameters and spatial contexts. This body of research highlights the complexities involved in formulating and solving these problems, especially the challenges in converting industry practices into effective cost functions. It indicates that explicitly modeling this problem relying solely on human knowledge presents significant challenges. Additionally, the complexities hinder further exploration of factors, such as the detailed spatial relationships between manholes and their surrounding environment, related to pipeline prediction.

2.3 Advantages of GNNs in Pipeline Network Completion

In the context of pattern recognition, learning-based methods can overcome the limitations of previous studies that struggled with explicitly modeling cost functions. With sufficient data, machine learning can quickly adapt to data from diverse practices.

Among the learning-based approaches, Deep Neural Networks (DNNs) [8,9] distinguish themselves from traditional machine learning methods by simultaneously learning features and objective functions. The advantages of using it for this problem lie in three aspects:

- 1. Alignment with Graphical Data Structures: Pipeline networks are inherently structured in a graphical format, with manholes serving as nodes and pipelines as edges. This naturally aligns with the architecture of Graph Neural Networks (GNNs), facilitating the integration of information into a unified network for discerning data correlations. Additionally, this problem can be formulated as linkage prediction in GNN studies [10], a wellestablished research area that is supported by a solid mathematical and statistical foundation.
- 2. Feature Extraction from Subgraphs: GNNs specialize in handling graph-structured data, enabling the extraction of comprehensive features from subgraphs [11]. These methods, known as graph embeddings, allow for the representation of pipeline networks by aggregating information not just from direct connections but also from the broader network context.
- 3. Discriminative Feature Learning: The concurrent learning of features and objective functions lead to

learn discriminative implicit representations [8]. In contrast to traditional methods that linearly model relationships between handcrafted features, DNNs excel in learning implicit feature representations that encode complex relationships within the data. These features are specifically optimized for the downstream task, enhancing the accuracy and effectiveness of pipeline network predictions [12].

4. Recent advances in using GNN models in Geospatial Artificial Intelligence (GeoAI): GNN models are particularly adept at handling geospatial challenges that involve analyzing points of interest, their spatial relationships, and non-grid topologies. GNNs have shown notable effectiveness in applications such as traffic flow [13] and PM2.5 level forecasting [14], where training and testing occur on the same nodes, referring to transductive learning. A significant challenge in GeoAI, however, is the application of models trained on one set of location data to completely new, unseen locations, known as inductive learning. To address this, significant advancements have been made in geospatial location encoding techniques [15]. These techniques transform location data, whether two- or three-dimensional, into a high-dimensional feature vector. This approach preserves relative distances and, optionally, directional relationships between locations, enhancing the model's ability to adapt to new locations not seen in the training phase.

Therefore, there is a need to explore the potential to overcome the limitations of current utility network completion methods using data-driven approaches. This study mainly introduces the framework of utility line prediction, addressing the following two challenges: (1) identifying which spatial and semantic contexts to include along with their encoding techniques; and (2) designing GNN models capable of efficiently propagating information across a heterogeneous graph such as nodes representing manholes and roads—and learn features for network topology prediction.

3 Methodology

The overall framework is illustrated in Figure 1. The process begins with building a relational data model to organize information on utility anchor points, lines, and ground facilities and their spatial relationships. Second, all the records in the relational data model are represented as graphs, with anchor points and facilities as nodes, and utility lines and their relationships as edges. Finally, a GNN model is developed to predict utility lines, which are the links between anchor point nodes.



Figure 1. Overall Framework

3.1 Relational Data Model Construction

Geospatial relational data modeling is a crucial step to present the properties and the relationships among different entities. It not only facilitates data extraction from existing databases but also aids in building the graph representations of the utility anchor points, lines and ground facilities. Figure 2 depicts the Entity-Relationship (ER) diagram.



Figure 2. Entity-Relationship Diagram

In this diagram, three entities are used:

- 1. Utility Anchor Point: These are visible utility line junctions, such as manholes and ground pumps, indicating the locations of underground lines. Attributes include ID, type, and geometry.
- 2. Road: Roads, as a typical ground facility, provide spatial contextual cues for utility line prediction. The alignment of utility lines along roads makes this data a potential indicator. Additionally, roads are accessible from satellite imagery and digital

road maps, which are widely available. Roads are characterized by their ID, length, and geometry.

3. Utility Line. These are typically buried utility lines that are the focus of prediction in this research. Data on these lines are used for model training and for validation and testing in the evaluation stage. As this study focuses on predicting the existence of lines, only ID and geometry information are utilized.

Three relationships are established based on spatial relationship analysis:

- 1. Utility Line-Anchor Point Connection: The connection between utility lines and anchor points is determined by merging two tables through point-line intersection analysis.
- 2. Anchor Point-Road Proximity: Anchor point-road proximity is identified by locating the nearest line to the anchor point, considering only those within a 100-meters radius as "close." Additionally, three attributes are extracted: the position of the nearest point on the road, the distance from the anchor point to this nearest road point, and the side of the road on which the anchor point is located. These attributes aid in predicting utility line placement, as most lines run parallel to, rather than across, roads. For instance, two connected manholes are likely to be on the same side of the road and in proximity.
- 3. Road-Road Intersection: The road-road relationship is built by merging road tables through line-line intersection analysis.

3.2 Graph Representation

Building the graph representation of the utility network and its surroundings, based on the geospatial data model, involves three main steps: (1) establishing relationships between anchor points through their connections with utility lines; (2) converting the relational data model into a graph data model; (3) encoding the data with numerical values.

3.2.1 Anchor Point to Anchor Point Relationship Establishment

This step transforms the utility line entity into relationships between anchor points. It is designed to align with the objective of predicting utility lines, which will be modeled as the edges between anchor point nodes in the graph network. Typically, in the graph data model, edges represent relationships in the relational data model.

The implementation process is straightforward. A list of anchor-point ID pairs is generated if they intersect with the same utility line segments. This action removes the utility line entity in the relational data model and establishes a many-to-many relationship between the anchor points themselves.

3.2.2 Relational Data Model to Graph Data Model Conversion

This step follows the typical process of transforming the relational database to graph database, including the following steps: (1) table to node label; (2) row to node; (2) column to node property; (3) foreign key to edge; (4) relationship attributes to edge properties.

	ASSETID	SUBTYPECD			geonetry						
105060	MH198430	22201MAINT	POINT (1	52.948	58 - 27.44766)	PH.	index1	MH_in	dex2	line_in	idex
104560	MH198431	22201MAINT	POINT (1	52.948:	26 -27.44747)	1 1	05060.0	1045	60.0		2.0
	index_MH	index_Road	NEAR_POS	SIDE	dist_bins		road_	index1	road	_index2	
125555	104560	21457	0.443777	1.0	4	17467	0	21457		5791	
125558	105060	21457	0.385538	1.0	4	63160		21457		21066	
						64063		21457		21361	
	- DV	NOTE THEFE Ch								C10	

 21457
 Neighbourhood / local
 170.465693
 LINESTRING (494339.784 6963903.639, 494850.038.
 26733

 5791
 Neighbourhood / local
 43.593375
 LINESTRING (494339.784 6963903.639, 494850.038.
 26733

 21060
 Neighbourhood / local
 68.806724
 LINESTRING (494740.400 6963963.744, 494736.744, 494735.72
 26011

 21361
 Neighbourhood / local
 188.480334
 LINESTRING (494760.400 6963966.744, 494731.928.
 26014

(a) Relational Data









Figure 3 presents an example illustrating utility anchor points and roads, along with their relationships, in three different formats: as relational data model representation, as graph data model representation, and as visualized data on an ArcGIS map.

3.2.3 Data Encoding

Data encoding is a step to transform the data in different formats into the numerical features fed into the neural networks to predict the links between anchor point nodes. Table 1 summarizes the features used in this research, along with data encoding methods.

Node / Edge	Attribute Name	Encoding Methods	
Utility Anchor	Location	Location Encoding	
Point	Туре	One-Hot Encoding	
	Centroid Location	Location Encoding	
	Туре	One-Hot Encoding	
Road	Length	Equal-Frequency Binning and One- Hot Encoding	
	Orientation	Equal-Width Binning and One- Hot Encoding	
Utility Anchor	Relative Position of Nearest Point on Road	None	
Point-Road Relationship	Distance	Equal Frequency Binning and One- Hot Encoding	
	Side	None	

Table 1. Data Encoding Methods for Attributes

3.3 Utility Line Prediction using Graph Neural Networks

This research develops a GNN model that consists of two main components: convolutional layers, and a classifier. Initially, it adopts a multi-scale location encoder [14] that applies sinusoidal functions of varying frequencies to transform location data. The convolutional layers include the GAT (Graph Attention Network [16]) and GraphSAGE (SAmple and aggreGatE [17]) as basic building blocks. GAT incorporates an attention mechanism, assigning importance weights to neighboring nodes that are learnable within the network. It processes node features, edge indices (indicating node connections), and edge attributes as inputs and generates updated node features and attention weights as outputs. GraphSAGE is a method of sampling neighboring nodes with specific weights and aggregating these neighboring node features into the weighted target node. Both layers

focus on feature aggregation at the graph nodes. The final component of the network is a binary classifier, designed to predict connections between node pairs through the multiplication of their feature vectors. The loss function used is cross-entropy loss function, commonly applied in binary classification tasks. Figure 4 presents a detailed visualization of the GNN model, including its inputs, outputs, and overall architecture.



Figure 4. Architecture of the GNN model and corresponding Inputs and Outputs

3.3.1 Architecture Variants

Since there are no existing GNNs for this application, several architectural variants are discussed, as illustrated in Figure 5. ReLU layers are not drawn for simplification.



Figure 5. Architecture Variants

The base model utilizes two GAT layers, which include dropout rates to prevent overfitting. The outputs from these GAT layers, which are the updated features of the nodes, along with the indices of the edges, are fed into two GraphSAGE layers. The first variant consists solely of four GraphSAGE layers. Unlike GAT layers, GraphSAGE layers do not process edge attributes, meaning that this model variant does not include information on road-anchor point spatial relationships beyond connectivity. The second variant examines the impact of alternating positions of GAT and SAGE layers. The third variant investigates the effective integration of edge attributes by introducing skip connections, with FC referring to fully connected layers.

4 **Experimentation**

4.1 Data Description and Preprocessing

This research utilizes two data sources: (1) The sewer network map provided by Urban Utilities, accessible at <u>https://services3.arcgis.com/ocUCNI2h4moKOpKX/arc</u> <u>gis/rest/services/UU_Sewer_OpenData/FeatureServer</u>. In the ArcGIS sewer network map, the manhole and pump feature layers are utilized to represent utility anchor points, and the gravity sewer main lines are used as the utility lines. (2) The road network from the Brisbane City Council, which is available at <u>https://services2.arcgis.com/dEKgZETqwmDAh1rP/arc</u> <u>gis/rest/services/Roads_hierarchy_overlay_Road_hierar</u> <u>chy/FeatureServer</u>. The road feature layer is employed as an example of ground facilities.

The raw data about manholes, pumps, gravity sewer main lines, and road networks were exported from ArcGIS Pro software as individual shapefiles. Subsequently, these files were processed using Python geospatial data analysis and network analysis packages. The proximity analysis between manholes and roads was conducted using the QueryPointAndDistance function in ArcGIS Pro Python API. This function identifies the nearest point on a polyline to a given point and calculates the distance between them. Additionally, it provides details about which side of the line the point is located on and the distance along the line, expressed as a percentage. The data was preprocessed in two steps. First, the data was cleaned by removing utility lines that lack connections with any manhole or pump points or are linked to only one point. This is because the method assumes that each utility line connects to a minimum of two anchor points. Second, roads located more than 100 meters from the manholes were filtered out, as roads not classified as "close" to the manholes do not contribute to link prediction. The statistics are summarized in Table 2.

Table 2. Data statistics before and after pre-processing

Name	Count	Count
Ttuille	(Before)	(After)
Utility Line	243,773	203,203
Utility Anchor Point	206,187	206,187
Road	41,753	32,080

4.2 Experiment Design

4.2.1 Training, Validation, and Testing Data Split

The data was divided into training, validation, and testing sets in three steps: (1) within the utility anchor point networks (excluding roads), connected components were identified, leading to a collection of subgraphs, each representing a distinct component; (2) the training, validation, and testing datasets were then randomly distributed in a 6:2:2 ratio from these subgraphs. (3) nodes representing roads were included in various datasets, determined by their connectivity to utility anchor points. Some road nodes might appear in multiple datasets if they are connected to anchor points belonging to different sets. This separation ensures that utility line edges and utility anchor point nodes from the training set do not appear in the validation or testing sets, and those from the validation set are excluded from the testing set.

This approach of using connected components for dataset division was chosen because the distribution of unknown utility lines typically concentrates in specific areas rather than being evenly spread throughout a city. Figure 6 illustrates the distribution.



Figure 6. Training, Validation and Testing Sets

4.2.2 Evaluation Metrics

The model outputs are numerical values representing classes assigned to each edge that connects two manhole nodes: " " 1 indicates the presence of a pipeline, " " signifies its absence. A common threshold of 0.5 is used to separate these two classes. These studies adopt the following evaluation metrics for experiments: (1) Precision. Precision is the proportion of true positive predictions, correctly predicted pipeline presence, out of

all positive predictions made. (2) Recall. Recall is the proportion of true positives, correctly predicted pipeline presence, out of all actual pipeline presences. (3) AUC (Area Under the Curve) ROC (Receiver Operating Characteristics) score. The ROC curve is a plot of the TPR (True Positive Rate or Recall) against the FPR (False Positive Rate) at various threshold settings. It is the measure of separability of two classes. (4) F1-Score. F1-Score is the harmonic mean of precision and recall. (5) Accuracy. Accuracy is the total number of correct predictions among all the cases. (6) MCC (Matthews Correlation Coefficient). MCC is a correlation coefficient between the observed and predicted classifications for imbalanced dataset. It returns a value between -1 and 1, where 1 indicates a perfect prediction, 0 means random prediction, and -1 indicates total disagreement between prediction and observations.

4.2.3 Hyperparameter Tuning

For model training, the number of epochs is determined using an early stopping approach. The maximum size of the epochs is 100, but once the validation loss does not decrease or decreases by less than 0.001 for five consecutive epochs, the training process will be stopped to prevent overfitting. Regarding optimization techniques, the Adam optimizer is used with a learning rate set at 0.001.

model fine-tuning The focuses on two hyperparameters: the size of the hidden layers and the dropout rates. This approach is chosen due to an overfitting problem observed during the experimentation process. The options for the hidden layer size are set at four specific values: 32, 64, 128, and 256. The dropout rates range from 0 to 0.6, with increments of 0.2. The hyperparameter tuning process is guided by various evaluation metrics on the validation set, and the testing data is used only for evaluating the optimal model. Due to space constraints, the detailed evaluation metrics corresponding to each model variant and hyperparameter combination are stored in the GitHub repository.

4.3 Experiment Results

The hyperparameter tuning of the model is driven by its performance on various evaluation metrics using validation data. The best-performing models on the validation set for each metric are summarized in Table 2.

For more detail, Variant 3a represents a model configuration with a hidden layer size of 32 and a dropout rate of 0; Variant 3b is configured with a hidden layer size of 128 and a dropout rate of 0; and Variant 3c features a hidden layer size of 32 with a dropout rate of 0.4. Variant 1a, on the other hand, corresponds to a model with a hidden layer size of 64, utilizing only SAGE layers.

The fine-tuned models, each selected for achieving the highest score for each evaluation metrics, are further evaluated using the testing data. The outcomes from these tests are compiled and presented in Table 3.

Overall, Variants 1 and 3 demonstrate the most robust performance. Variant 3 excels in ROC AUC scores, accuracy, and MCC metrics, indicating its superior capability in differentiating the presence and absence of pipeline connections. On the other hand, Variant 1, which focuses solely on node attributes and connectivity and overlooks edge attributes such as the manhole's location relative to the road, achieves the highest recall and F1 score. This outcome is reasonable since ignoring roadcrossing pipelines leads to more conservative predictions. This conservative approach is particularly advantageous in utility line detection scenarios, where the priority is to minimize the risk of missing lines.

Table 3. Optimal Model Architectures and Hyperparameter Combinations on Validation Set

Model	Var.3a	Var. 3b	Var. 3c	Var. 1a
ROC AUC	0.9619	0.9616	0.9608	0.9572
F1	0.8987	0.8980	0.8959	0.9001
Precision	0.9137	0.9215	0.9265	0.8992
Recall	0.8842	0.8756	0.8672	0.9010
Accuracy	0.9004	0.9005	0.8992	0.9000
MCC	0.8011	0.8020	0.8001	0.8000

Table 4. Testing Results on the Tunned Models

Model	Var.3a	Var. 3b	Var. 3c	Var. 1a	
ROC AUC	0.9520	0.9524	0.9488	0.9479	
F1	0.8855	0.8849	0.8790	0.8868	
Precision	0.8783	0.8884	0.8890	0.8687	
Recall	0.8927	0.8815	0.8692	0.9057	
Accuracy	0.8845	0.8854	0.8803	0.8844	
MCC	0.7692	0.7708	0.7608	0.7695	

5 Conclusion and Discussion

This research presents an effective method for completing utility networks. The approach includes three steps: (1) build a relational data model to arrange the data regarding utility anchor points, lines, ground facilities, and their spatial relationships; (2) convert all records in the relational data model to graphs, with anchor points and facilities as nodes, and utility lines and their relationships as edges. (3) develop a GNN model to predict utility lines. The experimental results demonstrate good performance, achieving a 95.2% ROC AUC score in inferring sewer lines between manholes. This novel approach offers advantages for utility owners and excavation contractors, providing a framework to deduce missing connections within utility networks.

However, a limitation of the model is its lack of explainability, which impacts user trust. Furthermore, applying the model directly to varied datasets presents challenges due to the necessity for: (1) aligning context features with standardized utility network criteria, and (2) considering diverse practices that vary by time and geography. Ensuring model adaptability to different utility networks requires accurate, complete, and regionspecific utility network training data. Future research will focus on assessing the impact of data quality on model performance. Additionally, expanding the model to include more spatial contexts, such as buildings and legal boundaries, could further improve its utility and accuracy in real-world applications. Lastly, considering potential consequences of false alerts and missed detections in utility strike prevention and flexibilities in pipeline network design, presenting the likelihood with uncertainty could further improve decision making.

6 Data and Code Availability

The code, data, and supplemental materials are available in the GitHub repository: <u>https://github.com/Yuxi0048/PipeNetworkCompletion</u>.

References

- [1] Damage Root Causes Remain Consistent. https://dirt.commongroundalliance.com/2022-DIRT-Report/Damage-Root-Causes-Remain-Consistent#mainContentAnchor, Accessed: 24/12/2023
- [2] Afshar, M. H., Partially Constrained Ant Colony Optimization Algorithm for the Solution of Constrained Optimization Problems: Application to Storm Water Network Design, *Advances in Water Resources*, Vol. 30, No. 4, 2007, pp. 954–965. https://doi.org/10.1016/j.advwatres.2006.08.004
- [3] Izquierdo, J., Montalvo, I., Pérez, R., and Fuertes, V. S., Design Optimization of Wastewater Collection Networks by PSO, *Computers & Mathematics with Applications*, Vol. 56, No. 3, 2008, https://doi.org/10.1016/j.camwa.2008.02.007
- [4] Bailly, J. S., Levavasseur, F., and Lagacherie, P., A Spatial Stochastic Algorithm to Reconstruct Artificial Drainage Networks from Incomplete Network Delineations, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 13, No. 6, 2011, https://doi.org/10.1016/j.jag.2011.06.001
- [5] F. B z ç ı, C., . H , K., M -Weight Spanning Tree Algorithms A Survey and

 Empirical
 Study,
 Computers
 & Operations

 Research,
 Vol.
 28,
 No.
 8,
 2001,

 https://doi.org/10.1016/S0305-0548(00)00007-1

- [6] Navin, P. K., and Mathur, Y. P., Layout and Component Size Optimization of Sewer Network Using Spanning Tree and Modified PSO Algorithm, *Water Resources Management*, Vol. 30, No. 10, 2016, https://doi.org/10.1007/s11269-016-1378-7
- [7] Chahinian, N., Delenne, C., Commandré, B., Derras, M., Deruelle, L., and Bailly, J.S., Automatic Mapping of Urban Wastewater Networks Based on Manhole Cover Locations, *Computers, Environment and Urban Systems*, 2019.
- [8] LeCun, Y., Bengio, Y., and Hinton, G., Deep Learning, *Nature*, Vol. 521, No. 7553, 2015, https://doi.org/10.1038/nature14539
- [9] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L., Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions, *Journal of Big Data*, Vol. 8, No. 1, 2021, https://doi.org/10.1186/s40537-021-00444-8
- [10] Zhang, M., and Chen, Y., Link Prediction Based on Graph Neural Networks, Vol. 31, 2018. *Advances in neural information processing systems*, 31.
- [11] Hamilton, W. L., Ying, R., and Leskovec, J., Representation Learning on Graphs: Methods and Applications, 2017. arXiv preprint arXiv:1709.05584.
- [12] Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.
- [13] Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C., and Yu, J., Traffic Flow Prediction via Spatial Temporal Graph Neural Network, New York, NY, USA, 2020. https://doi.org/10.1145/3366423.3380186
- [14] Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., and Gao, F., PM2.5-GNN: A Domain Knowledge Enhanced Graph Neural Network for PM2.5 Forecasting, New York, NY, USA, 2020. https://doi.org/10.1145/3397536.3422208
- [15] Mai, G., Janowicz, K., Yan, B., Zhu, R., Cai, L., and Lao, N., Multi-scale representation learning for spatial feature distributions using grid cells, 2020. *ICLR 2020*.
- [16] V čk v ć, ., C , . , C v , . , , A., Liò, P., and Bengio, Y., Graph Attention Networks, 2018. *ICLR 2018*.
- [17] Hamilton, W. L., Ying, R., and Leskovec, J., Inductive Representation Learning on Large Graphs, 2017. Advances in neural information processing systems, 30.

ADAPTIVE GEOLOGICAL UNCERTAINTY MODELING IN EXCAVATION

Hannu Juola, Sara Johansson, Olof Friberg and Rauno Heikkilä

Oulu University, Faculty of Technology, Civil Engineering Research Unit, P.O.Box 4200, FI-90014 Oulu, Finland hannu.juola@oulu.fi, Sara.Johansson@tyrens.se, Olof.Friberg@tyrens.se, rauno.heikkila@oulu.fi

Abstract –

The accurate detection of soil boundary levels is vitally important for creating the correct solutions for the earthwork phases of construction. In Building Information Modeling (BIM), subsoil models are interpolated from information obtained from ground investigations. Our study focused on the adaptive use of geological uncertainty modeling to derive additional information during excavation. The results show that the continuous mapping of excavated ground is an effective estimation tool for modeling boundary levels. This adaptive tool can significantly decrease uncertainty in earthwork, presenting new possibilities for productivity and sustainability. The results of this study enable comparisons of alternative options in adaptive uncertainty modeling during excavation, allowing for the development of uncertainty-based subsoil models for construction.

Keywords -

Soil boundary level; uncertainties; excavation; geological mapping; building information modeling; model-based design; ground investigation

1 Introduction

As the development of construction technology is focusing on questions related to productivity, safety, and sustainability, many are facing the challenge of adopting progression with the prevailing conditions of practice. In response to this challenge, building information modeling (BIM) tools were introduced, and they have played a crucial role in coordinating the planning and construction phases, addressing unforeseen problems, and facilitating information sharing between disciplines. These interdependent interactions between various engineering disciplines are built on information that is initially available.

In many cases, the problems that arise during construction are a result of the limited or partial information available during the planning phase. Because unforeseen problems are solved onsite, they tend to be reactive rather than proactive. Although many construction phases are flexible in the face of emerging challenges, certain planning disciplines, such as geotechnical engineering, rely on preliminary information to develop solutions for the entire construction process. The reactive nature of problemsolving in such cases creates challenges in terms of not only timetabling and costs but also overall planning solutions.

Proactive modeling addresses this issue by providing a more accurate estimation of ground conditions, but it requires updated soil models. Until now, achieving this has involved additional soundings, soil sampling, or boreholes to create new modeling conditions, simultaneously incurring additional costs due to extra ground mapping.

A recent study by Satyanaga et al. [1], which reviewed the current state of BIM applications for geotechnical engineering, focused on the integration of geological and geotechnical information into modeling the planning and construction phases using case studies. Their findings correlate with our study's hypothesis that there is a lack of flexibility to create adaptive models for implementation when applying preliminary information with finite element method calculations. There is a need for further research to create less intricate and calculation-dependent modeling tools when applying geotechnical information. Vanicek et al. [2] concluded that BIM modeling is behind other engineering disciplines in terms of the effectuation of geological and geotechnical information. This has much to do with the site-specific properties of soil parameter determination. Constitutive models of ground conditions require soil sampling and sounding results to create a condition map of an entire area. Since soundings and soil samples represent point-specific information, there is an element of uncertainty in interpolating these points. Similar findings regarding soil parameter uncertainty in BIM modeling have found support in the studies of Beaufils et al. [3] and Wu et al. [4].

According to Wiegel et al. [5], the development of geotechnical modeling with geostatistical information and uncertainty modeling improves overall decisionmaking and sustainability in construction. They highlighted that tools for incorporating geotechnical information into BIM models are still under development. In addition, they pointed out that, although different methods of uncertainty modeling are available, the integration of these methods with BIM models is not ready. We found no study in which the uncertainty of subsoil models was updated with adaptive soil boundary detection during excavation.

Pakyuz-Charrier et al. [6] concluded that overlooking the amount of uncertainty distorts the accuracy of the modeled area. The handling of uncertainties in geological information can either enhance or distort a 3D subsoil model, as pointed out in studies between 2001 and 2018 [7], [8], [9], [10], [11]. All of these studies were founded on the understanding that information fed into the uncertainty model is gathered beforehand and interpolated to create the modeled area. As preliminary information, the sounding, sampling, and borehole information has only limited room for flexibility during construction.

The risks of utilizing subsoil models with partial or inaccurate information can, in many cases, actualize economic loss. Estimations of cut and fill volumes, for example, can significantly impact tendering bids, transportation estimates, contract schedules, and emission calculations. This was the case in the Autostrada1 motorway construction project, with Skanska Poland estimating that each centimeter of excess excavation at depth would incur costs of EUR 1 million [12]. In this way, uncertainty modeling, together with the automatic implementation of geotechnical and geological BIM models, can prevent major errors in calculations and analyses [13].

Developments in construction safety and productivity have had an increasing impact on the digitalization of infrastructure construction. Automation has allowed worksite information management to incorporate BIM modeling. For this, accurate subsoil models are crucial (Fig. 1), and with the state of the art of geotechnical modeling today, they require intermediary software and investment in data gathering and storage, according to Hiltunen et al. [14].



Fig. 1 BIM data and subsoil models create the basis for automated excavation in earthwork construction.

The aim of our study is to investigate the potential of adaptive geological information derived during excavation by studying the effects of incorporating new information into a geological uncertainty model and then creating a verified soil boundary and geological model for an entire excavated area.

Based on the main aim, the following questions are addressed by our study:

- 1. How does the uncertainty of a geological model develop when excavation is conducted
 - a) systematically from one end of an uncertainty model to another
 - b) from the most uncertain area of the model outwards?
- 2. What is the potential for adaptive boundary detection in geological modeling?
- 3. What are adaptations of uncertainty modeling with updated geological information in the future?

2 Materials and methods

2.1 Uncertainty modeling

The method used in our study for modeling uncertainties in the determination of geological boundaries was based on kriging interpolation and Bayesian statistics [15]. This approach was theoretically inspired by the 3D geological uncertainty modeling of Wellmann et al. [16], although it was adapted to typical geotechnical conditions in infrastructure projects. In such projects, relatively few geological boundaries are typically expected, and geological conditions are only of interest within a limited area below the ground surface. For computational efficiency, the uncertainty modeling code was developed in 2D, with the resulting surface representing a specific geological boundary of interest. The uncertainty model used in this study was implemented in Sweden by Tyréns AB as part of its GeoBIM concept, with the company providing a direct connection to a geotechnical database. Uncertainty modeling was performed using Python code running on a server and controlled by the user through a Microsoft Teams interface.

The user was asked to evaluate the most probable vertical location of the geological boundary at each measurement point (e.g., geotechnical drilling and soundings). The minimum and maximum possible locations of the same boundaries were evaluated by the user, and these input data were used to construct a skewnormal distribution for each measurement point. These distributions represented all possible boundary-level values in the measurement points and their corresponding likelihoods. The uncertainty modeling was then, in short, based on an iterative approach, in which random boundary-level values were drawn from the uncertainty distributions of Svensson et al. [17]. This was followed by kriging interpolation of the boundary surface. Since the input boundary levels varied for each iteration, the approach resulted in an estimation of how uncertainty in the determination of the boundary level propagated from individual measurement points to the surface model as a whole.

In addition to this propagated uncertainty, the variance of the kriging interpolation, which is linearly proportional to the distance to the measurement points, was also taken into consideration. Both types of uncertainties (i.e., spatial variance and propagated uncertainty) were weighted together for the final quantification of the total measurement uncertainty of the modelled surface. In addition, the model output is also the most probable bedrock level as well as the minimum and maximum bedrock levels.

2.2 Study area and data collection

In an earlier study, Svensson and Friberg [18] successfully validated the algorithm used in the uncertainty model implemented at Tyréns AB to study the top boundary of bedrock in three separate infrastructure projects in Sweden. From these infrastructure projects, five uncovered and surveyed areas were mapped and studied alongside the uncertainty model created using preliminary information derived from Swedish JB-sounding (Soil-Rock soundings) investigation points.

Svensson and Friberg [18] produced a one-step verification system in which the preciseness of the uncertainty model was verified according to the results of the uncovered bedrock boundaries. The uncertainty model was not updated during the excavation of the bedrock boundary.

Using a three-step process, our study verified the development of the uncertainty model when new information was introduced into the algorithm during excavation. First, the estimated soil boundary that had been updated during excavation was verified with additional sounding points not used in the initial model. This produced additional data points for the algorithm with the same estimated uncertainty as the initial points. Second, the updated model was verified with soil samples, which produced additional data points with minimal estimated uncertainty. Third, the true soil boundary was mapped from the excavated ground.

This three-step method enabled us to verify the development of uncertainty with adaptive information derived during excavation using the prevailing conditions of practice. This broad adaption of data points enabled the generalization of the adaptive uncertainty modeling, since it was not dependent on certain informationgathering methods while the ground conditions were mapped.

Our study area site description represents a typical Scandinavian friction subsoil consisting of an aggregate sandy gravel layer on top of a moraine base layer. The sandy gravel layer's depth varied from 1.00 m to 13.48 m below the ground. On average, the depth was 12.40–12.60 m (Fig. 1). This site was chosen because of the varying depth of the soil boundary, which created an uncertainty variation around the site area (Fig. 2).

The geological uncertainty model was created for a 50 m \times 75 m area. The initial information for the uncertainty model was taken from nine measurement points. The soil boundary level was measured by static-dynamic penetration tests, producing an estimated soil boundary depth with a 20 cm difference in the evaluations of the minimum and maximum possible locations of the same boundaries. The actual boundary depth from these points was verified by soil sampling.

The test site was then divided into a 5 m \times 5 m mesh grid, from which the verified soil boundary level of each grid square was introduced into the uncertainty model. The true soil boundary level was verified from surface model measurements of the excavation pit floor, additional soil sampling, and static-dynamic penetration test results not used in the initial uncertainty model.



Fig. 1 Verified uncertainty model of the test site's sandy moraine layer.



Fig. 2 Initial geological uncertainty condition map of the test site. Uncertainty increases according to the shade of red. The initial measurement points are visualized with black and white "T" symbols.

2.3 Data processing and analysis

The data collected from the surface model measurements of the excavation pit floor, additional soil sampling, and static-dynamic penetration test results were processed into the uncertainty model to create two separate scenarios. These scenarios were simulated to study the adaptive properties of the geological uncertainty model—that is, how the algorithm adjusted the estimate of soil boundary levels—when the calculation parameters were refined.

Because, in most cases, soil boundaries vary in height, the uncertainty model's development is dependent on the weighted distance in which the data points are defined in contrast to one another. The longer the distance, the greater the uncertainty (traditional standard procedure). In our study, we compared the development of the uncertainty model when the distance was largest between the data points (i.e., the uncertainty was adjusted at the beginning of the excavation) with a case in which excavation took place with a traditional one-end-toanother approach.

Using the aforementioned three-step verification system, additional sounding data points and soil sampling were introduced as alternative sources of adaptive information. The estimated soil boundary uncertainty between the minimum and maximum possible locations was kept steady at 20 cm throughout the entire area.

In the first scenario, the uncertainty modeled area was systematically excavated from the western boundary toward the east (Fig. 3). In the second scenario, the excavation was modeled from the most uncertain area outward (Fig. 4). The development of the uncertainty model was reviewed in 20% increments as an estimation of the volume of sandy gravel layers.



Fig. 3 Development of the uncertainty model in Scenario 1, with excavation boundary levels in 20% increments.



Fig. 4 Development of the uncertainty model in Scenario 2, with excavation boundary levels in 20% increments.

3 Results

3.1 Systematic excavation from west to east

The results show that the uncertainty of the model decreased methodically as the excavation proceeded toward the eastern end of the test site (Fig. 5). The most probable level of volume estimated for the sandy moraine layer decreased below the end result as the excavation advanced from 40% to 60%. This was due to the most uncertain area of the site being located at the approximate center of the excavation. As the excavation modeled the correct level for the uncertain area, with 60% of the excavation completed, the estimation corrected itself back to the original trend of the development.

Even with largely varying uncertainty around the test site, with 60% of the excavation completed, the difference in the volume-level uncertainty was 29.46% for the completed excavation (Table 1). The uncertainty in the volume level, with the initial measurement point data alone, had a difference of 82.58% compared to the completed excavation.

Model	Upper level (m ³)	Most probable level (m ³)	Lower level (m ³)	Difference in volume-level uncertainty (%)	Model	Upper level (m ³)
0%	24,274	29,080	33,929	82.58%	0%	24,274
20%	24,909	29,178	33,473	72.39%	20%	26,523
40%	27,271	30,646	34,023	50.91%	40%	27,252
60%	26,392	29,088	31,790	29.46%	60%	27,854
80%	28,183	30,411	32641	10.53%	80%	28,275
100 %	28.282	30.286	32.294	0%	100%	28.282

Table 1. Results of uncertainty during systematicexcavation from west to east.

Table 2. Results of uncertainty in the excavation	from
the most uncertain area outward.	

Lower

level

 (m^3)

33,929

33,473

32,992

32,518

Most

probable

level (m³)

29,080

29.749

29,901

30,185

Difference in

volume-level

uncertainty (%)

82.58%

53.60%

35.44%

15.03%



Fig. 5 Development of the uncertainty model as excavation systematically proceeded from west to east.

3.2 Excavation from the most uncertain area outward

When initially mapping the most uncertain parts of the excavation area, and extending the excavation outward from there, the decrease in uncertainty was more effective and linear. The largest development in the uncertainty model, excavating with this approach, occurred in the first 20% of the excavation (Fig. 6). Here, uncertainty decreased by 35.09% (Table 2).

As the excavation progressed, the difference in volume-level uncertainty was 15.03% compared to the completed excavation, when 60% of the test site had been mapped. Compared to the systematic excavation from west to east, where the difference at the same percentage increment was 29.46%, the percentage difference between these two scenarios was 64.86%.



Fig. 6 Development of the uncertainty model as excavation proceeded from the most uncertain area outward.

4 Discussion

4.1 Potential and limitations of adaptive boundary detection in geological modeling

The use of adaptive boundary detection reduces the level of uncertainty when interpolating between measurement points and mapping boundary levels. In the future, this could have a large effect on the planning phases of geotechnical solutions, excavation, and transport. As the expected volumes of excavated soil become more precise, the planning of cut and fill volumes and possible aggregate consumption and emissions calculations will become more efficient and effective. The detection of soil boundary depths will allow for more precise geotechnical calculations in earthwork solutions, such as stabilization, preloading, and settlement, and for a more proactive approach to be taken to unforeseen problem-solving. The need for additional ground investigations after the initial planning can be reduced to a necessary minimum in cases where the detection of boundary levels during excavation is sufficient. This can help in cost- and timetable-effective construction, as ground condition mapping is one of the most costly phases of planning. These benefits are vital for contractors and clients to ensure that projects proceed according to the expectations of the tendering proposal. An algorithm continuously updated with the best estimate (± 2 std. dev., 95% uncertainty) [18] of boundary levels would be the best possible tool to achieve this goal.

Adaptive uncertainty models can potentially lead to precise, boundary-based autonomous excavation, with the excavation models dictating how the excavation adapts to changes in soil boundaries while in progress. This can create more dynamic uses for unmanned excavation through BIM–model interaction.

As the computational requirements of autonomous excavation do not seamlessly interact with the prevailing conditions of practice in excavation, a middle-ground solution utilizing modeling tools that can be incorporated in near real time into machine-control models can help with the interaction. Our proposed adaptive uncertainty model can create a tool that enables an autonomous excavator to excavate with precision, where earlier machine control models were more robust in their definitions. The uncertainty model could provide additional tools for contractors to create more precise and effective tendering offers while using traditional means of excavation.

As a tool, uncertainty modeling does not exclude the need for thorough ground investigations. Since the effectiveness of the model relies on the initial information fed into the calculations from measurement points, the ground investigation planning should be as detailed as possible. The initial uncertainty model provides vital information on the mapped area and, as our results show, allows for variable results depending on the execution of the excavation.

The evaluation of the minimum and maximum possible locations of boundaries plays an important role in total uncertainty. In our case, when the location difference was set at 20 cm, the degree of compaction in the soil played an important role. Due to the degree of compaction, static-dynamic penetration produced results per 20 cm increment. This is not the most precise sounding method available, and by choosing the most accurate method, depending on the soil type, the uncertainty can be minimized.

4.2 Adaptive geological uncertainty modeling in the future

BIM applications for geotechnical engineering are under widespread development, although one obstacle to their development concerns differences in the conditions of practice. To address this, uncertainty modeling allows individual uncertainties to be set for each investigation location, depending on the quality of the data to be obtained through various methods.

The same aspect of flexibility can be capitalized on in adaptive modeling since there are multiple methods for boundary detection. The integration of uncertainty modeling tools into BIM applications should be conducted with the same type of principles. To implement uncertainty modeling in everyday work, it is necessary to make the tool available without the need to install additional programs, learn new interfaces, and manage new data formats. An example of this is the uncertainty modeling tool used in our study.

This modeling method was incorporated into the Microsoft 365 Teams platform, which many users already know and can access. By making uncertainty modeling available on a platform that is already widely used in the industry, the threshold for starting to use the method was significantly reduced.

5 Conclusion

This study introduced an adaptive utilization method for geological uncertainty modeling. The results derived from the test site data indicate that adaptive updates of ground mapping information can significantly reduce the uncertainty of earthwork. Since the information utilized can be sourced from multiple information sources (i.e., the mapping of excavated ground), the method can be widely adapted to the prevailing conditions of practice. The main benefits of adaptive uncertainty modeling are the creation of precise and effective information during earthwork phases without additional work phases or significant costs.

At present, uncertainty modeling tools are not in widespread use. Further development of the algorithm depends on operational experience derived during field tests. Because the algorithm for calculating soil boundaries and the accompanying uncertainties depend on the individual uncertainty of each data point, the amount of information gathered from various test sites, which is effectively a ground condition database, can significantly advance the algorithm's development.

Such a database can help determine how various geological parameters and the distances between data points are weighted, depending on how much they should influence the calculation, thereby advancing the calculation of site-specific conditions by converging the predicted excavated soil boundaries with the individual uncertainties in each data point.

Acknowledgments

We thank Dr. Mats Svensson and his team at Tyréns AB for allowing us to study the adaptive properties of their geological uncertainty model [17]. This study was

funded by Business Finland (grant 253/31/2022) as part of the SWARM project. The text reflects only the authors' views, and the agency is not responsible for any use that may be made of the information it contains.

References

- Satyanaga, A., Aventian, G. D., Makenova, Y., Zhakiyeva, A., Kamaliyeva, Z., Moon, S. W., & Kim, J. (2023). Building Information Modelling for Application in Geotechnical Engineering. Infrastructures, 8(6), 103.
- [2] Vanícek, I., Jirásko, D., & Vanícek, M. (2021). Role of Geotechnical Engineering in BIM process modelling. IOP Conf. Ser. Earth Environ. Sci, 727, 1755-1315.
- [3] Beaufils, M., Grellet, S., Le Hello, B., Lorentz, J., Beaudouin, M., & Castro Moreno, J. (2020). Geotechnical data standardization and management to support BIM for underground infrastructures and tunnels. Tunnels and Underground Cities: Engineering and Innovation meet Archaeology, Architecture and Art, 655-664.
- [4] Wu, J., Chen, J., Chen, G., Wu, Z., Zhong, Y., Chen, B., ... & Huang, J. (2021). Development of data integration and sharing for geotechnical engineering information modeling based on IFC. Advances in Civil Engineering, 2021, 1-15.
- [5] Wiegel, A., Peña-Olarte, A. A., & Cudmani, R. (2023). Perspectives of 3D Probabilistic Subsoil Modeling for BIM. Geotechnics, 3(4), 1069-1084.
- [6] Pakyuz-Charrier, E., Giraud, J., Ogarko, V., Lindsay, M., & Jessell, M. (2018). Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo. Tectonophysics, 747, 16-39.
- [7] Bárdossy, G., & Fodor, J. (2001). Traditional and new ways to handle uncertainty in geology. Natural Resources Research, 10, 179-187.
- [8] Aug, C., Chilès, J. P., Courrioux, G., & Lajaunie, C. (2005). 3D geological modelling and uncertainty: The potential-field method. Geostatistics Banff 2004, 145-154.
- [9] Lindsay, M. D., Aillères, L., Jessell, M. W., de Kemp, E. A., & Betts, P. G. (2012). Locating and quantifying geological uncertainty in threedimensional models: Analysis of the Gippsland Basin, southeastern Australia. Tectonophysics, 546, 10-27.
- [10] Wellmann, J. F., & Regenauer-Lieb, K. (2012). Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. Tectonophysics, 526, 207-216.
- [11] Pakyuz-Charrier, E. J., Lindsay, M. D., Ogarko, V., Giraud, J. E. C. D., & Jessell, M. W. (2018). Monte

Carlo simulations for uncertainty estimation in 3D geological modeling, a guide for disturbance distribution selection and parameterization. Solid Earth, 9, 385-402.

- [12] Heikkilä, R., & Jaakkola, M. (2006, October). Automation of road construction-the state of the art in Europe. In 23th International Symposium on Automation and Robotics in Construction. ISARC.
- [13] Shi, C., Jin, Y., Lu, H., & Shi, J. (2023). A BIMbased framework for automatic numerical modelling and geotechnical analysis of a large-scale deep excavation for transportation infrastructures. Intelligent Transportation Infrastructure, 2, liad012.
- [14] Hiltunen, M., Heikkilä, R., Niskanen, I., & Immonen, M. (2023). Open InfraBIM for remote and autonomous excavation. Automation in Construction, 156, 105148.
- [15] Prästings A., 209, Managing uncertainties in geotechnical parameters: From the perspective of Eurocode 7, PhD thesis, KTH
- [16] Wellmann, J. F., Horowitz, F. G., Schill, E. and Regenauer-Lieb, K. (2010). Towards incorporating uncertainty of structural data in 3D geological inversion, Tectonophysics. Elsevier B.V., 490(3–4), pp. 141–151.
- [17] Svensson M., et al, 2022, Uncertainty models for optimal use of geo related resources in infrastructure planning, BeFo report 232, in Swedish, pp 96
- [18] Svensson, M., & Olof Friberg. (2022). GeoBIM for handling geological and geotechnical uncertainties in tunnelling. In WTC2022 and 47th General Assembly. Châtelaine; ITA-AITES.

Planning of Formwork Accessories in a BIM-based Aluminum Formwork Layout Planning System

Guan-Yong Xiong¹, Tzong-Hann Wu¹, Shang-Hsien Hsieh¹, Kuan-Yi Chen¹, Budy Setiawan², and Wen-Tung Chang²

¹Department of Civil Engineering, National Taiwan University, Taiwan ²FBC Formworks System Co., Ltd., Taiwan percy948@gmail.com, tzonghannwu@ntu.edu.tw, shhsieh@ntu.edu.tw, ian890711@gmail.com, huangwenren88@gmail.com, tonylook@ms52.hinet.net

Abstract –

Formwork accessory planning plays an important role in fastening and supporting the formwork panels for construction safety. This research aims to improve the automation of formwork accessory planning of a BIM-based aluminum formwork layout planning system currently developed at the National Taiwan University for automating the layout planning of formwork panels. It has been proven to save engineers time in conducting formwork panel planning in several case studies. However, the engineers still need to spend a lot of time conducting formwork accessories planning manually in the system. Therefore, this research proposes an approach based on accessories planning rules to support automation for engineers in conducting prop generation, brace planning, waler planning, double waler planning, and quantity estimation of the accessories including waler bracket, pull tie, wedge, and pin. The case study result shows that the proposed approach can save engineers 87.5% of the time compared to the original process of conducting formwork accessory planning.

Keywords -

Planning Automation; Formwork Accessories; Aluminum Formwork Planning System; Building Information Modeling

1 Introduction

The aluminum formwork system represents a significant advancement in construction technology and offers many benefits compared to the traditional timber formwork system. Notably, the stiffness of the aluminum formwork system surpasses that of the timber formwork system, providing a more robust and stable structure [1]. Moreover, the aluminum formwork system has longer durability, enabling multiple reuses and reducing the overall cost of construction [2]. Easily installing and dismantling aluminum formworks also contributes to

time efficiency, a critical factor in construction sites. Furthermore, the smooth surface of the aluminum formwork system often results in superior surface finishes on concrete structures, requiring no additional surface treatment. This advantage enhances the aesthetic and functional quality of the final product.

fully use the advantages То above, the implementation of the aluminum formwork system requires precise pre-planning, which is a time-consuming process for engineers. This formwork planning process includes panel planning and accessory planning. This research is focused on improving the automation of formwork accessory planning. The accessories play a crucial role in supporting and stabilizing the formwork, preventing significant displacement or even collapse during the concrete pouring process. Accessory planning must consider the type, size, and quantity of accessories, as well as their configuration. The primary objective of accessory planning is to ensure the stability and safety of the formwork system while optimizing construction efficiency.

However, few researches aimed at automating the accessory planning process. Furthermore, the prevailing practice in formwork accessory planning largely remains manual endeavour, with engineers dedicating considerable time and resources to the task. This reflects a notable gap in automation, a fact underscored by the industry's reliance on traditional methods and the corresponding absence of advanced technological integration in this area. This lack of automation in the accessory planning process motivated the authors to propose a semi-automatic approach for the configuration and quantity estimation of accessories after the panel planning process. This approach is implemented in a BIM-based aluminum formwork layout planning system [3]. The proposed semi-automatic approach includes prop generation, brace planning, waler planning, double waler planning, and quantity estimation of the accessories including waler bracket, pull tie, wedge, and pin.

This paper presents the literature review in Section 2, the detail of the proposed approach in Section 3, the result and discussion from a case study in Section 4, and the conclusions in Section 5.

2 Literature Review

The authors reviewed research related to the advancements in formwork planning automation, specifically focusing on the integration of BIM technologies. Hyun et al. [4] initiated the automation in formwork design using spatial analysis of BIM. Lee et al. [5] developed prototype software for automating 3D layout planning of formwork using BIM data. Chen et al. [3] introduced a BIM-based approach that combines a semi-automated generation module with a manual refinement module. These studies contribute to automating formwork planning, particularly in panel planning.

From the literature review, the authors noted that current research on formwork planning automation primarily emphasizes panel planning rather than accessory planning. However, formwork accessory planning also takes considerable time. Therefore, this research aims to automate accessory planning to address this gap. The following section outlines the approach proposed for automating accessory planning.

3 The Proposed Approach

The authors interviewed formwork accessory planning engineers to realize their planning principles and rules. Then the authors aimed to obtain the required information for planning automatically from a BIM model and manual inputs. Finally proposed a semiautomatic formwork accessory planning approach. The proposed approach consists of two parts: accessory planning and accessory quantity estimation. The first part focuses on arranging various accessories integral to the formwork system, such as props, braces, walers, and double walers. Figure 1 illustrates the placement of the accessories. The second part estimates the quantity of wedges, pins, pull ties, and waler brackets. The arrangement aims to provide clear guidance to construction workers on accessories placement for construction efficiency. Quantity estimation serves to accurately determine the quantity of accessories needed, preventing extra transportation and management costs from secondary production. Also, it minimizes waste resulting from overestimation.



Figure 1: The placement of accessories.

To appropriately place accessories, we extract BIM parameters such as dimensions, locations, and orientations from formwork or structural components as the basis for our analysis. For example, to position a waler on a panel, we first need to obtain the panel's dimension and location information to determine the waler's placement and size.

Figure 2 outlines a sequential process for accessory planning. Double waler planning occurs after waler placement, as the positioning of double walers depends on the walers' locations. After positioning walers and double walers, we calculate waler bracket numbers and place braces, ensuring large braces align with waler positions. Consequently, we perform pin and waler bracket estimations, paving the way for the calculation of the wedge requirements. This systematic approach ensures each component is accurately placed and quantified, maintaining the integrity of the formwork system.



Figure 2: The flowchart for the proposed approach.

3.1 Accessory Planning

proposed The approach leverages Building Information Modeling (BIM) technology to automate the planning of formwork accessories, including props, braces, walers, and double walers. By extracting and analyzing parameters from BIM components, such as location, dimensions, and elevation, the system plans the positioning and estimates the quantities of the necessary accessories. This process ensures that all elements are appropriately arranged and quantified based on precise data extracted from the BIM model, minimizing manual intervention and enhancing construction efficiency. The subsequent sections will delve into the specifics of how the approach utilizes BIM parameters for comprehensive accessory planning and quantity estimation, showcasing the method's reliance on accurate BIM data to inform and streamline the planning process.

3.1.1 Prop Generation

A prop is a vertically adjustable support. It would be placed beneath a panel called PB (Beam Prop Head). The prop generation of this research builds upon a previous study [3], which only focuses on the generation of props. In this research, we further calculate props' heights and categorize them into groups based on their application purposes. Some props are designed to directly support the floor, while others are positioned on the windowsill, as shown in Figure 3. The actual height requirement for a prop is determined by subtracting the elevation of the underlying object from the PB elevation. Basically, props are entirely dependent on the PB and remain fixed in their positions during accessory planning. Therefore, we place props first in the planning process, ensuring they serve as references for conflict checks. When other accessories are introduced, they must be carefully placed to avoid any conflicts with the props.



Figure 3: Props configuration.

3.1.2 Waler Planning

A waler is a square and elongated tube placed across several templates and secured by a waler bracket. Its purpose is to ensure the linearity and smoothness of the wall template after concrete pouring. The waler is configured on the wall formwork and the column formwork is located inside the building.

The planning rules of the waler are as follows:

- Adjacent walers need to be staggered and overlapped with a specific length determined by engineers.
- Walers come in standard sizes, e.g. L1, L2, L3, etc.

When planning the waler, the user needs to select all the formworks to be configured, and the software system will first group the formworks. Formworks that are connected along a horizontal line are grouped. Then, the waler is configured according to the total length of the formwork.

There are two types of formwork combinations. The first type is a panel combination in which the total length is greater than the longest standard length of a waler. In this situation, the longest possible waler is given priority in placement. The placement of the walers is staggered at two different heights. Moreover, the overlap of the two staggered walers needs to be larger than the specific length decided by engineers, as shown in Figure 4.



Figure 4: Walers configuration.

The second type is the panel combination less than the longest standard length and only one appropriate length of waler needs to be placed. There are three situations according to the corner conditions at both ends of the formwork combination, as shown in Figure 5. The corner conditions are divided into internal corners and external corners. The first situation is that both ends of the formwork combination are internal corners. In this case, if the length of the panel combination is not equal to the standard size of the waler, the waler needs to be cut to fit between the two internal corners. The second situation is that both ends of the panel combination are external corners. In this case, an appropriate length (the minimum standard size of the waler which is greater than the panel combination length) of the waler needs to be placed in the middle of the formwork combination, and there is no need to adjust the length of the waler. The last one is that the two ends are an internal corner and an external corner, respectively. In this case, a standard-size waler is used, but the waler must be aligned with the internal corner.



Figure 5: Three situations of walers. (1) Two interior corners. (2) Two exterior corners. (3) One interior corner and one exterior corner.

3.1.3 Double Waler Planning

Because exterior columns are difficult to secure with braces, a stronger double waler is used to maintain the stability of the column formwork. The example of double walers planning is shown in Figure 6.

The planning rules for double waler are as follows:

- To ensure the verticality of two horizontally adjacent panels at the corner where the column touches the wall panels, two double walers need to be welded into an L-shaped double waler combination.
- Double walers come in standard sizes, e.g. L₁, L₂, L₃, etc.

When configuring double walers, users only need to select all the formworks that need double walers and the software system would automatically place the L-shaped double waler at the interior corner of the external column and the wall, followed by prioritizing the configuration of the standard-size double walers on the remaining surfaces before the placement of non-standard-size double walers.



Figure 6: Double walers planning.

3.1.4 Brace Planning

A brace is used to control the verticality of a panel and the ground, and the panels on both sides of the same wall are tied by pull ties. Therefore, for a wall, only one side of the wall needs to be configured with a brace. The same applies to columns and only two adjacent faces need to be configured with braces. There are small braces and large braces. Small braces are placed in the gap between panels. Large braces are attached to the waler and connect the waler to the floor. Because the configuration rules of braces in contact with the column differ from those of the wall formworks, their planning rules for configuration are different.

The planning rules for braces for wall panels are as follows:

- The engineers should determine the specific distances (d₁, d₂, d₃) before planning. The maximum distance between small braces is d₁. The maximum distance between the endpoint and the large brace is d₂. The maximum distance between large braces is d₃.
- A small brace must be placed at both ends of each formwork combination, and the distance between each small brace needs to be less than d₁.
- A large brace needs to be placed less than d₂ from each end of each formwork combination, and the distance between each large brace in between needs to be less than d₃.
- There needs to be a floor near the wall to make sure the braces have some space to be fixed to the floor.
- A large brace needs to connect a waler to the ground, so the height of the large brace needs to be adjusted according to the height of the waler.

When configuring braces on a wall panel combination, users can choose the panels they want to configure with braces. The software system first configures small braces, chooses one end to start configuring, and places a small brace. Then, it places braces at intervals of d₁ until the remaining length is less than d₁. Then, it places a small brace at the other end. Next, it configures large braces, chooses one end to start configuring, and places a large brace d₂ from the endpoint. Then it places one brace at intervals of d₃ until the remaining length is less than d₃, Finally, it places a large brace d_2 from the other endpoint. Furthermore, the software system can automatically detect the waler nearby and adjust the large brace's height according to the height of the detected waler. An example of brace planning of wall panel combination is shown in Figure 7.



Figure 7: Brace planning of wall panel combination.

The planning rules for braces for column panels are as follows:

- The engineers should determine the specific length L for structural safety consideration before planning. If the total length of the column panel combination is larger than L, three small braces and two large braces need to be placed. If it is less than L, two small braces and one large brace need to be placed.
- There needs to be a floor near the wall to make sure the braces have some space to be fixed to the floor.
- The large brace needs to connect the waler to the ground, so the height of the large brace needs to be adjusted according to the waler above.

When configuring braces on column panel combination, users can choose the panels they want to configure with braces. When the total length is more than L, the software system places three small braces and two large braces at appropriate locations. Examples are shown in Figure 8. Furthermore, the software system can detect the waler nearby and adjust the large brace height according to the height of the detected waler.



Figure 8: Brace planning of column panel combination.

3.2 Accessory Quantity Estimation

In addition to arranging props, braces, and walers, we also need to estimate the quantity of wedges, pins, pull ties, and waler brackets. Since these accessories are numerous and their positions are fixed, we do not need to plan their arrangement positions but only need to estimate their quantity as accurately as possible.

3.2.1 Pin Quantity Estimation

A pin is used for connecting and holding two formwork panels together by passing through the holes of two adjacent formwork panels. It is used with a wedge inserted into the hole at one end of the pin to achieve fixation. All the holes of formwork panels are fitted with pins. Therefore, when estimating the quantity of pins, we can sum up the quantity of all the holes on the formwork panels and divide it by two.

3.2.2 Waler Bracket Quantity Estimation

A waler bracket is an accessory used to place walers and double walers. It has a structure like a pin and allows the entire waler bracket to be fixed on the hole of the formwork panels. The waler bracket is placed between the formwork panels. All positions with walers and double walers need to be fitted with waler brackets. Different lengths of waler and double walers require different quantities of waler brackets. When estimating the quantity of waler brackets, we need to multiply the quantity of waler of each size by the corresponding quantity of waler brackets and sum up the multiplications.

3.2.3 Wedge Quantity Estimation

A wedge is an accessory used to fix the pin and the waler bracket on the hole. All positions with pins or waler brackets need to be fixed by wedges. When estimating the quantity of wedges, we only need to sum the quantity of waler brackets and pins.

3.2.4 Pull Tie Quantity Estimation

A pull tie is a thin steel sheet. Its length is determined by the thickness of the wall or column. During concrete pouring, it tightly pulls the panels on the opposite sides together. This prevents the panels from expanding or bursting, ensuring the quality of the construction. To estimate the quantity of pull ties, we can count the number of pull tie holes of formwork panels. Also, through the thickness of the wall and column, we can know the type of pull tie used.

4 Case Study

Figure 9 shows the panel planning result from a previous study [3]. We utilized this real case to evaluate the efficiency of the proposed approach by recording the time engineers used during the planning process and comparing it with manual operations. The operation time for different accessory planning approaches is listed in Table 1.



Figure 9: A real case [3].

4.1 Evaluation

In the traditional manual process, engineers spend 400 minutes reviewing accessory placement rules and manually positioning BIM accessory components in the model, ensuring each piece aligns with structural requirements and project specifications. A preliminary automatic approach [3], which only supports the generation of prop, takes 365 minutes. It only saves 8.75% of the time. Our semi-automatic approach supports (1) generation of prop, (2) planning of brace, waler, and double waler, (3) quantity estimation of the abovementioned accessories, and (4) quantity estimation of the pin, waler bracket, wedge, and pull tie. With the proposed semi-automatic approach, the time is significantly reduced to only 50 minutes. This time includes the manual adjustment of conflicts between different accessories. The results indicate a timesaving of 87.5%

for engineers, a significant improvement in efficiency. Furthermore, our approach allows for more precise planning of aluminum formwork accessories in a semiautomatic manner. This precision is crucial in ensuring the quality and safety of the construction process. It also reduces the risk of errors that could potentially lead to costly and time-consuming corrections.

4.2 Discussion

The evaluation primarily measures the efficiency of various planning approaches through time spent. Engineers dedicated most of the 50 minutes saved by our semi-automatic method to adjusting the placement of double walers, with the remainder used for verification. Enhancing the automation level of double waler planning could further reduce the time required.

Although time serves as our main evaluation metric, quality and error rates are equally critical. Future assessments will quantify these aspects to provide a more comprehensive evaluation of our system. Additionally, comparing the results in different scenes, and recording the adjustments made by engineers after running the program could offer insights into the system's quality, suggesting a direction for future enhancements to reduce manual interventions and improve overall system effectiveness.

Table 1: The operation time for different accessory planning approaches.

Operation Approach	Time (Minutes)	Timesaving (%)
Manually	400	0
Preliminary Automatic [3]	365	8.75
Semi-automatic (this research)	50	87.5

5 Conclusions

This paper proposes an approach for accelerating the planning and quantity estimation of aluminum formwork accessories. Through a real case study, we validate the effectiveness of the approach in improving the efficiency of engineers' accessory planning. The results show a significant timesaving of 87.5% for engineers, and this could have a profound impact on the overall project timeline and cost. While our approach has been proven effective in improving efficiency, the need for manual adjustments to resolve conflicts between different accessories could be further addressed by the integration of an automatic conflict identification and resolution feature in future iterations.

Acknowledgement

The authors would like to express sincere gratitude for the received financial support. The funding from the project (112HT511004) conducted by National Taiwan University and FBC Formworks Systems Co., Ltd. has supported the authors in conducting this research. The financial support from the Computer Usage Fee of the Department of Civil Engineering at National Taiwan University (113KC009) has facilitated the authors in publishing and presenting this article at the conference.

References

- Zhang, L. Exploring the Design and Construction of Aluminum Formwork Based on BIM Technology. IOP Conference Series: Earth and Environmental Science, 474, 072078, 2020. DOI: 10.1088/1755-1315/474/7/072078
- [2] Rane, N. R., Achari, A., and Kadam, D. Evaluating the Selection Criteria of Formwork System (FS) for RCC Building Construction. International Journal of Engineering Trends and Technology, 71(3): 197-205, 2023. DOI: 10.14445/22315381/IJETT-V71I3P220.
- [3] Chen, K. Y., Wu, T. H., Setiawan, B., Tandri, C. C., Hsieh, S. H., and Chang, W. T. A BIM-Assisted Planning Tool for Facilitating the Application of an Aluminum Formwork System to Beam-Column Buildings. In Proceedings of International Conference on Civil and Building Engineering Informatics, pages 176–183, Bangkok, Thailand, 2023.
- [4] Hyun, C., Jin, C., Shen, Z., and Kim, H. Automated Optimization of Formwork Design through Spatial Analysis in Building Information Modeling. Automation in Construction, 95: 193–205, 2018. DOI: 10.1016/j.autcon.2018.07.023
- [5] 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. DOI: 10.1016/j.autcon.2021.103680

Development of 3D Digital-Twin System for Remote-Control Operation in Construction

Y. Mori¹ and M. Wada¹

¹ Research and Development Department, Hitachi Plant Construction, Ltd., Tokyo, Japan yoshihito.mori.vx@hitachi.com, masaomi.wada.rv@hitachi.com

Abstract –

This paper proposes a 3D digital-twin system for remote-control operation in construction. Prevalent remote operation using multiple cameras cannot eliminate blind spots completely due to the limitation of the number of installable cameras. Thus, to ensure various remote operations can be performed safely, the 3D digital-twin system applying is developed. This system can help remote operators to recognize the location and behavior of remote-control machines in virtual space simulating a work site. Experiments using a crawler robot with a manipulator verify that remote operators can grip an object with the gripper of the manipulator while avoiding collision between the robot and the robot's surrounding equipment. Furthermore, by using the system, remote operators can complement the field of view which camera images alone could not confirm. To apply the system in various work sites, the following research elements should be examined and verified. The first is to develop the collision-avoidance control with surrounding equipment for manipulators. The second is to improve the accuracy of self-localization and keep robots localized using 3D LiDAR. The third is to reflect the real behavior of objects in virtual space by using sensors such as LiDAR and stereo cameras.

Keywords –

Digital Twin; Remote Operation; Collision avoidance;

1 Introduction

Decreasing numbers of experienced workers and a lack of interest among the young have become a concern for the construction industry in recent years. As demand for construction work is nevertheless expected to grow strongly in coming years, this workforce shortage is a severe problem. Therefore, to improve the efficiency of on-site work, construction methods using Information and Communication Technology (ICT) have been studied.

In addition, various remote-control machines

working construction sites where people cannot enter have been developed. Hence, remote operation among ICT construction methods have become important. Here, remote operation is the work that operators from a remote location operate remote-control machines by using visual images or measurement data from sensors attached to them. Prevalent remote operation uses visual images from cameras in the work area together with remotecontrol machines, and various application examples at construction sites have been reported [1, 2].

However, these examples have the following issues. The first is that there is a limit to the number of locations where cameras can be installed, making it difficult to eliminate all blind spots during remote operation. Second, the difficulty to measure a quantitative distance between the remote-control machines and surrounding equipment has high risks of collision between them.

Thus, to eliminate the skill requirements of remote operation, one of our purposes is to develop a visual assistance system for viewing the remote-control machines on site from the multiple angles of view. Furthermore, the other is to develop a collisionavoidance control between them.

This paper is structured as follows. Section 2 introduces the scientific background. Section 3 gives the methods of the visual assistance system and the collision-avoidance control. Section 4 presents the results of experiments using a crawler robot with a manipulator. Finally, Section 5 concludes the paper.

2 Scientific Background

To develop the visual assistance system, it was the first thing to survey digital twin technology. A digital twin is a simulation technology that virtually reproduces and replicates phenomena and artifacts in physical space in digital space through large-scale data processing and phenomenon modelling based on data collected by advanced measurement and observation [3-5]. Research related to digital twins in the fields of engineering and computational science has increased rapidly in the past five years. In addition, the research is active in the United States, Germany, the United Kingdom, China. and



Figure 1. Configuration of 3D digital-twin system

research projects have been promoted in each country in collaboration with universities, public research institutes, and private companies.

Next, elemental technologies were defined to realize the visual assistance system using digital system as follows:

- (1) recognizing the self-position of remote-control machines,
- (2) recognizing the behaviour of remote-control machines, such as gripping, lifting and putting some objects by machines' manipulators,
- (3) Developing control function to prevent collisions with robots' surrounding equipment.

As famous technology related to (1), there is Simultaneous Localization and Mapping (SLAM) technology that simultaneously creates the maps of the surrounding environment of robots and estimates its own position from measurement data attached to the robots. SLAM technology estimates the self-position of the robots by comparing shape data (usually measured by 2D/3D LiDAR) around the robots with the map data acquired in advance. Some examples that display where robots are in work area on viewer have been reported [6-8]. In addition, these examples using SLAM utilize Robot Operating System (ROS), which is an OS for developing robot software. ROS is middleware that runs on existing OSs and can safely send and receive data in real time between control programs divided according to execution content. In recent years, some studies that control unmanned ground vehicles via ROS have been conducted [9-11]. As mentioned above, ROS is characterized by being able to acquire multiple sensor information via a local network and perform distributed

processing to control robots from various remote locations. On the other hand, regarding to (2), there is some examples that the behaviour of robots was synchronized in 3D virtual space [12]. Furthermore, by combining (2) and (3), the planning methods of optimal paths that avoid the collision between manipulators and surrounding equipment have been studied [13].

Thus, these individual studies on the elemental technologies (1)-(3) have been reported. However, no system has been developed that can collectively execute these elemental technologies. Hence, this paper proposes the following methods.

- (a) 3D digital-twin system for visual assistance that reflects *both* the location and the behaviour of remote-control machines including unmanned ground vehicles (UGV), manipulators and tip attachments of manipulators in real time.
- (b) collision-avoidance control as the function of the 3D digital-twin system that sends a control signal instructing it to slow down or stop to the remotecontrol machine.

3 Methods

3.1 3D digital-twin system

This section introduces the 3D digital-twin system mentioned in Section 2-(a). Figure 1 presents the detailed overview of the proposed 3D digital-twin system. Firstly, virtual (digital) space of the robot's surrounding environment built either as a point cloud or a Computer Aided Design (CAD) model is displayed on the viewer of the digital-twin system (shown at the right side of



Figure 2. Collision-avoidance control

Figure 1). Here, 3D models of the robots are superimposed onto the virtual space. For the development of the viewer, we have chosen Unity that is a commercially available software renowned for its ability to create 3D virtual environments using point clouds or CAD models. Subsequently, sensor data generated from devices in the real space, such as the joint angles of manipulators and the body position/posture of UGVs, are integrated into the 3D models of the robots in real time (shown at the left side of Figure 1). To facilitate the transmission and reception of various data between the devices in real space and the 3D models in virtual space, the digital-twin system uses Robot Operating System (ROS).

3.2 Collision-avoidance control

This section describes the 3D digital-twin system mentioned in Section 2-(b). Figure 2 presents an overview of the collision-avoidance control. As shown in Figure 2, the collision-avoidance control uses two cylindrical models centered around the robot's body axis: stop-range model, and deceleration-range model. The details of each model are described as follows.

When the 3D model of devices' surrounding equipment or point clouds measured by LiDAR fall within the yellow stop-range model, the system commands the actuators of devices to stop. By contrast, when these data falls within the blue deceleration-range model, it sends deceleration commands to the actuators. To detect whether these data has entered the stop (deceleration)-range model, the system uses Collider that Unity handles collision between Objects in virtual space with. In addition, as shown in Figure 2, an alarm on the



(a) Overview



(b) Enlarged view surrounded by red dotted line

Figure 3. Verification robot

viewer screen appears when the 3D model of devices' surrounding equipment move in on stop (deceleration)-range model.

3.3 Verification experiments

Figure 3 shows the overview of a verification robot using in experiments. The robot is assembled by a 6 Degree of Freedom (DOF) manipulator and a crawler UGV. As shown in Figure 3, A total of six encoders are attached to each joint of the manipulator. With the crawler, it has two drive wheels, and an encoder is attached to each drive wheel. In addition, an 2D LiDAR is attached to the tip of the UGV and an IMU sensor that measures the tilt and rotation angles of the UGV around the Earth's axis. Figure 4 shows an overview of the



(a) Overview



Plate-shaped object

(b) Enlarged view

Figure 4. Equipment for the experiments

equipment for the experiments. Here, Figure 4 (a) shows the overall view and (b) shows an enlarged view around a plate-shaped object. The equipment is configured with a tube pipe and plywood. Also, Network cameras for remote monitoring are also installed in a total of four positions as shown in Figure 4 (a).

The verification experiments were performed with the following procedure. First, a remote operator moves the robot from the starting position (drawn as yellow circle) to the front of the plate-shaped object. Next, control the manipulator to adjust the positional relationship between the gripper of the manipulator and the plate-shaped object. Finally, grip and lift the object using the gripper. In the experiments, two patterns of remote control were compared and verified: remote operation using network cameras, and remote operation using network cameras and the digital-twin system. To use the digital-twin system, the virtual space including the 3D models of the robot and the equipment was built in the viewer of the digital-twin system.

4 Results

4.1 Remote operation using network cameras

Figure 5 shows each camera image when the verification robot turned the corner of mockup equipment. From Figure 5, it was difficult to measure a quantitative distance between the robots and robot's surrounding equipment. Moreover, some operators could not accurately understand the current position of the robot within the equipment due to the lack of any overhead image of the equipment.

Figure 6 shows the camera images while lifting the plate-shaped object by the gripper. As with the robot

turning the corner, there was an issue with measuring a quantitative distance between them. In addition, while operating the robot, some operators collided the robot against the plywood by accident.

As mentioned above, remote operation using network cameras alone had many difficulties.

4.2 Remote control using network cameras and digital-twin system

4.2.1 Verification of digital-twin system

Figure 7 shows the top view of the virtual space when the verification robot turned the corner of mockup equipment. From Figure 7, the current position of the robot within the equipment could be determined even if a camera could not be attached to the ceiling.

Figure 8 shows the top/right-side view of the virtual space, and the right-side camera image of the robot while adjusting the position of the manipulator. From Figure 8 (a) and (b), the digital-twin system allowed operators to change the angle of view of the virtual space. By changing it, operators could understand the location and behavior of the robot more easily than by using network cameras. Moreover, Comparing Figure 8 (b), and (c), the manipulators were displayed in the same posture at the same time. In other words, the robot and the robot model in the virtual space were synchronized in real time.

Figure 9 shows the right-side view of the virtual space, and the right-side camera image of the robot while lifting the plate-shaped object by the gripper. From Figure 9, operators could control the robot without any collision with tube pipes. Furthermore, the digital-twin system could reflect the situation that the plate-shaped object was lifted by the gripper.


Figure 5. Camera images when the robot turned the corner



Figure 6. Camera images while lifting the plateshaped object

4.2.2 Verification of Collision-avoidance

Figure 10 shows the right-side view of the virtual space when the verification robot moves forward toward the plate-shaped object. Here, considering the dimensions of the robot, the deceleration range explained in Section 3.2 was set in a cylinder with a diameter of 1500 mm from the model center axis. In addition, the stop range was set in a cylinder with a diameter of 1000 mm.

As shown in Figure 10, the crawler UGV of the robot stopped immediately when the equipment entered the stop-range model. Moreover, the robot remained stopped even if a forward command was accidentally sent to the robot by operators.



Figure 7. Top view of the virtual space when the robot turned the corner



(a) Top view of the virtual space



(b) Right-side view of the virtual space



(c) Right-side camera image

Figure 8. Each view while adjusting the position of the manipulator

5 Conclusion

With the goal of ensuring the safety of remote operation, this paper proposes a 3D digital-twin system that can serve as a visual aid to remote-control operation. Prevalent remote operation using visual images from cameras have the issue that a limit to the number of locations where cameras can be installed make it difficult to eliminate all blind spots during remote operation. Moreover, the conventional examples using digital twin cannot reflect both the location of robots and the behavior of robots' manipulators in the virtual space.

Therefore, the 3D digital-twin system that can reflects both in real time was developed. The highly versatile system works by attaching sensors to remote-control machines and using a network connection to show their real-time movements and positions in a virtual space. In addition, the system also includes a collision-avoidance control based on analysis in virtual space. This was used to develop feedback control for collision avoidance that functions by sending a control signal to the remotecontrol machine, instructing it to slow down or stop if the collision risk model overlaps the model of the area surrounding the machine. From the result of experiments, by using the 3D viewer 3D digital-twin system, operators could grip and lift the plate-shaped object without any collision with the surrounding equipment. Furthermore, by using collision-avoidance control, the crawler slowed down when a nearby item of equipment comes within the deceleration region, and stopped when the model of the surrounding area comes within the machine stop region.

However, to apply the system in various construction sites, the following research elements should be examined and verified.

The first is to develop the collision-avoidance control with surrounding equipment for manipulators. Various remote-control machines have two main units: vehicle, and manipulator. As described in Section 3.2, the collision-avoidance control for vehicle is developed. However, to enhance the safety of remote operation, three-dimensional collision-avoidance control for will be required. As of now, we think that the same control for manipulators can be realized by setting the cylindrical deceleration stop (deceleration)-range model to each rigid body part of manipulators.

The second is to improve the accuracy of selflocalization and keep robots localized using 3D LiDAR. The system does not limit which self-position estimation methods are selected when updating the position of robots. In the verification experiments, the verification robot used the following algorithms: "gmapping" for making 2D occupancy grid maps, and "amcl" for 2D selflocation estimation. As well-known, using these algorithms need to define the initial pose of a robot and the definition is a time-consuming task. Therefore, we have verified 3D LiDAR SLAM in the same verification experiments. Experiments in progress verify that 3D LiDAR SLAM keeps robots localized more accurately than these 2D algorithms.

The third is to reflect the real behavior of objects in virtual space by using sensors such as LiDAR and stereo cameras. Since the system described in the paper is in the basic development stage, we assumed the following



(a) Right-side view of the virtual space



(b) Right-side camera image

Figure 9. Each view while lifting the plate-shaped object



Figure 10. Verification of collision-avoidance control conditions to reflect the position of objects:

- (1) The shape of objects is not varied,
- (2) The shape of objects is the same as the known data (such as drawings).

However, the work sites where (1) and (2) cannot be assumed should be considered. Hence, we need to study some approaches that automatically reflects site environment in virtual space.

References

- S.M. Moon et al. Development of tele-operation control station for Intelligent Excavator. In 2009 IEEE International Conference on Technologies for Practical Robot Applications, 123-128, Massachusetts, USA, 2009.
- [2] D.Zhu et al. "Moving to the centre": A gaze-driven remote camera control for teleoperation, *Interacting*

with Computers, 23(1): 85-95, 2011.

- [3] N. Kousi et al. Digital twin for adaptation of robots' behavior in flexible robotic assembly lines. *Procedia Manufacturing*, 28:121-126, 2019.
- [4] X. Wang et al. Interactive and Immersive Process-Level Digital Twin for Collaborative Human– Robot Construction Work. *Journal of Computing in Civil Engineering*, 35(6):1528-1535, 2019.
- [5] R. Sacks et al. Construction with digital twin information systems. *Data-Centric Engineering*, 1: e14, 2020.
- [6] M. Minos-Stensrud et al. Towards Automated 3D reconstruction in SME factories and Digital Twin Model generation. In 2018 18th International Conference on Control, Automation and Systems, 1777-1781, PyeongChang, South Korea, 2018.
- [7] X. Liu et al. Human-centric collaborative assembly system for large-scale space deployable mechanism driven by Digital Twins and wearable AR devices. *Journal of Manufacturing Systems*, 65:720-742, 2022.
- [8] M. Müller et al. Self-improving situation awareness for human-robot-collaboration using intelligent Digital Twin. *Journal of Intelligent Manufacturing*, open access:1-19, 2023.
- [9] Ilmir Z. Ibragimov et al. Comparison of ROS-based visual SLAM methods in homogeneous indoor environment. In 2017 14th Workshop on Positioning, Navigation and Communications, 1-6, Bremen, Germany, 2017.
- [10] A.F. Olalekan et al. Comparison of Two SLAM Algorithms Provided by ROS (Robot Operating System). In 2021 2nd International Conference for Emerging Technology, 1-5, Belagavi, India, 2021.
- [11] P. Phueakthong et al. A Development of Mobile Robot Based on ROS2 for Navigation Application. In 2021 International Electronics Symposium, 517-520, Surabaya, Indonesia, 2021.
- [12] S. Kim et al. Development of BIM-integrated construction robot task planning and simulation system. *Automation in Construction*, 127:1-8, 2021
- [13] M. Matulisa et al. A robot arm digital twin utilising reinforcement learning. *Computers & Graphics*, 95: 106-114, 2021.

Exploring the Potential of Reinforcement Learning in Pipe Spool Scheduling in Industrial Projects

Mohamed ElMenshawy¹, Lingzi Wu², Brian Gue³ and Simaan AbouRizk¹

¹Department of Civil and Environmental Engineering, University of Alberta, Canada ²Construction Management, University of Washington, USA ³PCL Industrial, Canada <u>elmensha@ualberta.ca</u>, <u>lingwu@uw.edu</u>, <u>bmgue@pcl.com</u>, <u>abourizk@ualberta.ca</u>

Abstract –

Pipe spools are key components in industrial projects. Usually, they are built off-site in a fabrication shop and then shipped to the project location for installation. The fabrication shop deals with numerous spools, each designed to specific requirements according to shop drawings. The nature of pipe spools being engineered to order, together with production constraints such as lead time of materials, different processing times, and availability of resources, render the scheduling process within the shop challenging and time-consuming. As such, this research aims to automate the scheduling process by developing a reinforcement learning model that includes an agent that is capable of handling the scheduling process. The proposed model is applied to an illustrative example to investigate the concept of automating the scheduling process. The construction professionals highlight the great potential of the proposed model in the fabrication scheduling process, and its ability to minimize manual intervention.

Keywords -

Pipe spools; Reinforcement learning; Industrial projects

1 Introduction

Industrial projects include facilities like power plants, nuclear plants, and oil/gas production sites. Pipe spools-a main element in these facilities-are often fabricated off-site in fabrication shops that receive numerous orders accompanied by shipping schedules to construction sites [1,2]. A pipe spool comprises components like pipes, flanges, and elbows, each having varying lead times. The pipe spool is considered a unique product, being custom-made based on the shop drawings, engineered-to-order, and subject to frequent modifications due to design alterations or priority changes, availability of resources, processing times, and due dates [3]. All these features and constraints render

the scheduling process challenging, laborious, and timeconsuming. As a result, previous research has attempted to address the challenges of the scheduling process.

Metaheuristic algorithms have been explored as solutions for pipe spool scheduling [4,5]. Methods applied include genetic algorithms, artificial bee colonies, and ant colony optimizations [6-12]. Yet, their static nature limits their applicability in the dynamic and complex shop environment [13]. Simulation-based approaches have also been investigated as potential solutions including simulation models of pipe spool fabrication to study the fabrication process [14-20]. These models were used to study cycle time, bottlenecks, and resource utilization, which provides a level of support to construction professionals but does not offer a direct solution to the time-consuming nature or need for manual intervention in the scheduling process. Based on the aforementioned research efforts, we concluded that research related to automating the scheduling process and minimizing manual intervention remains relatively unexplored.

This research aims to bridge this gap using a reinforcement learning model to automate the scheduling process and reduce human manipulation. The reinforcement model is comprised of an agent, actions, environment, states, and rewards. The agent employs a dueling deep Q-network and experience replay where the agent stores past experiences while interacting with the environment. The agent observes the states within the environment, takes an action from the available actions, and then learns through a reward system.

This study advances the body of knowledge by exploring the potential of reinforcement learning models in the domain of construction management, specifically scheduling.

The remainder of this paper presents background information on pipe spool fabrication and reinforcement learning. We then provide a brief review of the state-ofthe-art, identify research gaps, and discuss the potential of reinforcement learning in the scheduling process. The methodology section elaborates on the techniques and algorithms used, and then provides an illustrative example demonstrating our research's practicality.

2 Research Background

This section provides a brief review of the fabrication of pipe spools, and focuses on three main processes: fitting, welding, and inspection. Also, it presents an overview of reinforcement learning.

2.1 Pipe Spool Fabrication Process

The process of pipe spool fabrication encompasses multiple phases. This research focuses on the main fabrication processes: fitting, welding, and inspection. Pipes are the foundational element of spool assembly, being cut into specified sizes and shapes based on the shop drawings [20]. Once cut, the pipe moves to a fitting station where it is temporarily fixed with multiple components such as elbows, flanges, and reducers. Following this assembly stage, the semi-completed pipe spool moves to the welding station, either manually or lifted by machinery, based upon its weight and other handling requirements. The welding process ensures a permanent fixation of all components together [21,22]. Finally, the welded spool undergoes inspection, which represents the final stage in the process [3].

2.2 Reinforcement Learning

Reinforcement learning (RL) is a trending area of machine learning now used in many fields [23]. In RL, there are key components: agent, actions, environment, states, and rewards. An agent observes states in the environment, takes an action, and learns from interacting with its environment by receiving rewards or penalties, aiming to get more rewards and fewer penalties [23]. This learning process helps the agent make better decisions [24]. We can explain this using a Markov decision process (MDP), which includes states, actions, state transition probability matrix, reward functions, and discount factor [25,26]. This will be discussed in more detail in the methodology section.

3 Literature Review

This section explores two primary research areas: the application of RL in construction, and RL applications in job shop scheduling.

3.1 Reinforcement Learning in Construction

Scholars have explored the use of RL across multiple construction domains. Akanmu et al. [27] have presented a digital platform to train construction professionals encompassing wearable devices, RL, labor engagement, and monitoring tools. Mullapudi et al. [28] also developed a control strategy for stormwater systems that relies on an RL model to control the operations of valves, gates, and pumps.

3.2 Reinforcement Learning in Job Shop Scheduling

RL has been used to enhance dynamic scheduling in job-shop environments. Several researchers applied the Q-learning algorithm, where agents are taught to take action by choosing a dispatching rule to reduce tardiness in the process [29-32]. Yet, due to the large number of states in real-world production scenarios, Q-learning fails in practicality, as maintaining an extensive Q-table for such states becomes unfeasible.

Accordingly, there was a leap in estimating the Q-values by shifting towards deep reinforcement learning (DRL), which has demonstrated significant promise in job shop scheduling [33-35]. DRL, incorporating deep neural networks (DNNs), revolutionized the estimation of Q-functions [36].

A literature analysis reveals that there have not been any previous studies exploring the use of RL for scheduling in pipe spool fabrication. This finding aligns with the insights of Xu et al. [23], who assessed the current advancements in RL within construction engineering and management. The authors emphasize the limited number of studies incorporating RL in this domain, suggesting a need for increased focus [23]. Additionally, they highlight project scheduling and resource allocation as promising areas for future research.

The capabilities of RL make it a promising solution for pipe spool fabrication scheduling due to its strengths in the following areas: (1) navigating complex problems in changing environments with high dimensions; (2) adjusting to varied scenarios and states; (3) engaging in independent learning; and (4) predicting future action outcomes.

Building on this, our study introduces a pioneering RL-driven scheduling model tailored for pipe spool fabrication shops. The intention behind this model is to navigate the evolving landscape of such environments while addressing several inherent challenges in the scheduling area like resource allocation and adhering to due dates.

4 Model Development

Our model focuses on the fabrication process of pipe spools, specifically addressing the fitting stage. To develop the proposed model, the following information is required: start date, fitting processing time, required resources, and due date. As depicted in Figure 1, the model development involves two main stages: data preparation and RL. Each stage will be described in the coming subsections. The proposed model development process is designed to allow the integration of data related to pipe spools with an RL model. Once the RL model is trained on a training dataset, it can then be applied to an unseen dataset of pipe spools that need to be scheduled.



Figure 1. Methodology

4.1 Data Preparation

In this research, synthetic data is generated to mimic the spool information required to build the model. The generated dataset includes start date, fitting processing time, resources, and due date. The dataset includes approximately 1,900 records, and each record represents a spool. The dataset is then split into a training set with 80% of the records, and a testing set with the remaining 20% of the records.

4.2 Reinforcement Learning

This section describes the development of the RL model and formulates the scheduling process as a Markov decision process (MDP). Dueling DQN and prioritized replay are used in developing the agent.

4.2.1 MDP Formulation

As a general representation, the MDP can be described by five main components (S, A, P, R, γ) where S is the set of states, A is the set of actions, P is the probability transition matrix from one state to another, R is the reward function, and γ is the discount factor. Typically, in each time step t, the agent observes the states, takes an action a_t , then gets a reward or penalty r_t through the reward function R. It then moves to the next state s_{t+1} depending on P. While calculating the rewards, the effect of future rewards is taken into consideration using the discounting factor γ .

• State

The state is a crucial concept that represents the current status of the environment at a given time step, and based on the state, the agent determines its next action [37]. In our model, the state includes average processing duration for every spool, average slack, and average resources required for each spool.

• Action

Actions are crucial decisions made by an agent based on its observation of the current state. As the agent navigates the environment, it utilizes specific rules for its choices. In this context, the agent has two dispatching rules from which to select: first-come-first-serve (FCFS) or shortest processing time (SPT). The FCFS rule simply adheres to the order of spool arrival, processing spools in the sequence they were received, while the SPT rule prioritizes spools expected to be completed in the shortest duration. Given these choices, the agent dynamically identifies the optimal action for the current state during each decision-making time step.

Reward

The reward function plays a crucial role in guiding the agent's decisions. It provides feedback from the environment in response to the agent's actions, effectively serving as an indicator of the agent's performance. The reward function deals with minimizing the number of spools that are not completed before their due date as demonstrated in Equation (1):

$$R = \begin{cases} 0, & D_i \le t \\ -1, & D_i > t \end{cases}$$
(1)

where D_i represents the due date of the spool, and t represents the time step.

4.2.2 Deep Reinforcement Learning

In this model, we employ the Dueling Deep Q-Network (Dueling DQN) which builds upon the foundational principles of the Deep Q-Network (DQN). The Dueling DQN enhances the conventional DQN by

decoupling the state values and the action advantages, refining the policy learning process.

The RL model is developed through the features of PyTorch, a powerful open-source machine learning framework related to RL algorithms [41]. Additionally, OpenAI's gym serves as the training ground for the RL model [42]. Coupling PyTorch's modeling prowess with OpenAI gym's environmental interface and training capabilities results in a refined and highly effective RL solution.

• Model Training

The Q-value serves as a foundational metric in the MDP, quantifying the anticipated discounted future reward when a specific action is taken. The ultimate aim of the agent is to develop an optimal policy that increases the expected rewards, a principle outlined by Mnih et al. [34].

However, employing a non-linear function estimator-such as a neural network-to approximate the Q-function presents challenges. Specifically, it can lead to instability or divergence. Two predominant strategies, as described by Wang et al. [38], offer solutions to those challenges. First, the experience replay method involves storing the agent's experiences. This data is then randomized, reducing correlations, and sampled in mini-batches to train the Q-network. The second strategy introduces an iterative update mechanism. By integrating a target Q-network into the DQN with parameter θ_i^{-} , correlations with targets are diminished. This target Q-network synchronises with the primary Qnetwork's parameters at periodic intervals, as illustrated by Mnih et al. [34].

Enhancing the DQN's architecture, we adopted the Dueling DQN. This refined structure improves efficiency by addressing Q-value overestimations. Notably, the Dueling DQN separates its estimations, separately determining the state value function and the action's advantage. These assessments then combine to forecast the action's quality, a technique explained by Liang et al. [39] and Wen et al. [40].

In each time step, the agent observes the state of the environment and chooses an action from one of the two dispatching rules: FCFS or SPT. Once an action is taken by the agent, the environment starts to send feedback in the form of reward or penalty, so that the agent can learn from this interaction and make sure to take better actions that maximize rewards in future time steps.

The model is trained on the training dataset, which is approximately 1,500 records and 10 resources for fitting the spools. Additionally, the following hyperparameters (Table 1) are used while training the model:

Table 1. Model hyperparameters

Parameter	Value
No. of training episodes	100
Learning rate	0.0025
Minibatch size	32
3	0.9
γ	0.9

5 Model Results and Interpretation

After training, the agent was introduced to the testing dataset containing pipe spool records unseen by the agent. Consequently, the agent was evaluated on this dataset, making decisions to maximize rewards and concurrently reducing the number of spools that were not completed before their due date. Figure shows the average reward during the evaluation phase, based on the respective dispatching rule selected by the RL agent.



Figure 2. Average reward per episode (Exp #1)

Another experiment was conducted and the model was trained using 300 episodes—an episode is a complete cycle of scheduling all the pipe spools in the training dataset—and a learning rate of 0.001 in lieu of 100 and 0.0025, respectively. Figure 3 shows that the model's performance was relatively improved as the lowest average reward reached -50 when compared to the first experiment where the lowest average reward reached approximately -70.



Figure 3. Average reward per episode (Exp #2)

A third experiment was conducted and the model was trained using 500 episodes and a learning rate of 0.001. Figure 4 shows that the model's performance was impacted as the lowest average reward reached approximately -140 when compared to the first and second experiments where the lowest average reward reached -70 and -50, respectively. As such, these hyperparameters had a negative impact on the model performance. Future work will investigate the hyperparameters that improve the model performance.



Figure 4. Average reward per episode (Exp #3)

6 Contributions and Future Work

This research introduces an innovative method for automating the scheduling of pipe spool fabrication in industrial construction projects. It uses RL, specifically employing Dueling DQN with an experience replay buffer that stores the agent's previous experiences. This method proves that an agent can be developed to schedule tasks in the simulated fabrication shop, reducing human intervention. This work serves as an initial exploration into applying RL for pipe spool scheduling in a fabrication shop.

The research was then applied to an illustrative example that employed an artificial dataset that mimicked the data collected from the fabrication shop. The analysis highlights the significant potential of incorporating RL into scheduling, offering industry professionals a decision support tool for pipe spool scheduling.

These preliminary results show that the proposed RL model has the potential to outperform traditional methods, which often require human input to deal with continuous changes. This research also refines the automation of scheduling, emphasizing the main constraint of adhering to due dates, which is imperative to preventing onsite installation delays.

The study includes certain limitations that must be addressed in future research. First, the study focused on the fabrication process only, specifically the fitting operation. Second, only two of the basic dispatching rules were included as actions to be taken by the agent. Third, a simple straightforward reward function was used to calculate the rewards/penalties based on the agent's actions. Finally, the model should be trained and evaluated on a real-world dataset from a fabrication shop. Future efforts should expand this research by encompassing diverse shop operations, by comparing the performance of the agent against the dispatching rules, and by assessing the agent's ability to shift between those dispatching rules.

7 Acknowledgements

This project was supported by a Collaborative Research and Development Grant (572086-22) from the Natural Sciences and Engineering Research Council of Canada. We would also like to express our gratitude to PCL Industrial Management Inc. for their continued support throughout the projects, and Brann Munro for her assistance with paper proofreading.

8 References

- Mosayebi S. P., Fayek A. R., Yakemchuk L., and Waters S. Factors Affecting Productivity of Pipe Spool Fabrication. *International Journal of Architecture, Engineering and Construction*, 1(1):30–36, 2012.
- [2] Mohamed Y., Borrego D., Francisco L., Al-Hussein M., Abourizk S., and Hermann U. Simulation-based scheduling of module assembly yards: Case study. *Engineering, Construction and Architectural Management*, 14(3):293–311, 2007.
- [3] Mohsen O., Petre C., Mohamed Y. Machine-Learning Approach to Predict Total Fabrication

Duration of Industrial Pipe Spools Journal of Construction Engineering and Management, 149(2), 2022.

- [4] Safarzadeh S., Shadrokh S., and Salehian A. A heuristic scheduling method for the pipe-spool fabrication process. *Journal of Ambient Intelligence and Humanized Computing*, 9(6):1901–1918, 2018.
- [5] Moghadam A. M., Wong K. Y., Piroozfard H., Asl Derakhshan A., and Shanty Hutajulu T. Solving an industrial shop scheduling problem using genetic algorithm," in *Advanced Materials Research*, 845: 564–568, 2014.
- [6] Xing L. N., Chen Y. W., Wang P., Zhao Q. S., and Xiong J. A Knowledge-Based Ant Colony Optimization for Flexible Job Shop Scheduling Problems. *Applied Soft Computing Journal*, 10(3): 888–896, 2010.
- [7] Bagheri A. and Zandieh M. Bi-criteria flexible jobshop scheduling with sequence-dependent setup times - Variable neighborhood search approach. *Journal of Manufacturing Systems*, 30(1):8–15, 2011.
- [8] Zandieh M., Khatami A. R., and Rahmati S. H. A. Flexible job shop scheduling under condition-based maintenance: Improved version of imperialist competitive algorithm. *Applied Soft Computing*, 58: 449–464, 2017.
- [9] Xu Y., Wang L., Wang S. yao, and Liu M. An effective teaching-learning-based optimization algorithm for the flexible job-shop scheduling problem with fuzzy processing time. *Neurocomputing*, 148: 260–268, 2015.
- [10] Mokhtari H. and Hasani A. An energy-efficient multi-objective optimization for flexible job-shop scheduling problem. *Computers & Chemical Engineering*, 104:339–352, 2017.
- [11] Gao K. Z., Suganthan P. N., Pan Q. K., Chua T. J., Chong C. S., and Cai T. X. An improved artificial bee colony algorithm for flexible job-shop scheduling problem with fuzzy processing time. *Expert Systems with Applications*, 65:52–67, 2016.
- [12] Chang H. C. and Liu T. K. Optimisation of distributed manufacturing flexible job shop scheduling by using hybrid genetic algorithms. *Journal of Intelligent Manufacturing*, 28(8):1973– 1986, 2017.
- [13] Kardos C., Laflamme C., Gallina V., and Sihn W. Dynamic scheduling in a job-shop production system with reinforcement learning. In *Proceedia CIRP*, pages 104–109, 2020.
- [14] Lu M., Abourizk S. M., and Hermann U. Sensitivity analysis of neural networks in spool fabrication productivity studies. *Journal of Computing in Civil Engineering*, 15(4):299-308, 2001.
- [15] Wang, P., Mohamed, Y., Abourizk, S. M., & Rawa,

A. T. Flow Production of Pipe Spool Fabrication: Simulation to Support Implementation of Lean Technique. *Journal of Construction Engineering and Management*, 135(10):1027-1038, 2009.

- [16] Song, L., Mohamed, Y., & AbouRizk, S. M. Early Contractor Involvement in Design and Its Impact on Construction Schedule Performance. *Journal of management in engineering*, 25(1): 12-20, 2009.
- [17] Hu D. and Mohamed Y. Simulation-Model-Structuring Methodology for Industrial Construction Fabrication Shops. *Journal of construction engineering and management*, 140(5), 2014.
- [18] Taghaddos H., Hermann U., AbouRizk S., and Mohamed Y. Simulation-Based Multiagent Approach for Scheduling Modular Construction, *Journal of Computing in Civil Engineering*, 28(2):263–274, 2014.
- [19] Song L., Wang P., and AbouRizk S. A virtual shop modeling system for industrial fabrication shops. *Simulation Modelling Practice and Theory*, 14(5):649–662, 2006.
- [20] Sadeghi N. and Fayek A. R. A framework for simulating industrial construction processes. In *Proceedings of Winter Simulation Conference*, pages 2396–2401, 2008.
- [21] Wang P. and Abourizk S. M. Large-scale simulation modeling system for industrial construction," *Canadian Journal of Civil Engineering*, 36(9):1517–1529, 2009.
- [22] Hu D. and Mohamed Y. Pipe spool fabrication sequencing by automated planning. In *Construction Research Congress 2012: Construction Challenges in a Flat World*, pages 495–504, 2012.
- [23] Xu Y., Zhou Y., Sekula P., and Ding L. Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, 6, 2021.
- [24] Nguyen H. and La H. Review of Deep Reinforcement Learning for Robot Manipulation. In Proceedings - 3rd IEEE International Conference on Robotic Computing, pages 590–595, 2019.
- [25] Nian R., Liu J., and Huang B. A review On reinforcement learning: Introduction and applications in industrial process control," *Computers and Chemical Engineering*, 139, 2020.
- [26] Levine S., Kumar A., Tucker G., and Fu J. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems, http://arxiv.org/abs/2005.01643, 2020.
- [27] Akanmu A. A., Olayiwola J., Ogunseiju O., and McFeeters D. Cyber-physical postural training system for construction workers, *Automation in construction*, 117, 2020.

- [28] Mullapudi A., Lewis M. J., Gruden C. L., and Kerkez B. Deep reinforcement learning for the real time control of stormwater systems. *A Advances in water resources*, 140, 2020.
- [29] Aydin M. E. and Öztemel E. Dynamic job-shop scheduling using reinforcement learning agents. *Robotics and Autonomous Systems*, 33:169-178, 2000.
- [30] Wei, Y., & Zhao, M. A reinforcement learningbased approach to dynamic job-shop scheduling. *Acta Automatica Sinica*, 31(5), 2005.
- [31] Shahrabi J., Adibi M. A., and Mahootchi M. A reinforcement learning approach to parameter estimation in dynamic job shop scheduling. *Computers & Industrial Engineering*, 110:75–82, 2017.
- [32] Bouazza W., Sallez Y., and Beldjilali B. A distributed approach solving partially flexible jobshop scheduling problem with a Q-learning effect. *IFAC-PapersOnLine*, 50(1):15890–15895, 2017.
- [33] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. Playing Atari with Deep Reinforcement Learning, arXiv preprint arXiv:1312.5602, 2013.
- [34] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G. and Petersen, S. Human-level control through deep reinforcement learning," *Nature*, 518(7540):529–533,2015.
- [35] Li Y. Deep Reinforcement Learning: An Overview. arXiv preprint arXiv:1701.07274, 2017.
- [36] Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., & Kyek, A. Optimization of global production scheduling with deep reinforcement learning, In *Procedia Cirp 72*, pages 1264-1269, 2018.
- [37] Kim T., Kim Y. W., Lee D., and Kim M. Reinforcement learning approach to scheduling of precast concrete production. *Journal of Cleaner Production*, 336, 2022.
- [38] Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., & Freitas, N. Dueling Network Architectures for Deep Reinforcement Learning. In *International conference on machine learning*, pages 1995-2003, 2016.
- [39] Liang W., Xie W., Zhou X., Wang K. I. K., Ma J., and Jin Q. Bi-Dueling DQN Enhanced Two-stage Scheduling for Augmented Surveillance in Smart EMS," *IEEE Transactions on Industrial Informatics*, 2022.
- [40] Wen S., Lv X., Lam H. K., Fan S., Yuan X., and Chen M. Probability Dueling DQN active visual SLAM for autonomous navigation in indoor environment. *Industrial Robot: the international journal of robotics research and application,*

48(3):359-365, 2020.

- [41] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in neural information processing systems*, 32, 2019.
- [42] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J. and Zaremba, W. OpenAI Gym, arXiv preprint arXiv:1606.01540, 2016.

Current State and Trends of Point Cloud Segmentation in Construction Research

Samuel A. Prieto¹, Eyob T. Mengiste¹, Uday Menon¹, and Borja García de Soto¹

¹S.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 samuel.prieto@nyu.edu, eyob.mengiste@nyu.edu, um2038@nyu.edu, garcia.de.soto@nyu.edu

Abstract –

The construction industry is witnessing a transformative shift with the integration of advanced technologies, especially in the topic of 3D segmentation. This study underscores the current state and challenges of 3D segmentation, with special emphasis on construction research, and provides an insightful understanding of the latest research developments and trends. The study also looks at the performance metrics of the most relevant techniques, as well as the main limitations and research gaps, highlighting the need for further research in highly-performing techniques based on Deep Learning for point cloud segmentation in construction applications.

Keywords -

3D Segmentation; Systematic Literature Review; Construction Industry; Point Cloud; Deep Learning

1 Introduction

The abundance of 3D point cloud data is increasing with the availability and advancement of laser scanning equipment and the democratization of 3D semantic data generation. The availability of 3D data generally boosts the possibility and efficiency of developing deep learning algorithms to segment and classify 3D information, which heavily relies on the volume of available training data [1]. The manufacturing industry, amongst others, has greatly benefited from the extensive research over the past 15 years in 2D and 3D segmentation [2]. This was possible due to the static scenarios present in said industries, making them reliable semi-controlled environments. However, the construction industry has not developed at the same pace due to its dynamic and unstructured nature [3]. Some researchers and companies are already providing the means for the construction industry to take advantage of this revolution, where several applications and companies employ 3D data acquisition systems and processing algorithms [4], especially in the field of Scan-to-BIM [5]. Nevertheless, segmentation algorithms are still behind.

In this review article, the recent advancements in 3D point cloud processing are explored, particularly in the segmentation and classification domain, comparing the trending techniques in all industries with a focus on construction. From that, current challenges and research gaps are highlighted. The rest of the paper is structured as follows: Section 2 explains in detail the scope of this review. Section 3 presents the methodology used for the acquisition and analysis of the presented data. Section 4 analyses the obtained data. Section 5 provides a discussion of some of the findings from the analyzed data. Section 6 comments on some of the limitations of this study and future work. Finally, Section 7 summarizes the main takeaways of the study.

2 Scope of review

The review focuses on segmentation and classification algorithms. These can be applied to either 2D or 3D data. With 2D being extensively studied already, this study particularly focuses on 3D data, particularly in point cloud segmentation. The study also explores applications in the construction industry and non-construction industry (e.g., manufacturing, medical, robotics perception, etc.), with special emphasis on the construction industry. Point cloud segmentation techniques can be generalized into three different categories:

- 1. *Traditional techniques:* Traditional techniques of point cloud segmentation do not include training or clustering. These techniques include voxel cloud connectivity segmentation or super voxel-based segmentation, region growing, edge detection, and model fitting-based techniques.
- 2. Machine learning techniques: This category is comprised of techniques that include automated classification of points based on features (unsupervised machine learning), including K-Nearest Neighbors (KNN), Gaussian Mixture Models (GMM), and shallow learning models such as Support Vector Models (SVM).

3. *Deep learning techniques:* This is the most advancing category of techniques. They can be sub-categorized as point-based, projection-based, and discretization.

3 Methodology

The methodology used followed a Systematic Literature Review (SLR) [6] to conduct a methodical and comprehensive examination of existing research literature on a specific topic aimed at identifying, evaluating, and summarizing the findings of relevant studies to address a defined research question. Unlike traditional literature reviews, which can be more narrative and subjective, an SLR provides a comprehensive and unbiased overview of the current state of research. To ensure reproducibility and comprehensiveness, SLR involves defined steps, such as (1) defining clear research questions, (2) setting inclusion and exclusion criteria, (3) extracting and analyzing data, and (4) reporting and presenting findings in a structured manner. The following subsections delve into each of these steps.

3.1 Research Questions

Given the emphasis of the study on 3D data segmentation in the construction industry and its evolution, especially with the development of Machine Learning and Deep Learning approaches, the following research questions have been formulated for this review:

- 1. Which techniques are used in the construction industry to segment and/or classify point clouds from the construction sites?
- 2. What techniques are adopted to segment and/or classify non-construction site point clouds?
- 3. What are the common metrics used to measure the efficiency of the techniques?
- 4. What are the common challenges of the bestperforming techniques in the construction domain?

3.2 Databases and Search Engines

For this review, the Scopus (ScienceDirect) database has been chosen. Scopus is one of the largest abstract and citation databases, covering a broad spectrum of disciplines. Its extensive coverage of peer-reviewed scientific journals makes it a suitable choice for this review. Scopus includes papers published in reputable and relevant journals to the research topic evaluated in this paper, such as Automation in Construction, and proceedings of equally relevant and reputable conferences, such as the International Symposium on Automation and Robotics in Construction (ISARC). To ensure we did not miss any relevant ISARC publication, we also used the ISARC proceedings database available in the publication section of the IAARC website [7].

3.3 Search Strategy

For this review, the keywords shown in Table 1 were identified based on their relevance to 3D data segmentation in the construction industry and the previously established classification. The set of keywords is comprised of the different techniques and other pertinent terms that, based on the initial search and to the authors' knowledge, fall within the different categories.

Table 1.	Set of keywords used for literature search,
	based on the different categories.

Category	Sub-category	Keywords
1. Traditional	1.1 Supervoxel-	VCCS; Seed
	based techniques	Growing; Graph-
		based; Mean shift-
		based; Normalized
		cuts; Random
		walks;
		Hierarchical
		diffusion
	1.2 Region	Region growing;
	growing-based	Octree based; RG
	techniques	,
	1.3 Model fitting-	RANSAC; Hough
	based techniques	transform; Region
		growing with
		model fitting;
		Expectation-
		Maximization; EM
2. Machine	2.1 Unsupervised	KNN; GMM; K-
Learning	learning techniques	means
	2.2 Shallow	SVM; Decision
	learning techniques	trees; Random
		Forests; Density-
		based spatial
		clustering;
		DBSCAN
3. Deep	3.1 Point-based	PointNet;
Learning	techniques	PointNet++;
		PointCNN;
		DGCNN;
		KPConv;
		PointConv; Point
		Transformer;
		ShellNet;
		PointBERT;
		CurverNet; Self
		Organizing
		Network
	3.2 Projection-based	Spherical
	techniques	Projection; Voxel
		Grid projection
	3.3 Discretization-	3D CNN; VoxNet;
	based techniques	Submanifold
		Sparse CNNs

To maximize the scope of the search and ensure all relevant literature is captured, the keywords were combined using the Boolean operators "AND" and "OR". An example of the search query used in the database for subcategory 1.2 (region growing-based techniques) combining the keywords mentioned above with all the different inclusion and exclusion criteria is shown below:

TITLE-ABS-KEY ((("region growing" OR "octree" OR "RG") AND ("3D data") AND ("construction industry" OR "AEC"))) AND PUBYEAR>2007 AND PUBYEAR<2025 AND (LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "cp") OR LIMIT-TO(DOCTYPE, "re")) AND (LIMIT-TO(LANGUAGE, "English"))

In this typical structure of the query, *region growing*, *octree* and *RG* represent the list of keywords in Table 2. The terms *3D data*, *construction industry* and *AEC* were used to limit the search to construction and related domains. Moreover, *ar*, *cp*, and *re* are document types representing 'article', 'conference paper', and 'review', respectively. The use of these operators allows for a more refined search, ensuring that the results are closely aligned with the research questions. The results from this search criteria are available in [8].

3.4 Inclusion and Exclusion Criteria

To ensure the relevance and consistency of the literature selected for this review, the following criteria were established:

- Focus on techniques related to 3D data processing. The data processing is conducted in two separate streams for construction and non-construction domains.
- Papers that have appeared in recognized journals or conference proceedings.
- Papers published after 2008 for a broader overview of the topic, with a specific emphasis on those published after 2018 to evaluate the most recent and cutting-edge techniques.
- Papers written in English.
- Exclusion of non-peer-reviewed papers, such as opinion pieces, editorials, or news articles.
- Exclusion of papers not related to 3D data processing.
- To evaluate techniques, specific research is conducted by specifying the data as laser scan or LiDAR and Photogrammetry.

3.5 Data extraction and synthesis

For this review, the following data were extracted from each paper: Authors, Publication year, Journal or conference, Keywords, Abstract, Methodology, and Findings.

Following the extraction, the data were synthesized to provide a comprehensive overview of the current state of 3D data processing in the construction industry. The synthesis process involves analyzing the extracted data to identify patterns, trends, and key insights.

4 Analysis of the data

The number of publications found for each category is shown in Figure 1. The values shown indicate the number of papers obtained using the criteria specified in Section 3 and the classification and keywords provided in Table 1. The gradient provides a visualization of the number of papers per category (from low (light background) to high (dark background)). The search also distinguished among the different sub-categories (i.e., techniques) applied to the construction industry and those applied to non-construction industries. This distinction helps to identify how the construction industry is catching up with other fields.

From Figure 1, it can be seen that fitting-based segmentation techniques (Category 1.3), such as RANSAC, Hough transform, region growing with model fitting and EM, are the most frequently implemented and researched. Overall, the adoption of these techniques has been on a constant increase in the past 15 years. Figure 2 shows a significant increase in the use of these techniques in the construction industry in 2012 and 2017. However, in other industries, their usage has been steadily growing, with a slight decrease observed in 2020.



Figure 1. Number of publications for each subcategory.

Figure 1 also shows that the second most frequently used set of techniques is Category 2.2. This category encompasses shallow learning techniques like SVM, decision trees, Random Forests (RF), and DBSCAN. Until 2012, the application of these techniques was limited across various domains, including construction.

However, from 2012 to 2014, there was a significant increase in the number of construction industry publications. The usage of these techniques continued to rise with relative consistency until 2017, followed by a sudden decline. In the non-construction sector, a drop was observed in 2020.

Focusing on Category 3, the distribution of the different techniques and metrics from the 72 construction-related publications (66 from 3.1, 3 from 3.2 and 3 from 3.3) and 1,011(892 from 3.1, 30 from 3.2 and 89 from 3.3) from non-construction related are summarized in Figure 3.



Figure 2. Number of publications for each sub-category in (a) the construction industry and (b) nonconstruction industries versus time (from 2008 until November 2023).



Figure 3. Distribution of papers related to Category 3 for (a) techniques and (b) metrics used in construction publications and (c) techniques and (d) metrics used in non-construction publications.
Legend: P = Precision; R = Recall, OA = Overall Accuracy, Acc = Accuracy; IoU = Intersection over Union, CM = Confusion Matrix, MAE = Mean Absolute Error, RMSE = Root mean squared error.

Others for techniques include: PointConv, depth image estimation, 3DAGN, FPSnet, SOnet, Point Transformer, ShellNet, ResPointNet++, SEP network, GLSNet, DbNet, 3DLEB-Net, RandLA-Net, IAGMLP, RFFS-Net, TriangleConv, DSNet, SPGraph, PointCNN, IBPCS, CNN. Others for metrics include: Support, MAE, FLOPs, Success, R², CM In summary, it can be seen that the use of point-based techniques, part of Deep Learning (Category 3 in Table 1), is growing the most. Due to space constraints, this paper focuses on Deep Learning techniques as it is expected to be more relevant in future construction research.

5 Discussion

5.1 Techniques

As depicted in Figure 2, in the non-construction sector, despite limited mentions of Deep Learning techniques in earlier years since 2010, the number of publications incorporating these techniques began to increase significantly since 2016. Publications in the construction sector using Deep Learning techniques under Categories 3.1 and 3.3 began in 2018, while Category 3.2 started in 2021. The clear difference in the Category 3 popularity between the construction and nonconstruction sectors is affected by the fact that Deep Learning techniques heavily rely on the quality and quantity of the training data. The construction sector is challenging, dynamic, and often restricted by policies that do not allow the data to go public. This has prevented the widespread of training datasets related to construction, which has directly affected the usage of Deep Learning techniques in the construction field.

Among the Deep Learning techniques, Category 3.1, which includes point-based techniques such as PointNet, PointNet++, PointCNN, DGCNN, KPConv, PointConv, Point Transformer, ShellNet, PointBERT, CurverNet, and Self Organizing Network, is dominant compared to the other categories identified for Deep Learning. Since 2018, the adoption of these techniques has increased more significantly in non-construction sectors, making them the most widely used techniques in recent years, particularly since 2021. However, in the construction sector, Category 1.3 (model fitting-based techniques) and Category 2.1 (techniques like SVM, decision trees, random forests, and DBSCAN) continue to be predominant. Meanwhile, the frequency of publications mentioning Category 3.1 techniques has decreased by ten since 2022 in the construction domain. Figure 3(a) illustrates that from techniques grouped in Category 3, the field of construction research frequently employed PointNet, PointNet++, and DGCNN. Moreover, MVCNN, Spherical projection and 2D CNN were each utilized in more than 2 articles. The publications in the non-construction sector have a similar tendency to use PointNet and PointNet++ more frequently than the remaining techniques, with a greater usage of PointNet over PointNet++ and a 10% reduction in utilization of DGCNN. On the other hand, other techniques including PointConv, depth image estimation, 3DAGN, FPSnet, SOnet, Point Transformer, ShellNet, ResPointNet++,

SEP network, GLSNet, DbNet, 3DLEB-Net, RandLA-Net, IAGMLP, RFFS-Net, TriangleConv, DSNet, SPGraph, PointCNN, IBPCS and CNN were used in about 24% of reviewed publications in the construction field, and 17% for non-construction fields.

5.2 Metrics

Figure 2 (b) highlights that recent techniques have mostly employed metrics like Precision, Recall, F1-Score, and Intersection over Union (IoU) to assess performance. Accuracy, measured as a ratio of correct predictions to the total number of predictions, is a dominant metric used in the Category 3 papers published in the construction domain, where it is used in 30% of the reviewed publications. IoU was utilized to measure efficiency in 18% of these construction-related publications. Precision, Recall and F1 score were employed in an average of 14% of the publications for Category in the construction field.

In the case of the non-construction domain, there is an increased reliance on Accuracy, with 59% of the reviewed papers using this metric to evaluate the performance of their algorithms. Unlike the construction domain, research in non-construction fields utilized Recall, Precision and F1 scores in about 5% of the reviewed publications. In general, compared to the construction-related research, there is a similar utilization of the metrics that are grouped as others (Support, MAE, FLOPs, Success, R2, CM) in the non-construction.

5.3 Challenges

Despite providing highly accurate results, Deep Learning techniques have limitations. For instance, they heavily rely on the diversity and completeness of the initial training dataset for them to produce accurate results, as well as correctly labeled training datasets [9]. Hu et al. [10] demonstrated the negative impact of training a model on a dataset that contains geometries that are different from those in the scan location. Their segmentation technique utilizing ResPointNet++ was successful in scoring mean IoU (mIoU) values greater than 90% across all the classes. However, the mIoU for the segmentation of points belonging to chairs was 62.55%. The authors note that this low mIoU can be attributed to the dataset used for training the model. The dataset contained data for a specific kind of chair, whereas the scan site had chairs with different geometries. In order to account for this, the geometries of the building components could be identified before the scan, and the model could be trained on a dataset that contains information on the geometries of commonly occurring classes in a scan location. Alternatively, a more robust approach could include training datasets that are comprised of various generic chairs that have a higher probability of being found on a scan site.

Additionally, a common limitation that has been identified is the computational load required for segmentation [11]. The effects of high computational load can be mitigated through the utilization of voxelization [12] and octrees [13]. By reducing the access time for each data point and the number of data points, these techniques were capable of increasing computational efficiency.

Based on the construction-related publications that were considered, four key limitation categories have been identified in Table 2. First, some techniques' effectiveness is contingent on the quality of the point cloud; factors like point cloud density can significantly influence segmentation outcomes. The impact of noise and occlusion is another concern, as some algorithms are severely affected by these elements.

The second key limitation is the ability to handle complex geometry. While certain techniques perform well in segmenting planar objects or familiar shapes like furniture and vehicles, they are limited when segmenting objects on construction sites. The primary reason for this is that construction often involves irregular shapes, such as cylindrical forms or structures with unconventional geometry, like unfinished elements or formwork systems. The capability to accurately segment undefined shapes is crucial in construction applications. Many techniques are effective in specific scenarios but require extensive testing and fine-tuning when dealing with complex shapes.

The third limitation category is the dependence on synthetic point clouds, predominantly from BIM. Although BIM offers a detailed representation of element geometry, it usually does not align with real-world scenarios, especially in terms of texture and the geometric presence of unfinished or temporary objects on construction sites. For example, casting concrete requires formwork, and if the concrete element is elevated, it will require scaffoldings. These elements (i.e., formworks and scaffoldings) need to be accurately represented in the BIM to achieve an accurate representation of the construction site using a synthetic point cloud generated from the BIM. Moreover, this becomes even more pronounced when the specific surface texture of materials is relevant to explaining the status of the construction. This is because the renderings from the BIM are not sufficiently close to the texture of the real material at a given time. Therefore, a synthetic dataset might not be comprehensive enough to be used on the construction site.

The last category is a limitation related to the manual intervention requirements. Given the complexity of construction, complete automation would be beneficial to bypass time-consuming data processing steps. Hence, the need for manual intervention during segmentation is considered a significant limitation.

6 Limitations and future work

Despite having provided a comprehensive and systematic review of the different techniques used for point cloud segmentation, the methodology presents some limitations. Given the volume of handled data, the study had to be based on keyword extraction and evaluation of the abstract and author-provided keywords, which in most cases is sufficient to provide an overall idea of the study's objective and methodology. But in some cases, said keywords are not going to be well defined, and it could lead to mislabeling said studies. A more in-depth analysis would be needed for more accurate results.

Table 2. Summary of the 4 categories identified for the challenges.

No.	Categories	Challenges	References	
1		Dependence on point cloud density	[14], [15]	
	Point cloud quality	Handling and effect of noise	[16]	
		Handling and effect of occlusion	[17], [18]	
2		Limited to common geometric shapes	[10], [19]	
		Limited to planar surfaces	[12], [13], [20], [21]	
	Complex geometries	Not experimented with complex shaped elements	[22], [23], [24]	
		Technique trained for specific objective and set of elements	[25]	
3	Based on a	[26], [27]		
4	Manual inter	[26], [28]		

This study revealed the main challenges in current point cloud segmenting methods for applications in the construction domain. In practice, despite the increasing adoption of technology in construction sites, there is still a significant reliance on manual and hybrid (semiinformation automated) construction processing. Therefore, there is still more work to do in reshaping the proposed methods to be feasible for construction industry utilization. In addressing the challenges, recent advancements in generative algorithms, such as Large Vision Language modeling, and spatial computing technologies, such as Neural Radiance Fields and Gaussian Splatting, could offer solutions to many of the challenges outlined in Table 2. Future research should specifically investigate these methods for practical applications in construction sites, evaluate whether the limitations are adequately addressed, and recommend directions for future work.

Given the space limitations, not all the required data for a detailed comparison could be added to the study. For future work, a more extensive study will be done, providing a more detailed comparison and description of the different evaluated techniques, including a comparison of their effectiveness, especially in realworld construction scenarios and providing technical differences between the techniques and their mean performance/evaluation metrics, an objective comparison is provided.

7 Conclusion

This study provides a review of the literature on point cloud segmentation techniques with a focus on construction applications. The study examined the popularity of specific techniques over time and evaluated the implementation trends in the construction industry and non-construction sectors. A key observation is the rapid advancement of deep learning techniques in nonconstruction applications in recent years, while construction-related applications still predominantly rely on shallow learning or fitting-based techniques. Based on the reviewed construction publications, PointNet, PointNet++ and DGCNN are found to be dominant in deep learning-based techniques. This study also has revealed that out of the 72 reviewed construction-related articles, 30% employed Accuracy, 18% utilized IoU, 15% applied Recall, 14% used the F1 score, and 13% relied on Precision. Finally, the study presented a summary of limitations in the considered set of construction-related publications. The limitations were classified into four main areas such as working with lowquality point clouds, the necessity for manual intervention, reliance on synthetic data, and the capability to segment complex geometries, such as the elements in the construction site.

Acknowledgment

This work was partially supported by the Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES) Center, funded by Tamkeen under the NYUAD Research Institute Award CG013. This work also benefited from the collaboration with the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001.

References

- [1] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, Y. Bai, and A. Behnood, "3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review," *Advanced Engineering Informatics*, vol. 51, p. 101501, Jan. 2022, doi: 10.1016/j.aei.2021.101501.
- [2] Q. Wang, Y. Tan, and Z. Mei, "Computational Methods of Acquisition and Processing of 3D Point Cloud Data for Construction Applications," *Arch Computat Methods Eng*, vol. 27, no. 2, pp. 479–499, Apr. 2020, doi: 10.1007/s11831-019-09320-4.
- [3] 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*, vol. 39, pp. 306–319, Jan. 2019, doi: 10.1016/j.aei.2019.02.007.
- [4] "Matterport. Capture, share, and collaborate the built world in immersive 3D." Accessed: Dec. 01, 2023.
 [Online]. Available: https://matterport.com/
- [5] B. P. Maldonado, N. W. Hayes, and D. Hun, "Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) using Machine Learning," presented at the 40th International Symposium on Automation and Robotics in Construction, Chennai, India, Jul. 2023. doi: 10.22260/ISARC2023/0009.
- [6] B. Kitchenham, "Procedures for Performing Systematic Reviews".
- [7] "The International Association for Automation and Robotics in Construction (ISARC) publications database." Accessed: Mar. 11, 2024. [Online]. Available: https://www.iaarc.org/publications/search-php
- [8] "SMART@NYUAD Github database." Accessed: Dec. 06, 2023. [Online]. Available: https://github.com/SMART-NYUAD/ISARC2024 lit review data/tree/main
- [9] Y. Chen, S. Kim, Y. Ahn, and Y. K. Cho, "A Framework of Reconstructing Piping Systems on Class-imbalanced 3D Point Cloud Data from Construction Sites," presented at the 40th International Symposium on Automation and Robotics in Construction, Chennai, India, Jul. 2023. doi: 10.22260/ISARC2023/0058.
- [10] D. Hu, V. J. L. Gan, and C. Yin, "Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors," *Automation in Construction*, vol. 152, p. 104949, Aug. 2023, doi: 10.1016/j.autcon.2023.104949.
- [11] Y. Yajima, S. Kim, J. D. Chen, and Y. Cho, "Fast Online Incremental Segmentation of 3D Point Clouds from

Disaster Sites," presented at the 38th International Symposium on Automation and Robotics in Construction, Dubai, UAE, Nov. 2021. doi: 10.22260/ISARC2021/0048.

- [12] G. Cheng, J. Liu, D. Li, and Y. F. Chen, "Semi-Automated BIM Reconstruction of Full-Scale Space Frames with Spherical and Cylindrical Components Based on Terrestrial Laser Scanning," *Remote Sensing*, vol. 15, no. 11, 2023, doi: 10.3390/rs15112806.
- [13] F. Poux and R. Billen, "Voxel-based 3D Point Cloud Semantic Segmentation: Unsupervised Geometric and Relationship Featuring vs Deep Learning Methods," *IJGI*, vol. 8, no. 5, p. 213, May 2019, doi: 10.3390/ijgi8050213.
- [14] T. Xia, J. Yang, and L. Chen, "Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning," *Automation in Construction*, vol. 133, p. 103992, Jan. 2022, doi: 10.1016/j.autcon.2021.103992.
- [15] R. Pierdicca *et al.*, "Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage," *Remote Sensing*, vol. 12, no. 6, p. 1005, Mar. 2020, doi: 10.3390/rs12061005.
- [16] Y. Xu, S. Tuttas, L. Hoegner, and U. Stilla, "Voxelbased segmentation of 3D point clouds from construction sites using a probabilistic connectivity model," *Pattern Recognition Letters*, vol. 102, pp. 67– 74, Jan. 2018, doi: 10.1016/j.patrec.2017.12.016.
- [17] M. Kamari and Y. Ham, "Vision-based volumetric measurements via deep learning-based point cloud segmentation for material management in jobsites," *Automation in Construction*, vol. 121, p. 103430, Jan. 2021, doi: 10.1016/j.autcon.2020.103430.
- [18] Z. Chen, E. Che, F. Li, M. Olsen, and Y. Turkan, "Webbased Deep Segmentation of Indoor Point Clouds," presented at the 36th International Symposium on Automation and Robotics in Construction, Banff, AB, Canada, May 2019. doi: 10.22260/ISARC2019/0074.
- [19] J. Grandio, B. Riveiro, M. Soilán, and P. Arias, "Point cloud semantic segmentation of complex railway environments using deep learning," *Automation in Construction*, vol. 141, p. 104425, Sep. 2022, doi: 10.1016/j.autcon.2022.104425.
- [20] P. Wei, L. Yan, H. Xie, and M. Huang, "Automatic coarse registration of point clouds using plane contour shape descriptor and topological graph voting," *Automation in Construction*, vol. 134, p. 104055, Feb. 2022, doi: 10.1016/j.autcon.2021.104055.
- [21] W. Yue, J. Lu, W. Zhou, and Y. Miao, "A new plane segmentation method of point cloud based on mean shift and RANSAC," in 2018 Chinese Control And Decision Conference (CCDC), Shenyang: IEEE, Jun. 2018, pp. 1658–1663. doi: 10.1109/CCDC.2018.8407394.
- [22] J. Martens and J. Blankenbach, "VOX2BIM+ A Fast and Robust Approach for Automated Indoor Point Cloud Segmentation and Building Model Generation," *PFG -Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 91, no. 4, pp. 273–294, 2023, doi: 10.1007/s41064-023-00243-1.
- [23] R. Zeng, J. J. S. Shi, C. Wang, and T. Lu, "Integrating as-built BIM model from point cloud data in

construction projects," *ECAM*, Jul. 2023, doi: 10.1108/ECAM-12-2022-1196.

- [24] F. Poux, C. Mattes, Z. Selman, and L. Kobbelt, "Automatic region-growing system for the segmentation of large point clouds," *Automation in Construction*, vol. 138, p. 104250, Jun. 2022, doi: 10.1016/j.autcon.2022.104250.
- [25] J. S. Lee, J. Park, and Y.-M. Ryu, "Semantic segmentation of bridge components based on hierarchical point cloud model," *Automation in Construction*, vol. 130, p. 103847, Oct. 2021, doi: 10.1016/j.autcon.2021.103847.
- [26] H. X. Zhang and Z. Zou, "Quality assurance for building components through point cloud segmentation leveraging synthetic data," *Automation in Construction*, vol. 155, p. 105045, Nov. 2023, doi: 10.1016/j.autcon.2023.105045.
- [27] J. W. Ma, T. Czerniawski, and F. Leite, "Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds," *Automation in Construction*, vol. 113, p. 103144, May 2020, doi: 10.1016/j.autcon.2020.103144.
- [28] C. Fotsing, N. Menadjou, and C. Bobda, "Iterative closest point for accurate plane detection in unorganized point clouds," *Automation in Construction*, vol. 125, p. 103610, May 2021, doi: 10.1016/j.autcon.2021.103610.

Monitoring of 3D concrete printing quality through multi-view RGB-D images

Ahmed Magdy Ahmed Zaki¹, Marco Carnevale², Hermes Giberti² and Christian Schlette¹

¹SDU Center for large structure production, University of Southern Denmark, Odense Denmark ²Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, Pavia, Italy

ahza@mmmi.sdu.dk,marco.carnevale@unipv.it,hermes.giberti@unipv.it,chsch@mmmi.sdu.dk

Abstract -

Large-scale 3D printing using concrete is a promising technology in the construction industry. In recent years, with the introduction of new commercial products and some realworld building projects, 3D concrete printing is moving from experimental and lab-scale applications towards regular construction operations in the field. In this paper, a solution for quality monitoring of 3D concrete printed structures after printing is shown. The solution is based on the use of multiview RGB-D images captured using a low-cost stereo-depth camera. The objects considered in the paper are 3D printed using a gantry 3D concrete printer. The object scanning process is described. Then the 3D reconstruction of the printed structure 3D model is explained. Finally, a quality assessment technique to evaluate the accuracy of the printed structure is introduced. This technique to quantify the printing error, compares the reconstructed 3D model of the actual state of the print and the CAD model used in the printing planning.

Keywords -

3D printing, concrete additive manufacturing, 3D vision, multi-view reconstruction

1 Introduction

3D concrete printing (3DCP) is a highly growing trend in the construction field [1]. This is due to the many advantages of 3DCP in comparison to traditional techniques [2] as it increases construction speed, decreases material waste, and enhances flexibility to construct complex shapes without the need for specific and expensive formwork.

On-site, the printing quality is highly dependent on many factors. The most important of them is the operator's experience in setting up the machine and dialing the right fundamental printing parameters - i.e. the movement speed and material flow rate, which accurately match the used material mix and the current environmental conditions such as temperature and humidity. Additionally, the printing quality can be affected by basic machine accuracy and fluctuations in the quality of the extruded material and

external factors.

Post-printing quality monitoring is necessary to evaluate the mechanical characteristics of the print. The presence of extra material, material leakage, first layers tearing, or layers closure pattern can affect the object's geometrical accuracy.

In literature, many research works have focused on technological aspects related to the optimization of the printing process like the construction of the printing machine itself which could be a gantry system or robotic manipulators. Other investigated problems are the optimal material mix, finding printing parameters, and evaluation of mechanical characteristics of the printed parts at their wet and hardened states. While there is a lack of systematic approaches that defines the necessary tools and methods to monitor and evaluate the 3D concrete printed objects.

The goal of this paper is to introduce a methodology for quality inspection of 3D concrete printed objects. This is achieved by quantifying the printing error of cured prints by comparing their reconstructed 3D model "as-built" from multi-view RGB-D images against its initial CAD model "as-planned" which was used for the generation of movement commands for the printing process. In figure 1, an example of the multi-view images of a polygon-shaped 3D concrete printed object are shown. These images are used to reconstruct the 3D model shown in figure 2 that could be compared with the object CAD model for error assessment.



Figure 1. Example of multi-view images

The paper is structured as follows. In Chapter 2, related work is summarized. In chapter 3, the 3D reconstruction



Figure 2. Output: 3D model

technique and the 3D model elaboration are described. In chapter 4, the experimental setup and test results are shown. In chapter 5 the conclusions are shown.

2 Related work

In the field of 3D concrete printing, few systematic approaches were proposed for digital quality inspection of the printed objects. While in the construction field, digital measurement tools are becoming essential for quality evaluation. Several techniques [3, 4] were introduced for quality evaluation based on the use of multi-view images and point clouds of concrete buildings and roads.

The most used form of digital information is the point cloud [5], that are used for several purposes such as 3D model reconstruction, geometry quality inspection, and construction progress tracking. These point clouds are mostly captured using laser scanners or being generated elaborating images or videos of the considered object.

Laser scanner [6] and structured light sensors [7] are the most used in application requiring accurate measurements. These sensors present some limitations like high prices and low portability due to their big size.

Recently, computer vision-based quality monitoring techniques are gaining popularity thanks to the fast development of low-cost, accurate, and small vision sensors and the development of machine learning techniques. These machine-learning techniques are mostly based on the elaboration of color images.

Techniques based on the 3D reconstruction using multiview images are commonly used for quality monitoring in civil infrastructures, for example, roads [4, 8] and building [3, 9], having big dimensions where low accuracy margin is acceptable. While for more accurate quality inspection multi-view solutions based on laser scanners are preferred [10, 11].

In our previous paper [12], we have discussed two 3D reconstruction techniques based on the use of multi-view RGB-D images. These images are captured by a low-cost stereo depth camera (Realsense D415 camera) attached to the end-effector of a six-axis collaborative robot. One

of the introduced techniques, odometry-based technique, does not require previous pose knowledge at each viewpoint. It is based on the estimation of the 3D camera pose capturing each of the RGB-D images. These poses are then used for the 3D model reconstruction of the covered workpiece visible in the captured images. In the present work, a modified version of the odometry-based technique is applied to reconstruct the 3D model of concrete printed objects. With respect to the previously introduced work, only considering workpiece to be placed within the workspace of a robotic manipulator, in the present work the algorithm has been modified to be able to deal with bigger objects (typical of 3D printed concrete structure) and a higher number of images.

In this paper we propose a quality monitoring technique based on the comparison between the reconstructed 3D model of a 3D printed object and its original CAD model.

3 Multi-view 3D reconstruction and 3d model elaboration

In this section, the approach used for the 3D construction process of an unknown object is introduced. The solution is based on matching the RGB-D multi-view images to estimate the camera pose while capturing every image. Knowing the pose, color, and depth information at every viewpoint of the object, it is possible to integrate all this information to create a 3D model of the object. This technique was initially introduced for indoor scene reconstruction as shown in [13]. A variation of the algorithm has been introduced in our previous works [12, 14]. The proposed solution, in the previous papers, is used to reconstruct the 3D model of a workpiece considered in a contact-based robotic application. In the current paper, the algorithm is modified and applied in the field of 3D concrete printing considering relatively bigger objects that are scanned in a higher number of multi-view 3D images. The summary of the proposed approach is shown in figure 3. Step 1 is to capture multi-view RGBD images of the object. Step 2 is to elaborate the images to relate them to common reference. Step 3 is to integrate the images content to reconstruct the 3D model of the object. Step 4 is to extract only the object from the 3D model. Step 5 is to compare the reconstructed 3D model with the CAD model of the object to determine the printing error. The steps are described in details as follows.

To capture several images from different perspectives that cover all the objects or at least the interested area. In this work, the Realsense D415 [15] which is a low-cost 3D stereo depth camera is used.

The D415 stereo depth camera captures the RGB-D images using different sensors. The color image is captured by an RGB sensor having a resolution of up to 1920 X 1080 pixels and is able to capture up to 30 frames per



Figure 3. Proposed solution summary

second (fps).

In stereo cameras such as Realsense D415, the depth image is captured using two sensors, and the depth information is perceived by comparing the position of the pixels in the two images. The depth value of each pixel is calculated using triangulation methods considering the known physical distance between the two sensors. The depth image captured has a resolution of up to 1280 X 720 pixels at a frame rate of up to 90 fps.

Due to the fact that the RGB-D images are captured using different physical sensors, it is necessary to align the images. The alignment process consists of changing the parameters such as dimension and coordinate system origin of an image (color or depth) to match the other one. This step is necessary to have both images with the same dimension and refer to the same coordinate system. In such a way, a pixel in one image has the same exact position in the other image.

The multi-view RGB-D images are captured by a moving camera. That means that every image is referred to the camera coordinate system at that moment. The first step in the developed algorithm is the estimation of the camera pose of each image using the RGB-D odometry technique introduced in [16]. Every image is compared to the consecutive one to calculate the pose change.

The estimation of the pose change between two images, consists in the calculation of the homogeneous transformation matrix. If applied to one image, it matches it to the other one like it has been captured from the same position and same camera orientation.

The calculated camera motion matrices, are used to refer all the images to a common coordinate system. The common reference frame used is the frame of the first image. The referring process consists in the use of the camera motions found for every image to refer the content of the image to be with respect to the common reference frame.

3.1 Scanning process and 3D model reconstruction

The quality of the reconstructed 3D model is highly dependent on the input data. The color and depth sensors are set to the maximum resolution of 1920 X 1080 pixels and 1280 X 720 pixels respectively.

The depth measurement accuracy is dependent also on the distance between the observed object and the camera at the moment of capturing the image. For the resolution used, the optimal range for depth accuracy is between the minimum of 450 mm and the maximum of two meters. This range guarantees the accuracy of the readings with errors lower than 2% of the total distance between the object and the camera.

The 3D reconstruction technique used, is based on matching the images and comparing the overlapped parts in them. To guarantee that, the frame rate or the number of images that the camera can capture at every second and the camera movement have to be related to have the time necessary to capture the images covering all the parts of the object without large movement between sequential images. In this work, the camera is moved manually. Where the user moves the 3D stereo depth camera around the object to be scanned and capture RGB-D images of all the interested areas. The relationship between the camera movement speed and number of the RGB-D images captured is explained in our previous work [12].

The considered work-pieces in this paper have dimensions between one meter and two meters, in terms of length and width. The height is lower than 0.5 meters. To reconstruct the required 3D models, a dataset of 100 RGB-D images is used. In case of having a data set higher than the fragment size, the data set is partitioned in patches of 100 RGB-D images or less.

3.2 3D model elaboration and object extraction

The quality evaluation of a 3DCP object, which may not coincide perfectly with the starting CAD model used for generating the printing trajectory, could be done using the explained algorithm for the 3D model reconstruction of the object in its dry state.

The quality evaluation procedure, consists of the reconstruction of the 3D model of the observed scene containing the printed object. The 3D model is then cropped to extract only the 3D model of the object. The extracted part is then compared to the initial CAD model of the object to evaluate its quality to find the zones having similarities and zones having printing defects. The process in detail is explained in the following.

To extract only the object that we are interested in, it is necessary to make some assumptions about the scanning and the printing processes. The first assumption is that the object is positioned, in most of the RGB-D images, in the center of the covered area. The second assumption is that the printing process happened over a flat surface.

With the first assumption, the search algorithm could consider only the center of the constructed 3D model of the scene. That means only a part of the model has to be analyzed. The second assumption allows us to consider that in the central area could be found the 3D printed object and the floor on which it was printed. Searching in the model for a plane surface (points having the same height and connected together), it is possible to find all points combining the floor. By removing these points, it is possible to obtain the points of the object.

3.3 Registration of CAD and scanned point-cloud

To evaluate the reconstructed 3D model accuracy of the 3DCP object, it is compared with the CAD model. The comparison accuracy depends on the ability to compare an exact point in the scanned point cloud to the exact relative point in the 3D model. To guarantee good comparison accuracy the two point clouds have to be aligned.

Alignment of point clouds, known as the point cloud registration process, is a well-known problem in computer vision and is widely used in construction to compare and inspect buildings, roads, and civil infrastructures using laser scanners or similar scanning techniques. The 3D registration process consists of finding the transformation matrix that if applied to the second model aligns it to the first one and also refers both of the models with respect to the same reference coordinate system.

Different registration techniques can be applied that can be categorized into coarse and fine registrations [17]. Coarse registration is a feature-based technique that matches the features in the two point-clouds. The most used strategies are classified as point-based, line-based, and surface-based. These methods are very accurate but highly dependent on the similarities level of the two pointclouds that is dependent on the sensor noise, scanning point of view, point-clouds density, and geometrical nonconformity due to fabrication process accuracy. To guarantee high accuracy, complicated feature extraction tools may be necessary to develop.

Fine registration techniques are based on approximate iterative processes to find the optimal rigid transformation matrix between the two point clouds. The most used technique is Iterative Closest Points ICP [18, 19] and its variations. These techniques are based on the minimization of positional errors of relative point sets selected from the two point clouds. This implies the necessity for a good initial guess of the rigid body transformation to avoid local minimum problems. Elaboration time is dependent on the point-cloud size.

In the case of 3DCP objects, using the above techniques leads to registration failure for the following reasons. The low geometrical conformity of the printed object due to printing accuracy tolerance. Also difficult to identify features like lines or surfaces in the noisy scanned point cloud. Finally, lack of initial transformation guesses to be used for ICP techniques as the scanning process is done manually and the camera origin is not referred to a known reference frame.

To overcome these limitations, we propose an easy and fast-to-apply registration technique to align the two point clouds based on more general features considering the overall point clouds. These features are the boundary box and the three-dimensional center of the object. Aligning these features allows an alignment of the two 3D models.

4 Experimental setup and scanning results

In this section, the introduced 3D reconstruction algorithm is applied to reconstruct the 3D model of a 3D printed concrete object. The experimental setup used is a modular gantry system 3D concrete printer. As a first step of the analysis, the model reconstruction has been tested with images gathered by manual scanning: a 3D camera is moved around the object to capture the multi-view RGB-D images. Scanning results in terms of the reconstructed 3D models are shown together with an example of accuracy measurement of the reconstructed 3D model.

4.1 Gantry systems 3D concrete printer

Experiments done in this paper are done in collaboration with COBOD International A/S. It is a company providing innovative solutions in the construction field combining 3D printing and robotics technologies to automate the construction process. Printing solutions could be based on the use of robotic manipulators or gantry systems. The main product is BOD2 shown in figure 4 which is a modular gantry system that can have different sizes based on the size and shape of the building needed to construct. The smallest version BOD2 2-2-2 has a print area of 4.52 x 4.55 x 3.09 m. The biggest is BOD2 5-10-4 which has a print area of about 12.10 x 24.75 x 8.14 m.

4.2 3D reconstruction results and quality assessment

The 3D concrete printed object conformity with the design ("as-planned" CAD model) is an essential crite-



(a) Smallest printer BOD2 2-2-2



(b) Biggest printer BOD2 5-10-4

Figure 4. BOD2 models https://cobod.com/ bod2/

rion for its validation, i.e. are mechanical rigidity and the possibility of assembling the printed object to other components highly affected by accurate, conform 3DCP execution according to the planned geometry. To check the geometrical conformity, in this section a 3D printed concrete object is scanned and the proposed 3D reconstruction algorithm is used to reconstruct its 3D model that allows for error quantification.

Using RGB-D odometry technique previously described in [14], a set of 100 color and depth images, similar to those shown in figure 1 for the polygon shape object, are integrated to reconstruct the 3D model shown in figure 2. The developed 3D reconstruction pipeline is applied also to scan the ring shaped 3DCP object shown in figure 5. The ring-shaped object has an outer diameter of 0.8 meters and a height of 0.1 meters.

3D concrete printing process, similar to small dimension and traditional PLA 3D printing process, may lead to geometry deformation related to the first layers adhesion failure or to the layer closure area. The goal is to quantify the highlighted errors in correspondence to the layer closure and first-layer adhesion. The 3D reconstructed model of the object is shown in the sub-figure 6b.



Figure 5. Ring shaped object

Hausdorff Distance is used to compare the 3D reconstructed model of the print's actual result and the ground truth or the 3D CAD model of the object used to generate the trajectory of the 3DCP.

Hausdorff distance measures the magnitude of the biggest printing defect that may cause low resemblance between the printed object and the CAD model. To calculate Hausdorff distance it is necessary to calculate all the distances between relative points in the two point-clouds and find the maximum of them.

To calculate the Hausdorff Distance Meshlab [20] is used. Meshlab is an open-source 3D model elaboration software. It allows us to calculate, between two aligned and down-sampled point clouds, maximum distance (Hausdorff distance), minimum distance, distance mean value, and Root Mean Square error.

The two 3D models of the ring shaped object are shown in figure 6. The sub-figure 6a represents the CAD model or the ground truth to which the reconstructed 3D model shown in the sub-figure 6b is compared.

Comparison results are shown in the figure 7. The Hausdorff algorithm is applied over a total of 159039 relative points. The error is represented as a color-map. Lower values of errors, distances between relative points in the two 3D models, are represented in orange color. Points where higher error occurs are represented slight green while the maximum error is represented by a blue color. For better understanding of the error measurement results, on the left side of the figure, a histogram representation is used to show the quantity of points having different error values. Most of points, especially the points that are part of the upper surface are having the lowest error values. Considering all points, error mean value of 0.0057 meters and error Root Mean Square of 0.0073 meters are obtained. The printing defect in correspondence to the layer







(b) Constructed 3D model

Figure 6. Comparison between CAD model and output 3D model of the actual print



Figure 7. Hausdorff distance comparison applied to the ring case

closure is where the maximum error is obtained. Error value is 0.039 meters. A higher error values are obtained in correspondence to the first layer adhesion failure.

5 conclusion

In this paper, a vision-based approach for the quality monitoring of 3D printed concrete objects is proposed. The solution consists of the use of a low-cost stereo-depth camera to capture multi-view images to reconstruct its 3D model. To quantify the accuracy of the printed object, the 3D reconstructed model "as-built" is compared to its "as-planned" CAD model. The developed algorithm for the 3D reconstruction works in two configurations. In the first configuration, the 3D camera is moved manually around the object. This configuration could be used for either onsite or offsite quality monitoring since the camera is manually moved between viewpoints. The second configuration for automatic scanning is mostly used for onsite quality monitoring where the camera is attached to the printer and moved around it to scan the object without the need to move the object from its printing position. The viewpoint in this latter case is generated automatically knowing the position of the printed object. To quantify the accuracy, the reconstructed 3D model and the initial CAD model are compared using the Hausdorff Distance function. Future work is to exploit quality monitoring results to highlight the defects of the 3D printed concrete object and to generate instructions to adjust them, e.g. by controlling the printer to fix-up gaps with additional material or to direct workers to surface finishing and grinding operations where over-extrusion or extra material is present.

References

- [1] Guowei Ma, Richard Buswell, Wilson Ricardo Leal da Silva, Li Wang, Jie Xu, and Scott Z. Jones. Technology readiness: A global snapshot of 3d concrete printing and the frontiers for development. *Cement and Concrete Research*, 156:106774, 2022. ISSN 0008-8846. doi:https://doi.org/10.1016/j.cemconres.2022.106774. URL https://www.sciencedirect.com/ science/article/pii/S0008884622000655.
- [2] Ayesha Siddika, Md. Abdullah Al Mamun, Wahid Ferdous, Ashish Kumer Saha, and Rayed Alyousef. 3d-printed concrete: applications, performance, and challenges. *Journal of Sustainable Cement-Based Materials*, 9(3):127–164, 2020. doi:10.1080/21650373.2019.1705199. URL https://doi.org/10.1080/21650373.2019. 1705199.
- [3] Fei Dai and Ming Lu. Assessing the accuracy of applying photogrammetry to take geometric measurements on building products. *Journal of Construction Engineering and Management*, 136(2):242–250, 2010. doi:10.1061/(ASCE)CO.1943-7862.0000114.

URL https://ascelibrary.org/doi/abs/10. 1061/28ASCE29C0.1943-7862.0000114.

- [4] Laura Inzerillo, Gaetano Di Mino, and Ronald Roberts. Image-based 3d reconstruction using traditional and uav datasets for analysis of road pavement distress. *Automation in Construction*, 96:457–469, 2018. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2018.10.010. URL https://www.sciencedirect.com/ science/article/pii/S0926580518304047.
- [5] Qian Wang and Min-Koo 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. ISSN 1474-0346. doi:https://doi.org/10.1016/j.aei.2019.02.007. URL https://www.sciencedirect.com/science/ article/pii/S1474034618304683.
- [6] Daniel Huber, Burcu Akinci, Pingbo Tang, Antonio Adan, Brian Okorn, and Xuehan Xiong. Using laser scanners for modeling and analysis in architecture, engineering, and construction. In 2010 44th Annual Conference on Information Sciences and Systems (CISS), pages 1–6, 2010. doi:10.1109/CISS.2010.5464818.
- [7] Kwangwoo Wi, Vignesh Suresh, Kejin Wang, Beiwen Li, and Hantang Qin. Quantifying quality of 3d printed clay objects using a 3d structured light scanning system. *Additive Manufacturing*, 32:100987, 2020. ISSN 2214-8604. doi:https://doi.org/10.1016/j.addma.2019.100987. URL https://www.sciencedirect.com/ science/article/pii/S2214860419312023.
- [8] Jaeheum Yeon, Julian Kang, and Wei Yan. Spall damage repair using 3d printing technology. Automation in Construc-ISSN 0926-5805. tion, 89:266–274, 2018. doi:https://doi.org/10.1016/j.autcon.2018.02.003. URL https://www.sciencedirect.com/ science/article/pii/S0926580517307392.
- [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:https://doi.org/10.1016/j.autcon.2018.05.002. URL https://www.sciencedirect.com/ science/article/pii/S0926580518300165.
- [10] Jie Xu, Richard A. Buswell, Peter Kinnell, Istvan Biro, John Hodgson, Nikolaos Konstantinidis, and

Lieyun Ding. Inspecting manufacturing precision of 3d printed concrete parts based on geometric dimensioning and tolerancing. *Automation in Construction*, 117:103233, 2020. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2020.103233. URL https://www.sciencedirect.com/ science/article/pii/S0926580519311070.

- [11] Richard Buswell, Jie Xu, Daniel De Becker, James Dobrzanski, John Provis, John Temitope Kolawole, and Peter Kinnell. Geometric quality assurance for 3d concrete printing and hybrid construction manufacturing using a standardised test part for benchmarking capability. *Cement and Concrete Research*, 156:106773, 2022. ISSN 0008-8846. doi:https://doi.org/10.1016/j.cemconres.2022.106773. URL https://www.sciencedirect.com/ science/article/pii/S0008884622000643.
- [12] Ahmed Magdy Ahmed Zaki, Marco Carnevale, Christian Schlette, and Hermes Giberti. On the use of low-cost 3d stereo depth camera to drive robot trajectories in contact-based applications. *The International Journal of Advanced Manufacturing Technology*, pages 1–15, 2023.
- [13] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5556–5565, 2015. doi:10.1109/CVPR.2015.7299195.
- [14] Ahmed Magdy Ahmed Zaki, Marco Carnevale, Hermes Giberti, and Christian Schlette. Digital model reconstruction through 3d stereo depth camera: a faster method exploiting robot poses. *Procedia Computer Science*, 217:1542–1549, 2023. ISSN 1877-0509. doi:https://doi.org/10.1016/j.procs.2022.12.354. URL https://www.sciencedirect.com/science/article/pii/S1877050922024395. 4th International Conference on Industry 4.0 and Smart Manufacturing.
- [15] Intel® RealSenseTM Camera 400 Series (DS5) Product Family Datasheet. Intel Realsense, 1 2019. Revision 005.
- [16] Frank Steinbrücker, Jürgen Sturm, and Daniel Cremers. Real-time visual odometry from dense rgb-d images. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pages 719–722, 2011. doi:10.1109/ICCVW.2011.6130321.
- [17] Liang Cheng, Song Chen, Xiaoqiang Liu, Hao Xu, Yang Wu, Manchun Li, and Yanming Chen.

Registration of laser scanning point clouds: A review. *Sensors*, 18(5), 2018. ISSN 1424-8220. doi:10.3390/s18051641. URLhttps://www.mdpi.com/1424-8220/18/5/1641.

- [18] Jiaolong Yang, Hongdong Li, Dylan Campbell, and Yunde Jia. Go-icp: A globally optimal solution to 3d icp point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2241–2254, 2016. doi:10.1109/TPAMI.2015.2513405.
- [19] Hao Men, Biruk Gebre, and Kishore Pochiraju. Color point cloud registration with 4d icp algorithm. In 2011 IEEE International Conference on Robotics and Automation, pages 1511–1516, 2011. doi:10.1109/ICRA.2011.5980407.
- [20] Paolo Cignoni, Massimiliano Corsini, and Guido Ranzuglia. Meshlab: an open-source 3d mesh processing system. *ERCIM News*, 2008(73), 2008. URL http://dblp.uni-trier.de/db/journals/ ercim/ercim2008.html#CignoniCR08.

Robust High-Precision LiDAR Localization in Construction Environments

Andrew Yarovoi¹ Pengyu Mo² Yong Kwon Cho³

¹Woodruff School of Mechanical Engineering, Georgia Institute of Technology, USA ²School of Electrical and Computer Engineering, Georgia Institute of Technology, USA ³School of Civil and Environmental Engineering, Georgia Institute of Technology, USA

ayarovoi3@gatech.edu, pmo30@gatech.edu yong.cho@ce.gatech.edu

Abstract -

Accurate localization plays a crucial role in the effective operation of autonomous robotics systems, especially in dynamic environments such as construction sites. Simultaneous Localization and Mapping (SLAM) utilizing LiDAR sensors has emerged as a popular solution due to its ability to function without external infrastructure. However, existing algorithms exhibit significant shortcomings. Despite current methods achieving high accuracy over long trajectories, they struggle with precision and reliability in complex indoor environments. This paper introduces a novel feature-based LiDAR SLAM system designed to address these limitations and enhance short-term precision and overall robustness. The proposed system is evaluated using both existing datasets and a physical robot platform, addressing the limitations of current implementations and showcasing improved performance in challenging real-world scenarios, particularly in construction environments.

Keywords -

Simultaneous Localization and Mapping; SLAM; LiDAR, Localization; Navigation

1 Introduction

Localization is a critical component of nearly all autonomous robotics systems. Accurately understanding the robot's pose relative to its environment is often crucial for navigation and manipulation tasks. While GPS can sometimes be sufficient for outdoor environments, more complex approaches are typically needed for indoor environments due to significant signal attenuation by the building. In warehousing and manufacturing, autonomous robots are conventionally localized using visual markers placed in the environment or external camera systems with fixed positions. These approaches have enabled increased robotic automation which has lowered costs, increased workforce productivity, and improved efficiency [1]. However, these approaches are often not feasible in construction due to construction sites continuously evolving throughout the building process. This makes setting up external localization systems difficult.

One popular solution is to use onboard sensors to perform simultaneous localization and mapping (SLAM), allowing autonomous systems to localize in previously unmapped environments. SLAM can provide high-accuracy and precise positional estimates in indoor environments, without the need for any external infrastructure. One popular onboard sensor for performing SLAM is LiDAR, which typically uses the time of flight data of a laser to measure distances to nearby obstacles. These distance measurements are converted to a point cloud, providing a 3D representation of the environment. LiDAR has numerous advantages over cameras such as direct measurement of depth and invariance to lighting conditions. This eliminates scale ambiguities and allows for robust measurements, even in harsh or dim lighting conditions such as those commonly present in night-shift work operations. Additionally, LiDARs are not impacted by textureless or highly repetitive textured environments, which are common in construction (e.g. unfinished/unfurnished rooms, brick walls).

Over the last couple of decades, various algorithms and frameworks have been proposed for performing LiDARbased SLAM. One of the most popular and robust of the recent methods is LIO-SAM [2]. LIO-SAM uses a rangeimage generated from a point cloud to extract LOAM [3] features. It then uses frame-to-frame matching and frameto-global-map matching, along with GTSAM [4] to generate a 3D point cloud of the environment and localize the LiDAR within it. Additionally, LIO-SAM tightly couples the LiDAR and IMU, allowing for point cloud deskewing and full utilization of the IMU data in the factor graph optimization. Unlike other more recent point-based methods such as ART-SLAM [5], LIO-SAM is feature-based and thus more computationally efficient. This allows it to run on smaller robots without a GPU. As such, it has become a popular option for mobile robots and is used extensively in the research community [6, 7, 8, 9].

While LIO-SAM has been shown effective on multiple datasets, it has several limitations that make it sub-optimal for real-world localization of a navigating autonomous sys-



Figure 1. An overview of the proposed algorithm

tem. Despite achieving high long-term trajectory accuracy, LIO-SAM often provides poor short-term precision. This results in large short-term errors in the pose estimate that can lead to navigational issues and poor map quality. These issues are particularly evident when working with robotic systems that exhibit high vibrations (such as quadruped robots) or systems utilizing cheaper, lowerquality IMUs. Additionally, the algorithm lacks robustness in complex environments where limited visibility of the ground planes and LIO-SAM's lack of consideration for surface normals leads to weak constraints on the roll, pitch, and z-height. This can lead to complete loss of tracking in many confined indoor environments, such as stairways and small rooms. These limitations make current implementations unreliable in common construction environments, endangering the robot and the people around it.

To address these issues, we present a novel feature-based LiDAR-inertial SLAM system that significantly improves short-term precision and overall robustness. The proposed system is validated on existing datasets and a physical robotics platform.

2 Methodology

Figure 1 shows an overview of the different modules in the proposed algorithm. Unlike existing approaches, the proposed algorithm first organizes the cloud into a CxNarray (organized cloud), where C is the number of channels in the LiDAR, N is the number of samples taken per revolution of the LiDAR, and each element in the array is a point storing its x, y, z, and intensity values. This preprocessing step speeds up computations of later steps in the pipeline, enabling real-time performance. After organizing, the cloud is deskewed using IMU data to undo any distortions caused by rotations of the LiDAR sensor over the capture interval of the point cloud. This is necessary as LiDARs typically output their data as scans (one full rotation of the laser array) which contains points captured at different times over the scan period. Deskewing removes any distortions caused by rotations of the sensor over the scan period. The deskewed point cloud then undergoes

feature extraction where planar points (Figure 2) and edge points (Figure 3) are extracted from the cloud using novel proprietary feature extractors. Planar points are further clustered into individual planes. Unlike LIO-SAM which labels points as planar based on the local roughness of the range image, we consider the full 3D positions of each point and its neighbors to more reliably identify planar surfaces and remove non-planar points. Additionally, our approach efficiently estimates the normals of the points, providing us with richer features and additional information during feature matching and pose optimization. The individual planes, each containing a point cloud of their constituent points, along with an edge cloud comprising all the edge points, are subsequently assembled into a Frame object. This Frame is initialized with an initial pose derived from the last predicted pose and the IMU's orientation estimate.



Figure 2. Example of extracted plane points



Figure 3. Example of extracted edge points

An iterative process is then used to match the features and optimize the pose of the current frame with respect to the key frames. At each iteration, a 3-stage matching approach is used to first match the Frame to its nearest key frames, then match the Frame's planes to the selected key frames' planes, and finally match planar and edge points in the Frame to the planar and edge points stored in the matched planes and key frames' edge clouds. The feature matching step results in a set of planar and edge correspondences that are used to calculate multiple error metrics. These metrics are minimized with respect to the Frame's pose using Newton's method. The Frame's estimated pose is then updated and the iterations are repeated until we converge on a pose or have repeated a set number of iterations. The final pose of the LiDAR is returned as the final output of the system. If needed, the frame is added to the key frames to iteratively grow the map.

Ultimately, our approach differs from LIO-SAM in three main ways. First, we organize the cloud and use the full 3D positions of the points during feature extraction, instead of using the range image. The organization of the point cloud allows us to compute features in a comparable amount of time to LIO-SAM while utilizing more of the point cloud's information. Second, we do not use LOAM features, and instead use novel feature extractors that extract more reliable planes and edges. Our edge detector explicitly handles edges caused by occlusions and our plane detector efficiently estimates normals, providing more information for pose optimization. Our use of different feature extractors also enables us to utilize 3-stage feature matching which exploits the additional information captured by the features to improve matching accuracy. Lastly, our method uses key framing to efficiently represent the map, enabling frame-to-global-map matching for every frame. This results in a highly memory-efficient map representation and limits the short-term drift, improving the short-term precision of the localization estimate.

An additional difference in the current implementation is that we do not use a factor graph to optimize the global map or fuse IMU orientation estimates into our pose graph. This is a limitation of our current approach as it can lead to larger long-term drift for very long sequences. However, we plan to address this issue in future work by incorporating our design with GTSAM [4].

3 Experiments

To evaluate the performance of the proposed approach against a baseline, the algorithm is quantitatively compared to LIO-SAM on an existing construction site dataset, as well as qualitatively evaluated on physical robotics hardware. LIO-SAM provides a good baseline as it requires the same sensors (a 360-LiDAR and an IMU), is also featurebased, and requires similar computational resources. For numerical analysis, we primarily evaluate the accuracy of the localization estimates, as localization is the primary use case of real-time SLAM algorithms for most systems. Additionally, localization can be evaluated more directly than map quality and better localization directly leads to improved mapping.

3.1 Hilti Dataset

To evaluate the absolute positional accuracy, both algorithms were tested on some of the additional sequences provided by the Hilti SLAM Challenge 2022 dataset [10]. This dataset provides ROS bag recordings of LiDAR and IMU sequences collected from various construction and indoor environments using a Hesai Pandar XT-32 LiDAR. Vitally, the dataset also provides millimeter-accurate synchronized ground truth poses collected with a motioncapture system. This allows direct comparison of the predicted and ground truth trajectories.

To enable both algorithms to process the data, the point clouds were preprocessed to convert them into a Velodyne point format (does not change any data, but changes point format). Additionally, the IMU data was preprocessed as both algorithms expect the IMU to provide fused orientation estimates. These were generated using the opensource imu_filter_madgwick package [11]. To provide a wide range of difficulties and environments, 5 sequences from the dataset were tested. These are Exp04, Exp05, Exp06, Exp14, and Exp18.

Table 1. Dataset statistics for Hilti sequences calculated based on the ground truth poses

Dataset	Max Pitch / Roll (°)	Mean Ang Vel (°/s)	Mean Lin Vel (m/s)	Max Ang Vel (°/s)	Max Lin Vel (m/s)
Exp04	10.968	19.970	0.606	142.194	1.886
Exp05	17.324	17.110	0.561	146.696	1.616
Exp06	58.033	38.971	0.617	263.522	2.408

The first three scans were captured on three floors of a real-world indoor construction site with progressively more aggressive motions. The three datasets all include variations in Z-height, open and confined spaces, and tilting in both pitch and roll. Table 1 provides some statistics generated from the ground truth poses provided by the datasets. Exp06 in particular offers very fast motions and aggressive rotations. Both Exp04 and Exp05 were designated as easy difficulty by the dataset creators, while Exp06 was designated as medium difficulty due to the fast motions.

Captured in more demanding indoor settings, Exp14 and Exp18 both present challenging sequences with geometric ambiguity and confined spaces. Notably, Exp14 showcases a rectangular staircase, while Exp18 incorporates a spiral staircase at the beginning and end of the sequence, resulting in their classification as medium and hard difficulty, respectively.

Table 2. Results from the Hilti Sequences, N/A indicates an algorithm lost tracking and could not recover

Seq	Algorithm	Trans RMSE (m)	Trans SD (m)	Rot RMSE (°)	Rot SD (°)
Exp04	LIO-SAM Ours	0.1670 0.1147	0.0879 0.0566	1.486 0.966	0.355 0.493
E05	LIO-SAM	0.0945	0.0450	0.873	0.336
Exp05	Ours	0.1124	0.0564	1.818	1.055
Exp06	LIO-SAM Ours	0.3599 0.4825	0.2274 0.2578	2.258 4.098	1.053 1.707
Exp14	LIO-SAM Ours	N/A 0.7966	N/A 0.5292	N/A 3.985	N/A 1.559
Exp18	LIO-SAM Ours	N/A 0.7979	N/A 0.5713	N/A 10.433	N/A 3.854

The two algorithms were evaluated on all five sequences using a Ryzen 5600H processor, processing the data in real-time. The predicted poses were captured and stored in a text file, and then synchronized with the ground truth poses using timestamps. For evaluation, translation errors were calculated using the Euclidean distance between the estimated and ground truth trajectories. The rotation errors were calculated as the smallest angle between the estimated and ground truth orientations. The root-meansquare (RMSE) and the standard deviation (SD) of the translational and rotational errors are reported to evaluate the accuracy and consistency of the estimated trajectories in Table 2.



Figure 4. Predicted trajectory (red) versus ground truth (green) for Exp14 using our approach

As shown in Table 2, our approach achieved better results in 3 of the 5 sequences. Importantly, Exp14 and Exp18 show that our approach is more robust than LIO-SAM by completing the sequence. LIO-SAM on the other hand lost tracking in the first 5 seconds of both sequences due to starting in confined spaces and having to navigate stairways. Additionally, most of the errors in Exp14 occurred in the last few seconds due to the cloud being highly geometrically ambiguous as a result of the translational symmetry of the stairway (Figure 4). Exp18 also highlights the generalization capability of our approach to various built environments. While the other sequences primarily feature standard built environments, Exp18 was captured in a gallery with curved walls, ornate columns, and tight walkways. Despite there being few truly planar surfaces, our plane extractor was able to identify locally planar surfaces and outperformed LIO-SAM, completing the sequence.



Figure 5. Angled view of the predicted trajectory for Exp06 using our approach

Our algorithm also achieved better performance on Exp04. Upon initial inspection, it seems that LIO-SAM exhibited slightly better performance in Exp05 and Exp06. However, as shown in Figure 5 and Figure 6, our approach produced a smoother and more locally accurate trajectory, despite experiencing some additional drift in roll and pitch that resulted in comparable but slightly higher overall RMSE. This additional drift is due to how our algorithm integrates the IMU data into our pose estimates. Currently, our algorithm only utilizes the IMU data for initial pose estimation and point cloud deskewing. However, IMUs also generate attitude measurements, which provide absolute constraints on the roll and pitch of the system. Since our system does not fuse the IMU orientation estimates with our final predicted pose, our algorithm is more susceptible to long-term drift in the roll and pitch axes for longer sequences with aggressive motions. Therefore, even though our system provided better short-term estimates for Exp 05 and Exp06, the full trajectory errors were slightly higher. Future work will focus on reducing these errors by reintegrating the IMU data into the LiDAR pose estimate follow-



(a) LIO-SAM's predicted trajectory



(b) Our predicted trajectory

Figure 6. The predicted (red) and ground truth (green) trajectories for Exp06 (cloud generated from ground truth for reference)

ing Gauss-Newton optimization. Nevertheless, even with our current implementation, the preference for a smoother and locally accurate trajectory for navigation purposes may outweigh the marginal gain in long-term positional accuracy offered by LIO-SAM. A similar story was observed in Exp05.

3.2 Quadruped Robot Dataset

While the Hilti SLAM Challenge dataset provides realworld sequences collected via a high-quality hand-held system, it may not be fully representative of the types of trajectories followed and the vibrations produced by many robotics systems. To evaluate the algorithm on a robotics platform, our algorithm and LIO-SAM were evaluated on a dataset collected in an indoor environment using a teleoperated Unitree Go1 quadruped robot equipped with a Velodyne VLP-16 LiDAR and a YostLabs 3-Space Micro USB IMU (Figure 7). Compared to the Hilti Dataset, the data produced by this system is much noisier due to lower IMU precision, high vibrations due to the robot's walking gait, and a lower precision LiDAR. Additionally, the LiDAR has a very limited field-of-view (30°) and lower resolution (only 16 channels), providing an additional challenge to the SLAM system. Both algorithms took less than 50 ms to process each frame.



Figure 7. Lab robot used to collect dataset

Figure 8 shows a close-up of the resultant trajectory and generated map. While the overall trajectories produced by both algorithms were similar, the difference in short-term precision becomes apparent. During the test, LIO-SAM suffered from poor short-term precision in the predicted pose. This is likely due to their strategy of performing frame-to-global map matching at a slower rate than frameto-frame matching, as well as vibrations picked up by the IMU. Our approach did not suffer from these issues and produced a much smoother and more accurate trajectory. Additionally, our 3-stage matching approach can use normal information to distinguish between the two sides of a wall, accurately recovering the thicknesses of the walls. Meanwhile, LIO-SAM uses only proximity during matching, meaning both sides of the wall typically merge into a single plane, making it difficult to recover the geometry of the building. This has significant implications for building information modeling (BIM) as wall thicknesses are often of great interest for accurately modeling the building geometry.

As shown in Figure 9, the higher precision in predicted poses also leads to improved map quality and sharper point clouds. The difference is most evident when comparing the thin legs of the chairs, and the guard rail by the stairs. Improved map quality is of significant importance for tasks such as object detection, where sharper maps translate to lower noise in the object point clouds, capturing finer details of the object and improving the odds of accurate classification or segmentation. Additionally, the reduction in





(b) Our predicted trajectory

Figure 8. The predicted trajectory (yellow) and resultant map for the quadruped robot dataset (from same viewing angle)

noise in the object point clouds can improve the precision of measurements between object features, enabling more accurate object modeling.

Local stability in the estimated pose is also crucial for autonomous navigation. Poor short-term positional drift can cause the robot to temporarily assume it is too close or inside an obstacle, causing it to react unpredictably or sporadically. This can lead to jerky or dangerous recovery behaviors where the robot attempts to continuously re-plan its trajectory around obstacles. Our algorithm generates smoother, more accurate local trajectories than LIO-SAM, which can lead to less time spent re-planning and more consistent robot behavior.

4 Conclusion

In this study, we present a novel, computationally lightweight, LiDAR SLAM system for real-time localization and mapping. The proposed system was designed as a stand-alone C++ package to be used with 360 LiDAR scanners. The system was evaluated on a real construction dataset as well as on a quadruped robot. The system



(a) Map produced by LIO-SAM



(b) Map produced by our approach

Figure 9. Example maps recovered by LIO-SAM and our approach using the quadruped robot

achieved state-of-the-art performance on the majority of the datasets and demonstrated improvements in the pose estimate's robustness and precision.

While the proposed system provides clear advantages for the localization of autonomous systems, it does have some limitations. The current approach only uses the IMU for initial pose estimation and deskewing the cloud. This leads to eventual drift in roll and pitch as the IMU data is not utilized during pose estimation. Future work will integrate the final pose estimates with a factor graph to enable more flexible integration of additional sensors. This will also allow us to incorporate attitude or preintegrated IMU factors, improving the pitch and roll estimation over long trajectories.

5 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2222723. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- [1] Amandeep Dhaliwal. The Rise of Automation and Robotics in Warehouse Management. pages 63–72. 12 2020. ISBN 9781003032410. doi:10.1201/9781003032410-5.
- [2] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Daniela Rus. LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5135–5142. IEEE, 10 2020. ISBN 978-1-7281-6212-6. doi:10.1109/IROS45743.2020.9341176. URL https://ieeexplore.ieee.org/document/ 9341176/.
- [3] Ji Zhang and Sanjiv Singh. LOAM: Lidar Odometry and Mapping in Real-time. In *Robotics: Science and Systems X*, volume 2, pages 1–9. Robotics: Science and Systems Foundation, 7 2014. ISBN 9780992374709. doi:10.15607/RSS.2014.X.007. URL http://www.roboticsproceedings.org/rss10/p07.pdf.
- [4] Frank Dellaert and GTSAM Contributors. borglab/gtsam, 5 2022. URL https://github.com/ borglab/gtsam).
- [5] Matteo Frosi and Matteo Matteucci. ART-SLAM: Accurate Real-Time 6DoF LiDAR SLAM. *IEEE Robotics and Automation Letters*, 7 (2):2692–2699, 4 2022. ISSN 2377-3766. doi:10.1109/LRA.2022.3144795. URL https:// ieeexplore.ieee.org/document/9691876/.
- [6] Anton Koval, Christoforos Kanellakis, and George Nikolakopoulos. Evaluation of Lidar-based 3D SLAM algorithms in SubT environment. *IFAC-PapersOnLine*, 55(38):126–131, 2022. ISSN 24058963. doi:10.1016/j.ifacol.2023.01.144. URL https://linkinghub.elsevier.com/ retrieve/pii/S2405896323001519.
- [7] Juhyeon Kim, Jeehoon Kim, Sunwoong Paik, and Hyoungkwan Kim. Point cloud registration considering safety nets during scaffold installation using sensor fusion and deep learning. *Automation in Construction*, 159:105277, 3 2024. ISSN 09265805. doi:10.1016/j.autcon.2024.105277. URL https://linkinghub.elsevier.com/ retrieve/pii/S092658052400013X.
- [8] Yaoqiang Pan, Kewei Hu, Hao Cao, Hanwen Kang, and Xing Wang. A novel perception

and semantic mapping method for robot autonomy in orchards. *Computers and Electronics in Agriculture*, 219:108769, 4 2024. ISSN 01681699. doi:10.1016/j.compag.2024.108769. URL https://linkinghub.elsevier.com/ retrieve/pii/S0168169924001601.

- [9] Qipeng Li, Yuan Zhuang, and Jianzhu Huai. Multisensor fusion for robust localization with moving object segmentation in complex dynamic 3D scenes. *International Journal of Applied Earth Observation and Geoinformation*, 124:103507, 11 2023.
 ISSN 15698432. doi:10.1016/j.jag.2023.103507.
 URL https://linkinghub.elsevier.com/ retrieve/pii/S156984322300331X.
- [10] Lintong Zhang, Michael Helmberger, Lanke Frank Tarimo Fu, David Wisth, Marco Camurri, Davide Scaramuzza, and Maurice Fallon. Hilti-Oxford Dataset: A Millimeter-Accurate Benchmark for Simultaneous Localization and Mapping. *IEEE Robotics and Automation Letters*, 8(1):408–415, 1 2023. ISSN 2377-3766. doi:10.1109/LRA.2022.3226077. URL https:// ieeexplore.ieee.org/document/9968057/.
- [11] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan. Estimation of IMU and MARG orientation using a gradient descent algorithm. In 2011 IEEE International Conference on Rehabilitation Robotics, pages 1– 7. IEEE, 6 2011. ISBN 978-1-4244-9862-8. doi:10.1109/ICORR.2011.5975346. URL http: //ieeexplore.ieee.org/document/5975346/.

Buildability assessment of 3D printed concrete elements through computer vision

Shanmugaraj Senthilnathan^{1*} and Benny Raphael²

^{1,2}Civil Engineering Department, Indian Institute of Technology Madras, India

^{1*}<u>ce20d005@smail.iitm.ac.in</u>, ²<u>benny@iitm.ac.in</u>

Abstract -

Concrete 3D printing is a digital fabrication technology that has the potential to increase the level of automation in construction. However, getting consistent output quality is a challenge in concrete 3D printing because of the change in material properties with time and the influence of environmental parameters. A robust quality monitoring and control system is required to control the variations and obtain good-quality output. In this study, computer vision techniques are used to monitor the 3D printing process. Image features such as temporal variations in layer thickness and textural changes are used to assess the buildability properties. Two metrics have been developed for quantifying these features: entropy standard deviation and maximum layer thickness deformation. A significant correlation is found between the two metrics, and this relationship can be used to re-confirm the buildability assessment. For a given concrete mix, limiting values can be computed for the metrics to effectively classify an element into a stable type or one that is likely to collapse. This data can also be used as feedback to the printing system to make corrective actions to increase the quality of the print output. Thus, a real-time, nonintrusive buildability assessment system for concrete **3DP** elements is demonstrated in this study.

Keywords -

Concrete 3D printing; Computer Vision; Buildability,; Quality monitoring

1 Introduction and background

Concrete 3D printing (3DP) is a freeform technology that aims to realize the benefits of digital fabrication. It has been found to reduce construction time, material, and labor usage while improving sustainability. However, due to the dependency of material properties on multiple input parameters, 3D printed elements are found to have variations in the output quality. 3D printing challenges include proper extrusion and limiting dimensional changes with time [1]. One of the critical parameters of 3DP is buildability, which determines the number of layers that can be printed without significant dimensional changes in individual layers due to the weight of the top layers. The common buildability failures in 3D printing are plastic and buckling collapse [2]. It depends mainly on the workability of the concrete, and many studies have used fresh-state property tests to find a printable region.

There are very few real-time techniques that help in monitoring the variations during the printing process. Studies on quality monitoring and control techniques are increasing in the additive manufacturing industry [3]. However, studies on quality monitoring in concrete 3DP are relatively less. Monitoring using 2D camera images and computer vision techniques is gaining prominence in industry applications. In this study, a computer vision methodology for quality monitoring is developed that helps assess the buildability properties of 3D printed elements. Significant dimensional changes affect the print quality and long-term structural properties [4,5]; hence these are monitored in this methodology. In addition, new image features have been identified that can effectively assess the buildability properties of 3DP elements.

2 Methodology



Figure 1. Methodology of buildability assessment using image features extracted through computer vision techniques

Figure 1 shows the overall methodology for buildability assessment of 3D printed elements using Computer Vision (CV). It consists of the following three parts:

- Input data collection
- Image pre-processing
- Study of temporal image features

In the input data collection phase, 2D images of the 3D-printed elements are collected using a camera during printing. The input data is preprocessed in the second stage to remove all the background data, reducing the computational complexity and increasing accuracy. The final phase involves the usage of computer vision techniques to extract image features and understand the temporal variations of the image features. The temporal variations are used to assess the buildability and the dimensional changes in the printed elements.

3 Materials and Methods

3.1 Experimental procedure

To demonstrate the methodology for buildability assessment using computer vision, a series of elements

were printed in a laboratory setting. The experimental details are described in this section.

This study uses a new LC2-OPC mix, which uses Limestone Calcined Clay (LC2) and Ordinary Portland Cement (OPC) cement as binders with manufactured sand as aggregate. Super-plasticizers (SP) and viscosity modifying agents (VMA) are used to control the rheological properties of the printed elements. In general, buildability properties depend on the rheological properties, especially the workability of the printing concrete. The initial workability of concrete varies with super-plasticizer dosage. However, the workability during the printing process is dynamic, with variations occurring due to hydration or evaporation with time. Hence, elements are printed under different initial workability conditions and at different times of printing (time from the point of mixing water to the dry concrete mix) to understand the impact of the changes in rheological properties on the buildability of 3DP elements.

Table 1. 3DP elements printed with different Super-Plasticizer (SP) and Time Of Printing (TOP) as per experimental procedures

Category No.	SP (%)	TOP (min)	Status of Print
1		10-15	Collapsed
2	0.50	25-30	Collapsed
3	0.50	40-45	Good quality print with slight expansion/distortion in the middle
4		10-15	Good print with minor voids in the top layers
5	0.40	25-30	Fair print with many voids in the top layers
6		40-45	Non-extrudable
7		10-15	Good print with a fair amount of voids in the top layers
8	0.30	25-30	Non-extrudable
9		40-45	Non-extrudable

Table 1 shows the elements printed with different super-plasticizer values of 0.50%, 0.40%, and 0.30% to account for initial workability changes. For every super-plasticizer (SP) dosage, the Time of Printing (TOP) is varied in three ranges, 10-15, 25-30, and 40-45 mins. The different TOP values account for dynamic workability changes during the printing process. Nine different categories of printing were done as part of the experimental procedure. For every category, three prints were done to capture the variation in the data. For

Category 1 and 2, the super-plasticizer dosage was very high, resulting in collapse before reaching the target height of 500 mm. The super-plasticizer value was very low for categories 6, 8, and 9, making the concrete nonextrudable and non-printable. Since this is a buildability study, the non-extrudable categories are not considered.

A Canon 1300D 18MP camera was kept directly at the printing palette at a horizontal distance of 1000 mm from the printed element to capture the 2D video/images of the 3DP element. An illumination lamp was kept above the camera, which provides a constant illumination of 6500 (+/-100) Lux on the surface of the printed elements. The same camera and lighting setup were followed for all the print elements.

A stable element that retained shape and could be printed to a target height of 500 mm without collapse is designated as a good buildable element. If the element collapsed before the target height, it is classified as a collapsed or bad buildable element.

3.2 Image Data Collection and Preprocessing



Figure 2. Input image of a stable 3D-printed Element - A

Figure 2 shows the input images of the 3D-printed element that reached the target height of 500 mm (Element A). Individual snapshots were taken from the video after the completion of every layer (layer time instances). The layers are designated as L01, L02, and so on, numbered from bottom to top. The two bottom layers, L01 and L02, were intentionally compressed against the printing palette to act as the base for printing. Hence, the analysis is done from layers L03 to L10. Image 01 refers to the image instance captured after the completion of layer L03. Image 02 refers to the image instance captured after the completion of layer L04 and so on. The Image numbers – Image 01, Image 02, and so on refer to the image instances taken after the completion of every layer, and they indirectly represent the time of printing containing information about temporal changes in the overall printed element.

The input images are pre-processed to remove all the background data. It is first processed through Salient Object Detection (SOD) to segment only the objects in focus. All the objects outside the focus were removed and were converted into black pixels. Individual layers are cropped out, and their temporal changes are studied to extract useful image features.



Figure 3. Hough Transform output after preprocessing of images of Element A

Each layer is characterized by the top and bottom layer boundaries, which are effectively horizontal for linear elements. Hence, the Hough Transform technique is used to identify the layer boundary [6]. Then, additional rows of pixels are added to the top and bottom boundaries to account for undulations in layer boundaries to crop the individual layers. Figure 3 shows the preprocessing output where all the background data is removed, and only the printed element is cropped out. It also shows the Hough Transform output, where all the layer boundaries are identified as horizontal red lines.

The image of each layer is obtained by cropping the initial image of the layer instance using the boundaries identified through Hough Transform. The temporal changes in each layer after the printing of subsequent top layers are captured within the crop window.


Figure 4. Temporal images of layer L03 of a good, stable printed Element A



Figure 5. Temporal images of layer L03 of a collapsed Element B

Figure 4 and Figure 5 show the temporal changes of layer L03 within a crop window for Elements A and B, respectively. The individual layer images are processed through two different computer vision techniques - texture analysis and layer thickness extraction; these are discussed in later sections.

It is found that the dimensional stability and the buildability of the printed element can be evaluated by assessing the bottom ten layers of every 3D printed element. Hence, only the bottom ten layers are used in the analysis.

3.3 Texture extraction – Modified Histogram of Oriented Gradients (HOG)

The texture is the tangible feel on the surface of the material. In the case of an image, it is computed as the variations in the brightness intensity of the pixels. This study uses a modified version of the Histogram of Oriented Gradients (HOG) concept. The modified HOG computes the gradient changes in three directions. The details of the concept and methods of the modified HOG algorithm are discussed in [7].

3.4 Entropy value calculation

Entropy is a measure of uniformity in a distribution. It was first introduced by Shannon and is given by the following formula,

Entropy =
$$-\Sigma$$
 Pi * log2 Pi (1)

Where Pi is the probability of the i-th value in the distribution. Entropy can be used for quantifying the textural changes within an individual layer image [7,8]. The higher the entropy value, the higher the textural value, and vice versa.

3.5 Canny edge detection

Canny edge detection is an edge detection method that looks for sudden changes in the pixel brightness intensities.



Figure 6. Canny edge detection output of L03 of Element A

Figure 6 shows the Canny edge detection output of layer L03 of Element A at a particular instance. It shows that the boundary lines between two layers are captured. The layer thickness is the vertical distance between the top and bottom boundaries of the layer identified from the Canny edge detection output. Two edge lines are determined for every boundary because the pixel intensity changes on both sides of a layer boundary. The layer thickness is computed as the vertical distance between the top surface boundary of the target layer to the bottom layer. It is done to avoid the impact of voids on the bottom surface boundary of individual layers.

4 **Results and Discussions**

4.1 Temporal textural variation

Figure 7 shows the temporal changes in the texture of the layer L03 of Element A. It is found that layer L03 deforms minimally from Image 02 to Image 20. Beyond that, no significant dimensional changes are visible within the crop window. The black pixels in Figure 7 represent the gradient (change in pixel intensities) observed within the layer L03 crop window, as obtained from the modified HOG algorithm. The number of black pixels within the L03 crop window is found to be changing from Image 02 to Image 20. Beyond that, it is consistent. This trend matches the pattern found in the actual dimensional changes within the crop window.



Figure 7. Temporal changes in the textural output of layer L03 of Element A



Figure 8. Temporal changes in the texture of layers L03 to L10 of Element A

The variations in the pixel intensities (gradient/black pixels) are captured as a single entropy value for each image instance. Its temporal changes are given in Figure 8. It is seen that the entropy values have minor variations until Image 20. Beyond that point, the values are constant, indicating that layer L03 has become stable by achieving the initial stiffness/yield strength.



Figure 9. Temporal changes in the textural output of layer L03 of Element B

Figure 9 shows the modified HOG output of layer L03 of Element B. In contrast to Figure 7, the layer L03 of element B is continuously compressed with time. In Image 21, even the top layer has come into the crop window. This element eventually collapsed due to large deformations. The same is visible in the textural changes

in terms of the number of black pixels varying with time (image instances).



Figure 10. Temporal changes in the texture output of layers L03 to L10 of Element B

The temporal change in the texture of Element B is captured in terms of entropy value variations, as shown in Figure 10. Unlike Figure 8, the entropy values keep changing drastically and do not seem to consolidate.

Entropy Standard Deviation (ESD), or the standard deviation of the entropy values over time, is introduced in this study as a single measure of the temporal changes in entropy value for every individual layer. The standard deviation of entropy values (ESD) of layer L04 of element B is 0.1148, whereas the ESD value of layer L04 of element A is 0.0172. Similarly, ESD values for layers L03 to L10 of all the print elements. It is found that the ESD values are low for a stable print element and high for a collapsed print element. It proves that the ESD is a good measure of the dimensional stability or the buildability collapse assessment of the 3DP elements.

4.2 Temporal layer thickness variation

In continuation to section 3.5, the layer thickness is measured as the vertical distance between the two boundary edge lines identified from the Canny edge detection output. The layer thickness is found at thirty different sections along the line of printing. It is done to understand the layer thickness changes at different local sections rather than the overall average layer thickness. Figure 11 shows the layer thickness found in thirty different sections of layer L03 of Element A, identifying the layer boundaries and the vertical distance between them in terms of pixel points. Also, the changes in layer thickness over time are obtained by following the same steps on individual layer images taken at different time instances.



Figure 11. Layer thickness extracted at 30 different sections of layer L03 of Element A



Figure 12. Layer thickness extracted at different sections of layer L03 of Element A

The changes in layer thickness are captured and shown in terms of a graph in Figure 12. The x-axis shows the different section points along the line of printing (where layer thickness is measured), and the y-axis shows the layer thickness in terms of pixel point units of layer L03 of element A measured at different time instances (Image numbers). The figure shows that the layer thickness initially reduces with time (increase in Image numbers). However, the layers gain initial stiffness/ yield strength with time, and the thickness reduction stabilizes.



Figure 13. Layer thickness extracted at different sections of layer L03 of Element B

The layer thickness found for layer L03 of Element B is given in Figure 13. Since the printed elements have high workability, the layer thickness continuously

reduces with time, ultimately leading to the element's collapse.



Figure 14. Layer thickness extracted at different sections of layer L03 of Element B

The temporal layer thickness changes of layer L03 of element B are given in Figure 14. Unlike Figure 12, the layer thickness value reduces continuously with time. Beyond Image 19, there is a sudden decrease in the layer thickness, ultimately leading to the element's collapse beyond the time instance - Image 22.





To have a single metric for layer thickness reduction, the concept of maximum Layer Thickness Deformation (LTD) (%) is introduced, where LTD is measured using the following formula:

LTD (%) = (Layer thickness at initial instance – layer thickness at instance X) / Layer thickness at initial instance X 100. (2)

The LTD is calculated at every section and every time instance X. The maximum deformation value across the different sections and the time instances is considered the Maximum LTD (MLTD) value. The LTD values found at different sections and different time instances of layer L03 of Element B are given in Figure 15. It shows the LTD values and their temporal changes (image numbers) across different sections along the printing line. It is found that the LTD values increase even beyond Image 16 and reached the maximum value at section 07 (MLTD) before the element's collapse. The MLTD value for layer L03 of Element B is 45.47%, but in the case of layer L03 of Element A, it is only 12.50%. The MLTD values are found for layers L03 to L10 of all the experimental print elements. It is found that the MLTD values of layers from a stable printed element are low. And for the collapsed element, the MLTD values are drastically high. Hence, MLTD also serves as a good image feature to assess the buildability collapse or dimensional stability of 3D printed elements.

4.3 Limiting value for entropy standard deviation (ESD)



Figure 16. ESD vs. QOP plot showing the individual layer data points of good and bad quality printed element layers.

To find a limiting value that classifies a print element as a stable or collapsed element, each layer (L03 to L10) of all the experimental prints is considered as individual data points and analyzed. The layers L03 to L10 of a stable printed element are tagged as good-quality data points (Quality of Printing). In the case of collapsed print, all the individual layer data points (L03 to L10) are designated as bad-quality data points (Quality of Printing). Figure 16 shows the plot of the Entropy Standard Deviation (ESD) vs the Quality of Printing (QOP) designated based on the dimensional stability of every individual layer data point. The limiting value (discriminant) is found to be 0.065 (black dotted line) for ESD, which differentiates a stable and collapsed print element. The overall misclassification is 8.432% for the 130 individual layer data points assessed in this study.

4.4 Limiting value for maximum layer thickness deformation (MLTD)

Figure 17 shows the MLTD vs. QOP plot showing the maximum layer thickness deformation (MLTD) (%) of the individual layers (L03 to L10) of stable and collapsed printed elements. Similar to finding the limiting value for ESD, the layers of collapsed elements are designated as bad-quality data points, and the layers of stable elements are designated as good-quality data points. The limiting value is found to be <u>12.50%</u> for MLTD, with a misclassification of only 4.769%.



Figure 17. MLTD vs. QOP plot showing the good and bad quality individual layer data points

4.5 Relation between max layer thickness deformation (MLTD)(%) and entropy standard deviation (ESD)

Figure 18 shows the relation between the measured image feature metrics - MLTD (%) and ESD for all individual layer data points. All the individual layer data points from a stable printed element are given in green, and the collapsed printed element is given in red. A distinct region separation is noted from the plot. When the individual limiting values identified for each image feature are plotted, a discriminating boundary separating stable print data points and collapsed layer data points is found. It is to be noted that during the printing process, if any of the layers L03 to L10 falls in the collapse region or the rate of change of image feature values moves towards it, then there is a high chance of the 3D printed element collapsing. Hence, the correlation identified for ESD and MLTD can be used to supplement or re-confirm the buildability assessment from the individual image feature analysis. The developed limiting values of image features are valid for the current mix design and experimental procedures used in this study. They are not expected to vary drastically for other mixes as the

dimensional change might be in the same pattern for collapse in other mixes.



Figure 18. Plot showing the correlation between MLTD (%) and ESD

5 Conclusion

This paper presents a methodology for assessing the buildability of 3D printed concrete elements using computer vision techniques. Two new metrics have been developed in this research: Entropy Standard Deviation (ESD) and Maximum Layer Thickness Deformation (MLTD). These metrics are computed using image processing techniques and can be used for buildability assessment. The conclusions of the study are as follows:

- There exists a correlation between the two metrics. This relationship can be used to re-confirm the buildability assessment independently through two methods.
- For a given concrete mix, limiting values can be determined for the two metrics for discriminating between a stable print and one that could result in a collapse.
- Buildability or dimensional stability in terms of a stable/collapse print can be evaluated by monitoring and assessing the bottom ten printed layers of a single batch of concrete mix.

This is the first study to identify metrics that clearly predict properties of 3D printed concrete elements. This study paves the way for further research in computer vision on the use of image features to assess critical parameters like buildability. The image features can be used in a feedback loop to control the printing system. Based on the temporal variations in the image features, extrusion speed, and printing speed can be modified to give the bottom layers significant time to increase initial stiffness and yield strength to carry the weight of the top layers. It ensures sufficient buildability and avoids material wastage, increasing the sustainability of 3D printing technology. Thus, the study will help develop an autonomous, non-intrusive tool for the buildability assessment of concrete 3D-printed elements.

References

- A. Kazemian, X. Yuan, O. Davtalab, B. Khoshnevis, Automation in Construction Computer vision for real-time extrusion quality monitoring and control in robotic construction, Autom Constr 101 (2019) 92– 98. https://doi.org/10.1016/j.autcon.2019.01.022.
- [2] A.S.J. Suiker, R.J.M. Wolfs, S.M. Lucas, T.A.M. Salet, Elastic buckling and plastic collapse during 3D concrete printing, Cem Concr Res 135 (2020). https://doi.org/10.1016/j.cemconres.2020.106016.
- [3] A. Oleff, B. Küster, M. Stonis, L. Overmeyer, Process monitoring for material extrusion additive manufacturing: a state-of-the-art review, Progress in Additive Manufacturing (2021) 23–27. https://doi.org/10.1007/s40964-021-00192-4.
- [4] S. Lim, R.A. Buswell, P.J. Valentine, D. Piker, S.A. Austin, X. De Kestelier, Modelling curved-layered printing paths for fabricating large-scale construction components, Addit Manuf 12 (2016) 216–230. https://doi.org/10.1016/j.addma.2016.06.004.
- [5] B. Raphael, S. Senthilnathan, A. Patel, S. Bhat, A review of concrete 3D printed structural members, Front Built Environ 8 (2023). https://doi.org/10.3389/fbuil.2022.1034020.
- [6] International L. Chandrasekar, G. Durga, and Conference on Communications Signal Processing (ICCSP), in: 2014 International Conference Communication and Signal on Processing, IEEE, https://ieeexplore.ieee.org/document/6949962 (accessed March 8, 2024).
- [7] S. Senthilnathan, B. Raphael, Quality monitoring of Concrete 3D Printed elements using computer vision-based texture extraction technique, in: 2023 Proceedings of the 40th ISARC, Chennai, India, 2023:pp.474–481.

https://doi.org/10.22260/ISARC2023/0064.

[8] S. Senthilnathan, B. Raphael, Using Computer Vision for Monitoring the Quality of 3D-Printed Concrete Structures, Sustainability (Switzerland) 14 (2022). https://doi.org/10.3390/su142315682.

Automated Inspection Report Generation Using Multimodal Large Language Models and Set-of-Mark Prompting

Hongxu Pu¹, Xincong Yang^{1, 2, *}, Zhongqi Shi^{3, 4} and Nan Jin^{3, 4}

¹ School of Civil and Environmental Engineering, Harbin Institute of Technology (Shenzhen), China

² Guangdong Provincial Key Laboratory of Intelligent and Resilient Structures for Civil Engineering, China

³ Key Laboratory of Urban Safety Risk Monitoring and Early Warning, Ministry of Emergency Management, China
⁴ Shenzhen Technology Institute of Urban Public Safety, China

perrykelly779@gmail.com, yangxincong@hit.edu.cn, shizq@szsti.org, jinnan@szsti.org

Abstract

In the context of the increasing expansion and complexity of civil engineering projects, construction inspection plays a crucial role in ensuring project quality and safety. The traditional construction inspection report writing process mainly relies on the manual records of on-site inspectors. This process is not only time-consuming but also easily affected by personal subjective judgments. In the current rapidly evolving construction environment, there are obvious limitations to this traditional method, especially in terms of the accuracy and timeliness of the reports. In view of this, this study proposes an innovative approach that combines the Set-of-Mark (SoM) prompting technology and the multimodal Large Language Models (LLMs), aiming to automate the construction inspection report generation process and improve the efficient and effectiveness of the onsite inspection. The case study shows that the method can fulfill the basic requirements of construction inspection reports and further improves the quality of the report in complicated scene through SoM prompting. The core of this method is to conduct a more accurate analysis of the conditions of the construction site by overlaying marks on key areas of the construction inspection images and using the multimodal LLMs to capture the region of interest (ROI), and then automatically generate detailed construction inspection reports. This technological innovation not only significantly improves the efficiency of construction inspection report writing, but also greatly enhances the quality and credibility of the report content through in-depth image analysis and text generation.

Keywords

Construction inspection, multimodal large language model, Set-of-Mark prompting, automated report generation

1 Introduction

Construction inspection work is a key component to the success of civil engineering projects, especially in today's increasingly large-scale projects and complex construction environments. To obtain a comprehensive understanding of the construction sites, engineers have to monitor the entire construction procedures and associated factors all the time, which is a process often hindered by subjectivity and inefficiency. To address this issue, this study proposes an innovative approach that leverages the Set-of-Mark (SoM) prompting in conjunction with multimodal LLMs to automate the generation of construction inspection reports.

The proposed approach entails processing construction site images using SoM prompting [1], a visual prompt technique that segments images into distinct, semantically meaningful regions, each marked with identifiable symbols. These enhanced images are then delivered to multimodal LLMs, such as GPT-4 [2], which are adept at interpreting complex visual and textual data. This collaboration has resulted in the generation of construction inspection reports that are not only detailed and accurate, but also largely avoid the problem of subjective bias that tends to occur when reports are prepared manually.

The novel approach of combining SoM prompting with multimodal LLMs in construction inspection represents a significant advancement over traditional methods. It utilizes the cutting-edge capabilities of image processing technology and language models to provide high quality construction inspection reports. This method promises to enhance the overall efficiency of construction management and quality control processes.

The subsequent sections of the paper will elaborate on the methodology of applying SoM prompting for inspection image processing, the integration of these images with multimodal LLMs, and the impact of this integration on the quality and reliability of construction inspection reports. We will also present a case study that demonstrates this method's application in a real-world engineering project and provide an analysis of the results. The conclusion will summarize the findings and explore potential future research directions.

2 Methodology

This section describes in detail the two core technologies used in our study: the SoM prompting and the multimodal LLMs. First, we introduce the SoM prompting to mark the construction inspection images, and then utilize the multimodal LLMs to convert images and texts to automatically generate construction inspection report.



Figure 1. The workflow of the proposed approach

2.1 SoM Prompting

SoM prompting is a novel visual prompting method designed for multimodal LLMs. Its main purpose is to enhance the visual localization ability of the model through visual marks. The mathematical formulation of SoM prompting can be expressed as follows:

$$T_o = F(SoM(I), T_i) \tag{1}$$

where I is the input image, T_i is the textual query, and T_o is the textual output of the model. The key to the SoM prompting strategy is the ability to divide an image into a series of semantically meaningful regions and to impose auxiliary information on these regions to make them both interpretable and describable by multimodal LLMs.

2.1.1 Image Segmentation

Image segmentation is the first step in SoM prompting, aiming at automatically or semiautomatically extracting semantically aligned regions. For this purpose, a range of image segmentation tools have been adopted, such as SEEM [3], Semantic-SAM [4], and SAM [5]. These tools need to have the following properties, firstly robustness to ensure that the tools can segment regions with precision and convey fine-grained spatial layout information to the multimodal LLMs, secondly an open vocabulary to recognize objects outside of the predefined vocabulary, and lastly richness of granularity to support not only the segmentation of complete objects, but also the segmentation of part of an object's region.

In this study, these image segmentation tools can be used to delineate different areas of the construction site such as construction equipment, worker locations, and building materials. As shown in Figure 2, a construction inspection image *I* of size $A \times B$ is divided into *N* regions $R = [r_1, r_2, \cdots r_N] \in \{0, 1\}^{N \times A \times B}$. Each region can be represented by a binary mask, where each mask corresponds to a key region in the image. For example, one mask may represent the location of an air conditioning duct, while another mask may represent the area where construction materials are stacked.



Figure 2. Image segmentation using SAM

2.1.2 Mark Generation

After completing the image segmentation, marks that contribute to multimodal LLM localization need to be generated for each region. The type of marks depends on whether they can be interpreted by the multimodal LLMs. In this study, considering the complexity of the construction site scene, alphanumeric marks should be used because they not only take up less image space, but also can be recognized by most multimodal LLMs using their OCR function [6]. In addition to this, the polygon and box marks used in image segmentation will be used as auxiliary marks.

In addition to selected mark types, assigning optimally located marks to each significant region in the inspection image is critical for the generation of subsequent inspection report. There is a significant problem in current mark assignment methods: placing the marks directly at the center of the mask leads to overlapping between marks. To solve this problem, an optimization algorithm is employed as shown in Figure 3, which are based on the principle of reducing the overlap between marks by calculating the area of the identified regions in the image and sorting them in order of size to ensure that smaller regions are processed first before larger ones when assigning marks. The algorithm further excludes overlaps with processed regions to obtain an independent mask, and then applies a distance transformation algorithm to find the optimal mark locations.

def	Find_Optimal_Mark_Location (region): optimal_location = calculate_optimal_location (region) return optimal_location
def	Assign_Marks_Optimally (regions): sorted_regions = sort_by_area (regions) mark_locations = {}
	for region in sorted_regions: if not is_overlap (region, processed_regions): mark_location = Find_Optimal_Mark_Location (region) processed_regions.append (region) mark_locations [region] = mark_location

return mark locations

Figure 3. Pseudocode of mark allocation algorithm

By using the image segmentation tools and the mark allocation algorithm, the inspection image processed by SoM prompting can be obtained as shown in Figure 4.



Figure 4. Inspection image with SoM prompting

2.2 Multimodal LLMs

Multimodal LLMs represent an important advancement in the field of artificial intelligence (AI), and their core capability is to concurrently process and understand data from different modalities, such as images, text, and sound [7]. These models are typically based on deep learning architectures, in particular neural networks such as convolutional neural networks (CNNs) [8] and vision transformer (ViT) [9] for image processing, and some transformer models [10] for processing sequential data such as text.



Figure 5. The process of extracting image features by vision transformer

2.2.1 Selection of Multimodal LLMs

the process of automatically generating In construction inspection reports, multimodal LLMs not only extract information from inspection images, but also text them to generate inspection reports. Therefore, the selection of an appropriate multimodal LLM is an important factor in the quality of the report. The choice of model relies on a comprehensive evaluation of the model in multiple aspects [11]. First, the model must demonstrate strong perception ability, which includes accurately identifying specific objects and details on the construction site. For example, the model needs to be able to identify construction equipment, engineering materials, and even safety signs during construction. Furthermore, cognitive capability plays a crucial role. The model is expected to interpret visual data effectively and translate it into a logically coherent and comprehensive construction inspection report. This requires that the model is not just a simple information processing tool but has certain reasoning and logical analysis capabilities.

Secondly, the model should have good instruction following ability. In the process of generating construction inspection reports, the model needs to perform tasks according to specific instructions, such as generating specific project status descriptions based on construction images. Therefore, the selected multimodal LLM should be able to accurately understand and respond to these instructions to avoid misunderstandings or biases.

In addition, preventing hallucination problems [12] is also an important criterion for selecting multimodal LLMs. When dealing with complex construction scenarios, the model should not produce incorrect object

recognition, such as incorrectly identifying engineering hazards or equipment that do not exist. Therefore, models with higher accuracy and reliability are more appropriate.

In practical applications, considering the particularity of construction inspections, the training data and instruction design of the model are also very critical. The training data of the model should cover a wide range of construction scenarios to ensure its effectiveness and accuracy in practical applications. Meanwhile, the instruction design should be as concise and clear as possible so that the model can accurately interpret and execute it.



Figure 6. The criteria and weights of model selection

Taking the above criteria into consideration, when selecting a multimodal LLM suitable for construction inspection scenarios, it is recommended to choose those models that perform well in perception and cognitive tests, can accurately follow instructions, and have good performance in object recognition and logical reasoning. In addition, considering the complexity and variability of construction inspections, it will be more advantageous to choose a model that can quickly adapt to new scenarios and tasks. Through careful selection, multimodal LLMs can effectively improve the quality of automated generation of construction inspection reports, thereby improving overall project management efficiency and safety.

2.2.2 Construction Inspection Report Generation

In multimodal LLMs, the image and text inputs are first processed through their respective preprocessing modules and then embedded into the same dimensional space. For image data, this may include steps such as feature extraction, object recognition, and image segmentation, which are designed to extract useful visual information from the image. For text data, word embedding is a key step. Word embedding converts each word into a vector in a high-dimensional space, and these vectors are usually learned by training algorithms such as Word2Vec [13] on large text datasets. Mathematically, this can be expressed as a mapping function as follows:

$$f: word \to \mathbb{R}^d \tag{2}$$

where *word* is a word in the vocabulary and \mathbb{R}^d is a ddimensional vector space. The vector representation of each word not only captures the semantic information of the word, but semantically similar words are mapped to locations close to each other in this high-dimensional space.

The processed data is then fed into the core of the model, which typically contains multiple levels of network structure for in-depth analysis and fusion of information from different modalities. In this process, the model learns how to correlate and interpret data from different sources. Considering that the data is different for each modality, the fusion process needs to combine this information efficiently. For example, if there are an image feature vector $v_i \in \mathbb{R}^m$ and a text feature vector $v_t \in \mathbb{R}^n$, a common fusion strategy is to use weighted sum [14] or concatenation [15]. The mathematical principle of weighted sum is as follows:

$$v = \alpha v_i + (1 - \alpha) v_t \tag{3}$$

where α is a weight parameter to balance the importance of the two models. Another common approach concatenation is to directly concatenate the two types of feature vectors to form a new vector:

$$v = [v_i; v_t] \tag{4}$$

where [.;.] represents the concatenation of vectors. In both approaches, the key is to select and tune the parameters to combine information most effectively from different sources. In multimodal learning, this is usually achieved by optimizing an objective function that quantifies the degree of match between the fused data and the expected output. With such a fusion approach, the model can synthesize information from different modalities and improve its ability to understand complex data.



Figure 7. The flowchart of construction inspection report generation

On the output side of a multimodal LLMs, the model can generate a comprehensive output based on the input data, which may be a report that combines image and textual information, or a comprehensive interpretation and analysis of the input data. This capability makes the multimodal LLMs particularly suitable for complex tasks that require the concurrently understanding of visual and textual information, such as the automated generation of construction inspection reports.

In the construction inspection scenario, the multimodal LLMs can use its ability to understand images to identify specific features of the construction site, while using its text processing capabilities to generate a detailed report. This combination of visual and textual analysis makes multimodal LLMs ideal for generating accurate and comprehensive construction inspection reports.

3 Case Study

This case study focused on a large-scale construction project under construction in Pingshan District, Shenzhen City, Guangdong Province. This project provides a valuable practical scenario for in-depth analysis and verification of the application of the multimodal LLMs in the automated generation of construction inspection reports. The project covers an area of 0.2 square kilometers, providing diverse construction activities and rich visual data for multimodal LLMs. Since the MiniGPT-4 multimodal LLM [16] has demonstrated excellent performance in several benchmark tests, we chose MiniGPT-4 for the case study.

MiniGPT-4 is designed to maintain the original efficient processing power of GPT-4 while optimizing its size and resource consumption to make it more suitable for application scenarios that require faster speeds and fewer computational resources. It is specifically designed for multimodal tasks, including image and text processing. It utilizes the state-of-the-art LLM Vicuna [17] as a decoder, combined with the pre-trained visual component BLIP-2 containing ViT and Q-Former [18]. MiniGPT-4 aligns the encoded visual features with the Vicuna language model through a single projection layer and fixes the other visual and language components. The initial training of the model covers a large image-text alignment dataset, which is then fine-tuned with high-quality, accurately aligned image-text pairs in a second phase, a step that is critical to improving the reliability and overall usability of the model. MiniGPT-4 demonstrates a variety of GPT-4-like capabilities, including detailed image description generation and website creation based on handwritten text instructions.



Figure 8. The architecture of MiniGPT-4

To meet the computational requirements of the project, a high-performance desktop computer with an RTX 4080 16GB was used to ensure the efficiency and stability of the process of using the MiniGPT-4 to generate construction inspection reports.

The case study first explored the ability of the MiniGPT-4 model to rely on original inspection image which is not combined with SoM prompting to generate construction inspection report.



Figure 9. Construction inspection report generated by MiniGPT-4 for inspection image

without SoM prompting

By analyzing the inspection report shown in Figure 9, it can be found that the inspection report comprehensively and accurately describes the contents of the construction inspection image. Although there are errors in the identification of objects such as crane and dump truck marked in red in the report, the generated report still adequately fulfills the standards required for construction inspection report.

Further case study introduced SoM prompting, which aimed to enhance the model's ability to identify and describe construction details by annotating key areas of construction inspection images.



Figure 10. Construction inspection image marked by SoM prompting

With the assistance of SoM prompting technology, the MiniGPT-4 processed the annotated images and generated detailed construction inspection report. The inspection report generated by MiniGPT-4 for the image annotated by SoM prompting is shown in Figure 11.

Describe this image from the perspective of construction inspection.



This image shows a construction site with a large excavator in the foreground. The excavator is digging up the ground and appears to be in the process of building a foundation. There are several other pieces of construction equipment in the background. The image is taken from a bird's eye view, looking down on the construction site. The excavator appears to be yellow and has several attachments on its arm, including a shovel and a bucket. The dump truck is also yellow and has a large cargo bed. The image is well-lit, with the sun shining down from the top right corner. The background is a mix of green and brown, with some trees visible in the distance. The image appears to be taken in the morning, with the sun still low in the sky.

Figure 11. Construction inspection report generated by MiniGPT-4 for inspection image with SoM prompting The results shown that although the MiniGPT-4 can generate basic construction inspection reports by analyzing construction inspection images that have not been processed by SoM prompting, the introduction of SoM prompting has significantly improved the detail and accuracy of the report, which verifies the value of the multimodal LLMs and SoM prompting for the application in real construction scenarios.

4 Conclusion and Future Work

In this study, we successfully introduced the MiniGPT-4 to a large-scale construction project in Shenzhen and realized the automated generation of construction inspection reports that obtain а comprehensive understanding of the field scene. This result demonstrates the potential of multimodal LLMs in the construction industry. Multimodal LLMs can effectively analyse complex visual data at construction sites and successfully capture key factors of construction activities, thereby generating comprehensive and accurate construction inspection reports. Furthermore, we also improved the accuracy of reports by introducing SoM prompting, effectively improving the efficiency and quality of construction scene data analysis.

However, this method also has certain limitations. Due to the huge number of parameters of the multimodal LLMs, the automated generation of construction inspection reports places high demands on computing resources. The running and processing of the current multimodal LLMs requires powerful hardware support, which may limit its feasibility in some practical application scenarios. In addition, the effectiveness of the report generated by the multimodal LLMs depends on the quality and diversity of the training data used to pre-train the models, which also places higher requirements on the source of training data for the multimodal LLMs.

In the future, we aim to harness the specialized knowledge within the construction field to fine-tune pretrained multimodal LLMs, and by leveraging cloud computing resources, we can distribute the computing load and reduce the time and local resource requirements. Additionally, the exploration of more lightweight LLMs architectures would be beneficial. These architectures can remain the essential features necessary for accurate report generation while minimizing the model size and complexity. Moreover, we also plan to further develop tools and user-friendly interfaces for real-time analysis of construction inspection images so that construction managers can more easily use and understand the reports generated by the model. Through these measures, we look forward to advancing the automation process within the construction industry and providing new ideas and solutions for the practical application of multimodal LLMs in the field of civil engineering. As technology

continues to develop, models like MiniGPT-4 will play an increasingly important role in the digital transformation of the construction industry.

5 Acknowledgements

This research is supported by Key Technologies R&D Program (2022YFC3801203), supported by National Natural Science Foundation of China (Grant No. 52108286), supported by Shenzhen Science and Technology Programs (Grant No. GXWD20220818002513001, Grant No. RCBS20221008093128076).

References

- Yang J, Zhang H, Li F, Zou X, Li C, Gao J. Set-of-Mark Prompting Unleashes Extraordinary Visual Grounding in GPT-4V [J]. arXiv preprint arXiv:2310.11441, 2023.
- [2] OpenAI. GPT-4 Technical Report[J]. arXiv preprint arXiv:2303.08774, 2023.
- [3] Zou X, Yang J, Zhang H, et al. Segment everything everywhere all at once[J]. arXiv preprint arXiv:2304.06718, 2023.
- [4] Li F, Zhang H, Sun P, et al. Semantic-sam: Segment and recognize anything at any granularity[J]. arXiv preprint arXiv:2307.04767, 2023.
- [5] Kirillov A, Mintun E, Ravi N, et al. Segment anything[J]. arXiv preprint arXiv:2304.02643, 2023.
- [6] Liu Y, Li Z, Li H, et al. On the hidden mystery of ocr in large multimodal models[J]. arXiv preprint arXiv:2305.07895, 2023.
- [7] Yin S, Fu C, Zhao S, et al. A Survey on Multimodal Large Language Models[J]. arXiv preprint arXiv:2306.13549, 2023.
- [8] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern recognition, 2018, 77: 354-377.
- [9] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [11] Fu C, Chen P, Shen Y, et al. MME: A Comprehensive Evaluation Benchmark for Multimodal Large Language Models[J]. arXiv preprint arXiv:2306.13394, 2023.
- [12] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 1-38.
- [13] Church K W. Word2Vec[J]. Natural Language

Engineering, 2017, 23(1): 155-162.

- [14] Marler R T, Arora J S. The weighted sum method for multi-objective optimization: new insights[J]. Structural and multidisciplinary optimization, 2010, 41: 853-862.
- [15] Zhang C, Yang Z, He X, et al. Multimodal intelligence: Representation learning, information fusion, and applications[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(3): 478-493.
- [16] Zhu D, Chen J, Shen X, et al. Minigpt-4: Enhancing vision-language understanding with advanced large language models[J]. arXiv preprint arXiv:2304.10592, 2023.
- [17] Chiang W L, Li Z, Lin Z, et al. Vicuna: An opensource chatbot impressing gpt-4 with 90%* chatgpt quality[J]. See https://vicuna. lmsys. org (accessed 14 April 2023), 2023.
- [18] Li J, Li D, Savarese S, et al. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models[J]. arXiv preprint arXiv:2301.12597, 2023.

Assisting in the identification of ergonomic risks for workers: a large vision-language model approach

Chao Fan¹, Qipei Mei², and Xinming Li¹

¹Department of Mechanical Engineering, University of Alberta, Canada ² Department of Civil Engineering, University of Alberta, Canada <u>cfan3@ualberta.ca</u>, <u>gipei@ualberta.ca</u>, <u>xinming1@ualberta.ca</u>

Abstract -

In the construction industry, due to workers frequently engaging in highly physically demanding tasks and using various tools, workers are often exposed to ergonomic risks and safety hazards. Various observation-based traditional or computer vision-based artificial intelligence methods have been applied in the field of construction to assess ergonomic risks. However, the method of assessing ergonomic risks using Generative Pretrained Transformers (GPT) based visual language models has not been thoroughly explored. This study explores its unique ability in visual-text interaction to extract ergonomic risk information from images and generate corresponding human-like language descriptions. To test the feasibility and performance of the proposed method, two datasets were created. Each dataset contained 100 different scenarios with ergonomic risk information for finetuning and testing. Performance after finetuning the vision-language model with the finetuning dataset outperformed the model before finetuning; the results showed that the fine-tuned model achieved an accuracy of 81%, while the model before finetuning only achieved 28% accuracy. Therefore, the proposed method offers an automated, real-time, non-traditional artificial intelligence approach for identifying ergonomic risks and providing human-like language descriptions. This expands the perspective of health and safetyrelated problem-solving and promotes the prevention of work-related musculoskeletal disorders (WMSD) in the construction industry.

Keywords –

Ergonomic Risks Identification; Work Safety; GPT; Vision Language Model; Construction Safety

1 Introduction

The timely recognition and mitigation of workplace hazards play a pivotal role in safeguarding the safety, health, and productivity of the working environment. Job tasks in construction often expose workers to conditions that entail repetitive movements, strenuous labor, and awkward body postures, subtly but adversely affecting their well-being. The consequential impact of these operations frequently manifests in the development of WMSD. Based on data from the European Agency for Safety and Health at Work, WMSDs exhibit the highest prevalence within the construction sector when compared to other industries. [1] In 2018, 52% of individuals employed in construction reported instances of backaches, while 54% experienced WMSD in the upper limbs and 41% in the lower limbs. The construction and manufacturing sectors demonstrate a high percentage of sick leave attributed to WMSD due to the physically demanding nature of the work. [1] Notably, the United States Bureau of Labor Statistics reported that WMSD accounted for 30% of occupational injuries and illnesses in 2018. Moreover, injuries and illnesses related to WMSD in the construction industry ranked fifth highest among all industries. [2] Consequently, proactive identification and prevention of WMSDs and associated health risks prove to be highly constructive.

Conventional approaches to managing worker safety and health heavily depend on human observation, selfreporting, and direct measurements. [3]-[6] In these traditional methods, ergonomists utilize manual procedures to detect ergonomic risks, frequently incorporating the observation and interviewing of workers. In essence, the subjective and time-consuming aspects inherent in the traditional identification of ergonomic risks impede its efficacy in promptly analyzing and preventing such risks. [7], [8] Consequently, it is imperative to explore methodologies that can objectively and swiftly pinpoint ergonomic risks.

To address the inherent subjectivity and timeconsuming limitations associated with traditional ergonomic risk identification, researchers have shown a growing interest in automated approaches. Current research in automated ergonomic risk identification predominantly centers around computer vision-based methods [4], [9]-[14], with vision-language models yet to receive widespread attention. These computer visionbased methods directly train on visual data to learn patterns, particularly focusing on human joints for ergonomic risk identification. For instance, these computer vision methods implement ergonomic risk identification by training on visual data related to human joints. Thus, they predict ergonomic risks by estimating the joints of workers and calculating joint angles based on these estimations.

On the other hand, vision-language model approaches, employing large language models (LLM) and 'zero-shot learning,' are gaining recognition for their ability to generate human-like descriptions. [15] They are pretrained on LLM/large text corpora and subsequently finetuned for vision tasks using image-text pairs [15]-[17]. Unlike conventional computer vision methods, vision-language model-based approaches offer a unique perspective by generating human-like descriptions instead of numerical representations, thereby providing a more comprehensive understanding of ergonomic risk identification. GPT-4V [18] by OpenAI and MiniGPT-4 [15] are two approaches based on vision-language models. Despite both incorporating LLM for initial pretraining and refining their models through subsequent finetuning with image-text pairs, it is noteworthy that the specific LLM and image-text pairs employed in each approach are unique. Furthermore, the unavailability of the source code and model for GPT-4V poses a limitation, as it hinders the ability to finetune its model using personalized image-text pairs for the identification of ergonomic risks. This lack of accessibility underscores the challenges associated with adapting the method for specific applications when critical components are not openly accessible. This study employed the MiniGPT-4, incorporating a vision encoder with a pre-trained Vision Transformer and Q-Former, a single linear projection layer, and the Vicuna LLM [15]. It is important to highlight that MiniGPT4 incorporates image-text pairs for finetuning pre-trained models, but the images within these pairs are not related to construction workers. Furthermore, the text captions associated with these images do not provide any descriptions of ergonomic risks related to the individuals depicted in the images. To equip the vision-language model finetuned with pertinent visual knowledge related to ergonomic risks, the imagetext pair dataset offered by MiniGPT-4, limited to daily scenes or common objects, proves inadequate for effectively addressing scenarios associated with the identification of ergonomic risks in the context of workers. To address this challenge, this study curates datasets containing images of construction workers along with corresponding text descriptions pertaining to ergonomic risks.

During the quest for engineering-related studies, an inquiry into the terms "large language model" and "construction" on Scopus produced 25 results, yet merely 3 were pertinent to applications in the construction domain. A similar search combining "vision language model" and "construction" revealed only 1 relevant outcome within the construction field. Regrettably, no results were obtained for the query pairing "large language model" and "ergonomic" in the Scopus database. Prior studies have utilized GPT models for tasks such as question answering, extracting information from Building Information Modeling (BIM) datasets, and optimizing scheduling and sequencing in engineering contexts. Zheng et al. (2023) presented a prompt-driven virtual assistant framework aimed at bolstering natural language-centric BIM search by integrating GPT technologies. This framework autonomously interprets users' natural language inquiries, retrieves pertinent information, and provides succinct natural language responses alongside corresponding 3D visualizations via a user interface. You et al. (2023) introduced a methodology harnessing the capabilities of ChatGPT to realize automated sequence planning in robotic assembly for construction. The efficacy of this method was demonstrated through its ability to decrease reliance on manual intervention, shorten planning durations, and enhance the overall efficiency of robot-driven assembly processes within the construction sector. Prieto et al. (2023) assessed the suitability of GPT for aiding in the creation of an automated construction schedule using prompts expressed in natural language. Chen et al. (2024) research presents an interactive query system based on Augmented Reality and Deep Learning for delivering real-time safety information through the analysis of onsite images. Table 1 shows the search results on Scopus of relevant applications in the construction sector.

Table 1. Relevant applications in the construction sector

Study	Application
Zheng et al. [19]	Natural language-based
	BIM information
	retrieval
You et al. [20]	Automated sequence
	planning in robotic
	assembly for
	construction tasks
Prieto et al. [21]	Automated construction
	schedule
Chen et al. [22]	Augmented reality-based
	safety information
	retrieval

Current vision-language model methodologies in construction predominantly concentrate on tasks such as BIM information retrieval, automated construction schedule generation, sequence planning, and safety information querying. To the best of our knowledge, even with the presence of vision-language models such as GPT-4V and MiniGPT-4, there is currently no

specialized vision-language model approach explicitly developed for the identification of ergonomic risks. In other words, without the presence of ergonomic experts, accomplishing real-time identification of ergonomic risks and generating human-like descriptions of the risks for timely prevention of WMSDs on construction sites or in front of surveillance screens becomes challenging. To address the identified gap of lacking methods for identifying and generating human-like language descriptions of the ergonomic risks, this study aims to introduce a vision-language model-driven approach. This method facilitates users without specialized expertise to obtain human-like language descriptions of ergonomic risks occurring on construction sites, with the objective of mitigating WMSDs. Additionally, the study seeks to investigate the feasibility of implementing such methodologies. The code, finetuned model, pre-trained model, and the text descriptions in the image-text pair datasets can be accessed here:

https://github.com/xinmingliUofA/ErgoGPT

2 Methodology

This study proposes a large vision-language modelbased ergonomic risk identification and seeks to assess the viability and performance of employing the visionlanguage model approach in the domain of ergonomics risk identification. To fill the research gap of lacking image-text pair datasets for ergonomic risk identification, finetuning and testing datasets comprising image-text pairs that portray scenarios relevant to ergonomic risks faced by construction workers were generated. In the first step, we conducted finetuning on the pre-existing Vicuna LLM using the finetuning dataset we curated, consisting of image-text pairs depicting ergonomic risks. Subsequently, we evaluated the performance of the finetuned vision-language model using the distinct dataset of image-text pairs designed specifically for testing its capability in identifying ergonomics risks. Subsequently, this study conducted a performance comparison between the model finetuned using the curated data for ergonomic risk identification and the model finetuned using generic data supplied by MiniGPT-4.

The MiniGPT-4 establishes a connection between the visual encoder and the LLM through the integration of a linear projection layer. MiniGPT-4 employs the open-sourced Vicuna as its language decoder, an LLM built on the foundation of LLaMA, capable of executing diverse and intricate linguistic tasks. For visual perception, it utilizes the open-sourced visual encoder in BLIP-2, incorporating a ViT backbone paired with their pre-trained Q-Former. The pre-trained model obtained through MiniGPT-4 is employed to acquire vision-language knowledge from an extensive dataset of image-

text pairs. The pretraining process involved the utilization of approximately 5 million image-text pairs sourced from the Conceptual Caption, SBU, and LAION datasets. [15]

The methodology section of this study is structured into three main steps. Initially, an image-text pair dataset was curated by a professional ergonomist, who provided text descriptions highlighting ergonomic risks associated with each photograph depicting construction workers' activities. These descriptions were subsequently transformed into an annotation file using a Python script. Secondly, the study involved finetuning a Vicuna LLM vision-language model utilizing the image-text pair dataset and a framework built upon MiniGPT. Additionally, a separate vision-language model exclusively employing Vicuna LLM was trained. Thirdly, the performance of both models derived from the second step was assessed. Specifically, the text descriptions generated by each model were compared to the ground truth using the testing dataset.

2.1 Generating alignment data for the visionlanguage approach

As models undergo pre-training in the initial phase, they sometimes face challenges in producing coherent human language output. Consequently, additional finetuning is necessary to address issues like repetitive phrases, fragmented sentences, and irrelevant content that may arise in models that are solely pre-trained. Similar issues are present in other approaches utilizing large language models, like GPT-3. GPT-3.5 mitigates these issues through a combination of instruction finetuning and reinforcement learning. [15], [23] Hence, to make the output results fluent and more akin to human language, finetuning the pre-trained model is indispensable.

Datasets for finetuning instruction and conversations are abundant in the field of natural language processing/large language models, yet they remain scarce in the realm of vision language, particularly in the context of ergonomic risk identification within ergonomic risk identification. As the primary objective of this study is to investigate the feasibility and proficiency of large visionlanguage models in responding to questions related to ergonomic risk assessment, the datasets are rooted in real-world situations involving construction workers. The dataset employed for finetuning incorporated 100 images featuring real-world construction workers engaged in various tasks, accompanied by captions detailing ergonomic risks associated with the depicted scenarios. Likewise, the dataset used for testing the finetuned vision-language model comprised similar images and captions focusing on ergonomics risks. The models used for testing include one finetuned with generic image-text pairs unrelated to ergonomic risks,

provided by MiniGPT-4, and the other finetuned using the dataset created in this study for finetuning purposes.

The image component of the generic data supplied by MiniGPT-4 for the finetuning of models does not pertain to construction workers or the identification of ergonomic risks-many of the images solely depict mere objects. Furthermore, the text component in the data provided by MiniGPT-4 lacks descriptions related to ergonomic risks or workers' postures that might cause ergonomic issues. Consequently, the approach employed in this study for the generation of image-text alignment data differs from that of MiniGPT-4. In the process of generating alignment data, MiniGPT-4 generates the description of chosen images using a pre-trained model and subsequently employs ChatGPT to refine these descriptions, addressing issues such as repetitive words and fragmented sentences. In contrast to MiniGPT-4, the text component of the datasets in this study was curated with input from an experienced ergonomist. In particular, the text descriptions aligned with the images in the image-text pairs of the finetuning and testing datasets were created by an ergonomist with three years of professional expertise. The internet-sourced images portraying construction workers in action, utilized as the dataset for finetuning and testing, were supplied to the ergonomist. Subsequently, the ergonomist assessed each image, responding to the question 'Describe the workers and their postures in the image and tell me if they are exposed to ergonomic risks due to their postures?' The ergonomist's responses were documented in a CSV file as ground truth.

This study chooses to adopt a methodology different from that of MiniGPT-4 for generating text descriptions corresponding to images in the dataset. The rationale lies in the fact that the generic datasets (Conceptual Caption, SBU, and LAION) employed for model pretraining were not crafted by ergonomists and lacked comprehensive ergonomic-related descriptions for the images within the dataset. Simply put, in light of the requirement for domain-specific knowledge in ergonomics within the pre-trained model, rather than employing the pre-trained model method in MiniGPT-4 to generate the text component for images, this study opts for input from ergonomists to guarantee the reliability of the relevant knowledge. Consequently, employing a pre-trained model to generate descriptions for selected images related to ergonomic risks is not deemed an optimal approach in this study.

The dataset creation process in this study unfolds through three stages. Initially, images are procured by conducting a search for 'construction worker' in a search engine, and relevant images depicting workers in action are selected from the search results. In this process, a total of 200 images were carefully chosen, with 100 allocated to the dataset later employed for finetuning the pretrained model and the remaining 100 set aside for evaluating the models' performance. Subsequently, text descriptions corresponding to the images are composed in the second step, a task accomplished by an experienced ergonomist. These text descriptions encompass an assessment of whether the workers depicted in the images are exposed to ergonomic risks, along with the factors contributing to these risks. The final step involves the construction of image-text pairs, wherein images are named in an incremental numerical format with a specific file extension (e.g., .jpg). Notably, a Python program was devised in this study to automatically integrate the text descriptions created by the ergonomist into an annotation file, denoted by the .json file extension.

Figure 1 shows an image sample from the finetuning dataset, and the ground truth text description of this image from the ergonomist is 'The image shows a worker wearing a white safety hat pouring cement on the road. The worker is exposed to awkward working posture because of the leaning forward posture, which is an ergonomic risk. The awkward posture may lead to the development of work-related musculoskeletal disorders.' Figure 2 shows an image sample from the testing dataset, and the ground truth text description of this image from the ergonomist is 'The image shows a worker standing in an awkward working posture on a black metal frame or a scaffold. The worker is exposed to ergonomic risks due to his overhead work posture, which may lead to workrelated musculoskeletal disorders. The worker may be exposed to fall hazards, and it is recommended to have a safety harness on him.'



Figure 1. An image sample from the finetuning dataset.



Figure 2. An image sample from the testing dataset.

2.2 Finetuning the pre-trained visionlanguage model

Following the initial phase, the pre-trained model underwent additional finetuning in the second phase to achieve coherent linguistic output. The finetuning phase involved the application of the Python script provided by MiniGPT-4, which was specifically designed for finetuning pre-trained models. The workstation's configuration for finetuning comprised an Intel Xeon Gold 6242 CPU, NVIDIA RTX A6000 GPU, and 128 GB of DDR4 memory. The workstation operated on the Ubuntu 22.04.1 LTS 64-bit operating system.

2.3 Testing the performance of the finetuned vision-language model

To validate the proposed approach in this study for identifying ergonomic risks using a vision-language model and to evaluate the performance of the finetuned models, the test dataset outlined in Section 3.1 was employed. Initially, the finetuned models were configured to serve as the vision-language models utilized by MiniGPT-4 for executing vision-language tasks. Subsequently, following the instructions provided by MiniGPT-4, the program was executed, and the finetuned models generated text descriptions corresponding to the images in the test dataset. The prompt used to generate text descriptions was 'Describe the workers and their postures in the image and tell me if they are exposed to ergonomic risks due to their postures?'

This study utilized a uniform prompt across the entire test dataset to maintain consistency in both the format and content of text descriptions across all test cases. Employing a uniform prompt for each test case also ensures fairness in the evaluations. Of greater importance, the ground truth text descriptions of the test dataset comprise descriptions of workers' actions and whether the workers are exposed to ergonomic risks based on these actions. Hence, the prompt is designed to focus explicitly on capturing descriptions of workers' actions and whether they are exposed to ergonomic risks. This uniform prompt aims to minimize unrelated factors when conducting a comparison between the ground truth text descriptions and the text descriptions generated by the finetuned model.

To quantitatively assess the enhancement attained by the model trained on the ergonomic risk dataset proposed in this study, this study additionally finetuned a model using the generic image-text pairs provided by MiniGPT-4. Following that, this study evaluated the performance of this model using the test dataset and the identical prompt. Subsequently, the accuracy of the two models was compared.

The prompt is framed as a yes-or-no question, inquiring whether the workers depicted in the images are subject to ergonomic risks. Considering that the objective of this study does not involve assessing or improving the accuracy of computer vision algorithms/models for human pose estimation, only the performance of whether the finetuned models can correctly identify ergonomic risks in the images was tested. The finetuned model's judgment is considered correct only when there is an agreement between ground truth and the finetuned model's judgment regarding whether workers in the images are exposed to ergonomic risks. The accuracy of both models on the test dataset can be expressed by equation 1.

$$Accuracy = \frac{number of correct cases}{number of all cases}$$
(1)
* 100%

3 Results

In contrast to the model finetuned with generic data from MiniGPT-4 (referred to as Model 1), the model finetuned with the dataset tailored for ergonomic risk identification among construction workers proposed in this study (referred to as Model 2) exhibited superior accuracy in recognizing ergonomic risks within the images of the test dataset. In terms of ergonomic risk identification, model 1 achieved an accuracy of 28% in generating text descriptions for the 100 test cases within the test dataset, whereas Model 2 achieved an accuracy of 81% for the same test cases, indicating a noteworthy accuracy obtained by testing Models 1 and 2 using the test dataset. Figure 3 illustrates an actual model 1 test case in the testing dataset featuring only one worker. Figure 5 shows the result from model 2 for the same test case. The ground truth of this test case is 'The image shows a worker standing on a ladder using a power tool to fix pipes on the roof of a pavilion. He is exposed to safety hazards because he is standing on the ladder without wearing a safety harness. He is exposed to ergonomic risk because he is leaning his head forward for a long time, which can strain the neck and shoulders, leading to discomfort and potential neck pain.' Meanwhile, Figure 4 illustrates an actual model 1 test case in the test dataset featuring multiple workers. Figure 6 shows the result from model 2 for the same test case. The ground truth of this multi-worker test case is 'The image shows workers laying bricks. They are wearing hard hats and gloves. They are exposed to ergonomic risks because they are squatting and bending forward, and their squatting and bending forward postures may lead to the development of work-related musculoskeletal disorders.'

Table 2. Obtained accuracy by testing Models 1 and 2 using the test dataset.

Model	Accuracy (%)	
Model finetuned with generic data	28	
Model finetuned with our data	81	



Figure 3. An example of model 1 generating a text description for an image containing a worker.

overreach.

Describe the workers and their postures in the image and tell me if they are exposed to ergonomic risks due to their postures?



In this image, several construction workers are seen working on a sidewalk. They appear to be in the process of laying bricks or some other type of construction material on the ground. One worker is wearing a hard hat while others are not. All of the workers are standing in a symmetrical position with their feet close together and their bodies straight. None of them appear to be experiencing any ergonomic risks due to their postures, as they are all maintaining proper posture and alignment. The image does not show any tools or equipment that could cause harm to the workers if used improperly. Overall, this image appears to depict a safe and secure construction site with no apparent ergonomic risks to the workers.

Figure 4. An example of model 1 generating text description for an image containing multiple workers.



Figure 5. An example of model 2 generating text description for the same image in Figure 3.





Figure 6. An example of model 2 generating text description for the same image in Figure 4.

4 Conclusions and future work

The dataset with a specific focus on identifying ergonomic risks among construction workers was created to finetune a GPT-based vision-language model. The proposed methodology showcased robust generalization capabilities when applied to the testing dataset derived from real-world scenarios. The model, which underwent finetuning using image-text pairs featuring ergonomic risk information, demonstrated an 81% accuracy in real-world test cases. This performance surpasses that of the model finetuned with generic image-text pairs lacking ergonomic risk information by 53%. These findings underscore the effectiveness of finetuning GPT-based vision-language models in achieving robust performance for the identification of ergonomic risks.

As the objective of this study is to explore the feasibility and ability of visual question answering related to ergonomic risk identification by finetuning vision-language models using data related to ergonomic risks, the correctness of the text descriptions generated by the models is entirely dependent on the correctness of the text descriptions for postures. The study's objective does not prioritize the precision of computer vision algorithms in recognizing the content of images. For instance, if the model identifies ergonomic risks for different reasons compared to the ground truth, it is considered a correct identification.

While computer vision algorithms have garnered significant attention for assessing ergonomic risks based on workers' actions in images, the unique and powerful capabilities of vision-language models in managing interactions between visual and textual elements have not been harnessed for ergonomic risk assessment. The validation experiments conducted in this study, which encompass scenarios featuring both multiple construction workers and a single construction worker, illustrate the efficacy of the proposed approach. This method, involving the finetuning of vision-language models with ergonomic-related data, demonstrates superior performance in ergonomic risk identification compared to models finetuned with generic image-text pairs data provided by MiniGPT-4. We aspire for the GPT-based vision-language approach presented in this study to play a role in identifying ergonomic risks for upcoming construction workers, thereby augmenting the well-being of workers and the safety of their work environments. Moreover, it is expected that this research will inspire further research of vision-language models for the identification of ergonomic risks and safety measures.

As interest in this field grows, the potential applications of data acquired by surveillance cameras at construction sites are expanding, moving beyond simple video storage and traditional computer vision tasks like object recognition. With large vision language models capable of extracting human-like language descriptions from surveillance data, they can aid or potentially supplant the role of ergonomists and safety experts in real-time video analysis and alerting workers or safety personnel about safety concerns on construction sites. Furthermore, individuals without specialized expertise can leverage this technology to produce injury or safety reports based on construction activities, leading to improved construction practices and streamlining the process of filing insurance claims.

5 Limitations

Despite the remarkable accuracy demonstrated by the proposed method, akin to the utilization of visionlanguage models in other fields, it has its limitations. These limitations revolve around issues such as language hallucination and insufficient perception capacities. As the vision-language model is constructed upon LLMs, its limitations in terms of language hallucination primarily result from inheriting the unreliable reasoning ability and hallucinating non-existent knowledge of LLMs. The inadequacy in perception capacities is predominantly linked to the limited visual perception capability of MiniGPT-4. Future research endeavors could mitigate language hallucination by further leveraging high-quality image-text pairs containing ergonomic information and refined LLMs. Addressing limited perception capacities may involve incorporating multiple layers in the projection layer, implementing a more robust visual perception model, and utilizing data that is more wellaligned.

In general, the mentioned limitations can be mitigated by integrating a wider range of diverse image-text pairs and refining the accuracy of textual descriptions detailing ergonomic risks depicted in the images. Expanding the pool of images and refining the accuracy of textual descriptions/annotations linked with these images has the potential to alleviate these limitations.

References

- European Agency for Safety and Health at Work. Work-related musculoskeletal disorders: prevalence, costs and demographics in the EU. Online: https://osha.europa.eu/en/publications/msdsfacts-and-figures-overview-prevalence-costs-anddemographics-msds-europe, Accessed: 14/12/2023.
- [2] U.S. Bureau of Labor Statistics. Injuries, Illnesses, and Fatalities. On-line: https://www.bls.gov/iif/factsheets/msds.htm, Accessed: 03/12/2023.
- [3] David G. C. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine* (*Chic Ill*), 55(3):190-9, 2005.
- [4] MassirisFernández M., Fernández J. Á., Bajo J. M., and Delrieux C. A. Ergonomic risk assessment based on computer vision and machine learning. *Computers & Industrial Engineering*, 149:106816, 2020.
- [5] Plantard P., Shum H. P. H., Le Pierres A.-S., and Multon F. Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied ergonomics*, 65:562-9, 2017.
- [6] Vignais N., Bernard F., Touvenot G., and Sagot J.-C. Physical risk factors identification based on body sensor network combined to videotaping. *Applied* ergonomics, 65:410-7, 2017.
- [7] Li X., Han S., Gul M., and Al-Hussein M. Automated Ergonomic Risk Assessment based on 3D Visualization. In 34th International Symposium on Automation and Robotics in Construction (ISARC 2017), Vol. 34, Taipei, Taiwan, 2017.
- [8] Guo S. Y., Ding L. Y., Luo H. B., and Jiang X. Y. A Big-Data-based platform of workers' behavior: Observations from the field. *Accident Analysis & Prevention*, 93:299-309, 2016.
- [9] Fan C., Mei Q., Yang Q., and Li X. Computervision based rapid entire body analysis (REBA) estimation. In *Modular and Offsite Construction* (MOC) Summit Proceedings, pages 90–97, Edmonton, Canada, 2022.
- [10] Jeong S. and Kook J. CREBAS: Computer-Based REBA Evaluation System for Wood Manufacturers Using MediaPipe. *Applied Sciences*, 13(2):938, 2023.

- [11] Fang W., Ding L., Love P.E., Luo H., Li H., Pena-Mora F., Zhong B., and Zhou C. Computer vision applications in construction safety assurance. *Automation in Construction*, 110:103013, 2020.
- [12] Barberi E., Chillemi M., Cucinotta F., Milardi D., Raffaele M., Salmeri F., and Sfravara F. Posture Interactive Self Evaluation Algorithm Based on Computer Vision. In *International Joint Conference on Mechanics, Design Engineering & Advanced Manufacturing*, pages 1516-1526, 2022.
- [13] Nayak G. K. and Kim E. Development of a fully automated RULA assessment system based on computer vision. *International Journal of Industrial Ergonomics*, 86:103218, 2021.
- [14] Seo J., Yin K., and Lee S. Automated Postural Ergonomic Assessment Using a Computer Vision-Based Posture Classification. In *Construction Research Congress 2016*, pages 809–818, Reston, United States, 2016.
- [15] Zhu D., Chen J., Shen X., Li X., and Elhoseiny M. MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models. arXiv preprint, arXiv:2304.10592, 2023.
- [16] Driessen T., Dodou D., Bazilinskyy P., and de Winter J. Putting ChatGPT Vision (GPT-4V) to the test: Risk perception in traffic images. 2023.
- [17] Yang Z., Li L., Lin K., Wang J., Lin CC., Liu Z., and Wang L. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). *arXiv preprint*, arXiv:2309.17421, 2023.
- [18] OpenAI. ChatGPT can now see, hear, and speak. On-line: https://openai.com/blog/chatgpt-can-nowsee-hear-and-speak, Accessed: 15/12/2023.
- [19] Zheng J. and Fischer M. Dynamic prompt-based virtual assistant framework for BIM information search. *Automation in Construction*, 1;155:105067, 2023.
- [20] You H., Ye Y., Zhou T., Zhu Q., and Du J. Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT. *Buildings*, 13(7), 2023.
- [21] Prieto S. A., Mengiste E. T., and García de Soto B. Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings*, 13(4):857, 2023.
- [22] Chen H., Hou L., Wu S., Zhang G., Zou Y., Moon S., and Bhuiyan M. Augmented reality, deep learning and vision-language query system for construction worker safety. *Automation in Construction*, 157:105158, 2024.
- [23] See A., Pappu A., Saxena R., Yerukola A., and Manning CD. Do massively pretrained language models make better storytellers?. arXiv preprint, arXiv:1909.10705, 2019.

Knowledge Graph-based Deconstruction Planning of Building's Products

Amr S. Allam and Mazdak Nik-Bakht

Department of Building, Civil, and Environmental Engineering, Concordia University, Montréal, Québec, Canada amr.allam@mail.concordia.ca, mazdak.nikbakht@concordia.ca

Abstract –

The transition from a linear economy (take-makedispose) to a circular economy in the construction industry (i.e., circularity in construction) necessitates adopting deconstruction instead of demolition. Deconstruction planning methods need to be investigated to facilitate this transition. Successful deconstruction planning requires considering various information to capture the destiny of the extracted building's product. Therefore, there is a dire need to propose a deconstruction planning method that is interoperable, easily integrated with various data sources, and conducive to stakeholder collaboration. To this end, this paper aims to propose a deconstruction planning method using Knowledge Graph (KG) technology. Firstly, the authors extracted the characteristics of the facility to be deconstructed, including hosting and hosted relations of components. Secondly, the characteristics of the facility's products were transformed into a KG. Finally, disassembly rules were defined, and new knowledge was inferred via automated reasoning. The developed method was tested on two case studies, involving two-dimensional and three-dimensional representations. The results aligned with other methods in the literature, requiring fewer inputs.

Keywords -

Deconstruction planning; Disassembly sequence; Knowledge graph; Circularity in construction

1 Introduction

The construction industry has been identified as one of the largest consumers of virgin materials and a major contributor to landfills through Construction, Renovation, and Demolition (CRD) waste [1]. It is estimated that the construction industry is responsible for more than 30% of the world's total waste, and it consumes around 50% of the world's virgin materials [2]. The End-of-Life (EoL) phase of the built facility is the most critical phase regarding the amount of waste generated from the construction industry; more than half of the generated waste is caused by the implemented scenario to handle the facility at the EoL [3]. Two EoL scenarios can be implemented, namely, demolition and deconstruction [4]. Demolition is the act of destroying a built facility regardless of the recoverability of its products (materials/components/subsystems); most of the generated waste is landfilled with little consideration for recycling [5]. The resource-friendly scenario is deconstruction, which is a planned disassembly of products from the built facility. The output of this scenario can serve several purposes such as building relocation and repurposing, product reuse, and recycling [4].

Demolition, with its short-term economic benefits and quicker process, remains the prevalent EoL scenario compared to deconstruction, which offers environmental, social, and long-term economic benefits. In response to this, the deconstruction planning research line has gained prominence. Deconstruction planning is divided into strategic and operational planning [6], [7]. Strategic planning is more high-level (e.g., minimizing the total duration of the deconstruction project), while operational planning is more in-detail and tactical (e.g., minimizing the duration of a single activity). Both planning methods are crucial and complement each other; strategic planning objectives may serve as the set of constraints to be followed in operational planning.

Deconstruction planning requires integrating various pieces of information to make informed decisions on how to execute deconstruction most efficiently, including (i) building type; (ii) building age and condition; (iii) Bill of Quantities (BoQ); (iv) the existence of hazardous materials; (v) health and safety considerations; and (vi) legal requirements [6]. We refer to the information required for deconstruction planning as 'disassembly information'. Achieving the main goal of deconstruction, circularity in construction, necessitates three main pillars: deconstructability, capability, and marketability [8]. Deconstructability involves the ability of a facility's products to be deconstructed, handled, loaded, and transported, encompassing information on type of the facility, products' type, quantity, and condition, as well as working space, and lifespan. The capability pillar

focuses on stakeholder and infrastructure readiness, collecting information such as contractor resources, infrastructure capacity, and legal requirements [9]. In terms of the marketability pillar, engaging potential customers and stakeholders early on is essential to quantify the demand for reused products.

Although the labor-intensive nature of deconstruction yields environmental and social benefits [7], [8], it remains one of the primary reasons stakeholders opt for demolition to expedite the removal of the built facility [10]. With the imminent large-scale retirement of aging facilities, there is a dire need to shorten the required duration for implementing deconstruction. To achieve this objective, it is essential to plan each deconstruction activity before its execution [6]. To this end, this paper aims to propose a deconstruction planning method that is interoperable, easily integrated with various data sources, and conducive to stakeholder collaboration to facilitate objective decision-making. The proposed method adopts a bottom-up approach, specifically focusing on operational deconstruction planning.

2 Existing Operational Deconstruction Planning Methods

In practice, operational deconstruction planning is based on professionals' experience, which leads to suboptimal workflow due to the subjective nature of decision-making. To provide an objective method for deconstruction planning Sanchez and Haas (2018) developed a single-target Sequence Disassembly Planning for Buildings (SDPB). The proposed method was inherited from the Disassembly Sequence Structure Graph (DSSG) model theory for manufactured products. The disassembly graph of this method consisted of 11 constraint matrices; 9 of them representing the physical, functional, and interdependence between components and fasteners, while the other two related to the environmental impacts and the cost of the disassembly works. The output of this method is an inverted tree graph of the chosen path of disassembly; root nodes in the inverted tree represent target components, leaf nodes represent parts that constrain the target components, and the links between them represent constraints. This research line that Sanchez and Haas (2018) started was the keystone for other works to explore the field of operational deconstruction planning.

In this sense, Sanchez et al. (2019) extended the SDPB to include multiple targets. The multiple-target SDPB was obtained by merging the all single-target SDPB, by matching the identical components that have the same extraction direction. To cut the computational time, they introduced limits of design, which represents the unmodified components in the models. the components within the limits of design were not included

in the disassembly model but however they were considered as physical and motion constraints. The output of the multiple-target SDPB was exported to Microsoft Project (MS) to develop the deconstruction baseline schedule. Yet, the developed SDPB methods rely only on a single deconstruction method. In this, another extension of SDPB was considering different deconstruction methods such as selective demolition, destructive disassembly, and perfect disassembly [13]. In the same vein, Mahmoudi Motahar and Hosseini Nourzad (2021) proposed a hybrid method for disassembly planning for buildings that support sequential and parallel approaches. Instead of using expert rules to cut the computational cost, they utilized the Non-dominated Sorting Genetic Algorithm (NSGA-II).

All the previously mentioned methods are mainly based on physical, interdependency, accessibility, and motion constraints. However, as mentioned in the introduction section, disassembly information includes more attributes than the ones mentioned in the previous studies. In this context, Denis et al. (2018) proposed a Disassembly Network Analysis (DNA) method that plans for disassembly by considering several product's attributes such as accessibility, transportability, condition, weight, reversibility of connection, disassemble time, demolition time, and sequential dependence. Based on the values of these parameters, a flowchart of four steps has been followed to (i) check the possibility of disassembling the target component; (ii) identify potential paths to access the target component; (iii) determine the recoverable elements and lost ones during disassembling; and (iv) make decisions based on recovered elements, lost elements, and disassembly time for each path.

Deconstruction is a significant step towards the adoption of circularity in construction. To ensure reaching this goal, post-deconstruction (i.e., the destiny of the extracted building's products) needs to be considered [16]. This necessitates integrating various information beyond the building level such as technical feasibility, market feasibility, and legal environment in the region. Given these complexities, there is a pressing need to conduct deconstruction planning using methods that are interoperable, easily integrated with various data sources, and conducive to stakeholder collaboration. This is where semantic web technologies can play a crucial role. Recently, in the manufacturing industry, studies are modeling disassembly information using Knowledge Graphs (KGs) to describe manufactured products [17]. Knowledge graphs are capable of modeling, consolidating, and deducing insights from intricate, diverse data originating from various sources, offering scalability, expressiveness, and extensibility [18]. Its strength resides in its semantic processing and

interconnected organizational abilities, forming the foundation for intelligent information applications [19].

3 Disassembly Information Modeling

To construct the KGs in this study, three sequential steps were followed by the authors. Firstly, extracting the characteristics of the facility to be deconstructed including hosting and hosted relations of components. Secondly, the characteristics of the facility's products were transformed into knowledge graph. Finally, disassembly rules were defined, and a new knowledge was inferred via automated reasoning.

To test the proposed method for deconstruction planning, a simplified typical building frame assembly was used as a case study in both two and threedimensional representation, as shown in Figure 1. This case study was introduced by Sanchez and Haas (2018a) and has been utilized in the literature to prove the concept of other deconstruction planning methods [14], [15], [20]. Figure 1 illustrates the components and connections of the assembly in both dimensions. This information was translated into two matrices: a hosted component constraint matrix (HC) and a liaison constraint matrix for components (LC). Due to space limitations, the HC and LC matrices of the two-dimensional case study are shown in Equation 1 and Equation 2, respectively. HC records the individual relationship between host and hosted components, while LC documents the fasteners physically attaching the hosted components to the hosting component under analysis. Both HC and LC were utilized as inputs for the selective disassembly sequence planning method proposed by Sanchez and Haas (2018a).

1	ך <i>HC</i> 1 ך		r C4	1	(1)
	HC2		С5		
	НСЗ		С6		
	HC4		С7		
	HC5	_	<i>C</i> 7, <i>C</i> 8		
	HC6	_	С8		
	HC7		C9,C10	1	
	HC8		0		
	HC9		0		
	HC10		0	J	
	[LC1	1	[^{f1}]		(2)
	LC2		f4		
	LC3		f7		
	LC4	1	<i>f</i> 2		
	LC5	_	f3,f5		
	LC6	-	<i>f</i> 6		
	LC7		f8, f9		
	LC8		0		
	LC9		0		
	LLC10	J			

The information extracted in the previous stage was

represented by Resource Description Framework (RDF) triples to form the Deconstruction Knowledge Graph (DKG). RDF serves as a versatile and universal data model employed for the representation and amalgamation of data through directed labeled graphs [21]. Each triple consists of two nodes (i.e., subject and object) connected with an edge (i.e., predicate) that defines the relationship between them. In this research, Blazegraph Database was utilized to construct the knowledge graph of the case study [22]. RDF triples were encoded using turtle format. Figure 2 shows part of the DKG of the 2D case study that includes all the in-flow and out-flow edges of nodes C7, C9, and C10. Two kinds of nodes were utilized in the DKG: Internationalized Resource Identifier (IRI), represented by circular borders, and literal nodes, represented by no borders. IRI nodes are used to describe entities that will have a unique addresses or references, while literal nodes are used to represent values such as strings, numbers, or dates. The relationships between the nodes were described using 6 predicates as depicted in, Table 1. Two of these predicates, indicated by red arrows in Figure 2, were inferred after applying specific rules. More details about the inferred knowledge will be explained in the next paragraph.



Figure 1. The prototype building in 2D and 3D - adapted from Sanchez and Haas (2018a)



Figure 2. Part of the disassembly knowledge graph for C7, C9, and C10 of the 2D case study (Processed by Arrows.app)

Table 1. Predicates used to define relationships.

Predicate	Description	Reciprocal
		relation
:host	Records hosting	:is_hosted
	hosted relationship	
	between components	
:block	Records what	:is_blocked
	components block the	
	accessibility to the	
	target component	
:has_to_be	Records the	:disassembled
_disassem	predecessor(s) that	_after
bled_befor	need to be	
e	disassembled before	
	disassembling the	
	target component	
:connected	Records how the	NA
_by	hosted component is	
	connected to the	
	hosting component	
:has_mater	Describes the	NA
ial	material of the	
	component	
:has_type	Describes the	NA
	component and	
	connection type	

Some rules will be used to infer new knowledge that will be used to develop the deconstruction plan. In the manufacturing industry, disassembly task planning (i.e., deconstruction planning) refers to the generation of the sequence of tasks, disassembly direction, and the tool to be used [23]. Table 2 lists the rules which were covering two main aspects, structural stability and accessibility. The structural stability was defined based on the host hosted relationship, while the accessibility was defined based on the theory of building layers, which describes six classes according to their life expectancy, stuff, space plan, services, skin, structure, and site [24]. It should be noted that other rules were encoded to define the inverse relationship between node (when applicable). The rules were applied utilizing SPARQL query language that is capable of retrieving and updating data stored in RDF format.

Table 2. Rules used to infer new knowledge

Rules	Description
?c1 :host ?c2.	If a component (C1) is
2c1 :host $2c3$	hosting other
	acomponents (C2 and
	C_{2} Components (C2 and
?c3 a :Space_Plan >	C3; $C2$ is either
?c3 :block ?c2	plumping, electrical,
	mechanical, or
	hydraulics; and C3 is
	either internal wall
	partitioning finish or
	furniture, then the space
	furniture, then the space
	plan layer (C3) block the
	accessibility to the
	fasteners of the service
	layer C2.
?c1 :host ?c2 >	If a component (C1) is
2c2 thas to be disassem	hosting another
hlad hafara 2a1	a = 100000000000000000000000000000000000
bled_belore ?e1	component $(C2)$, then the
	nosted component $(C2)$
	needs to be disassembled
	before the hosting
	component (C1) to
	ensure the stability of the
	structure.
2c1 ·block 2c2>	If the fastener of
201 that to be disassem	component (C1) needs to
bled_before ?c2	be accessed and another
	component (C2) is
	restricting its
	accessibility, then the
	barrier component (C2)
	needs to be disassembled
	to reach the fastener of
2-1-1	C1.
?c1 :nost ?c2 >	If a component (C1) is
?c2 :1s_hosted ?c1	hosting another
	component (C2), then the
	C2 is hosted by C1.
?c1 :block ?c2 >	If a component (C1) is
?c2 is blocked ?c1	restricting the access to
	the fastener of another
	and a state of a state of a state of the s
	C2 = 11 = 01
	C2 is blocked by CI.
?c1:has_to_be_disassem	If a component (C1)
bled_before?c2 >	needs to be disassembled
?c2 :disassembled after?	before another
c1 _	component (C2), then the
	• ` ''

C2 can be disassembled after C1.

4 Model Solution and Deconstruction Plan

To assess the effectiveness of the proposed knowledge graph-based deconstruction planning method, two case studies were employed. As illustrated in Figure 1, the 2D example comprises 10 components and 9 connections, whereas the 3D example involves 21 components and 22 connections. SPARQL query language was used to retrieve the deconstruction plan from each example, as shown in Figure 3. The retrieved deconstruction plan is divided into two main parts. The first part records the phase in which the deconstruction activities take place; activities within the same phase can be performed in parallel. The second part provides a description of the component to be disassembled, including its connections to the hosting component. The proposed method in this study was tested by developing the deconstruction plan to extract the target components (C5) and (C19) in the two- and three-dimensional case studies, respectively. These two components were chosen because their deconstruction plan was proposed using the method developed by Sanchez and Haas (2018a).

Figure 4 shows a visual representation of the results of the 2D and 3D case studies. Four phases were required to extract the target component (C5) in the 2D example. In the first phase, two components can be disassembled in parallel. Then, starting from phase B and all the way to phase D one component can be disassembled per each phase. In the three-dimensional case study, the deconstruction plan of the target component (C19) was sequential, i.e., only one component per each phase. The two knowledge graph-based deconstruction plans are aligned with the plans proposed by Sanchez and Haas (2018a). What sets the knowledge graph-based deconstruction planning method apart from the previous methods is that it needs fewer inputs to develop the plan. The proposed method did not require inputs indicating physical and motion constraints for fasteners and components, and still it provided the same results with 5 matrices less from the existing methods in the literature.

It should be noted that the proposed KG-based deconstruction planning method serves as a proof of concept. In the upcoming stages of this ongoing project in deconstruction planning, various pieces of information will be integrated, including both deconstruction and post-deconstruction phases.

```
WHK - SPAROL Query

PREFIX : <http://disassembly-sequence-example2d.com/>

SELECT ?s ?Component_Name ?f ?Connection ?Phase
Where {
    ref = refut to the disassembled_before* :C5;
    iconnected_by ?f;
    itype ?ConnectIon
    BIN0 {
        COALESCE(
        If EXISTS { ?s a :Stuff } && EXISTS { ?s :host ?c}, "B", 1/0},
        If EXISTS { ?s a :Stuff } && EXISTS { ?s :host ?c}, "B", 1/0},
        If EXISTS { ?s a :Stuff } && EXISTS { ?s :host ?c}, "B", 1/0},
        If EXISTS { ?s a :Stuff } && EXISTS { ?s :host ?c}, "B", 1/0},
        If EXISTS { ?s a :Stuff } && EXISTS { ?s :host ?c}, "B", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Plan } && EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Ref & EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Ref & EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Ref & EXISTS { ?s :host ?c}, "C", 1/0},
        If EXISTS { ?s a :Space Ref & EXISTS { ?s :host ?c}, "C", 1/0},
```

Figure 3. A snapshot of the applied SPARQL query to plan for deconstruction operations of the 2D case study

5 Conclusion

The transition towards circularity in the construction industry is gaining momentum, primarily to achieve the sustainable development goals, especially SDG12 (Responsible Consumption and Production). In this context, deconstruction should be adopted instead of demolition at the end-of-life (EoL) stage. Unlike demolition, deconstruction is a labor-intensive task that requires detailed planning. Therefore, this paper developed a deconstruction planning method based on knowledge graphs. The developed method was tested on two case studies, involving two-dimensional and threedimensional representations. The results aligned with other methods in the literature, requiring fewer inputs. The developed method was able to infer new knowledge with minimum inputs possible.

The main contribution of this work was the proof of concept that knowledge graphs can be used to develop deconstruction plans in the construction industry. However, the work has the following limitations: (i) deconstruction performance criteria, such as deconstruction cost, environmental performance, and the duration of each activity, should be considered while developing the plan; and (ii) the two case studies mainly focused on structural components.



Figure 4. The deconstruction plan for components 5, and 19

6 References

- Ajayebi, A., P. Hopkinson, K. Zhou, D. Lam, H. M. Chen and Y. Wang, "Spatiotemporal model to quantify stocks of building structural products for a prospective circular economy," *Resour. Conserv. Recycl.*, vol. 162, no. June, p. 105026, 2020.
- Miller, N., "BBC Future," 2021. [Online]. Available: https://www.bbc.com/future/article/20211215the-buildings-made-from-rubbish. [Accessed: 18-Feb-2023].

- [3] Akanbi, L. A. *et al.*, "Salvaging building materials in a circular economy: A BIM-based whole-life performance estimator," *Resour. Conserv. Recycl.*, vol. 129, no. May 2017, pp. 175–186, 2018.
- [4] Akinade, O. O. *et al.*, "Waste minimisation through deconstruction: A BIM based Deconstructability Assessment Score (BIM-DAS)," *Resour. Conserv. Recycl.*, vol. 105, pp. 167–176, 2015.
- [5] Panizza, R. O., A. S. Allam, A. Kasliwal and M. Nik-Bakht, "Labeling Construction, Renovation, and Demolition Waste through Segment Anything Model (SAM)," in *Construction Research Congress 2024*, 2024.
- [6] Hübner, F., R. Volk, A. Kühlen and F. Schultmann, "Review of project planning methods for deconstruction projects of buildings," *Built Environ. Proj. Asset Manag.*, vol. 7, no. 2, pp. 212–226, 2017.
- [7] Allam, A. S., and M. Nik-Bakht, "From demolition to deconstruction of the built environment: A synthesis of the literature," J. Build. Eng., vol. 64, no. 15679, pp. 1–18, 2023.
- [8] Allam, A. S., R. O. Panizza and M. Nik-Bakht, "A SWOT Analysis for Deconstruction of the Canadian Built Environment," in *Proceedings of* the Canadian Society of Civil Engineering Annual Conference 2023, 2023, pp. 1–13.
- [9] Panizza, R., and M. Nik-Bakht, "Building stock as a future supply of second-use material – a review of urban mining methods," *Waste Manag. Bull.*, vol. 2, no. 2, pp. 19–31, 2024.
- [10] Allam, A. S., and M. Nik-Bakht, "Barriers to Circularity in Construction: An analysis of experts' perspectives," in 2023 European Conference on Computing in Construction, 2023.
- [11] Sanchez, B., and C. Haas, "A novel selective disassembly sequence planning method for adaptive reuse of buildings," *J. Clean. Prod.*, vol. 183, pp. 998–1010, 2018.
- [12] Sanchez, B., C. Rausch and C. Haas, "Deconstruction programming for adaptive reuse of buildings," *Autom. Constr.*, vol. 107, no. August 2018, p. 102921, 2019.
- [13] Sanchez, B., C. Rausch, C. Haas and R. Saari, "A selective disassembly multi-objective optimization approach for adaptive reuse of building components," *Resour. Conserv. Recycl.*, vol. 154, p. 104605, 2020.
- [14] Mahmoudi Motahar, M., and S. H. Hosseini Nourzad, "A hybrid method for optimizing selective disassembly sequence planning in adaptive reuse of buildings," *Eng. Constr. Archit. Manag.*, 2021.

- [15] Denis, F., C. Vandervaeren and N. De Temmerman, "Using network analysis and BIM to quantify the impact of Design for Disassembly," *Buildings*, vol. 8, no. 8, pp. 1–22, 2018.
- [16] Allam, A. S., and M. Nik-Bakht, "Supporting circularity in construction with performancebased deconstruction," *Sustain. Prod. Consum.*, vol. 45, pp. 1–14, 2024.
- [17] Wu, H., Z. Jiang, S. Zhu and H. Zhang, "A Knowledge Graph Based Disassembly Sequence Planning For End-of-Life Power Battery," *Int. J. Precis. Eng. Manuf. - Green Technol.*, no. 0123456789, 2023.
- [18] Kebede, R., A. Moscati, H. Tan and P. Johansson, "Circular economy in the built environment: a framework for implementing digital product passports with knowledge graphs," in 2023 European Conference on Computing in Construction, 2023.
- [19] Peng, J., and X. Liu, "Automated code compliance checking research based on BIM and knowledge graph," *Sci. Rep.*, vol. 13, no. 1, pp. 1–12, 2023.
- [20] Mahmoudi Motahar, M., S. H. Hosseini Nourzad and F. Rahimi, "Integrating complete disassembly planning with deconstructability assessment to facilitate designing deconstructable buildings," *Archit. Eng. Des. Manag.*, pp. 1–18, 2023.
- [21] Kebede, R., A. Moscati, H. Tan and P. Johansson, "Integration of manufacturers' product data in BIM platforms using semantic web technologies," *Autom. Constr.*, vol. 144, no. March, p. 104630, 2022.
- [22] "Blazegraph Database." [Online]. Available: https://blazegraph.com/. [Accessed: 11-Dec-2023].
- [23] Yu, J., H. Zhang, Z. Jiang, W. Yan, Y. Wang and Q. Zhou, "Disassembly task planning for end-oflife automotive traction batteries based on ontology and partial destructive rules," *J. Manuf. Syst.*, vol. 62, no. 947, pp. 347–366, 2022.
- [24] Brand, S., *How Buildings Learn: What Happens After They're Built.* 1995.

Development of Online Course for Open Infra Built Environment Information Model

Annika Kemppainen¹, Tanja Kolli¹ and Rauno Heikkilä¹

¹Civil Engineering Research Unit, Department of Technology, University of Oulu, Finland <u>Annika.kemppainen@student.oulu.fi</u>, tanja.kolli@oulu.fi, rauno.heikkila@oulut.fi

Abstract –

Infra Built Environment Information Modeling (infraBIM) enables efficient and accurate planning, construction, and maintenance of complex infrastructure systems. InfraBIM has spread worldwide and is used in various countries for infrastructure projects. InfraBIM experts are essential to take full advantage of the benefits of digital information modeling.

The goal of the article is to find the structure of the infraBIM online course. For this reason, a total of 19 online courses related to building information modeling (BIM) were explored based on their available title level structuality. The online courses were then analyzed and tracked with 17 subject occurrences in the courses. The use of artificial intelligence (AI) was tested to get information and its correctness as well as to prepare the test task material for the course. International interviews were done to study the current situation of InfraBIM. The course module was piloted during summer 2023 and the feedback was collected and analyzed. As a result, the open InfraBIM online module course was developed to promote international utilization and the learning process.

Finally, the general structure of the developed online module course serves as a good starting point for the further development of an open and multilingual website for learning more about open InfraBIM.

Keywords -

Open InfraBIM; online course; marketing; elearning; AI

1 Introduction

The level of use of building information modeling (BIM) varies from country to country on an enormous scale [1]. In 2010, the implementation rate of information modeling in the Western European construction sector was only slightly more than a third (36%) [2]. In Europe, the implementation of BIM in construction companies was only 29 % in 2016 [3]. However, the estimation of

BIM implementation is exceedingly difficult due to the fact each country has a different BIM maturity Level [4]. One reason could be the variety of technologies, which makes open infrastructure BIM complex and challenging. According to Salzano et al. (2023), one solution is to develop and increase the know-how of technological aspects when solving case study problems [5].

Increasing the use of the BIM in industry also requires that students and professionals develop the necessary skills and competencies to succeed in the modern infrastructure industry. Therefore, multiple construction and consulting companies are offering paid education. According to the NATSPEC Report (2023), 22 countries have BIM education in higher education and technical training institutions (vocational education) with a focus on Architecture, Engineering, and Construction (AEC) students. However, only 10 countries, such as Finland, Sweden, and Germany, have also BIM education both physical and/or virtual courses, and training for civil engineering. In addition, road construction-related projects have also been conducted between different countries, for example ERASMUS+ project inROAD. [6] The knowledge of educator's base and skills in information technology tools and software, constantly changing or lacking instructions or requirements, students of different levels, availability of both financial and physical resources, and resistance to change among educational institutions and their staff are the main reasons why for example the education of infraBIM is not spread worldwide. [6]

The introduction of infrastructure modeling has been promoted in countries by many parties when the benefits of its use have come to the attention of governments. At the European level, the use of infrastructure modeling is a requirement in public sector projects, in addition to Finland, for example in Sweden, Norway, and Great Britain. The Nordic countries have been at the forefront of introducing open infrastructure modeling. In Finland, a lot of work has been done since 2010 [7] to develop an open infra built environment model (infraBIM) to present information about infrastructure objects (such as roads) that is suitable for digital communication, interpretation, and processing by people and computer applications. The concept of open InfraBIM includes Common InfraBIM Requirements (YIV2021), classification, and open format. YIV2021 supplies general requirements and guidelines for model-based projects as well as more detailed model technical requirements [8]. Classification presents the numbering and naming conventions covering the life cycle of infrastructures and models. InfraModel (IM), which is based on the international LandXML standard, is recommended for use as an open format to transfer infrastructure data.

Globally, there are still a limited number of countries that are pioneers in infrastructure modeling [9,10]. In Central Europe, in several countries, the public sector has recommended the use of infrastructure modeling and has taken development steps to promote its use by creating standards, implementing programs, and conducting pilot projects [9,10]. Germany is a potential export country for InfraBIM technologies due to the ongoing development of infra modeling there, which is why the country was selected as one of the examples countries for a closer examination. Germany has already seen potential savings with the "selective use" of infrastructure modeling in small, medium, and large operations [11].

However, the need to harmonize and update the requirements has already been found in Germany, and the vision of the Standardization of Data Modelling of the Built Environment project program launched in 2022 is to unify the data modeling requirements of the built environment, which includes, for example, the standardization of guidelines, formats, and nomenclature by 2025 [8].

In this article, the online pilot course for InfraBIM was developed and tested. For this purpose, the study of other online courses related to the topic was done and analyzed. The material was selected based on the interview of experts and using modern tools, such as artificial intelligence.

2 Development of the online course

2.1 Study of BIM-related online courses

A total of 19 online courses [12-30] were selected in this paper. The following criteria were to choose 1) must be online courses that are openly available, 2) can be completed independently online, 3) are at the basics level, 4) course materials must be available in English, and finally 5) the course must be found on the Google search. Search phrases that were used were BIM for the infrastructure course, InfraBIM, and course free. BIM for infrastructure course, InfraBIM course free, InfraBIM course EU (European Union), InfraBIM e-learning, and variations of those. The phrases were selected based on what regular information seekers would be expected to use to find information about the topic.

2.2 Determination of the content of the course

A methodical analysis was utilized to examine the content frameworks of the identified Building Information Modeling (BIM) courses. The evaluation was based on the criteria outlined in section 2.1. Initially, each course was scrutinized based on the available information, without any payment or login constraints. The table of contents or main headings were reviewed because of their accessibility and functioning as a decision-making tool for undergoing the course. Subsequently, data was systematically collected in Excel from the table of contents, outlining the inclusion of construction information modeling topics within these BIM courses at a fundamental level, along with their sequence. A color-coded system was used to identify similarities among topics across different courses. Finally, 17 topics were chosen based on shared themes and the required content.

2.3 Interview study of the experts

In the interview study, the point of view of experts about infrastructure modeling in Finland, Germany, and Denmark was investigated, and ideas about the content and structure of the website were mapped. The interviews were done by Teams meetings during spring 2023.

 Table 1. Profile of BIM experts from Finland, Germany, and Denmark.

Country	Profile of expert
Finland	Software industry
Finland	Development Manager in Bridge and
	Concrete Structures, Product and
	Equipment Development for a
	Contractor
Finland	Development Manager for BIM
Germany	BIM Coordinator and Consultant
Germany	Software Industry
Denmark	Senior BIM Specialist and Project
	Manager

2.4 Use of artificial intelligence

ChatGPT is a chatbot developed by OpenAI based on artificial intelligence (AI) technology, which is trained using a large language model. ChatGPT in dialogue format can produce human-like text and communicate interactively, as well as learn from earlier conversations. The browser-based ChatGPT was released to the public in November 2022 and is currently freely available in a research phase gathering feedback. [31]. In this paper, the use of ChatGPT was briefly investigated for creating multiple-choice tests and summarizing entities. The free version of ChatGPT was presented with natural language tasks and questions in Finnish and English, and its ability to translate a multiple-choice survey into German was tested. ChatGPT's summarizing skills were tested with three questions presented in different ways: "How to summarize BIM or Building Information modeling in the infrastructure industry", "What is InfraBIM" and "What does Open InfraBIM mean" to see how ChatGPT's answers differ according to the layout of the questions and how it interprets the concepts of BIM in the infrastructure industry, InfraBIM, and Open InfraBIM.

2.5 Pilot course

The pilot course aimed to test the idea of online InfraBIM. The focus of the pilot course was decided to be a summary of the basics of the world of InfraBIM, model-based construction, and benefits. The Microsoft Sway application was used as a tool in the course publication to implement the course. The content of the course included text, pictures, a text-to-speech video, Teams-interview video, a case study, and a quiz. The language of the course was decided to be English, since one of the target groups is Germany.

The link to the pilot course can be found on the Ouluzone+ web pages. [32] The link to the course was distributed by e-mail to the persons involved in the topic, selected experts, and contacts from Germany. Permission was granted to share the link freely. A total of sevenquestion survey was done by Microsoft Forms to receive feedback. The answers were analyzed and used to develop the general structure for an InfraBIM online course.

3 Results

3.1 The availability of infraBIM courses

When doing some background research with the Google search engine, it soon became clear that despite several standard search terms such as BIM for infrastructure course, BIM for infrastructure online course, InfraBIM course, InfraBIM course free, InfraBIM course EU, etc., no website or online course dealing with the basics of infrastructure modeling could be found freely online at hand. With the selected criteria like free access, independent online, basic level, in English, finding online courses was challenging even for construction information modeling, as even most free courses require the creation of credentials to view the content. After expanding the criteria, the lists of 19 BIM courses were collected. Some of the founded courses were online courses offered by universities such as Coventry University [13], National Taiwan University [15], Purdue University [16] using for example edX.org or Coursera.org platforms which may have included

lecture teaching, too. Many of the BIM courses are produced by industry-specific company such as TÜVSÜD and HOCHTIEF ViCon [17, 26]. Si!BIM course is the ERASMUS+ project which is focused on vocational training [30]. In addition, many of the courses can be found on LinkedIn Learning (formally known as Lynda.com) [12], Udemy online learning platform [14, 18-24] as well as Alison.com [25] or Reed.co.uk [27-29]. The names of the 19 studied BIM-related online courses are listed in Figure 1, and the number of subjects is presented.



Figure 1. A total of 19 BIM online courses [12-30] were analyzed related to specific 17 subjects.

3.2 Content of the studied BIM courses

Since one of the criteria was free access to the course, only the available material, e.g. the titles, were analyzed. A total of 17 BIM-related subjects was selected to study in more detail and the occurrence in the course was calculated (see Figure 2). The studied subject had to come out clearly from the title. The lengths of the courses varied so a more precise location of the subjects' determinations was not made.



Figure 2. The occurrence of BIM subjects in 19 online courses [12-30].

What the courses had in common was to get started by somehow defining BIM, what BIM is, or why it is needed in the construction industry. The benefit aspect was therefore placed at the top in several courses, but it also appeared at the very end of the courses. The category that appeared the most in the courses was "implementation/use of BIM" in a total of 13 courses. Other most popular topics were "Benefits/why BIM" and "Communication/Collaboration", which appeared in the bodies of nine (9) courses, and "Models/Crash Analysis" and "Tools/technology/software" which were in the bodies of seven (7) courses. Not a single course had those topics appearing at the same time. The clearest positions can be seen with the categories "Future of BIM", which without exception came last in the course, if it was included in it, and "BIM's dimensions", which was the second first topic in both of its courses. For several topics, it is possible to find whether they are most often found before or after the middle of the course. Titles related to sustainable development, or the infrastructure sector appeared in only one course. BIM in design was included in the courses a little more often than BIM in construction. The terms model-based construction, machine control, machine automation, open BIM, or open data did not appear in the title of any course.

3.3 The result of expert interviews

The first query revolved around the current state of BIM implementation in the interviewee's respective country. In the Finnish interviews, significant attention was drawn to the term "infrastructure information modeling." Interviewees expressed concerns over its lack of a clear or universally understood definition within the infrastructure sector, complicating its application. While it was acknowledged that the adoption of information modeling in the infrastructure sector lagged behind that of the construction industry, there was optimism about its potential to surpass the latter in development.

In contrast, in Germany, the discussions dealt with varying degrees of maturity of the infrastructure sectors, ranging from 0 to 3, depending on the sector. In particular, engineering firms demonstrated a more comprehensive understanding and use of infrastructure modeling compared to other sectors. The design sector appeared as the primary focal point around 2014-2015 when discourse on infrastructure modeling started. Larger design firms swiftly embraced infrastructure modeling, reaching a commendable maturity level between 2 and 3. Medium-sized firms fell within levels 1 and 2, with smaller entities typically at level 1. Interest among contractors in infrastructure modeling burgeoned around 2020, possibly catalyzed by pressure from the design sector. Larger contracting companies approached a maturity level nearing 1, while smaller contractors largely refrained from model-based construction due to perceived redundancy. The spectrum of maturity levels among infrastructure owners appeared notably broad, spanning from 0 to 3. For major infrastructure stakeholders, such as Germany's railways, early endeavors related to infrastructure modeling signify considerable potential for digitalization advancement.

According to a Danish BIM expert, Denmark occupies an intermediate position, akin to Sweden, rather than being at the forefront of infrastructure modeling use. Although infrastructure modeling is still relatively uncommon, there is a gradual ascent in maturity within the sector. The adoption of certain Finnish open infrastructure modeling practices underscores this Denmark faces evolution. However, challenges, particularly in aligning with the pace of the construction industry. The longitudinal nature of infrastructure objects poses a distinct hurdle, complicating their portrayal as 3D Insufficient object requirements impede entities. progress, with the existing two-part Industry Foundation Classes (IFC) geometry proving inadequate. Consequently, there exists a palpable necessity for an IFC 4.3 infrastructure extension update to ease seamless data exchange, potentially spurring software companies to adapt to open infrastructure modeling standards.

Subsequently, discussions centered on the essential components of a learning environment. Experts encapsulated the concept of InfraBIM, emphasizing the use of 3D models enriched with metadata and information for each structural element. Standardized data and information structures were considered imperative for streamlining the modeling process. Central to the narrative was the critical role of standardization in enhancing information flow and collaborative functionality. Furthermore, deliberations touched upon data transfer formats, advocating for the predominance of open standards. However, recognition was given to the practicality of native formats for internal modeling purposes. The inclusion of IFC or Inframodel in online courses was deliberated, considering the varying subscriber preferences for open formats across different countries. Significantly, the potential future standardization of IFC's infra extension as an ISO standard was highlighted as a significant development, with the potential to catalyze international adoption and compel software platforms to incorporate IFC support.

3.4 AI as a tool to supply course material

Based on the response produced by ChatGPT, the general concept of BIM is recognized, as well as the benefit perspective and life cycle management, in the consolidation experiments of the entities, but the response to the definition of "BIM in infrastructure" did not take a position on the perspective of the infrastructure sector and ChatGPT did not know how to connect the InfraBIM concept to the response. When asked about InfraBIM, ChatGPT was able to distinguish the term InfraBIM from BIM, which is the application of BIM for infrastructure projects. The answer brought up the same benefits as in the case of the first compaction, and no other differences in infra modelling compared to information modelling of building construction have been brought up, other than the target's infra projects. Linguistic problems were noticeable in the form of repetition, the answer listed "collaboration, efficiency, accuracy, and sustainability" twice, and a water treatment plant is included in the infrastructure items. Regarding Open InfraBIM, ChatGPT produced the longest answer and found that the use of open standards and data transfer formats should be encouraged and listed mostly achievable benefits. There is no mention of the use of nomenclature or common modelling requirements in the answer.

For the course material testing, ChatGPT produced multiple-choice surveys when handed the topics. There were some errors in the questions and answers, but after a human inspection, they could be used if there is a need for efficient material production. What ChatGPT could be more usable for is in terms of translating the course materials for different languages, which it did successfully from English to German.

Text-to-speech video sound recording for the pilot course was made with the free version of Amazon Polly according to the text materials.

3.5 The results of the pilot course

For the pilot module, the most important topics had to be defined as the basics of infra modeling, for which expert interviews, and observations from existing BIM courses' were used. Based on the analyses of the other BIM online courses, the learning goals in this tested module were decided to have three goals (Figure 3).



Figure 3. The learning outcomes for the tested pilot course.

The material of the course was BIM vocabulary, statistics, text-to-speech video, summary after the video, Teams-interview, case study, and quiz. An example of the material presented in the pilot course is among other things the difference between information exchange using a Common Data Environment (CDE) and traditional information exchange is explained in the text, but also in the picture (Figure 4.).



Figure 4. Comparison of data exchange between traditional and Common Data Environment (CDE).

The table of the pilot course had decided to have the following 11 topics:

- 1. Why BIM in the infrastructure industry?
- 2. Vocabulary
- 3. What are the benefits of using InfraBIM?
- 4. 3D modelling vs. BIM
- 5. Roles and new responsibilities
- 6. Digital processes change the way we work
- 7. Why so many different models?
- 8. What are the requirements for InfraBIM?
- 9. Case study
- 10. Ouiz
- 11. Learning more and references.

Even though the course link had been opened more than 200 times, only nine participants gave feedback to the survey at the end of the course. Users were allowed to select their top 3 favorite topics from the list mentioned above. A total of 18 votes were cast. According to feedback the most useful topic in the course was the part where the benefits of using infraBIM (28% of the answers) were described. The roles and new responsibilities (22% of the answers) were the second useful topic. The other topics were evaluated between 5 to 11 %, as can be seen in Figure 5.



Figure 5. Topics in the course that were considered the most useful for learning based on the feedback.

In the feedback survey, we were also asked to evaluate what materials were useful for the learning aspect. Users were allowed to select their five most interesting materials. A total of 36 votes were given. The results can be seen in Figure 6.



Figure 6. What types of material on the course were considered most useful for learning based on the feedback of the final survey.

The materials of the pilot course ended up emphasizing a practical solution, whereby customers, designers, and contractors can all have their databases, and the most important thing is how to ensure a functioning open data transfer between the databases. The challenge of the pilot module was finding a suitable level of accuracy in summarizing the topics, with the goal being as concise a presentation as possible, for example, summarizing the role of machine control in model-based construction and whether breakdown into 2D and 3D machine control systems is carried out. In the pilot, no breakdown was made, but the processing of machine control was carried out on a more general level, and the section included a link to Novatron Oy's website for other information. It was decided to use three examples of open data transfer formats in the materials throughout the module, IFC, LandXML, and InfraModel, the first two because of their international recognition and the last one from the perspective of Finnish infrastructure modelling and its international growth potential.

During the preparation and translation of the teaching materials, it was discovered that the terms translated directly from Finnish to English are not necessarily as well understood or in the same way. Based on the correction suggestions received, the English version of the YIV instructions could not be used straightforwardly.

As a result of the pilot course and feedback, a general structure for an InfraBIM online course module was developed (Figure 7). The general structure aims to perfect the conveyance of the fundamentals of the ideology of open infrastructure modeling using different teaching resources in the most efficient order.

1. Introduction
2. Vocabulary
3. Text-to-speech-video, Summary of the video
4. Past vs. InfraBIM
5. Theory
6. Implementation of InfraBIM
7. Case Study, Interview video
8. Quiz
9. Learn more, References
10. Feedback survey

Figure 7. The general structure of the basic level InfraBIM online course module.

4 Conclusion

Based on observations, Open InfraBIM still needs more international visibility and recognition as a term and a process. Infrastructure modeling has been overshadowed within the existing BIM online course offerings and content, focusing primarily on building information modeling (BIM) in construction. This supports the need for an independent website dedicated to infrastructure modeling. In addition, it is justified to cover the fundamental principles of open infrastructure modeling in its education and marketing efforts because the expansion of its adoption is freshly beginning in several countries.

Considering the feedback received, the general structural model of the InfraBIM online course is a good starting point for the further development of an open and multilingual website. An open infrastructure modeling website would be beneficial both internationally and domestically because the accessibility of easily available information within infrastructure modeling is a common issue.

There is a need for educational materials to start from the very basics, distinguishing infrastructure modeling from mere 3D modeling. 3D modeling is already a familiar topic within the infrastructure sector and has been in use for quite some time. Therefore, the emphasis on materials should shift from generating information to consuming and managing information using an open model-based process and infrastructure models. Open infrastructure modeling is a vast domain, but it is possible to narrow down the model-based process by focusing on a particularly essential aspect: the renewed information consumption within infrastructure modeling.

Acknowledgments

The authors would like to acknowledge all the participants who tested the module. Authors are grateful to the experts from the construction industry, Destia Oy, Infrakit Oy, Pointscene Oy, Sandvik Oy, and Novatron Oy for supplying feedback on their experiences and for bringing their technology. This Open Ecosystems2 project was financed by Business Finland (2021-2023).

References

- Mwila, L., Rautiainen, J. & Lehtoviita, T., 2022. Benchmarking BIM maturity level in various European countries to develop BIM competence in Finland. LAB RDI Journal. Online: https://www.labopen.fi/en/lab-rdijournal/benchmarking-bim-maturity-level-invarious-european-countries-to-develop-bimcompetence-in-finland/, Accessed: 13.12.2023.
- [2] Bradley, A., Li, H., Lark, R., Dunn, S. BIM for infrastructure: An overall review and constructor perspective, *Automation in Construction*, 71(2): 139-152, 2016.
- [3] European Construction Sector Observatory Trend Paper - Building Information Modelling in the EU

construction sector – March 2019. Online: https://ec.europa.eu/docsroom/documents/34518. Accessed: 7.12.2023.

- [4] Luomala, P. The different phases of BIM adoption in Europe. Online: The different phases of BIM adoption in Europe. Available: https://www.magicad.com/en/blog/2020/03/bimadoption-europe/, Accessed: 8.12.2023.
- [5] Salzano, A., Intignano, M., Mottola, C., Biancardo, S.A., Nicolella, M., Dell'Acqua, G. Systematic Literature Review of Open Infrastructure BIM. Buildings 2023, 13, 1593. Online: https://doi.org/10.3390/buildings13071593, Accessed: 13.12.2023.
- [6] NATSPEC: BIM education Global 2023 update report. International Construction Information Society. 2023 Online: https://www.icis.org/wpcontent/uploads/2023/05/BIM-Education-Global-2023-Update-Report-V10.0.pdf. Accessed: 16.2.2024.
- [7] Heikkilä R., Kolli T., and Rauhala T. Benefits of Open InfraBIM – Finland Experience. 39th International Symposium on Automation and Robotics in Construction (ISARC 2022). Pages 253-260, Bogota, Columbia, 2022.
- [8] BuildingSMART Finland. Common InfraBIM Requirements 2021. Online: https://wiki.buildingsmart.fi/fi/04_Julkaisut_ja_Sta ndardit/YIV, Accessed: 23.11.2023.
- [9] Noeskoski J., 2021. Tietomallintamisella ja integroidulla rakentamisella kohti infra-alan yhteistoiminnallisuutta. Tampereen yliopiston väitöskirjat 446. Online: https://urn.fi/URN:ISBN:978-952-03-2043-0, Accessed: 13.12.2023.
- [10] Alsina-Saltarén S., Ponz-Tienda J.L., Gutiérrez-Bucheli L. & Sierra-Aparicio M. Implementation of BIM in infrastructure: the need to address it from the public sector. Building & Management, vol. 2(3), pp. 62-72, 2018. Online: http://dx.doi.org/10.20868/bma.2018.3.3843, Accessed: 13.12.2023.
- [11] Meister A., Scholz F. & Banemann S., 2022. Federal Trunk Roads BIM Masterplan. Federal Ministry for Digital Transport. Germany, Berlin. Online: https://bmdv.bund.de/SharedDocs/DE/Artikel/StB/ masterplan-bim-bundesfernstrassen.html, Accessed: 13.12.2023.
- [12] 1. Free Basic BIM Training: Introduction To BIM, National Institute of Building Technology (NIBT) E-Learning. Online: https://www.udemy.com/course/introduction-tobim/, Accessed: 20.4.2023.
- [13] 2. BIM: Building Information Modelling. Online:

https://www.reed.co.uk/courses/bim-buildinginformation-modelling/383084#/courses/bim, Accessed: 20.4.2023.

- [14] 3. BIM Building Information Modelling. Online: https://www.reed.co.uk/courses/bim-buildinginformation-modelling/428573#/courses/bim, Accessed: 20.4.2023.
- [15] 4. BIM Introduction, Construction BIM -Construction process. Online: https://www.udemy.com/course/bimintroconstruction-bim-construction-process-part-1of-6/, Accessed: 20.4.2023.
- [16] 5. BIM Basic course (e-learning). Online: https://www.hochtief-vicon.com/vicon_en/7.jhtml, Accessed: 20.4.2023.
- [17] 6. Udemy BIM essentials. Online: https://www.udemy.com/course/certificate-inbuilding-information-modelling-bim-level-2/, Accessed: 20.4.2023.
- [18] 7. Purdue University, BIM for construction. Online: https://www.classcentral.com/course/bim-forconstruction-52892, Accessed: 20.4.2023.
- [19] 8. BIM Training & Management Learn the BIM Process today! Online: https://www.udemy.com/course/bim-managementprocess-planning-implementation-part-1/, Accessed: 20.4.2023.
- [20] 9.TÜVSÜD, BIM Basics e-learning course. Online: https://www.tuvsud.com/en-ae/services/training/elearning-courses/building-information-modellingbim-basics, Accessed: 20.4.2023.
- [21] 10. Coventry University, An Introduction to building information modelling. Online: https://www.classcentral.com/course/buildinginformation-modelling-10604, Accessed: 21.4.2023.
- [22] 11. Si!BIM. Online: https://www.sibim.eu/course-material, Accessed: 21.4.2023.
- [23] 12. BIM The Methodology to coordinate construction. Online: https://www.udemy.com/course/bim-themethodology-to-coordinate-construction/, Accessed: 21.4.2023.
- [24] 13. Udemy BIM training. Online: https://www.udemy.com/course/bim-training/, Accessed: 21.4.2023.
- [25] 14. Jargon busting BIM theory course, Udemy BIM made Easy. Online: https://www.udemy.com/course/bim-made-easy/, Accessed: 21.4.2023.
- [26] 15. BIM Application for Engineers, National Taiwan University. Online: https://www.classcentral.com/course/bimapplication-12117, Accessed: 21.4.2023.

- [27] 16. Basics of building information modelling. Online: https://alison.com/course/basics-of-buildinginformation-modelling-bim, Accessed: 22.4.2023.
- [28] 17. BIM The methodology complete course. Online: https://www.udemy.com/course/the-completecourse-of-the-bim-methodology/, Accessed: 22.4.2023.
- [29] 18. LinkedIn Learning, BIM Foundations. Online: https://www.linkedin.com/learning/bimfoundations-2, Accessed: 22.4.2023.
- [30] 19. Building Information Modelling (BIM) for Manager. Online: https://www.reed.co.uk/courses/buildinginformation-modelling-bim-formanager/360383#/courses/bim, Accessed: 22.4.2023.
- [31] Rahman, M., and Watanobe Y. ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences* 13(9):5783, 2023.
- [32] Kemppainen A. InfraBIM e-learning website. Online: https://www.ouluzoneplus.com/infrabim, Accessed: 9.4.2024.

Integration of BIM and RFID-Sensing for Automated Prefabrication and Progress Monitoring in Modular Construction

Maggie Y. GAO¹, Chengjia HAN², Yaowen YANG^{3*}, Robert L.K. TIONG^{4*}

¹PhD Researcher, School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798. E-mail: <u>yuan008@e.ntu.edu.sg</u>

²Research Fellow, School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798; Email: chengjia.han@ntu.edu.sg

³Professor, Associate Chair (Faculty), School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798. E-mail: cywyang@ntu.edu.sg

⁴Associate Professor, School of Civil and Environmental Engineering, Nanyang Technological University,

Singapore 639798. E-mail: clktiong@ntu.edu.sg

*Corresponding Authors.

Abstract

Modular construction represents a significant advancement in the architectural, engineering, and construction (AEC) domain, bolstering productivity by integrating factory-based manufacturing and efficient assembly techniques. This study presents a holistic approach, merging Building Information Modelling (BIM) with technological automation interoperability, to facilitate seamless prefabrication and automatic monitoring within modular construction projects. It underscores the importance comprehensive management across the of construction lifecycle, from prefabrication to final assembly. The research advocates for an integrated framework encompassing information identification, data mapping between Industry Foundation Classes (IFC) and ProgressXML (PXML), and semantic enrichment for Radio Frequency Identification (RFID) tracking. This approach is aimed at streamlining data exchange, automating fabrication processes, enhancing quality control, and optimizing precast supply chain management, thereby improving overall information management among project stakeholders.

Keywords -

Building Information Modeling, Progress Monitoring, Modular Construction, Industry Foundation Classes (IFC), ProgressXML (PXML), Radio Frequency Identification (RFID)

1 Introduction

Modular construction integrates mass production and automated techniques to refine construction workflows, enhance workforce productivity, and promote environmental sustainability [7]. Building Information Modeling (BIM) functions as a digital tool that encapsulates both the physical and functional attributes of a building. It provides a comprehensive and evolving model that adapts continuously over the construction lifecycle. It has attracted considerable attention for its application in multiple fields of modular construction, including but not limited to design automation [3], automated manufacturing processes [2], the use of robotics in assembly [21], optimization of supply chain logistics [6], and the facilitation of a circular economy [1]. BIM functions as an object-oriented model that stores critical data for managing a building's lifecycle, covering aspects such as scheduling, analysis, and assessment.

In addressing the increasing demand for digital transformation within the modular construction sector, this study introduces an approach to bridge the digitalphysical divide through the mapping of Industry Foundation Classes (IFC) and ProgressXML (PXML). IFC serves as the cornerstone for BIM interoperability, enabling seamless data exchange and management across various platforms. Meanwhile, PXML is tailored to represent the geometric and semantic details of prefabricated components, enhancing the flow of information for fabrication and logistics. This integration aims to automate and digitize the construction workflow, improving efficiency, quality control, and supply chain management. By ensuring data integrity and fidelity in the mapping process, this framework advances the digital management of information in modular construction, highlighting the potential of BIM and digital fabrication technologies to revolutionize construction processes, achieving higher efficiency and sustainability.

Managing the construction progress of precast components with BIM and sensing technologies streamlines the process from design to end-of-life.
Unique identifications in BIM model correspond to Radio Frequency Identification (RFID) tags on components, allowing real-time tracking through fabrication, storage, transportation and assembly [5]. Post-installation, real-time data in BIM supports maintenance scheduling and asset management. At the component's end-of-life, it assists in sustainable decommissioning practices. This integration enhances efficiency, reduces errors, and improves lifecycle traceability.

The conventional approach to construction progress monitoring, traditionally focusing on individual processes (including fabrication, temporary storage, fitting out, delivery, and installation), fails to align with the evolving global trends in modular construction. There is a growing need for an approach that integrates BIM to oversee the entire construction process, from factorybased production to site construction. Therefore, this paper introduces an innovative BIM methodology, enhanced by RFID-sensing technology, to facilitate automated construction progress monitoring, spanning from factory prefabrication to on-site assembly. We provide an illustrative example to demonstrate the effectiveness of these methods.

To assess the effectiveness of the suggested framework, field trials were conducted on a Prefabricated Prefinished Volumetric Construction (PPVC) residential project using RFID. Results confirmed the accurate onsite placement of precast units and the reliability of sensor-derived data.

2 Literature Review

2.1 Modular Construction

Modular construction, also known as offsite construction, prefabrication, or industrialized building system in literature, is commonly defined as a process in which building components or modules are produced in a controlled environment [9]. Different apart from traditional cast-in-place construction methods, modular construction is characterized by of standardization, industrialized production processes, and assembly-based construction [18]. Standardization in modular construction streamlines automation and improves construction progress. A notable obstacle is encountered insufficient when there is coordination and communication between the contractors and the precast suppliers [16]. Moreover, the complexity of managing congested construction sites with limited storage capacity underscores the importance of having access to itemlevel data for effective logistical coordination [20].



Figure 1. Integrated BIM and Sensing for Modular

Construction

Figure 1 provides an overview of the integration of BIM and sensing application in modular construction, with illustrative contents to demonstrate the studies in different construction stages.

2.2 PXML For Automated Prefabrication

In the rapidly evolving field of modular construction, the integration of digital technologies has become paramount for enhancing efficiency and accuracy. Industry Foundation Classes (IFC), a universally recognized data schema integral to BIM, plays a crucial role in exchanging and managing construction and facility management data [17].

This study conducted interviews with three prefabricators from the Singapore Integrated Construction and Prefabrication Hub (ICPH). It was found that the PXML format, utilized for fabrication and enterprise resource planning, along with steel fabrication machine codes (BVBS) for reinforcement fabrication, are the predominant data formats employed in the contemporary precast fabrication industry. Liu, et al. [12] integrated BIM and steel fabrication machine codes (BVBS) using openBIM standards to streamline steel reinforcement design and prefabrication automation.

In IFC framework, the relationship and structural details of components are essential [14]. By converting this information into a unified data expression mode based on the IFC standard, the model's relevance and consistency are ensured [10]. This work focuses on enhancing the IFC framework specifically for precast entities. It involves refining attribute definitions and represent association links to accurately the characteristics and relationships inherent in prefabricated structures. This enhancement is key to achieving a detailed and precise description of the prefabricated model within the IFC standard system.

On the other hand, PXML offers a specialized format for capturing the geometrics and semantics of prefabricated components [8]. The strategic alignment of IFC and PXML data structures paves the way for a more streamlined and cohesive workflow in precast fabrication. This alignment involves meticulous mapping of geometrical data, ensuring the physical characteristics of components are accurately represented, and semantic data integration, which encapsulates the functional attributes of these components. This study explores this data mapping of IFC and PXML in precast fabrication discipline, driving forward the construction industry towards a more digitized and efficient future.

2.3 Data acquisition for automatic construction progress monitoring

Problems with inadequately monitoring, finding, and recognizing highly tailored prefabricated components in supply chains for the construction industry include erroneous installations, double handling, and misplacement of parts, which cause delays in schedules and higher labor expenses [4]. Based on the previous research works, the data collection method for construction progress monitoring is categorized into enhanced information technology (IT) communication tools (emails, messages, etc.), Geo-spatial, 3D sensing and imaging technologies [13]. Since enhanced IT is not conducive to automated construction progress monitoring, this study summarized prior research and conducted the comparison of data acquisition methods as follows:

Technologies	Features	Advantages	Disadvantages
Geospatial technologies Barcodes and QR codes, RFID, UWB, GNSS	Automated and efficient monitoring of construction sites; Efficient tracking of materials, inventory, construction progress and labour tracking;	 High capacity of data content and type; High reliability; Ease of production; 	Initial investment; Manpower in installation, scanning, and maintenance; Cannot efficiently indicate the progress of partially completed or operation-level tasks;
3D Sensing technologies LIDAR Scanning	Can avoid potential errors and can be adopted in structural health monitoring;	Accurate method to capture real condition in construction sites;	 Slow warm-up time; A clear line of sight, noise from moving moving machinery and personal; Expensive device; Lack of portability;
Imaging technologies Digital images, videos, depth images	Examples for damage detection and safety evaluation; Automatically calculate the percentage completion and measure the progress;	Cost-effective; Moderate accuracy and quality;	Considerable time of computation; Sensitivity due to different lighting conditions; Line of sight;

Figure 2. Comparison of Data Acquisition Methods

Geo-spatial sensing is chosen for its robust data capacity, reliability, and potential for seamless integration with precast supply chain management systems. Moreover, the collecting of real-time scheduling data can be enabled by BIM, and the monitoring of precast components during the whole processes is made possible by RFID and GNSS [11]. Recent research studies use IFC, a standard data schema for BIM model, to express monitoring data [19]. The differences between RFID and QR / Bar code [15] are summarized in Table 1.

Table 1 Comparison of RFID and QR/Bar Code

RFID	QR / Bar Code	Maintenance	schedules.
Around several meters	Close proximity	Information	
Writable, allowing data to	Limited amount of data	Health and	Safety procedures for handling and

be updated	
More durable and can be	Less durable in harsh environments
used in harsh environments	and require a clear, unobstructed
like construction sites.	view to be scanned.

These technologies mainly assist in identifying materials. Barcodes function through the line-of-sight scanning of specific patterns on materials. In contrast, passive RFID operates by using electromagnetic fields to detect and monitor tags affixed to items, eliminating the need for direct line-of-sight. Passive RFID system technology has been selected to as an essential connector, bridging BIM elements with the real-world components or modules of precast structures, owing to its costefficient nature, characterized by affordability, durability, and reusability/recyclability.

3 Methodology

This study initially identifies the critical information required for BIM-enabled progress monitoring in modular construction. Following this identification, the research proceeds with data mapping between IFC and PXML for precast prefabrication, and then integrates RFID technology for the tracking of precast components.

3.1 Information requirement

A BIM model for modular construction management should encompass various stages of the building process, from initial design to final decommissioning. Key information is categorized into different categories, including design information, manufacturing specifications, construction updates, maintenance records. The details are summarized in Table 2.

Table 2 Identification of information requirements

Category	Description
Design	Detailed architectural, structural, and MEP
Information	designs for each modular component.
Manufacturing	Information about materials, fabrication
Specifications	processes, and quality control for module
	prefabrication.
Transportation	Details on module transportation from
and Logistics	manufacturing site to construction site,
Data	including size, weight, handling.
Assembly and	Instructions for on-site module assembly and
Installation	installation, including assembly and connection
Guidelines	details.
Integration	Information on interaction between modular
with	components and traditionally cast in-situ
Traditional	elements.
Construction	
Methods	
Sustainability	Data on environmental impact, energy
Features	efficiency, and waste reduction strategies.
Operational	Details on building operation and maintenance,
and	including maintenance access and replacement
Maintenance	schedules.
Information	
Health and	Safety procedures for handling and assembly

Safety	modular components.		
Protocols			
Cost and Time	Detailed budgeting and scheduling information		
Estimates	for the construction process.		
Future	Strategies for eventual module disassembly or		
Disassembly	recycling.		
and Reuse			
Plans			
Regulatory	Ensuring designs and processes comply with		
Compliance	relevant building codes and standards.		
Stakeholder	Contact information and roles of all parties		
Information	involved in the project.		
Change	Records of any changes made during the		
Management	project execution.		
Data			
As-Built	Accurate as-built drawings and models		
Documentation	reflecting the final state of the construction.		

3.2 Schematic Enrichment of BIM Data

(1) Mapping with PXML

According to IFC4.3.1.x (IFC4X3_ADD1), the majority of PXML-relevant data can be gathered from IFC files. Therefore, a mapping chart (Table 3) have been created to show the relation more clearly and accurately between the two data sets.

Table 3 Data Mapping for IFC and PXML

PXML Eleme nt	PXML Attributes (Version 1.3)	IFC Entity (IFC4X3_ADD1)	Missing IFC Entity
Global	Unique identifier for each element	IfcGloballyUniqueld	
DocInf o	GlobalID, Document Version, Comment, ConvertConventions, Mode	IfcDocumentInforma tion	
Order	Order Information, Import Source Information, ApplicationName, ApplicationGUID, ApplicationVersion	IfcWorkOrder, IfcApplication	Missing specific IFC entity for Import Source Information
Produ ct (Elem ent)	ElementNo, ProductType, PieceCount, Comment, RotationPosition, Stacking Information, Project Coordinates, Supplementary Product Information	lfcElement, lfcBuildingElement	Missing specific entities for Double Walls and Project Coordinates
Eleme ntInfo	Fields of ElementInfo entries, Predefined ElementInfo types, ElemInfoVal	IfcElement, IfcPropertySet, IfcElementCompone nt, IfcBuiltElement	No direct IFC entity for Predefined ElementInfo types; custom property sets may be required
Slab (Elem ent Part)	PartType, Geometric Slab Placement, Slab Production Directives, Geometric Placement and Production Directives for Double Walls, Multi-Layer Elements, Legacy Slab Fields, Simplified geometry representation	IfcSlab, IfcSlabType, IfcProductDefinition Shape, IfcShapeRepresenta tion	Missing entities for Slab Production Directives and Multi-Layer Elements
Outlin e	Geometric Outline Placement, Height, Name, GenericInfo, MountingInstruction, MountPartArticle, MountPartArticle, MountPartProperties, Concrete Properties, Layer, ObjectID, Shape SVertex	IfcObjectPlacement IfcLocalPlacement IfcMaterial, IfcBuildingElementP roxy	Missing entity for Simplified geometry representation
Steel	Geometric Steel Placement, ToTurn, StopOnTurningSide, Name, MeshType, WeldingDensity, BorderStrength, Generic Steel Info, Steel Production Directives, Layer, ObjectID	IfcReinforcingBar, IfcReinforcingBarTy pe, IfcReinforcingMesh, IfcShapeRepresenta tion	No direct IFC entity for MeshType and WeldingDensity; possibly handled by IfcMaterialPropertie S
Bar	ShapeMode, ReinforcementType, SteelQuality, PieceCount, Diameter, XYZ, RotZ, ArticleNo, NoAutoProd, ExtIronWeight, Bin, Pos, Note, Machine, BendingDevice, Spacer, WeidlingPoint, Segment, Canonical Bar representation	IfcReinforcingBar, IfcReinforcingBarTy pe IfcElementQuantity, IfcShapeRepresenta tion	Missing detailed specification for ShapeMode and SteelQuality in IFC

Girder	PieceCount, XYZ, GirderName, Length, AngleToX, NoAutoProd, Height, TopExcess, BottomExcess, Weight, TopFlangeDiameter, BottomFlangeDiameter, GirderType, MountingType, ArticleNo, Machine, Period, Width, AnchorBar, GirderExt, Section	IfcGrid, IfcGridPlacement	No equivalent for GirderExt; custom extensions may be needed
Custo m PXML Elomo	Custom attributes		No direct IFC entity; custom mapping or extension required
nt			extension required

The openBIM standards together with some common BIM specifications are introduced in this research, and the parameters in the PXML specification are listed and parsed using the latest version of the IFC4.3.1.x schema in terms of the semantic content of the precast components. The data mapping between PXML standard and IFC schema is made to address missing information in IFC based on existing entities of building elements and incomplete conversion in order to make it easier to input information prefabrication machines and tracking system. The efficiency of fabricating will be greatly improved with the use of data mapping, which will also hasten the process' automation.

(2) Mapping with RFID



Figure 3. Association between physical object, IFC Unique ID and RFID EPC

After fabrication, each component is assigned a unique identifier which will be linked to its RFID tag. An algorithm has been developed to link the IFC Unique ID of building model objects with a corresponding Electronic Product Code (EPC) used in RFID systems. This integration allows for the automatic conversion between these two identifiers. When a task involves an object with an RFID tag, the system detects the EPC and translates it into the IFC Unique ID. This process enables precise identification and management of physical objects in the whole life cycle, ensuring real-time synchronization with their digital counterparts in the building information model, enhancing efficiency and accuracy in building management.

The RFID-enabled digital inventory and logistics management system, detailed in Table 3, is designed for efficient tracking and management of construction materials. It utilizes an RFID handheld scanner, adept at tracking materials from prefabrication to on-site delivery and installation, adapting to various logistic scenarios like temporary storage or return for repairs. Crucial for the dynamic construction environment, the scanner's mobility and versatility align well with site unpredictability.

Table 3 RFID-enabled Tracking Operations

Phase	Operations
I. Production	Attach RFID tag after precast component is fabricated, associate RFID with BIM elements and select produced
II. Transportation (to warehouse or construction site)	Scan RFID tag at factory gate, transport precast component to construction site Scan RFID tag at factory gate, transport precast component to warehouse
III. Arrival (at warehouse or	Scan RFID tag on site, stockpiled on site storage area Scan RFID tag on warehouse, processed in fitting out yard
construction site)	Scan RFID tag on construction site/warehouse, return immediately to prefabricator due to quality issues
IV. Assembly	Deliver precast component to the assembly location on site, and scan RFID tag

4 Case study

This study utilized a BIM Platform, serving as a centralized communication center for project communities, enabling collaboration and data exchange across construction processes and supply chains. It acts as the data backbone for the system, allowing access to object-based data and API, as demonstrated in Figure 4. In the context of BIM-enabled prefabrication and progress monitoring, an integrated BIM and RFID system is developed to track, monitor, and manage the flow of building materials. This system updates project stakeholders on material status, ensuring quick, simple, and trustworthy identification, recording, tracking, and sharing of material logistic information.



Figure 4 Smart Tracking System Framework

Project stakeholders in the construction supply chain can provide and obtain real-time information and project updates, enhancing budget and schedule management. The system reduces project failure costs and time through real-time monitoring, effective project management, and early schedule violation notifications. It coordinates planning errors, production, logistics, and assembly operations, preventing misunderstandings and avoiding rework and site readjustment.

Figure 5 illustrates the practical on-site RFID tagging and the association of RFID tags to precast elements. Passive RFID technology is chosen for cost-effective digital inventory and logistics management, supported by the BIM Platform. The system electronically manages construction inventory, reducing delays, increasing productivity, and preventing errors in deliveries.



Figure 5 Actual on-site implementation

The progress monitoring process involves attaching RFID tags to precast components after fabrication, associating them with BIM elements, and marking them as produced. Then the precast component will be shipped to warehouse or construction site, RFID tags are scanned at the factory gate, and components are transported to the construction site or warehouse. Upon arrival, RFID tags are scanned on-site or in the warehouse, with components either stockpiled on-site or processed in the fitting-out yard. Quality issues may prompt an immediate return to the prefabricator. During assembly, precast components are delivered to the assembly location on-site, and RFID tags are scanned. The locations and processes are illustrated in Figure 6.



Figure 6 RFID tracking process and locations

This comprehensive process integrates RFID technology with BIM, allowing for real-time tracking and monitoring of precast components throughout their lifecycle. It enhances efficiency in logistics, quality control, and assembly operations, contributing to streamlined construction processes and improved project management. The seamless combination of RFID and BIM ensures accurate data exchange and facilitates effective communication across the construction supply chain, ultimately optimizing the overall construction workflow.

The dataset provides detailed information on construction assemblies, including identification, naming, location, status, production and installation dates, classification, and type, facilitating tracking and management within the construction project. The PPVC components are under IFCELEMENTASSEMBLY, indicating as integrated model for construction. The analysis of the precast modular production and installation dataset has revealed some critical insights and potential bottlenecks:

- The mean installation delay, calculated as the difference between the actual installation date and the planned production date, is approximately 280 days. This indicates a significant gap between when modules are planned to be produced and when they are actually installed.
- The standard deviation for installation delays is about 80 days, suggesting variability in how different modules experience delays.

To mitigate bottlenecks in construction projects, it is recommended to enhance project management by fostering improved coordination among production, logistics, and installation teams, thereby minimizing delays. Further, the supply chain should be analyzed and optimized to ensure the efficient delivery of modules from production facilities to installation sites.

Based on the analysis of the precast modular production and installation dataset, potential areas for future research include developing dynamic planning strategies to adapt to changing conditions, conducting uncertainty analysis to assess variability in processes, optimizing production workflows to minimize delays, enhancing supply chain integration, implementing realtime monitoring for proactive intervention, and devising risk management strategies to mitigate disruptions.

5. Conclusions and future work

This study demonstrates the interoperability of BIM with automation technologies, enabling prefabrication, and progress monitoring in modular construction. The incorporation of BIM enables the automated exchange of geometric and product data from the model, streamlining the prefabrication process. Additionally, BIM integration with an RFID-enabled digital inventory and logistics management system is implemented to oversee and trace the precast component's journey until its assembly on-site.

One limitation of this study on data mapping between PXML and IFC for precast fabrication is the potential complexity and variability of data structures between the two formats, which may require extensive effort and expertise to ensure accurate and comprehensive mapping. Additionally, discrepancies in data granularity and semantics could pose challenges in achieving seamless interoperability, potentially leading to inconsistencies or loss of information during the mapping process. Because different geometrical formulations are used for various forms of precast components, there are still some issues with the conversion of the geometry information in PXML. These issues might be addressed and resolved in the future along with further information about the IFC extension. The study plan also calls for testing the PXML code produced by the created Dynamo program in a nearby precast factory to ensure that it is valid.

To assess the effectiveness of the suggested framework, field trials were conducted on a PPVC residential project using RFID. Results confirmed the accurate on-site placement of precast units and the reliability of sensor-derived data. The research adopted an integrated methodology, enabling the real-time transmission of data to engineers for decision-making, thus optimizing workforce and resource distribution in modular construction. It is observed that integrating RFID data seamlessly into BIM platforms may require significant technical expertise and could pose challenges in terms of data synchronization and interoperability, especially when catering BIM design changes.

ACKNOWLEDGEMENT

This research is supported by the Ministry of National Development (MND), Singapore under the MND Research and Innovation Fund. Any opinions, findings and conclusions or recommendations expressed in the material are those of the author(s) and do not reflect the views of Ministry of National Development Singapore.

References

- L.A. Akanbi, L.O. Oyedele, K. Omoteso, M. Bilal, O.O. Akinade, A.O. Ajayi, J.M. Davila Delgado, H.A. Owolabi, Disassembly and deconstruction analytics system (D-DAS) for construction in a circular economy, Journal of Cleaner Production 223 (2019) 386-396.
- [2] W. Anane, I. Iordanova, C. Ouellet-Plamondon, BIM-driven computational design for robotic manufacturing in off-site construction: an integrated Design-to-Manufacturing (DtM) approach, Automation in Construction 150 (2023).
- [3] H. Dibowski, J. Ploennigs, K. Kabitzsch, Automated design of building automation systems, IEEE Transactions on Industrial Electronics 57 (11) (2010) 3606-3613.
- [4] A. Ergen, B., & Sacks, R., Tracking and locating components in a precast storage yard utilizing radio frequency identification technology and GPS, Automation in Construction, 16(3), 354–367. (2007).
- [5] Y. Fang, K. Cho Yong, S. Zhang, E. Perez, Case Study of BIM and Cloud–Enabled Real-Time RFID Indoor Localization for Construction Management Applications, Journal of Construction Engineering and Management 142 (7) (2016) 05016003.
- [6] S. Fosso Wamba, Achieving supply chain integration using RFID technology, Business Process Management Journal 18 (1) (2012) 58-81.
- [7] R. He, M. Li, V.J.L. Gan, J. Ma, BIM-enabled computerized design and digital fabrication of industrialized buildings: A case study, Journal of Cleaner Production 278 (2021).
- [8] R. He, M. Li, V.J.L. Gan, J. Ma, BIM-enabled computerized design and digital fabrication of industrialized buildings: A case study, Journal of Cleaner Production 278 (2021) 123505.
- [9] E. Iacovidou, P. Purnell, K.D. Tsavdaridis, K. Poologanathan, Digitally enabled modular construction for promoting modular

components reuse: A UK view, Journal of Building Engineering 42 (2021) 102820.

- [10] Y.S. Jeong, C.M. Eastman, R. Sacks, I. Kaner, Benchmark tests for BIM data exchanges of precast concrete, Automation in Construction 18 (4) (2009) 469-484.
- [11] L. Liu, X., Chen, J., & Jin, R., Real-Time Optimization of Precast Concrete Component Transportation and Storage., Advances in Civil Engineering, 2020, 1–18. (2020).
- [12] Y. Liu, M. Li, B.C.L. Wong, C.M. Chan, J.C.P. Cheng, V.J.L. Gan, BIM-BVBS integration with openBIM standards for automatic prefabrication of steel reinforcement, Automation in Construction 125 (2021) 103654.
- [13] T. Omar, M.L. Nehdi, Data acquisition technologies for construction progress tracking, Automation in Construction 70 (2016) 143-155.
- [14] M. Venugopal, J. Teizer, C.M. Eastman, Formal Specification of the IFC Concept Structure for Precast Model Exchanges, 2012.
- [15] D.Y. Wang, Fashion supply chain traceability: RFID vs barcode, Routledge, United States, 2019, pp. 54-64.
- [16] Z. Wang, H. Hu, W. Zhou, RFID Enabled Knowledge-Based Precast Construction Supply Chain, Computer-Aided Civil and Infrastructure Engineering 32 (6) (2017) 499-514.
- [17] Z. Xu, J. Abualdenien, H. Liu, R. Kang, An IDM-Based Approach for Information Requirement in Prefabricated Construction, Advances in Civil Engineering 2020 (2020).
- [18] Z. Yuan, C. Sun, Y. Wang, Design for Manufacture and Assembly-oriented parametric design of prefabricated buildings, Automation in Construction 88 (2018) 13-22.
- [19] L. Zhao, Z. Liu, J. Mbachu, Development of Intelligent Prefabs Using IoT Technology to Improve the Performance of Prefabricated Construction Projects, Sensors 19 (19) (2019) 4131.
- [20] Z. Zheng, Z. Zhang, W. Pan, Virtual prototyping- and transfer learning-enabled module detection for modular integrated construction, Automation in Construction 120 (2020) 103387.
- [21] A. Zhu, T. Dai, G. Xu, P. Pauwels, B. De Vries, M. Fang, Deep Reinforcement Learning for Real-Time Assembly Planning in Robot-Based Prefabricated Construction, IEEE Transactions on Automation Science and Engineering 20 (3) (2023) 1515-1526.

Scan-to-BIM: Unlocking current limitations through Artificial Intelligence

Queruel Maxime^{1,2}, Bornhofen Stefan², Histace Aymeric², Ducoulombier Laure¹

¹Bouygues Construction - Research and Development Department ²ETIS, UMR 8051, CY Cergy Paris Université, ENSEA, CNRS, F-95000

m.queruel@bouygues-construction.com

Abstract -

This paper discusses the current methods of Scan-to-BIM, a process which allows creating a digital representation of existing buildings for a planning methodology called Building Information Modeling (BIM). The study covers all stages of the process, from point cloud generation and pre-processing to BIM modeling and formatting. We review the work already done in this area both conventionally and with the addition of Artificial Intelligence approaches which have significantly improved the efficiency and accuracy of the process. With a particular focus on Artificial Intelligence, we explore how these advanced technologies transform and optimize every step, offering innovative insights and significant improvements over conventional methods. Through this investigation, we aim to provide insights into the capabilities and constraints of the Scan-to-BIM workflow, and to shed light on academic advancements and industrial perspectives.

Keywords -

AEC, BIM, 3D Scan, Scan-to-BIM, 3D Point Cloud, Artificial Intelligence

1 Introduction

In the Architecture, Engineering, and Construction industries (AEC), Building Information Modeling (BIM) relates to the creation of a digital representation of physical and functional characteristics of a building. BIM accelerates the digital transformation as a knowledgesharing, collaborative platform among all stakeholders throughout the entire building life cycle. Over the last two decades, the adoption of BIM for building projects has been continuously growing thanks to many advantages and opportunities using the approach, such as automatic quantity estimation, swift responses to design changes, improved construction schedule visualization, and enhanced design coordination [1]. However, the implementation of BIM for existing buildings presents significant challenges. The primary issues encompass (1) technological, (2) financial, (3) managerial, (4) social, and (5) legal aspects [2].

One promising way to address the challenges associated with BIM for existing structures is the concept of Scan-to-BIM. Scan-to-BIM streamlines the process of gathering real-world data and converting it into a BIM-ready format. By capitalizing on advanced technologies such as laser scanning and photogrammetry, it eliminates the need for time-consuming manual data collection, ensures up-to-date information and reduces the associated costs.

Although already successfully applied in industrial use cases, Scan-to-BIM remains at the heart of current research due to many challenges such as manual intervention, lack of interoperability, algorithmic demands, technological limitations and the significant cost of these solutions [3][4][5]. Research has been intensifying with the recent integration of Artificial Intelligence (AI) approaches, thanks to their ability to speed up the process and even to remove some of the barriers by simplifying the processing steps [6][7][8]. However, there are still a considerable number of obstacles to be overcome before the process can be fully automated at both academic and industrial levels.

In this context, our paper presents a short and comprehensive state of the art on the Scan-to-BIM workflow, by inspecting its different stages and providing insight into their capabilities and current challenges integrating AI technologies. We highlight the advances made in the academic world and the state of industry through the scope of the general contractor Bouygues Construction.

2 State-of-Art

To conduct this state-of-the-art review, we analyzed 106 scientific papers that contribute to the body of knowledge surrounding Scan-to-BIM through all the stages of the process. The review was guided by the following top five keywords: "Point Cloud", "BIM", "Artificial Intelligence", "Laser Scanning", and "3D Reconstruction". These keywords were instrumental in filtering the vast



Figure 1. Scan-to-BIM differents steps

array of literature to focus on the most relevant and impactful studies in the field. A first observation is the gap between purely academic papers and those involving an industrialist in the field. In fact, in our review only 21 papers indicated that they had been written in partnership with industry or had received industrial funding. Most of these were software companies and not necessarily specialized in the construction field.

The Scan-To-BIM process represents the set of operations required to obtain a BIM model, and is now well defined in the literature. The process can be segmented into (1) generating the point Cloud from the existing structure, (2) pre-processing the data, and (3) creating the model [3][9]. Furthermore, these stages can themselves be further divided, as shown in figure 1, including steps like data acquisition and registration for the first, various filtering algorithms for the second, and geometry modeling, or link assignment for the last stage [10]. However, not all of these steps are equivalent in terms of complexity and know-how, and some of them are currently the subject of in-depth research.

2.1 Point Cloud generation

The first essential step in creating a BIM model consists in collecting accurate data from the existing structure. Several methods are commonly employed for this purpose, but the two main ones are photogrammetry and lasergrammetry, which result in a set of point clouds representing the geometry of the capture [9]. Photogrammetry is often favored for its speed and cost-effectiveness, while laser scanning, particularly terrestrial laser scanning (TLS), excels in delivering high accuracy [3][11]. It is however important to note that these observations are subject to change due to the rapid evolution of equipment in this field.

Laser scanning encompasses various categories, including terrestrial laser scanning (TLS), mobile laser scanning (MLS), and airborne laser scanning (ALS), each of them offering specific advantages. For instance, TLS is ideal for capturing large areas, whereas ALS is preferable for large-scale data acquisition. A number of scanning solutions for each category are already available on the market from companies such as Trimble, Faro and Leica. The last brand offers all three types of scanner, namely BLK360 (TLS), BLK2GO (MLS), and BLK2FLY (ALS) [12]. The choice may depend on the typology of the building and the given accuracy requirements [13].

The subsequent step is registration. This operation involves the alignment of point cloud scans taken at different stations, and their assembly into a unique point cloud. Some studies use the AI possibility to enhance this part and make it more efficient and quicker by reducing the time and resources used by identifying and eliminating noisy points [14] or improving position matching [15]. Beyond the academic research, the registration is often integrated into the software solution by the editing companies because it's important for industry to have a complete sequence from the point cloud acquisition to the point cloud deliverable.

The file format used for point cloud is typically binary such as PCD (Point Cloud Library) or LAS. Alternatively it can be an ASCII format like XYZ, or a hybrid format containing both, such as E57 [3]. All formats allow for the storage of point information, images, as well as metadata like timestamps.

In industry, this stage is often mastered and readily applied. Companies do not hesitate to acquire point clouds by surveyors, specialized external service providers or on their own, to produce valuable data which can be used either for a full Scan-to-BIM or for simpler tasks such as comparison or visualization.

2.2 Point Cloud Pre-Processing

Once the point cloud is acquired, the raw data should be refined, compressed and simplified in order to reduce the following processing costs as illustrate in the **figure** 2. To do so, different algorithms have already been studied and suggested in the literature.



Figure 2. Pre-processing Filtering: (a) raw data, (b) outlier removal, (c) downsampling with different voxel sizes [16]

Outliers are data points that considerably deviate from the expected model, typically signifying errors or extreme Many algorithms and techniques have been values. developed to solve these issues. Traditional approaches dealing with outliers in point clouds include radius-based statistical suppression and mean-shift clustering [17] as well as the use of Z-scores with adaptive thresholds AI has also become increasingly important in [18]. outlier detection, thanks to statistical approaches based on density, clustering, learning and ensemble methods [19]. PointCleanNet uses deep neural networks to identify outliers and noise, and corrects them to preserve essential surface details, demonstrating remarkable efficiency even in dense and varied point clouds [20].

Noise can be described as erratic fluctuations in the data, often originating from measuring inaccuracies or other stochastic factors. To mitigate noise, Edge-Aware filtering or Bilateral filtering can be applied, as well as more recent approaches such as a high-performance algorithm preserving sharp features [21]. The use of AI, notably through approaches such as PointCleanNet, has marked a significant advance in denoising. Moreover, recent studies, including one introducing graph-convolutional representations [22] and local surface estimation via a deep neural network [23], show how AI can improve the accuracy and efficiency of denoising, effectively tackling even complex structured noise.

Downsampling is crucial for reducing data volume and facilitating further processing. Different approaches, such as voxelization, subdivide three-dimensional space into cubes (voxels), enabling a discrete representation of 3D points as volumetric data [24][25]. Although these methods have been widely used to simplify point clouds, they can lead to a loss of precision and data, limiting their effectiveness in demanding applications [26]. The introduction of AI into the field of downsampling has marked a turning point, with the development of sophisticated techniques capable of dynamically adapting the data reduction process while preserving essential features for downstream tasks [8]. Recent studies illustrate important progress in this field by evaluating different downsampling strategies such as 3D Edge-Preserving Sampling (3DEPS) [27], and by introducing frameworks capable of handling arbitrary cloud point sizes [28].

2.3 Point Cloud Processing

Segmentation plays a crucial role for Scan-to-BIM in order to classify and distinguish building elements such as walls, openings and floors. This process is crucial for the transformation of the scanned data to a high-level representation of the environment. In the previous steps, the algorithms did not differentiate between interior and exterior environments because they are the same type of data, but for the segmentation and modeling steps, these environments are treated differently. As a matter of fact, the treatment of the exterior (e.g. facade) or interior (e.g. room) will not be the same due to the different structural and architectural elements at different levels of complexity and involvement. Some algorithms such as RANSAC or similar methods are involved in the treatment of both typologies but employed differently.

In interior reconstruction, segmentation begins with the separation of floors, before moving on to delineating rooms, then walls and slabs, finally addressing the remaining elements if necessary [4]. This progressive approach is essential for establishing the overall geometry of a floor, advising that walls should be segmented first before rooms are defined [29]. A commonly employed method for floor segmentation is the use of z-histograms which generate horizontal slices of the interior space, facilitating the separation of the structure into individual floors [30]. This segmentation is followed by that of the rooms, which progresses from the base of the structure, i.e. the floor, and extends towards the elevations [31][32]. However, when reconstructing interior spaces, it is essential to take into account wall occlusions which can obstruct the view of certain elements. To overcome this challenge, specialized algorithms have been designed to deal with occluded elements, ensuring that no critical details are missed during the reconstruction process [33][34].

Exterior reconstruction follows a similar approach, segmenting buildings according to their relevant components: facades, roofs and ground surfaces. Facade segmentation, in particular, can be subdivided to improve information extraction, by dividing the facade into storeys, then into uniformly sized tiles, and finally into elements such as doors and windows [35]. Another approach considers the entire facade, using feature lines and cell complexes to determine division boundaries while addressing some occlusion issues [36]. In addition, segmentation of exterior point cloud can be achieved using images rather than scans, as demonstrated by contour and aperture detection methods [37][38].

The evolution of Machine Learning and Deep Learning represents a watershed in the efficiency of the Scan-to-BIM process, automating complex operations from occlusion detection to large-scale reconstruction. However, despite its transformative potential, Deep Learning faces challenges such as the need for manual design and heavy dependence on available data [6]. The contribution of Machine Learning to the field is illustrated by the improvement in semantic segmentation, in particular with the Random Forest (RF) algorithm, which led to the possibility of semantic segmentation [39], enabling precise recognition of building elements and automated generation of models based on predefined templates [40]. In the specific context of exterior facade analysis, Deep Learning applied to 2D orthoimages has facilitated automatic segmentation, combining initial segmentation and 3D back-projection to obtain a semantically segmented point cloud. This method, although efficient, could benefit from images taken from more advantageous angles and the exploration of new data sources to improve overall accuracy [41]. At the heart of the innovations in 3D point cloud processing, PointNet and PointNet++ offer effective solutions to overcome the challenges of unstructured and disordered data (figure 3). By providing accurate segmentation and classification through a globally invariant representation, these algorithms play a key role in the detailed analysis of local features, essential for semantic segmentation [42][43].

Following these AI algorithms, more recent architec-



Figure 3. Example of PointNet classification [42]

tures such as Graph Convolutional Networks effectively exploit the graph structure of point cloud data to better capture local relationships [44]. Dynamic Graph CNN build a neighbourhood graph at each layer to capture richer local features [45]. PointCNN uses a convolution approach to learn a hierarchical representation of the data [46].

Despite these new and promising algorithms, AI still needs to overcome a number of challenges, including the sheer amount of data required for training. Only few valuable training datasets such as S2DIS (Stanford 2D-3D-Semantics) [47], ScanNet [48] or Paris-Lille 3D [49] are publicly available. Current research is therefore turning to innovative approaches using synthetic point cloud generation in order to enrich the availability of varied and representative training data [50]. However, this type of method has its limits in terms of geometric representations, whether volumetric or in terms of detail accuracy. As discussed in [6], the success of Deep Learning highly depends on the relevance of its input data, and the use of small or synthetic data sets represents a major handicap for obtaining a stable solution.

As far as industry is concerned, some start-ups, companies and scanner publishers offer solutions based on artificial intelligence to process the point cloud and extract all the previous steps. However, this process has not yet reached maturity, and in most cases, these steps are performed manually, as they are considered more precise and important.

2.4 Modeling

The last step of the Scan-to-BIM process is the modeling into usable data such as 3D or BIM. The Level of Development (LOD) is an industry standard that defines among other things the degree of refinement for the 3D geometry of a BIM model shown in figure 4. LOD is an important key to completing Scan-to-BIM, as it defines the accuracy of detail required during acquisition and is a crucial input for obtaining a consistent BIM model.



Figure 4. Example of LOD in a BIM model [51]

There are several different methods for obtaining a model from a point cloud that has been segmented and labelled. The first is parametric modeling, which offers great flexibility in the manipulation of architectural forms, particularly useful for facade renovation [52][53]. This approach emphasizes the importance of adaptability and reusability in the design process. On the other hand, semi-automatic methods combine human intervention with automation, enabling a balance between precision and adaptability [31] [54] [55]. These techniques aim to improve the classification and reconstruction of specific architectural elements, while offering superior geometric quality. Finally, full automation promises efficient creation of BIM and parametric 3D models, significantly reducing the time and effort required for modeling [33][56]. These approaches focus on accurate reconstruction of walls and their topology, and on optimizing the overall connectivity of interior spaces. These global methods use approaches that can be grouped into three categories: (1) planar primitive detection, (2) volumetric primitive, (3) mesh-based reconstruction [3].

The papers therefore typically focus on one type of typology such as exterior or interior, which corresponds to a specific industry use case. Also, the type of reconstruction will depend on the needs, as a mesh reconstruction will be less accurate than a full BIM, but quicker to obtain for elementary visualization. Modeling in industry is mainly done manually, using appropriate software such as Autocad Revit. This can be explained by the fact that existing algorithms are not yet mature enough, due to a lack of adaptability, and that industry, through its use cases, does not necessarily require a complete BIM model. For the integration of AI in this area, an important future approach could be the automation of repetitive tasks and the point cloud quality control, which will be an important issue in ensuring the accuracy of the data created through the Scan-to-BIM process.

3 Conclusion

The state-of-the-art analysis attests promising advancements in point cloud generation methods, pre-processing techniques, segmentation and 3D modeling approaches. Most literature on Scan-to-BIM targets finished structures or historic buildings, overlooking its potential in managing the construction phase. Applying Scan-to-BIM for monitoring work progress and as-built verification introduces new challenges in harnessing construction site data beyond traditional practices. Furthermore, only few work is dedicated to the overall Scan-to-BIM process. The different stages presented in figure 1 do not get the same amount of attention in the literature, which is understandable for a relatively mature stage such as data acquisition, but less so for BIM formatting, i.e. translating the 3D model to a BIM model. As a matter of fact, the challenges involved in producing a complete BIM model are different from those for creating a 3D geometry.

The Scan-to-BIM process needs to be improved to become a complete and automatic process. This includes addressing existing challenges such as occlusion and clutter in the point cloud [3], improving the AI in its design, its training, and in controlling the result [8], and implementing a complete path among all stages so that they are no longer independent. In addition, involving industry to provide data sets and test sites would significantly speed up the development.

The industry, through the example of Bouygues Construction, is currently not using the solutions proposed in the literature, for several reasons. The absence of readily usable off-the-shelf solutions capable of managing the entire process, and the lack of resources to engage themselves in research are major issues. It is important to recognize that complete solutions, even addressing only specific environments or building types would be preferable to the current manual methods. Our next work will therefore focus on the implementation of a complete Scan-to-BIM process for a particular use case, namely the facade in the context of energy renovation for French real estates.

References

- F. Araya. State of the art of the use of BIM for resolution of claims in construction projects. *Revista ingeniería de construcción*, 34:299–306, 2019. doi:10.4067/S0718-50732019000300299.
- [2] Chengshuang Sun, Shaohua Jiang, Miroslaw J. Skibniewski, Qingpeng Man, and Liyin Shen. A literature review of the factors limiting the application of BIM in the construction industry. *Technological and Eco*-

doi:10.3846/20294913.2015.1087071.

- [3] Nuno Abreu, Andry Pinto, Aníbal Matos, and Miguel Pires. Procedural Point Cloud Modelling in Scanto-BIM and Scan-vs-BIM Applications: A Review. ISPRS International Journal of Geo-Information, 12: 260, 2023. doi:10.3390/ijgi12070260.
- [4] C. Gourguechon, H. Macher, and T. Landes. Automation Of As-Built BIM Creation From Point Cloud: An Overview Of Research Works Focused On Indoor Environment. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2022:193-200, 2022. doi:10.5194/isprs-archives-XLIII-B2-2022-193-2022.
- [5] David J. Griffiths, David Griffiths, and Jan Boehm. A review on deep learning techniques for 3d sensed data classification. Remote Sensing, 2019. doi:10.3390/rs11121499.
- [6] M. Buyukdemircioglu, S. Kocaman, and M. Kada. Deep Learning For 3D Building Reconstruction: A review. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2022:359-366, 2022. doi:10.5194/isprs-archives-XLIII-B2-2022-359-2022.
- [7] Tianzhen Hong, Zhe Wang, Xuan Luo, and Wanni Zhang. State-of-the-art on research and applications of machine learning in the building life cycle. Energy and Buildings, 212:109831, 2020. doi:10.1016/j.enbuild.2020.109831.
- [8] Zifeng Ding, Yuxuan Sun, Sijin Xu, Yan Pan, Yanhong Peng, and Zebing Mao. Recent Advances and Perspectives in Deep Learning Techniques for 3D Point Cloud Data Processing. Robotics, 12:100. 2023. doi:10.3390/robotics12040100.
- [9] Viorica Pătrăucean, Iro Armeni, Mohammad Nahangi, Jamie Yeung, Ioannis Brilakis, and Carl Haas. State of research in automatic as-built modelling. Advanced Engineering Informatics, 29:162–171, 2015. doi:10.1016/j.aei.2015.01.001.
- [10] 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, 19:829-843, 2010. doi:10.1016/j.autcon.2010.06.007.

- nomic Development of Economy, 23:764–779, 2017. [11] Jisang Lee, Seunghwan Hong, Hanjin Cho, Ilsuk Park, Hyoungsig Cho, and Hong-Gyoo Sohn. Accuracy Comparison Between Image-based 3D Reconstruction Technique and Terrestrial LiDAR for As-built BIM of Outdoor Structures. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 33:557-567, 2015. doi:10.7848/ksgpc.2015.33.6.557.
 - [12] Leica Geosystems Shop Directory Buy Leica BLK, 2023. URL https://shop.leica-geosystems. com/gb.
 - [13] A. Dlesk, K. Vach, J. Šedina, and K. Pavelka. Comparaison Of LEICA BLK360 And LEICA BLK2GO On Chosen Test Objects. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-5-W1-2022:77-82, 2022. doi:10.5194/isprs-archives-XLVI-5-W1-2022-77-2022.
 - [14] Yang Ai and Xi Yang. A Dynamic Network for Efficient Point Cloud Registration. 2023. doi:10.48550/arXiv.2312.02877.
 - [15] Yue Wang and Justin Solomon. Deep Closest Point: Learning Representations for Point Cloud Registration. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 3522–3531, 2019. doi:10.1109/ICCV.2019.00362.
 - [16] Shengjun Tang, Yunjie Zhang, You Li, Zhilu Yuan, Yankun Wang, Xiang Zhang, Xiaoming Li, Yeting Zhang, Renzhong Guo, and Weixi Wang. Fast and Automatic Reconstruction of Semantically Rich 3D Indoor Maps from Low-quality RGB-D Sequences. Sensors (Basel), 19:533, 2019. doi:10.3390/s19030533.
 - [17] Faisal Zaman, Ya Ping Wong, and Boon Yian Ng. Density-Based Denoising of Point Cloud. In 9th International Conference on Robotic, Vision, Signal Processing and Power Applications, pages 287–295, 2017. doi:10.1007/978-981-10-1721-6_31.
 - [18] Hafsa Benallal, Ilyass Abouelaziz, Youssef Mourchid, Ayman Al Falou, Hamid Tairi, Jamal Riffi, and Mohammed El Hassouni. A new approach for removing point cloud outliers using the standard score. In Pattern Recognition and Tracking XXXIII, volume 12101, pages 56-62, 2022. doi:10.1117/12.2618835.
 - [19] Md Nazmul Kabir Sikder and Feras A. Batarseh. 7 -Outlier detection using AI: a survey. pages 231-291, 2023. doi:10.1016/B978-0-32-391919-7.00020-2.

- [20] PointCleanNet: Learning to Denoise and Remove Outliers from Dense Point Clouds - Rakotosaona -2020 - Computer Graphics Forum - Wiley Online Library, 2020. URL https://onlinelibrary. wiley.com/doi/10.1111/cgf.13753.
- [21] Shuaijun Chen, Jinxi Wang, Wei Pan, Shang Gao, Meili Wang, and Xuequan Lu. Towards uniform point distribution in feature-preserving point cloud filtering. *Comp. Visual Media*, 9:249–263, 2023. doi:10.1007/s41095-022-0278-4.
- [22] Francesca Pistilli, Giulia Fracastoro, Diego Valsesia, and Enrico Magli. Learning Graph-Convolutional Representations for Point Cloud Denoising. 2020. doi:10.48550/arXiv.2007.02578.
- [23] Chaojing Duan, Siheng Chen, and Jelena Kovacevic. 3D Point Cloud Denoising via Deep Neural Network based Local Surface Estimation. 2019. doi:10.48550/arXiv.1904.04427.
- [24] Tomas M. Borges, Diogo C. Garcia, and Ricardo L. de Queiroz. Fractional Super-Resolution of Voxelized Point Clouds. *IEEE Trans Image Process*, 31: 1380–1390, 2022. doi:10.1109/TIP.2022.3141611.
- [25] Y. Xu, L. Hoegner, S. Tuttas, and U. Stilla. Voxel-And Graph-Based Point Cloud Segmentation Of 3D Scenes Using Perceptual Grouping Laws. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1-W1:43–50, 2017. doi:10.5194/isprs-annals-IV-1-W1-43-2017.
- [26] Yusheng Xu, Xiaohua Tong, and Uwe Stilla. Voxelbased representation of 3D point clouds: Methods, applications, and its potential use in the construction industry. *Automation in Construction*, 126:103675, 2021. doi:10.1016/j.autcon.2021.103675.
- [27] Dawei Li, Yongchang Wei, and Rongsheng Zhu. A comparative study on point cloud down-sampling strategies for deep learning-based crop organ segmentation. *Plant Methods*, 19:124, 2023. doi:10.1186/s13007-023-01099-7.
- [28] Peng Zhang, Ruoyin Xie, Jinsheng Sun, Weiqing Li, and Zhiyong Su. AS-PD: An Arbitrary-Size Downsampling Framework for Point Clouds. 2023. doi:10.48550/arXiv.2211.01110.
- [29] Srivathsan Murali, Pablo Speciale, Martin R. Oswald, and Marc Pollefeys. Indoor Scan2BIM: Building information models of house interiors. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 6126–6133, 2017. doi:10.1109/IROS.2017.8206513.

- [30] Kate Pexman, Derek D. Lichti, and Peter Dawson. Automated Storey Separation and Door and Window Extraction for Building Models from Complete Laser Scans. *Remote Sensing*, 13:3384, 2021. doi:10.3390/rs13173384.
- [31] Hélène Macher, Tania Landes, and Pierre Grussenmeyer. From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings. *Applied Sciences*, 7: 1030, 2017. doi:10.3390/app7101030.
- [32] C. Gourguechon, H. Macher, and T. Landes. Room Point Clouds Segmentation: A New Approach Based On Occupancy And Density Images. *ISPRS Annals* of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-M-1-2023:93–100, 2023. doi:10.5194/isprs-annals-X-M-1-2023-93-2023.
- [33] Maarten Bassier and Maarten Vergauwen. Unsupervised reconstruction of Building Information Modeling wall objects from point cloud data. *Automation in Construction*, 120:103338, 2020. doi:10.1016/j.autcon.2020.103338.
- [34] Xuehan Xiong, Antonio Adan, Burcu Akinci, and Daniel Huber. Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction*, 31:325–337, 2013. doi:10.1016/j.autcon.2012.10.006.
- [35] Pascal M"uller, Gang Zeng, Peter Wonka, and Luc Van Gool. Image-based procedural modeling of facades. ACM Trans. Graph., 26:85–es, 2007. doi:10.1145/1276377.1276484.
- [36] X. Ning. Structural Wall Facade Reconstruction of Scanned Scene in Point Clouds. *Adv. Electr. Comp. Eng.*, 21:11–20, 2021. doi:10.4316/AECE.2021.04002.
- [37] Habib Fathi, Fei Dai, and Manolis Lourakis. Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. *Advanced Engineering Informatics*, 29:149–161, 2015. doi:10.1016/j.aei.2015.01.012.
- [38] Antoine Fond, Marie-Odile Berger, and Gilles Simon. Model-image registration of a building's facade based on dense semantic segmentation. *Computer Vision and Image Understanding*, 206:103185, 2021. doi:10.1016/j.cviu.2021.103185.
- [39] V. Croce, M. G. Bevilacqua, G. Caroti, and A. Piemonte. Connecting Geometry And Semantics Via Artificial Intelligence: From 3D Classification Of Heritage Data To H-BIM Representations.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2021:145–152, 2021. doi:10.5194/isprsarchives-XLIII-B2-2021-145-2021.

- [40] Valeria Croce, Gabriella Caroti, Andrea Piemonte, Livio De Luca, and Philippe Véron. H-BIM and Artificial Intelligence: Classification of Architectural Heritage for Semi-Automatic Scan-to-BIM Reconstruction. *Sensors (Basel)*, 23(5):2497, 2023. doi:10.3390/s23052497.
- [41] A. Murtiyoso, C. Lhenry, T. Landes, P. Grussenmeyer, and E. Alby. Semantic Segmentation For Building Façade 3D Point Cloud From 2D Orthophoto Images Using Transfer Learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021:201–206, 2021. doi:10.5194/isprsarchives-XLIII-B2-2021-201-2021.
- [42] 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 Conference on Computer Vision and Pattern Recognition (CVPR), pages 77–85, 2017. doi:10.1109/CVPR.2017.16.
- [43] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, 2017. URL http:// arxiv.org/abs/1706.02413. arXiv:1706.02413 [cs].
- [44] Yawei Li, He Chen, Zhaopeng Cui, Radu Timofte, Marc Pollefeys, Gregory Chirikjian, and Luc Van Gool. Towards Efficient Graph Convolutional Networks for Point Cloud Handling. 2021. doi:10.48550/arXiv.2104.05706.
- [45] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic Graph CNN for Learning on Point Clouds. 2019. doi:10.48550/arXiv.1801.07829.
- [46] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. PointCNN: Convolution On X-Transformed Points. 2018. doi:10.48550/arXiv.1801.07791.
- [47] Iro Armeni, Sasha Sax, Amir R. Zamir, and Silvio Savarese. Joint 2D-3D-Semantic Data for Indoor Scene Understanding, 2017.
- [48] 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.

- [49] Xavier Roynard, Jean-Emmanuel Deschaud, and François Goulette. Paris-lille-3d: A large and highquality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6):545– 557, 2018. doi:10.1177/0278364918767506.
- [50] Jong Won Ma, Thomas Czerniawski, and Fernanda Leite. Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds. *Automation in Construction*, 113:103144, 2020. doi:10.1016/j.autcon.2020.103144.
- [51] STRUCTURE magazine | Beating Chaos and Achieving Profits in BIM with LOD 350, 2013. URL https://www.structuremag.org/?p=558.
- [52] Conor Dore and Maurice Murphy. Semi-automatic generation of as-built BIM façade geometry from laser and image data. *Journal of Information Technology in Construction (ITcon)*, 19:20–46, 2014.
- [53] Oscar Gámez Bohórquez, William Derigent, and Hind Bril El-Haouzi. Parametric point cloud slicing for facade retrofiting. *International Journal of Architectural Computing*, 2021. doi:10.1177/14780771211029747.
- [54] Valeria Croce, Gabriella Caroti, Livio Luca, Kevin Jacquot, Andrea Piemonte, and Philippe Véron. From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning. *Remote Sensing*, 13: 461, 2021. doi:10.3390/rs13030461.
- [55] Sungchul Hong, Jaehoon Jung, Sangmin Kim, Hyoungsig Cho, Jeongho Lee, and Joon Heo. Semiautomated approach to indoor mapping for 3D as-built building information modeling. *Computers, Environment and Urban Systems*, 51, 2015. doi:10.1016/j.compenvurbsys.2015.01.005.
- [56] 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. doi:10.1016/j.isprsjprs.2019.03.017.

Assessing the Viability of Robotic Disassembly of Building Components for Resource Recovery

Christopher Rausch^{1*}, Seungah Suh¹, Nikiforos Repousis¹, Nathan Titterington¹, Han Nguyen¹

¹ Digital Harvest Lab, Fariborz Maseeh Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin, TX, USA c.rausch@utexas.edu, sasuh@utexas.edu, nrepousis@utexas.edu, npt392@utexas.edu, han317734@utexas.edu

Abstract

The transition towards a circular economy will, in large part, necessitate the disassembly and recovery of components from existing building stock. Robotic disassembly has emerged as a technique in other (nonbuilding) industries as a method for efficient and scalable resource recovery. Since robotic disassembly has yet to achieve a similar level of maturity in the building industry, this paper presents an assessment framework towards this aim. This framework harmonizes the demonstrable capabilities of robotic systems (via literature synthesis and currently available hardware) with suitable deconstruction applications (using a case study of a large institutional building). The results yield strategic paths forward for enabling robotic disassembly of building components.

Keywords -

Robotics; Demolition; Circular Economy; Computing; Literature Review; Deconstruction; Machine Vision; Resource Recovery

1 Introduction & Background

With global raw material consumption set to double by 2060 [1], and the fact that the building sector is already the largest consumer of materials globally [2], it is imperative to find alternative ways to source and conserve materials for building construction. Recovery and reuse of building materials is an alternative to the current supply chain feedstock of new materials; for which disassembly is a fundamental operation to perform.

In contrast to manual disassembly, automated disassembly has emerged as a viable method in many applications for its ability to increase disassembly efficiency, be implemented at scale, combat labour market concerns, minimize costs for retrieving highvalue items, and to perform complex disassembly processes (e.g., bespoke fixturing, multiple disassembly tools and very small geometric conditions and interfaces), among others [3]. Robotic disassembly has been successfully applied to numerous products including automotive (e.g., batteries, electronic control units, printed circuit boards, etc.), electronics (e.g., cell phones, circuit boards, LCD monitors, computers), and mechanical products (e.g., aerospace components, chassis, industrial manifolds, etc.) [4], [5]. The societal value created by robotic disassembly includes landfill diversion, material conservation for new products, creation of new local jobs [6], and functions as a key step in the growing the re-manufacturing industry.

Robotic disassembly in the building sector has yet to emerge at a scale similar to consumer and industrial product sectors. If realized however, it can overcome several key impediments: deconstruction often has heavy impacts to humans for the noise, dust, vibration, contaminated materials (e.g., asbestos) and debris generated, safety concerns, labor shortages, and inefficient resource recovery rates [7]. While currently ambitious, the ability for robots to initiate or augment current disassembly activities presents significant opportunities.

1.1 Robots for Demolition

Single-task construction robots that are deployed for dismantling structures are called demolition robots [8]. They are "all-terrain" machinery with a relatively small size, which allows them to operate efficiently in the challenging conditions of a demolition site. Demolition robots are not yet autonomous, due to the complexity of their work environment. However, they can be guided with remote controls from on-site or off-site personnel [9]–[11]. This function is beneficial both from an economic and safety perspective. To be specific, demolition works are hazardous and highly repetitive, thus utilizing robots minimizes the risks for the personnel and at the same time increases productivity [9], [11], [12].

Demolition robots often have a hydraulic-powered arm mounted on wheels, crawlers, or tracks to which shears, breakers, crushers, drills, buckets, cutters, grapples, and even high-pressure water jets can be attached [8], [10]. The first one launched was by PE Holmgren and Rivteknik in Sweden, back in 1976. In 1981, Brokk produced the first mass-produced model (Brokk 250), which was remote-controlled. Brokk still continues to manufacture similar models. TopTec has specialized in demolition robots since the late 1980s. Aquajet and CONJET have developed models for hydrodemolition since the late 1980s and early 1990s respectively. Husquarna has also been offering various options for high ambient temperatures since 2009.

The presented demolition robots are some of the currently available products in the industry. However, the research is still ongoing, with the goal being to develop fully autonomous ones. This might be possible in the future as shown by the use of autonomous vehicles for excavation, transportation, and finish grading [8], [13]. Researchers are still exploring ways to make this happen by combining sensors and lasers with sophisticated algorithms [9], [12], [14]. The potential autonomy of demolition robots would greatly enhance the feasibility of deconstruction techniques, thereby facilitating the reuse of materials such as concrete [15].

2 Proposed Methodology

This paper assesses the viability of robotic disassembly of buildings as a method for supporting greater reuse and resource recovery of building components.



Figure 1. Active demolition project

The proposed methodology uses a three-tier assessment to determine the viability of robotic disassembly for building components (Figure 2). First, we review prevalent robotic disassembly applications and tasks (focusing outside the construction industry, where such work is more mature). Next, we assess current mobile robot hardware capabilities (using manufacturer catalogues) and software requirements specifically in the context of on-site selective disassembly of building components. Finally, we evaluate the viability of specific building components that could be subject to on-site robotic disassembly within a local geographic context (we analyse materials from an institutional building).



Figure 2. Methodology for assessing the viability of robotic disassembly of buildings

2.1 Robotic Disassembly Applications/Tasks

In their comprehensive review of robotic disassembly applications (which includes digital assistance e.g., sequence planning, decision making, vision systems, and physical assistance e.g., robotic systems and humanrobot collaboration), Poschmann et al [3] identified 41 unique robotic disassembly applications. These include robotic disassembly of vehicles and electronics; of which, the majority of robotic tasks focus on handling, removing and separating components. Vision systems emerge as a key aspect of robotic disassembly for identifying target components, fasteners or conflicts along the disassembly path. Another key trend in robotic disassembly is the setup and deployment of entire cells which bring assemblies (e.g., waste electrical and electronic equipment) into a controlled environment to perform disassembly operations. It is typically in this context that human-robot-collaboration (HRC) systems are most prevalent. HRC is often favored even when a-priori geometric data is available for assemblies, since a key challenge centers on the uncertainty of other conditions (whether items are clean, have deviations, or if they are still in good working condition) [16]. Across existing applications of robotic disassembly, the following trends are noted in the literature: (a) maximizing the practical use of robots for disassembly occurs in HRC workflows [17], (b) the vast majority of applications (>90%) rely on accurate and up-to-date a-priori information on the assemblies, coupled with high-repetition across tasks [3], (c) robotic disassembly is optimized when planned upfront in product design [6], and (d) one of the most common robot tasks centers on removal of fasteners [18].

Reviewing the state of existing robotic disassembly

applications, the following insights can be made for application to building components. First, HRC is likely required, since buildings possess more variability than manufactured products, greater constraints for robots (in terms of reach and payload), and less up-to-date digital information (e.g., 3D models) for disassembly planning. Second, void of accurate a-priori information on buildings, additional considerations are required for to program sensors and to perform learning in order to aid in efficient disassembly operations.

2.2 Mobile Robot Hardware Considerations

Following the guidance for construction robots outlined in Dritsas et al. [19], mobility criteria are considered feasible when robot weight (including platform) is less than 1 ton (1000kg), and when the robot reach is equal to or greater than 1m (in the horizontal axis direction). From this criterion, we analysed off-the-shelf mobile robots (using a library outlined by RoboDK© software) which fall into the following categories: 5 DOF (degree of freedom) robots, 6 DOF robots, 7 DOF robots, Delta robots and Palletizing robots. Without delving into the feasible disassembly tasks that each of these robots (and robot typologies) can perform, we identified 352 unique payload vs. reach datapoints from which to identify potential disassembly tasks (Figure 3). While the reach of each robot examined ranges up to 3280 mm, the use of a mobile platform can provide extended mobility on-site where needed for certain activities (e.g., mobile co-bots, compound robots).



Figure 3. Mobile robot hardware capability assessment of 352 unique off-the-shelf robots

The next step in assessing robot hardware capability for disassembly is understanding (classifying) which types of tasks can be used. Based on general robot motion primitives for construction tasks as outlined by [20] (e.g., grasping, cutting, drilling and screwing), we propose a series of distinct disassembly operations that these robot motion primitives can be used for, as outlined in Table 1. While these robot primitives require custom end-of-arm tooling and programming for unique disassembly operations, it can be shown that many primitives already used in non-disassembly operations can also be used specifically for disassembly operations. For instance, screwing primitive could be used for removal of fasteners with operations including de-screwing and unbolting.

Table 1. Robot motion primitives for disassembly

Robot Motion Primitives	Robot Disassembly Operations
Grasping	Holding/supporting/prying/pulling
Cutting	Selective destruction for removal
Drilling	Access for cutting tool
Screwing	De-screwing, unbolting
	(removing fasteners)

2.3 Software Architecture Considerations for Robotic Disassembly in Construction

Construction robots generally need to be highly adaptable and agile due to greater environmental uncertainty and complexity as compared with manufacturing applications [21]. This section discusses the requirements of the high-level decision-making layer of their software [22], and the first two workflows of common software programming: environment perception and planning [23], as shown in Figure 4.

For environment perception, robots use various sensors such as cameras, laser scanners, and radiofrequency identifiers (RFID) to collect data. Building information modeling (BIM) serves as an additional valuable information source for understanding the built environment [24]. Robots need to interpret, sense and localize based on collected data, detecting objects or estimating motions through deep-learning models or predefined algorithms, for instance [25]. [26] also has proposed a graph-based multi-modal sensor data integration approach to enhance real-time state and location awareness. Considering how unstructured and dynamic construction and demolition sites typically are, real-time sensing is crucial, accordingly heightening the need for efficient data processing models.

In terms of the planning and decision-making process, construction robots should efficiently update their pretrained and predetermined motion plans in response to

changing (and sometimes unknown) environments. Software requirements are influenced by various factors such as autonomy level, the number of tasks, and required functionalities. For example, fully autonomous robots, driven without human interventions, often rely on artificial intelligence [8], while non-autonomous robots, including pre-programmed and teleoperated robots, are operated by deterministic software modules or human operation. Regarding the number of tasks, multi-task robots demand higher flexibility and learning efficiency than single-task robots [27]. Likewise, robots require different system configurations based on task types (e.g., on-site monitoring, on-site assembly, off-site prefabrication) and mobility types (e.g., gantry systems, aerial, and terrestrial) [28]. Building disassembly tasks, in particular, present additional challenges due to significant variations in target components' conditions, type, and geometry compared to the assembly stage [29].

To enhance the planning process, previous studies have introduced various solutions. BIM has been actively explored for task and motion planning, as shown in the studies utilizing it to provide detailed task descriptions, spatial information, and point calibration for tasks such as brick assembly, structural components assembly, and indoor wall painting [30]-[32]. Similarly, leveraging rich project information from BIM will assist in automating detailed disassembly sequence planning. A hierarchical reinforcement learning training strategy has been developed for more generalized control policies [33]. The efficiency of a demonstration-based motion sequence learning module in multi-task motion sequencing has also been demonstrated [27]. For the teleoperation of construction robots, a multi-user immersive environment has been explored for interaction during excavator teleoperation [34], and a brainwavebased teleoperation system has been proposed for workers in limited movability environment [35].



Figure 4. Software hierarchy and programming workflow of construction robots

2.4 Robotic Disassembly Feasibility Scale

In light of the high-level hardware and software considerations for robotic disassembly of building components, we posit a Likert-based robot feasibility scale for assessing initial feasibility. This scale uses five unique categories: RF0 (no robot feasibility), RF1 (limited robot feasibility, i.e., 1 task can be supported), RF2 (some robot feasibility, 2+ tasks can be supported), RF3 (strong robot feasibility, but requires human intervention or collaboration) and RF4 (strong robot feasibility, requiring little-to-no human support, e.g., a fully autonomous solution is probable). To help define each of these robot feasibility levels, we consider key constraints for disassembly tasks, drawing on experience the lead author has from industry work overseeing robot cells for fabrication in construction and experience with disassembly operations. The considered constraint categories are organized into those related to components (torque (T) and payload (P)), and those related to environment or hardware (reach (R) and spatial (S)), with examples defined below:

- (T): rusted/broken fastener, stripped head on fastener, overtightened bolt
- (R): end effector required location exceeds reach capability. For instance, commercial and residential buildings often have a floor-to-ceiling height between 8-12 feet (2.44 m to 3.66 m). In the case where a robot needs to reach the ceiling for disassembly, a custom platform might be required for current robots.
- (P): object being moved is too heavy, and or task moment (force*distance) exceeds capacity
- (S): robot arm and end effector hardware does not have ability to navigate in confined spaces adequately to perform disassembly task

It is important to note that the proposed robot disassembly feasibility scale is based on a specific material in isolation from its overall quantity within a building. As such, determining overall viability of resource recovery potential for a given material may also need to account for a minimum threshold of quantity (e.g., for supply-demand mapping of specific reuse items).



Figure 5. Overview of specific robotic disassembly constraints for resource recovery of building materials

3 Case Study

3.1 Background

The Erwin Center is an events center located in Austin, Texas, built in 1977, and is comprised of a steelframed structure, with a precast concrete panel enclosure (Figure 6). In addition to its structure, there is a wide range of materials and components which can be potentially reused including doors, windows, chairs, fixtures (electrical, plumbing, mechanical, HVAC), gypsum board, electrical equipment, etc. It underwent deconstruction and demolition starting in 2023, during which time, the authors participated in a building material reuse audit with a third-party consultant.

As part of conducting this audit, priority was given to items which were found to have the highest potential for reuse based on local market factors, quantity and condition of components. Examples of the highest reuse valued materials include interior doors, ceiling tiles, carpet floor tiles, cabinets, mechanical and electrical equipment, light fixtures, network infrastructure and newer-condition appliances. Additionally, during the demolition phase, selective demolition and disassembly were performed to separate and sort many building materials for the purpose of recycling, including auditorium seating, piping, conduit, plumbing fixtures, and light gage wall framing.



Figure 6. Aerial view of the Erwin Center

3.2 Robotic Disassembly Assessment

The authors assessed robotic disassembly feasibility of specific materials using the proposed framework, identifying potential disassembly operations, known constraints and an estimate of overall quantities (we adopted a simple low (0-50 items), medium (50-500 items), and high (500+ items) approach, where overall robot feasibility was best for a high quantity of materials recovered since the return on investment of a robotic system is highest where it can be deployed in a highlyscalable manner). In cases where there is high value for specific material recovery but existing in a low quantity, it is assumed that manual (non-robotic) extraction is more cost-effective and preferred. The result of this overall assessment is shown in Table 2, where the best candidate materials for robotic disassembly are highlighted in green, candidates with moderate potential are shown in yellow, and candidates with poor potential are shown in red.

Table 2. Assessment of robotic disassembly feasibility

Salvage Component	Disassembly	RF*	Known	Quantity
Interior doors	De-screwing, supporting/ handling, pulling	RF2	T, P: Hollow core: 25lbs, Solid core: 50lbs	M
Exterior door	De-screwing, prying, supporting	RF2	T, R: 80" height to upper screws, P: 100lbs	М
Decorative lighting	De-screwing, prying, supporting	RF2	T, R: ceiling height (8-12')	Н
Carpet tile	Handling, prying	RF3	Т	Н
Ceiling tile	Handling	RF3	R: ceiling height (8-12'), S: framing for tiles	Н
Cabinets	De-screwing, cutting, supporting	RF2	T, P: 200+lbs, S cabinet geometry	M

Sink top with	Cutting	RF1	S: sink geometry	М
fixtures intact				
Light fixtures	Supporting	RF0	S: ceiling system geometry	Н
2x4 wood	Cutting, De- screwing	RF2	T, P: 20lbs	Н
Mechanical pump	De-screwing, supporting	RF0	T, P: >500lbs	L
Motors	De-screwing, supporting	RF0	T, P: >500lbs	М
Controllers	De-screwing, supporting	RF0	Т	М
Transformers	De-screwing, supporting	RF0	T, P: >500lbs	М
Plumbing	Cutting	RF1	S: plumbing	Н
fixtures	-		geometry	
Fire doors	De-screwing, prying, supporting	RF2	T, P: 100lbs	М
Electrical	De-screwing,	RF1	T, S: panel	М
panel box	handling		geometry	
Exit signs	De-screwing, handling	RF1	R: ceiling height (8-12')	М
Wood fibre	Handling,	RF1	T, R: ceiling	Η
panel	prying		height (30'+)	
Auditorium seating	De-screwing	RF1	Т	Н

*RF = Robot Feasibility Factor

**T is listed generally for a range of conditions included stripped or rusted fastener, unknown tactile movements, etc.

4 Discussion

Based on the robotic disassembly feasibility assessment, carpet tiles, ceiling tiles and auditorium seating were found to have the best overall potential for resource recovery supported by robots (Figure 7). Of the building components reviewed, ceiling tiles and carpet tiles had the highest robot feasibility (RF) factor. Both of these materials incorporate achievable disassembly operations as they are predominately gravity-installed (ceiling tiles being supported by a supporting frame, and carpet tiles being mounted using a special removable double-sided tape). Moreover, the vast majority of robots reviewed in Figure 3 can be used to support the removal because these components do not have a notable payload constraint: ceiling tiles are 2.2kg and carpet tiles are 0.63kg. The quantity of both materials also plays a significant role in its robotic disassembly feasibility (there were approximately 22,200 carpet tiles and 41,000 ceiling tiles). Although both building components are increasingly being supported by take-back programs (either through the original manufacturer or by third parties), current programs typically require the owner or onsite contractor to palletize tiles (which includes the labor for removal and stacking). This implies that robotic disassembly can even contribute to promoting readily

available resource recovery programs by reducing the need for substantial manual disassembly.

The third building component identified as being viable for robotic disassembly was auditorium seating. While this component had a very low robotic feasibility (RF) factor, and despite its potential for reuse being low (due to poor item quality), extraction of auditorium seating was found to be a significant source of manual labor during demolition. Given the necessity to recycle the steel in the seating (i.e., many local jurisdictions require a minimum waste diversion rate for C&D projects), during demolition, 100 workers were brought on site for 5 continuous days to remove fasteners and grind the base connection of each seat. The labor cost for this is estimated at USD \$100,000 (based on local market labor rate and contractor markup fees). In this case, the use of a robotic system to support disassembly not only could be used to address the high cost of fastener removal, but also reduces safety risk on site for the contractor.

The study demonstrated the potential of the proposed viability assessment framework, grounded in existing research and industry resources, to identify the feasibility of robotic disassembly and the most suitable building components. The framework was initially applied to one institutional building but can be utilized to evaluate the various other building components with different contexts.



Figure 7. Carpet tiles, ceiling tiles and auditorium seating

5 Conclusions

In summary, this paper explored the potential to incorporate robots to support the resource recovery of building materials at their end of life through disassembly. Current approaches for resource recovery in the construction industry are driven by either manual labor tasks (which include the manual operation of various mechanical equipment). Other industries are beginning to see the prevalence of robots to aid in disassembly operations, however the same level of maturity is not currently seen in the construction industry. To fill this gap, we proposed a conceptual framework that looks at broad applications for disassembly, understanding both hardware and software requirements and finally looking at the specific constraints posed by different building material removal processes. Since the purpose of this framework was only to assess high-level feasibility for robotic resource recovery, additional design and engineering are required to develop robotic systems. A case study of a large institutional building found that three potential building components could be supported by robots for disassembly tasks. In all three cases, manual labor was found to be the primary driver for adopting robotic support.

5.1 Recommendations

A challenge for robotic disassembly in the construction industry is that compared with other applications (e.g., consumer electronics, automotive assemblies, batteries, etc.) the quantity of like-products is not as vast and given the physical size and bespoke nature of construction techniques, geometric variability can be more prolific. As a result, developing fully robotic solutions for disassembly of building components is likely too prohibitive to pursue from an economic and technical standpoint. In fact, fully robotic disassembly approaches are also considered to be prohibitive in many other manufactured product applications [17]. For this reason, human-robot collaboration (i.e., co-bot) solutions are often pursued as a pragmatic approach for many disassembly applications. Based on our assessment of robotic disassembly for buildings, we also recommend such collaborative approach. While there are many potential applications of robots to aid in disassembly planning, it is necessary to consider the practical constraints regarding robot torque, reach, payload and spatial maneuverability when selecting and designing systems. A key limitation in this paper was the robot feasibility based on quantity was only based on a single building. Given the case study explored was a very large structure, with select building materials existing in large quantities, one could potentially argue that justification of a new robot for disassembly support could be based on this building's material alone. While this of course does

not factor in numerous prototyping and deployment feedback, we recommend that surveying building material quantities should be done across a wide range of building stock rather than exploring a single building at its end of life. Such review will necessitate that specific building materials have a high degree of standardization in order for the development of robotic systems to be used across multiple buildings. Other practical considerations need to be factored into the deployment of robotic disassembly solutions which we did not explicitly cover such as how robots can navigate in dynamic/cluttered environments, that need to be considered in a full-scale solution.

Acknowledgments

The authors would like to acknowledge the financial support of the University of Texas at Austin Green Fund initiative. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Green Fund committee.

References

- [1] OECD. Global material resources outlook to 2060: Economic Drivers and Environmental Consequences. On-line: https://www.oecd.org/development/globalmaterial-resources-outlook-to-2060-9789264307452-en.htm, Accessed: 29/10/2023.
- [2] B. C. Guerra, F. Leite, and K. M. Faust. 4D-BIM to enhance construction waste reuse and recycle planning: Case studies on concrete and drywall waste streams. *Waste Mgmt*, 116:9–90, 2020. doi: 10.1016/J.WASMAN.2020.07.035.
- [3] H. Poschmann, H. Brüggemann, and D. Goldmann. Disassembly 4.0: A Review on Using Robotics in Disassembly Tasks as a Way of Automation. *Chemie Ingenieur Technik*, 92(4):341–359, 2020. doi: 10.1002/CITE.201900107.
- [4] W. J. Tan, C. M. M. Chin, A. Garg, and L. Gao. A hybrid disassembly framework for disassembly of electric vehicle batteries. *Int J Energy Res*, 45(5):8073–8082, 2021. doi: 10.1002/ER.6364.
- [5] S. Vongbunyong, S. Kara, and M. Pagnucco. Application of cognitive robotics in disassembly of products. CIRP Annals, 62(1):31-34, 2013. doi: 10.1016/j.cirp.2013.03.037.
- [6] 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. *Intl J Adv Mfg Tech*, 124:2973–3000, 2023. doi: 10.1007/s00170-021-08389-1.
- [7] M. Zabek, L. Hildebrand, M. Wirth, and S. Brell-

Cokcan. Used building materials as secondary resources – Identification of valuable building material and automized deconstruction. *J of Facade Design and Eng*, 5(2):25–33, 2017. doi: 10.7480/JFDE.2017.2.1684.

- [8] M. Casini. Advanced building construction methods. *Construction 4.0*, 405–470, 2022. doi: 10.1016/B978-0-12-821797-9.00006-4.
- [9] J. Huang, Y. Cen, N. Xie, and X. Ye. Inverse calculation of demolition robot based on gravitational search algorithm and differential evolution neural network. *Int J Adv Robot Syst*, 17(3), 2020. doi: 10.1177/1729881420925298.
- [10] N. Melenbrink, J. Werfel, and A. Menges. On-site autonomous construction robots: Towards unsupervised building. *Autom Constr*, 119, 2020. doi: 10.1016/j.autcon.2020.103312.
- [11] M. O. Anderson and D. C. Wadsworth. The Modified Brokk Demolition Machine with Remote Console. *IFAC Proceedings Volumes*, 34(9):221-225, 2001. doi: 10.1016/S1474-6670(17)41709-5.
- [12] Z. Mu, L. Liu, L. Jia, L. Zhang, N. Ding, and C. Wang. Intelligent demolition robot: Structural statics, collision detection, and dynamic control. *Autom Constr*, 142, 2022. doi: 10.1016/j.autcon.2022.104490.
- [13] A. Stentz, J. Bares, S. Singh, and P. Rowe. A Robotic Excavator for Autonomous Truck Loading. *Auton Robots*, 7:175–186, 1999. doi: 10.1023/A:1008914201877.
- [14] F. Corucci and E. Ruffaldi. Toward Autonomous Robots for Demolitions in Unstructured Environments. Advances in Intelligent Syst and Computing, 302, 2015. doi: 10.1007/978-3-319-08338-4 109.
- [15] H. J. Lee, C. Heuer, and S. Brell-Cokcan. Concept of a Robot Assisted On-Site Deconstruction Approach for Reusing Concrete Walls. In *Proceedings of the ISARC*, pages 442–429, Kitakyushu, Japan, 2020.
- [16] S. Hjorth and D. Chrysostomou. Human–robot collaboration in industrial environments: A literature review on non-destructive disassembly. *Robot Compt Integr Manf*, 73:102208, 2022. doi: 10.1016/J.RCIM.2021.102208.
- [17] S. Parsa and M. Saadat. Human-robot collaboration disassembly planning for end-of-life product disassembly process. *Robotics and Compt-Integr Manf*, 71:102170, 2021. doi: 10.1016/j.rcim.2021.102170.
- [18] D. P. Brogan, N. M. DiFilippo, and M. K. Jouaneh. Deep learning computer vision for robotic disassembly and servicing applications. *Array*, 12:100094, 2021. doi: 10.1016/J.ARRAY.2021.100094.

- [19] S. Gim, and S. Soh. Building robotics design for construction Design considerations and principles for mobile systems. *Const Robotics*, 3:1–10, 2019. doi: 10.1007/s41693-018-0010-1.
- [20] 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. *Autom Constr.*, 155: 105071, 2023. doi: 10.1016/j.autcon.2023.105071.
- [21] K. S. Saidi, T. Bock, and C. Georgoulas. Robotics in Construction. *Springer Handbooks*, 1493–1520, 2016. doi: 10.1007/978-3-319-32552-1_57.
- [22] M. Luckcuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher. Formal Specification and Verification of Autonomous Robotic Systems. *ACM Computing Surveys*, 52(5), 2019. doi: 10.1145/3342355.
- [23] MATLAB, Robot Programming MATLAB & Simulink. On-line: https://www.mathworks.com/discovery/robotprogramming.html, Accessed: 09/11/2023.
- [24] B. Xiao, C. Chen, and X. Yin. Recent advancements of robotics in construction. *Autom Constr*, 144:104591, 2022. doi: 10.1016/J.AUTCON.2022.104591.
- [25] N. Pereira da Silva and S. Eloy. Robotic Construction: Robotic Fabrication Experiments for the Building Construction Industry. In Sustainability and Autom in Smart const, 2021.
- [26] J. Nubert, S. Khattak, and M. Hutter. Graph-based Multi-sensor Fusion for Consistent Localization of Autonomous Construction Robots. In *Proceedings* of *IEEE Int Conf Robot Autom*, pages 10048–10054, Philadelphia, USA, 2022.
- [27] 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. *Autom Constr*, 155:105071, 2023. doi: 10.1016/J.AUTCON.2023.105071.
- [28] N. Melenbrink, J. Werfel, and A. Menges. On-site autonomous construction robots: Towards unsupervised building. *Autom Constr*, 119:103312, 2020. doi: 10.1016/j.autcon.2020.103312.
- [29] J. Li, M. Barwood, and S. Rahimifard. Robotic disassembly for increased recovery of strategically important materials from electrical vehicles. *Robot Comput Integr Manuf*, 50:203–212, 2018. doi: 10.1016/J.RCIM.2017.09.013.
- [30] L. Ding, W. Jiang, Y. Zhou, C. Zhou, and S. Liu. BIM-based task-level planning for robotic brick assembly through image-based 3D modelling. *Adv Eng Informatics*, 43:100993, 2020. doi: 10.1016/J.AEI.2019.100993.
- [31] S. Kim, M. Peavy, P. C. Huang, and K. Kim. Development of BIM-integrated construction robot

task planning and simulation system. *Autom Constr*, 127:103720, 2021. doi: 10.1016/J.AUTCON.2021.103720.

- [32] 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. *Autom Constr*, 140:104370, 2022. doi: 10.1016/J.AUTCON.2022.104370.
- [33] L. Huang, Z. Zhu, and Z. Zou. To imitate or not to imitate: Boosting reinforcement learning-based construction robotic control for long-horizon tasks using virtual demonstrations. *Autom Constr*, 146:104691, 2023. doi: 10.1016/J.AUTCON.2022.104691.
- [34] D. Liu, J. Kim, and Y. Ham. Multi-user immersive environment for excavator teleoperation in construction. *Autom Constr*, 156:105143, 2023. doi: 10.1016/J.AUTCON.2023.105143.
- [35] Y. Liu, M. Habibnezhad, and H. Jebelli. Braincomputer interface for hands-free teleoperation of construction robots. *Autom Constr*, 123:103523, 2021. doi: 10.1016/J.AUTCON.2020.103523.

Robotic platform for the (semi-) automated assembly of façade panels

Christoph Heuer¹^(b), Aaron Lentes¹^(b), Zhongqian Zhao¹ and Sigrid Brell-Cokcan^{1,2}^(b)

¹Construction Robotics, Campus-Boulevard 79, 52074 Aachen, Germany ²Chair of Individualized Production (IP), RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany

{heuer,lentes,brell-cokcan}@construction-robotics.de; brell-cokcan@ip.rwth-aachen.de

Abstract

48% of total energy consumption in Europe is used in buildings and homes. These are therefore accountable for 35% of Europe's total CO₂ emissions.

To solve this problem in a sustainable way, the building stock would have to be thermally refurbished further. Currently, the annual thermal refurbishment rate is only 1% of the total building stock, but the sector is already at its production limit. One reason for this is the ongoing shortage of skilled labour. Robotic systems can help here. In collaboration with Fundermax GmbH, St. Veit/Glan, Austria, and KUKA AG, Augsburg, a concept was developed for a (semi-) automated assembly platform that relieves workers of the task of assembling heavy façade panels. Systems that have already been developed are dependent on a complicated setup. This makes it very difficult to use these systems on the construction site.

This research project investigates how existing processes can be expanded so that the threshold for using robotic systems on the construction site can be lowered further. To this end, this work analyses existing processes, derives a concept and then evaluates it using simulations.

Keywords –

Façade robot; Robotic refurbishment; Automation

1 Introduction

As a result of the energy transition, the topic of energy-efficient refurbishment of existing buildings is moving further into the focus of public discussion. Looking at the potential savings, the building sector accounts for 35% of Europe's CO_2 requirements. However, only just under 1% of the building stock is currently thermally refurbished each year, although a gradual increase in the refurbishment rate to 4% could save 14.4% of Europe's gas demand by 2025 [1].

However, modernisation is a cost-intensive measure [1]. There are various options for the subsequent thermal

modernisation of a façade [2]. Solutions for the automated installation of ventilated rainscreen facades are being researched as part of the Enable Refurbishment 4.0 research project. Ventilated rainscreen facades are based on a combination of supporting structure, insulation and rain protection/optical covering [3]. The first step is to examine the extent to which the external façade panels, which serve as visual and rain protection [2] can be installed (semi-) automatically on the façade.

The individual façade panels, though, weigh up to 105 kg in some cases (see 2.1). Scaffolding currently has to be erected for the installation of the façade elements. The façade elements are then lifted to their destination by specialised personnel. The panels must be positioned very precisely, as deviations are easily recognisable in the later appearance. Assembly is currently still carried out exclusively by hand.

Working on the façade in particular involves a high risk for construction site personnel. Risks include: Danger of falling from great heights, increased risk of injury due to lifting heavy loads, changing weather conditions. Automation can minimise these risks [4] and counteract the ongoing shortage of skilled workers [5].

Automated activities in the area of façades include building maintenance (see 2.2). They can perform valuable work, particularly when cleaning large façades. Developed systems are guided to the desired position by a cable system, among other things. The system for this, however, requires a complex installation [4].

As part of the *Enable Sanierung 4.0* research project, an alternative approach is being developed with the aim of enabling the (semi-) automated assembly of façade panels. The concept developed utilises a tower crane as a lifting device, unlike existing solutions such as the *Hephaestus CDPR prototype* [6]. The project is based on a patented gyroscope development, which was developed at the *Centre Construction Robotics*, Aachen, Germany [7]. The process is exemplary for a range of activities in the field of façade modernisation and can be applied to other fields. In the first step of this work, existing research work and its limitations are presented and the reader is given an insight into the functional structure of ventilated rainscreen facades. The requirements and the developed concept are then explained and subjected to an initial simulation using *Gazebo*. The results provide an initial assessment of the extent to which the concept can be implemented in a prototype.

2 State of the art

2.1 Conventional montage of the façadesystem



Figure 1. 1) Bracket for fastening the rail 2) Tshaped rail is screwed to the brackets 3) Fastened T-shaped rail 4) Rock wool is inserted into the intermediate areas 5) Visible panel is connected to the T-shaped rail (fastening method: screws, rivets, adhesive)

A variety of methods are available for the modernisation of façades [2]. This paper only considers the ventilated rainscreen facade system. The system can react flexibly to the respective requirements, e.g. insulation thickness [8]. This is a three-stage system consisting of a support layer, insulation layer and weathering cladding [3].

The façade system used in the research project is described below (see Figure 1):

- Aluminium brackets are attached to a structural load-bearing layer (e.g. concrete or stone) at a fixed grid spacing (see Figure 1. point 1)
- The insulation can now be applied to the façade (see Figure 1. point 4). This can be done using a bolt setting tool, for instance
- An aluminium T-shaped rail is attached to the brackets. The rails can be used to level out any unevenness in the façade. The aim is to create a flat

surface for attaching the cover panels [9] (see Figure 1. point 2)

The façade panels are then attached to the T-shaped rails. Common methods of attaching the façade panels to the T-shaped rails are stapling, nailing, screwing and bonding [10]. In order to achieve a clean joint pattern, the components must be installed precisely. The required tolerance is an accuracy of $\pm 20\%$ of the joint width [11]. The weight of the façade panels is heavily dependent on the actual size of the components. Due to the relatively high density of 1450 kg/m3 of the façade panels, a common format of (280cm x 130cm) results in a weight of 105kg (see Figure 1. point 5)

2.2 Existing façade robots

The field of façade robotics covers a wide range of different tasks. These range from inspection tasks to assembly works. In general, the following four functional principles can be derived according to Kepa Iturralde [12].

2.2.1 Cable-guided robots

A cable-guided parallel robot is guided along the façade using actuators. An example of this is the Hephaestus Robot (see Figure 2). The cable robot, developed for the assembly of façade elements, has a high load-bearing capacity. However, it requires a complex set-up, as the developers highlight in [4]. The time required to set up the system is currently the biggest obstacle to market readiness.



Graphic derived from: [6]

Figure 2. 1) Façade structure 2, 5) Tension point assemblies guide the cable system 3) Platform connected by the cable system 4) Cables for lifting the platform [6]

2.2.2 Crane-guided systems

Powerful lifting machines lift the robot to the desired position on the façade [12]. As part of the research project to develop a cable robot for the installation of curtain walls. Tests have shown that the developed cable robot has a positional deviation of 22 mm. However, this varied within the robot's working range. By recording the position with a total station, the error could be further reduced [4].

2.2.3 Cartesian moving systems

An assembly platform is lifted to the desired position using a linear system installed on the façade. Vertical overhead cranes are included in this group.

2.2.4 Other systems

This category includes systems and individual solutions [12]. One example is the wall-climbing robot developed, which can apply a maximum load of 35 kg using an electrically activated adhesive [13].

2.2.5 Conclusion

An overview of the solutions developed for façade robot systems reveals limitations in the handling of heavy panel components. The field of façade robotics covers a wide range of different tasks. These range from inspection tasks to assembly works.

The cable-guided robots in particular require a particularly complex assembly. Restrictions in the accessibility of the platform also complicate its use in existing building modernization [4]. The challenges mentioned above emphasize the need for further research and development in this area in order to develop more efficient and practicable solutions.

3 Robot-Assisted montage approach

3.1 **Process requirements**

The most important aspects resulting from the catalogue of requirements developed are explained in the following chapter.

3.1.1 Lifting devices

The maximum weight of the platform is determined by the permissible lifting capacity of the crane. A tower crane (L1-24) from *Liebherr-Components Biberach GmbH, Biberach an der Riβ, Deutschland,* is used to lift the façade plattfom. With a maximum outreach of 25 metres, the crane can lift 950 kg at the top. The load is picked up via a crane hook block. The advantage of the crane is its quick assembly on site and the general availability of finished crane systems [14].

The dynamic forces and the inertia properties of the construction crane generate oscillations in the vertical direction when a load is moved. The oscillation varies depending on the weight, wind, trolley position and acceleration (see 4.1).

3.1.2 Rough and fine positioning

The principles of rough and fine positioning should serve as the basis for the development of the façade robot, which was already developed in previous project at the *Centre Construction Robotics*, Aachen, Germany. According to that the platform should first be navigated to its rough position on the façade. Once this predefined position has been reached, the *KUKA robot (KR 10)* should carry out fine positioning and place the façade panel at its destination. The integrated force-torque sensor can help to find the desired alignment. The selected robot initially limits the panel weight to less than 10 kg.

3.1.3 Assembly process of the panels

For an initial test of the façade robot, a fixed panel format of horizontal: 125cm vertical: 62.5cm was selected. The façade element was then glued to the Tshaped rail. The advantage of gluing as a fastening method is that the panels can be prepared in advance with the adhesive and can be applied directly to the T-shaped rails by the robot without any further intermediate step. The fastening method involves two adhesives. A doublesided adhesive tape from Innotec GmbH & Co. KG is applied in such a way that it is later centred on the Tshaped rails. An even strip of silicone adhesive is applied to each side. The purpose of the adhesive tape is to hold the façade panel in position until the silicone adhesive has cured. The bonding process requires compliance with fixed process boundary conditions such as temperature, humidity and surface cleanliness [8].

With a centre-to-centre distance of the T-shaped rails of 62.5 cm, each façade panel rests on three T-shaped rails. This results in a shear load (0.28 N/mm^2) of 6,300 N and a breaking load (0.28 N/mm^2) of 8,640 N for each façade panel [15]. In addition, the manufacturer of the façade system specifies a permissible horizontal compressive force of 2.0 kN/m², which results in a maximum load limit per façade panel of 1.5 kN [8].

3.2 Concept description



Figure 3. Display of the degrees of freedom of the robotic system. The axes that are greyed out cannot be controlled. The development is based on the patent [7]

The following section explains the overall system concept and how it fulfils its function of façade panel positioning under these constraints. The basic elements of the system and their task as well as a description of the assembly process are then presented.

3.2.1 General parameters



Figure 4. Function diagram of the facade robot. 1) T-profile rail 2) Upper gripper assembly 3) End effector 4) Lower vacuum grippers 5) Frame of the coarse positioning unit 6) KUKA robot (KR 10) 7) Adjustable load fastening point 8) Rough-positioning-unit 9) Fine-positioningunit

When positioning an object in space using a construction site crane alone, all six degrees of freedom are in principle indeterminate with regard to the accuracy required here, as will be explained below. The concept of how these degrees of freedom are controlled is then presented. The crane cable results in two pendulum axes, which results in an approximate linear degree of freedom in the X and Y directions for the object on the crane hook. In addition, high loads result in spring pendulum-like movements in the Z direction (see Figure 4). A construction site crane controls the Z-position of the load basically by the length of the crane rope (see Figure 3), but as described in the tests in (see 4.1), it was shown that this positioning in the Z-direction does not meet the requirements defined by the process. The vibration measurements showed that a sinusoidal carrier frequency is found here, which is overlaid by smaller vibrations that represent a disturbance variable of the positioning. Initial measurements showed that a sinusoidal beam frequency is present here, which is superimposed by smaller oscillations (see 4.1)

There is also a rotational degree of freedom around the Z axis and minor influences due to rotational movements around the X axis and the Y axis, resulting from the unguided pendulum motion of the mass, which are not considered further in the following analysis due to their minor influence on positioning (see Figure 3). To summarise, there are three linear degrees of freedom and one rotational degree of freedom that need to be controlled.

One of the main aspects of this project is to utilise existing structures of the façade in order to reduce the number of degrees of freedom and the complexity of the task while maintaining the required high positioning accuracy. To this end, the T-shaped rails used in this façade system are utilised as linear guides. By means of two sliding grippers (see Figure 4. point 2), the entire rough positioning unit slides along the T-shaped rails in the Z direction, which are already firmly bolted to the wall, while the weight is still almost entirely taken by the crane.

This concept determines all open degrees of freedom with the exception of the translational degree of freedom in the Z direction, which is affected by the sinusoidal oscillation of the crane. This disturbance is detected and compensated for with high positioning accuracy by the KUKA robot (KR 10) and a specially developed end effector.

By determining the degrees of freedom in this way, the weight force and thus the main load of the system continues to be borne by the crane and only lateral forces caused by wind or vibration superimposition are absorbed by the T-shaped rails.

3.2.2 General design

The following section explains the individual assemblies and how they work.

Upper grippers: The main task of this assembly is to secure the rough positioning unit to the T-shaped rails while maintaining the mobility of the entire system with low friction in the Z direction in order to minimise the load caused by the weight of the system on the profile rails. This assembly also ensures lateral guidance of the system on the façade. (Figure 4. Point 2)

Vacuum grippers: This assembly uses vacuum grippers to ensure permanent contact with the already installed elements of the lower row of façades during the installation process, whereby it should be noted that the following façade elements are always installed in the positive Z direction. It is installed in the rough-positioning-unit so that it can move in the Z direction by means of a spindle with a movement thread, thus allowing the crane to oscillate periodically without losing contact with the wall surface and still remaining controllable (Figure 4. Point 4). By controlling the torque of the stepper motor that drives the linear module, the weight of the gripper is compensated as far as possible so that the force required to move the carriage along the linear guide is minimised.

Adjustable load fastening point: The adjustable load fastening point of the system is actively controllable along the X and Y axes and, in combination with a force sensor at the contact points with the façade and an inclination sensor on the outer frame of the system with the façade in the form of a closed control loop, ensures that any load on the substructure that could occur due to uneven loading or dynamic forces is minimised by moving the adjustable load fastening point so that the entire system is aligned parallel to the façade and the forces in the contact points are minimised. (Figure 4. Point 7).

Robot: The KUKA robot (KR 10) has the task of compensating for the movement of the permanent spring-pendulum-shaped oscillation of the crane and installing the façade elements at the desired location with the required accuracy. (Figure 4. Point 6)

3.2.3 Process description

For the process presented here, it is necessary that the substructure of brackets and aluminium T-shaped rails is already attached to the wall. Furthermore, at this stage of the project, it is necessary to assemble both the bottom row and a lateral vertical row of façade elements in advance, which serves as a reference point for the system shown here and as the first contact point for the two lower vacuum grippers (see Figure 5).



Figure 5. Flowchart of the installation process

The assembly process begins with the movement of the rough positioning unit by the construction site crane to the starting position. The façade robot is guided to the starting position by the site personnel. To ensure the lowest possible load on the substructure, the entire rough positioning unit is aligned using the adjustable load fastening point, as shown in the module description.

The first contact with the façade substructure is established by aligning and closing the upper grippers. The lower vacuum grippers are then activated, thus reaching the starting position for assembly. Next, the KUKA robot (KR 10) starts the assembly process of an element. As soon as this is complete, first one vacuum gripper of the lower assembly is released and moves upwards in the Z direction by means of a linear guide within the rough positioning unit, while the other gripper continues to ensure a fixed connection to the façade and the rough positioning unit is connected to the wall at at least three points at all times in order to minimise process disruptions caused by influences such as wind. When the suction pad has reached its upper end position, it is reactivated and the second suction pad moves upwards in the same way as the first. The entire rough positioning unit is now moved up one element height by the construction crane and the lower vacuum pads can remain attached to the façade thanks to the linear guide, so that the fine positioning unit can continue the assembly process with a new façade element.

4 Simulation

The developed rough-positioning-unit now needs to be tested for functionality. For this purpose, the developed system was implemented in *ROS* and *Gazebo*. *Gazebo* [16] is a powerful physics simulator that can be used in combination with *ROS* (Robot Operation System). *ROS* is a framework in which complex and robust robot behaviours can be implemented [17]. In *ROS*, the individual axes were virtually equipped with force and acceleration sensors. This allows the corresponding data to be recorded in the *Gazebo* simulation for later evaluation and prototype design. It was also possible to include the recorded oscillation data from the preliminary tests in the simulation

4.1 **Recording the oscillation amplitude**



Figure 6. Results of the vibration tests. The vertical movement was recorded via the distance sensor. The frequency was recorded via a video capturing

In order to be able to estimate the oscillation amplitude of the platform for the simulation, a weight of approx. 1000 kg was attached to the construction crane. A distance sensor was attached under the component, which measured the vertical movement of the component over time (see Figure 7 and). It can be seen that the component has a oscillation amplitude of 50 mm and an oscillation frequency of approx. 0.5 Hz.



Figure 7. 1) Component weight \sim 1000 kg 2) Distance sensor, which measures the distance to the floor, to record the oscillation

4.2 Implementation

A URDF file (Universal Robot Description Format) was then generated from the CAD model. The properties of the components and joints can be precisely defined. In addition, the required sensors for recording the process forces can already be implemented [18]. The joints were implemented in ROS with the help of ros_control.

Gazebo also enables the implementation of vacuum grippers [19]. The two surfaces of the suction pads could be provided with the corresponding function in the *URDF* file.

For the simulation, the ventilated rainscreen facades also had to be integrated into the simulation environment. For this purpose, the original dimensions of the T-shaped rails and façade panels used in the project were modelled. The façade panels were installed according to the specified dimensions. The façade system was successfully integrated into the simulation environment.



Figure 8. 1) Placement of the vacuum gripper 2) Stabilisation of the platform 3) Simulation of the platform with the determined oscillation amplitude

4.3 Simulation results

The simulation was performed using the internal physics engine of the Gazebo simulation software. The interpretation of these results is therefore purely indicative of the system design specification. The results of the simulation provide an initial assessment of the system. The following issues were identified by the simulation:

Compensation of crane oscillations: A key aspect is the ability of the system to compensate for the oscillations in the Z direction caused by the crane. Both gripping systems were able to keep the rough positioning unit constantly controlled. Due to the linear guide on the vacuum gripper, the forces were not transferred to the gripping system (see Figure 8. 1) Placement of the vacuum gripper 2) Stabilisation of the platform 3) Simulation of the platform with the determined oscillation amplitude. Point 1). As assumed, it can be seen that the torque control can compensate for the static weight force and the remaining forces are absorbed by the facade.

Precise positioning of the vacuum grippers: The precision of the system in positioning the vacuum grippers is a critical quality factor (see Figure 8. 1) Placement of the vacuum gripper 2) Stabilisation of the platform 3) Simulation of the platform with the determined oscillation amplitude. Point 3). The vacuum grippers could be released individually from the wall and moved to the next position accordingly. It is important to precisely coordinate the vacuum activation and movement of the motors so that the grippers do not shear off the façade panel in an uncontrolled movement.

Gripping the T-shaped rails: Gripping the T-shaped rails is a complex challenge. The grippers must be moved precisely to the target position and then closed firmly so that the subsequent movement process along the T-shaped rail runs smoothly. In the simulation, the T-shaped rails were successfully gripped and the rough positioning unit was guided along them.

Balance control using the adjustable load fastening point: The simulation showed that the

adjustable load fastening point was able to successfully shift the centre of mass relative to the attachment point as described in chapter 3.2.2 in order to minimise the load on the gripping systems.

5 Conclusion and future work

The results achieved in this study show promising results. The solution approach of vertical motion decoupling clearly stands out from previously developed systems and offers a simplification of existing cable robots. In further steps, the results of the simulation can be analysed in greater depth in a future publication of the project results. This becomes relevant when the right components are assembled for the first prototype.

Particularly in the field of building envelope renovation, significant growth potential is forecast for the coming years [20], and the (semi-) automated solutions developed could represent an important addition to the construction industry, which is still heavily characterised by manual labour [21].

The next steps in this project will be

- Incorporating the simulation results into the prototype
- Simplification of the gripper kinematics
- Coordination of the individual subsystems of the platform, because only if the systems such as the vacuum gripper and actuators work in coordination with each other can it be guaranteed that the platform moves safely on the façade
- Construction of a first prototype

In addition to the development of the rough positioning, the development of the fine positioning is crucial. As described under (see chapter 3.2.3), the Kuka robot should detect and compensate for vertical movements and positioning inaccuracies with the help of forcetorque sensors. Accordingly, an end effector must be developed for the robot, which grips the façade panels and brings them safely to the façade while working in coordination with the robot controller.

Acknowledgements



The research project *Enable Sanierung 4.0* is funded by the state of NRW in the funding programme *Digitalisation of the construction industry and innovative construction*.

References

- [1] S. M. Behr, M. Küçük, and K. Neuhoff. Energetische Modernisierung von Gebäuden sollte durch Mindeststandards und verbindliche Sanierungsziele beschleunigt werden. Online: https://www.diw.de/documents/publikationen/73/ diw_01.c.868217.de/diw_aktuell_87.pdf, Accessed: 14/12/2023.
- [2] H. M. Danaci and N. Akin. Thermal insulation materials in architecture: a comparative test study with aerogel and rock wool. Environmental science and pollution research international, vol. 29, no. 48, pp. 72979–72990, 2022, doi: 10.1007/s11356-022-20927-2, 2022.
- [3] BauNetz. Vorgehängte hinterlüftete Fassaden. Online: https://www.baunetzwissen.de/fassade/ fachwissen/fassadenarten/vorgehaengtehinterlueftete-fassaden-vhf-2341005, Accessed: 14/12/2023.
- [4] K. Iturralde et al.. Cable-driven parallel robot for curtain wall module installation. Automation in Construction, vol. 138, p. 104235, 2022, doi: 10.1016/j.autcon.2022.104235, 2022.
- [5] L. Malin and R. Köppen. KOFA Kompakt Fachkräftemangel und Ausbildung im Handwerk. Online: https://www.kofa.de/daten-und-fakten/ studien/fachkraeftemangel-und-ausbildung-imhandwerk-2023/, Accessed: 05/12/2023.
- [6] K. Iturralde et al. A Cable Driven Parallel Robot with a Modular End Effector for the Installation of Curtain Wall Modules. Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), Kitakyushu, Japan, 2020.
- [7] D. Haarhoff, S. C. Stumm, and S. Brell-Çokcan. Entwicklung einer Control-Moment-Gyroscope(CMG)-basierten Plattform. WO2020208263A1.
- [8] Fundermax. PR.102_Exterior Technik_Broschüre_DE. Online: https:// www.fundermax.com/de/Service/Downloads? category=52463, Accessed: 14/12/2023.
- [9] Hilti AG. Vorgehängte hinterlüftete Fassaden. Online: https://www.hilti.de/content/hilti/E3/DE/ de/business/trades/metal-fabrication/lightventilated-facades.html, Accessed: 14/12/2023.

- [10] Fundermax. Befestigungsempfehlung Funderplan. Online: https://www.fundermax.com/de/Service/ Downloads, Accessed: 06/12/2023.
- [11] Fachverband Baustoffe und Bauteile. VHF Planung und Ausführung von Vorgehangenen Hinterlüfteten Fassaden (VHF). 2017.
- [12] K. I. Lerchundi. Study on Automated and Robotic Renovation of Building Façades with Prefabricated Modules. Online: https://mediatum.ub.tum.de/doc/ 1600234/1600234.pdf, Accessed: 14/12/2023.
- [13] D. Graham-Rowe. Building a Better Wall Climber: A new kind of robot can cling to walls and relax its grip. Online: https://www.technologyreview.com/ 2008/05/29/220398/building-a-better-wallclimber/, Accessed: 14/12/2023.
- [14] Liebherr. L1-24 data sheet. Online: https:// www.liebherr.com/de/deu/produkte/ baumaschinen/turmdrehkrane/schnelleinsatzkrane/ l1-krane/details/70429.html #!/content=table_module_downloads_1, Accessed: 14/12/2023.
- [15] Innotec. data sheet Fixation Tape 1. Online: https:// www.innotec.at/produkte/innotec-produkte/ kleben-sichern/fixation-tape.html, Accessed: 14/12/2023.
- [16] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In Proceedings of 2004 IEEE/RSJ Internationel Conference on Intelligent Robots and Systems, pages 2149–2154, Sendai, Japan, 2004.
- [17] M. Quigley et al. ROS: an open-source Robot Operating System. ICRA Workshop on Open Source Software, vol. 3, 2009.
- [18] Tesseract, URDF. Online: https://tesseractdocs.readthedocs.io/en/latest/_source/intro/urdf_ doc.html, Accessed: 14/12/2023.
- [19] ROS. Class GazeboRosVacuumGripper. Online: https://docs.ros.org/en/ros2_packages/rolling/api/ gazebo_plugins/generated/classgazebo_plugins_ 1_1GazeboRosVacuumGripper.html, Accessed: 14/12/2023.
- [20] J. Anders, A. Dahms, R. Fleischhauer, and T. Hinz, Innovationen im Handwerk. Online: https:// www.handwerk-magazin.de/produkt/studieinnovationen-im-handwerk-226921/, Accessed: 14/12/2023.
- [21] BauPortal. Digitalisierungspotenziale in der Baubranche. Online: https://bauportal.bgbau.de/ bauportal-12020/thema/meldungen/zukunft-desbauens/studie-zeigt-in-der-baubranche-liegennoch-grosse-digitalisierungspotenziale-brach, Accessed: 14/12/2023.

Automating the Tower Crane: Integrating the Development and Simulation of Path Planning and Trajectory Tracking of Tower Crane in ROS Framework

Muhammad Muddassir¹, Mohamed A. A. Abdelkareem², Tarek Zayed¹ and Zoubeir Lafhaj³

¹Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong ²Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong ³Laboratoire de Mécanique Multiphysique Multiéchelle, LaMcube, UMR 9013, Centrale Lille, CNRS, Université de Lille, F-59000 Lille, France

> muhammad.muddassir@polyu.edu.hk, mohamed.a.ali@mu.edu.eg, tarek.zayed@polyu.edu.hk, zoubeir.lafhaj@centralelille.fr

Abstract -

In modern construction sites, tower cranes play a crucial role but often rely on multiple human operators. Despite the advancements of the Construction 4.0 era, a comprehensive framework for automated tower crane operations is currently lacking. This study proposes a framework that integrates a real-scaled construction site and tower crane into a physics-based simulation in ROS (robot operating system) framework to enable collision-free motion planning and control. Specifically, we develop time-varying linear quadratic regulators (LQR) for trolley and jib control while employing a proportion-integrated-derivative (PID) method for hoisting control. Additionally, we utilise 5th-order quintic spline trajectories to plan the desired pose of the payload, reducing acceleration discontinuities. The framework's effectiveness is validated through simulations of a real-scaled tower crane and construction site equipped with LiDAR sensors. The results demonstrate that higher-order trajectories effectively minimise oscillations in unactuated systems. Our scalable framework holds promise for real-scale operations in the field of tower crane automation.

Keywords -

Tower Cranes; Automation in Construction Sites; Timevarying LQR; Robot Operating System (ROS); Underactuated Systems

1 Introduction

Tower cranes (TCs) are indispensable in modern construction sites, serving vital purposes such as lifting and handling heavy construction materials, precise placement of concrete, installation of essential equipment and machinery, and efficient site logistics during construction processes. In addition, TCs are valuable in various highhazard environments, such as shipyards, nuclear plants, and factory floors, where they facilitate the safe transportation of high-risk materials. Efficient operation of TCs in construction and logistics sites minimises operation time per load and ensures high safety standards for workers and operators. In this sense, TCs automation and control can enhance efficiency, improve safety, and optimise productivity in construction and logistic sites where multiple TCs are also considered. Automating the TC's operations can minimise human error, precise load control, real-time monitoring, streamline operations, and reduce risks.

In practice, most industrial TCs are manually controlled and still rely on various human operators (TC operator, signalperson and lifting supervisor) despite the availability of advanced sensing and monitoring systems. The onsite working environment poses many uncertainties and variations, challenging efficient TC operations. Failure in TC operation control can lead to material damage and pose risks to on-site workers, with an uncertain amount of time waste and operation cost [1, 2]. Generally, gripping the load, lifting, transporting from one location to another, lowering, and releasing are the main TC operations [3]. The most time-consuming operation is load transport between two locations, and it requires a skilled operator to minimise the swings and collision of the load during transportation. Automation of TC operations has been a research interest in the construction industry to improve efficiency, mitigate risks, and decrease dependence on human operators.

Numerous studies have delved into various aspects of automated tower crane control [4, 5, 6, 7, 8, 9]. An adaptive control method for tower cranes was developed by [10, 11, 12, 13] to improve outdoor efficiency by addressing parameter uncertainties and disturbances. The robust control methods for controlling jib, trolley, and hoisting of TC were developed by [8, 14]. Similarly, the energy-based methods to develop a regulator controller for achieving the desired state while minimising the swings in hanging payload can be found in [13, 15, 16] In a separate study, He et al. [17] proposed an anti-collision system for tower cranes that considers the lifting of heavy objects, specifically focusing on the dangerous area concept of cylindrical heavy objects to simulate real working situations. For monitoring and planning the collision-free lifting operations, [18, 19, 20, 21, 22] proposed various path planning methods using CAD and BIM information, specifically focusing on time efficient and obstacle avoidance while transporting or lifting from blind spot of crane operators.

The application of computer-aided design and modelling in tower crane operations has facilitated convenient modelling, simulation, and control via physics-based simulations. This approach enhances tower cranes' understanding and motion analysis in full-scale construction sites. Kang and Miranda [23] developed a physicsbased dynamic model of TCs to improve the understanding of crane dynamics and develop computer-aided training methods for crane operators to reduce accidents and enhance overall safety.

In this study, a physics-based simulation module for a tower crane situated in a full-scale construction site was developed utilising the ROS framework. Subsequently, a time-variant Linear Quadratic Regulator (LQR) control system was proposed to govern the trolley and jib movements, while a Proportional-Integral-Derivative (PID) method was employed for controlling the hoisting action. 5th-order quintic spline trajectories were utilised to facilitate planning the payload's desired pose and reduce acceleration discontinuities. While previous studies have explored various aspects of tower crane automation, our work presents a comprehensive framework that addresses collision-free motion planning, trajectory planning, and control algorithms.

2 Why ROS?

Robot Operating System (ROS) is a popular opensource framework for developing robotic systems. It provides a collection of libraries and tools to help developers create complex robotic systems. ROS has a wide range of capabilities, including:

- Hardware Abstraction: ROS provides a hardware abstraction layer that allows one to write code that can run on different types of robots with minimum modifications.
- Message Parsing: ROS provides a message-parsing system that allows different parts of a robot system to communicate. This makes it easy to develop modular systems where various components can be developed independently. This capability is scalable to multi-agent configurations, where each agent can communicate with the master or other agent.

- Simulation: It provides a simulation environment called Gazebo [24] that allows one to test their code in a virtual environment before deploying it on a real robot. Gazebo can simulate the dynamics of a robot, including the effects of gravity, friction, and other physical forces like wind and magnetic field. Gazebo supports various sensors, including cameras, lidars, sonars, and IMU. This makes it possible to integrate sensor data into the control algorithms of a robot for offline development, testing, and validation.
- Scalability: ROS is highly scalable, which means it can be used to develop robotic systems of different sizes and configurations. This makes it a versatile platform for developing automated systems for various applications on a small scale and then porting it to a more extensive system.
- Flexibility: ROS is a flexible platform that allows developers to create custom modules and libraries that can be integrated into similar robotic systems. This makes developing complex robotic systems that can perform various tasks uncomplicated.

Despite its many capabilities, tower crane integration in the ROS framework has some challenges. For example, the tower crane system's complexity, size, and scale differ from most systems for what ROS tools are available. Furthermore, the dynamics of the tower crane contain flexible cable, unactuated swing in radial and tangential directions, and a complex pulley mechanism. Despite this, the integration of tower cranes in ROS, presented here, shows the potential of robotising the tower crane system just like other robotic systems. Finally, ensuring the safety of the crane and its surroundings is of utmost importance, and this requires careful consideration, evaluation and validation of the control algorithms and sensor configurations used in the system.

3 Dynamics Model of a Tower Crane

A tower crane (more specifically, a hammerhead TC) is a five-degree-of-freedom (5-DOF) nonlinear dynamic system, as shown in Figure 1. 3-DOF are actuated, namely jib, trolley and hoisting cable. The generalised coordinates of the jib, trolley and hoisting cable will be γ , x and l, respectively. Whereas 2 of them are unactuated: radial and tangential swings. ϕ and θ denote generalised coordinates of radial and tangential swings, respectively. The jib is the rotating arm housed over a tower (known as a mast) and powered by an electric motor with a gear mechanism. The trolley is housed on the jib and can travel in both inward or outward radial directions. A cable-driven mechanism powered by the second electric motor pushes or pulls the

trolley. A pulley system powered by the third electric motor enables a tower crane for lifting/hoisting motion.

In this paper, we have adopted the dynamical system reported by [25]. They used the Lagrangian approach to study the dynamics of a tower crane. The position vector of trolley position and payload position is given by:

$$\mathbf{x}_{P} = [x - l\cos(\theta)\sin(\phi), \ l\sin(\theta), \ -l\cos(\theta)\cos(\phi)]^{T}$$
(1)

$$\mathbf{x}_T = [x, \ 0, \ 0]^T \tag{2}$$

 \mathbf{x}_P is the position vector of the payload and \mathbf{x}_T the position vector of the trolley. The linear velocity of the trolley and payload can be calculated as:

$$\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} + \dot{\gamma} \times \mathbf{x} \tag{3}$$

Now, the kinetic and potential energies can be written as:

$$T = \frac{1}{2}m\langle \dot{\mathbf{x}}_P, \dot{\mathbf{x}}_P \rangle + \frac{1}{2}M\langle \dot{\mathbf{x}}_T, \dot{\mathbf{x}}_T \rangle + \frac{1}{2}J_o\dot{\gamma}^2 \qquad (4)$$

$$V = -mgl\cos(\theta)\cos(\phi) \tag{5}$$

Here, *m* and *M* are the mass of the payload and trolley, respectively, whereas *g* is the gravitational constant. The generalised displacement vector is $\mathbf{q} = [x, \phi, \gamma, \theta, l]^T$ and generalised forces correspond to \mathbf{q} is $\mathbf{F} = [F_x, 0, T_\gamma, 0, F_l]^T$.

Lagrangian is the difference of kinetic and potential energies $\mathcal{L} = T - V$ and the Lagrangian equation will be:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right) - \frac{\mathcal{L}}{\dot{q}_j} = F_j, \quad j = 1, 2, 3, 4, 5 \tag{6}$$

This will provide the equation of motion (EOM) similar to the one reported in [25]. The derived equations are complex and highly coupled. To devise the controllers from these equations, we have linearised the equation of motion and divided the system into two subsystems for designing two LQR controllers. During the linearization of the system, we assumed small swing angles, considered the parameters of each subsystem as time-varying and ignored the derivative of coupled variables in each subsystem, resulting in 2 simplified subsystems. The controller \mathbf{K}_T to position the trolley and minimise the radial swing angle will be derived from Subsystem 1 (S_1). Where S_1 is defined as:

$$\ddot{x} + m_t \phi = \bar{F}_x \tag{7}$$

$$l\ddot{\phi} + g\phi - \ddot{x} = 0 \tag{8}$$

Here, $m_t = m/M$ and $\bar{F}_x = F_x/M$. The controller \mathbf{K}_{γ} to rotate the jib and minimise the tangential swing angle will be derived from Subsystem 2 (S_2), which is defined as:

$$(1 + M_r x^2)\ddot{\gamma} - m_r g x \theta = \bar{T}_{\gamma} \tag{9}$$

$$l\ddot{\theta} + g\theta - x\ddot{\gamma} = 0 \tag{10}$$

Here, $m_r = m/J_o$, $M_r = M/J_o$ and $\bar{T}_{\gamma} = T_{\gamma}/J_o$. The subsystems S_1 and S_2 will be utilised in the next section to derive the time-varying LQR controller for the trolley and jib positioning.



Figure 1. Dynamics of a tower crane with the 5 degrees of freedom.

4 Tower Crane Control Methods

4.1 Problem Statement

Here, we aim to design a control strategy that can bring the tower crane system to the desired states χ_d , Γ_d and \mathbf{L}_d in finite time whilst respecting the hardware limitations such as F_x , T_γ and F_l . Two LQR controllers, namely K_χ and K_γ and a PID controller, K_l are designed to bring the states χ , Γ and \mathbf{L} to χ_d , Γ_d and \mathbf{L}_d , respectively.

4.2 Controller Design

Figure 2 shows the architecture of the proposed control method. The real-scaled tower crane in the physics simulation (Gazebo [24] acts as a plant and provides all states of the system. Three controllers K_{Γ} , K_{χ} , and K_{L} are input with current states and the desired states, and they output forces F_x , F_l and torque T_{γ} required to apply on the trolley, payload, and jib for regulating the desired states. All outputs from the controller pass from the saturation function to ensure input to the tower crane system couldn't exceed the physical limits of the mechanical and electrical systems. In the proposed control method, the gain matrices for slew and trolley control are recomputed in each control cycle as the parameter of the subsystems S_1 and S_2 are varying with respect to the time. This kind of LQR implementation is also known as time-varying LQR.

4.3 Trolley Controller

A full-state feedback LQR controller is designed to control the trolley motion of the tower crane using the S_2 . The state space model of the S_2 can be written as, $\dot{\chi} = \mathbf{A}_{\chi\chi} + \mathbf{B}_{\chi} F_x$ and the state vector will be



Figure 2. Flow diagram of the control methods.

 $\chi = [x, \dot{x}, \phi, \dot{\phi}]$. Where:

$$\mathbf{A}_{\chi} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -m_r g & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{g}{l(m_r+1)} & 0 \end{pmatrix} , \qquad (11)$$
$$\mathbf{B}_{\chi} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1/l \end{pmatrix} \qquad (12)$$

The input force on the trolley F_x will be computed as:

$$F_x = -G_\chi(\chi - \chi_d) \tag{13}$$

Here, G_{χ} is the gain matrix and χ_d the desired states. G_{χ} is computed using the LQR method, where an optimisation function is minimised to compute the optimal gain matrix G_{χ} . The optimisation function for the LQR method is defined as:

$$\min \int \left(\chi^T \mathbf{Q}_{\chi\chi} + \mathbf{R}_{\chi} F_{\chi}^2 \right)$$
(14)

Where \mathbf{Q}_{χ} is a weight matrix to penalise the system states in order to tune the control performance and \mathbf{R}_{χ} the weight for input.

4.4 Slew Controller

For deriving the slew motion control, the full-state feedback system will be $\dot{\Gamma} = \mathbf{A}_{\Gamma}\Gamma + \mathbf{B}_{\Gamma}F_{\gamma}$ and the state vector will be $\Gamma = [\gamma, \dot{\gamma}, \theta, \dot{\theta}]$. \mathbf{A}_{Γ} and \mathbf{B}_{Γ} are:

$$\mathbf{A}_{\Gamma} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{m_r g x}{1 + M_r x^2} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{g}{l} \left(\frac{1 + x^2 m_r}{1 + x^2 M_r} \right) & 0 \end{pmatrix} \quad , \tag{15}$$

$$\mathbf{B}_{\Gamma} = \begin{pmatrix} 0\\ \frac{1}{1+M_{r}x^{2}}\\ 0\\ -\frac{x}{L(1+M_{r}x^{2}} \end{pmatrix}$$
(16)

Where the slew controller generates the input torque T_{γ} for the mast and jib joint. The control signal to regulate the slew motion, the T_{γ} is:

$$T_{\gamma} = -G_{\Gamma}(\Gamma - \Gamma_d) \tag{17}$$

Here Γ_d is the desired state for the jib and tangential swings. The optimal gain matrix G_{Γ} for the slew controller is computed using the LQR method, as computed for G_{χ} .

4.5 Hoisting Controller

The hoisting controller is responsible for regulating the height of the payload during the transporting or lifting operation. The hoisting motion is not directly responsible for inducing motion in unactuated joints (although variation in cable length can change the natural frequency of hanging load. But this coupling is not considered here). A standard PID controller is implemented to regulate the height of the payload.

$$F_l = G_p e_l + G_i \int_0^t e_l d\tau + G_d \frac{de_l}{dt}$$
(18)

Here, the G_p , G_i , G_d are the proportional, integral and derivative gain of the hoisting controller. These values were manually tuned to obtain optimal performance. e_l denotes the error between the current and desired length of the cable $e_l = l - l_d$

5 Trajectory Generation

The MoveIt provides the collision-free path from the current state of the TC to the desired state. It ensures that the planned path avoids the surrounding obstructions and minimises the mechanical effort of a complete path. When followed, the planned path is a function of an interpolated set of joint positions, ensuring collision-free plan execution. When each set of joint positions in the planned path is assigned a timestamp, it is called trajectory, and the controllers are required to bring the state of the system to the planned state at a particular timestamp. Here, the trajectory is obtained from a quintic spline, where each segment between interpolated path points is a polynomial function of order 5. This means that the trajectory is differentiable till the fifth order. Thus, the resulting velocities and accelerations of the trajectory will be a continuous function which leads to smoother transition of payload from stationary state to moving state and vice versa, as shown in Figure 5(g) and 5(h), and Figure 6(g) and 6(h).

6 Results

We have implemented the proposed controller on a realscaled version of a tower crane in a physics-based simulation, namely, Gazebo [24]. Gazebo is a highly integrated



Figure 3. Flowchart of various modules used in the methodology.

simulation environment with the ROS framework and supports the ROS's message parsing mechanisms. This enables the development and validation of robot controllers and motion planning on a near-to-real system before porting the software to the robots in the real world. Figure 4 shows the planning screen for the physics-based simulated tower crane. Figure 4(a) illustrates the planning scene of MoveIt [26]. The tower crane in orange is the desired state, whereas grey shows the current state of the tower crane. The coloured voxel displays the occupancy grid, which is a way to notify the path planner about obstructions in the space during the planning paths.

Figure 4(b) shows the real-scaled tower crane in Gazebo, which can incorporate all the inertial and collision properties of a real tower crane, as well as the friction and damping values between each joint. It was assumed that the cable was non-stretchable and that the tower crane structure was non-deformable (or rigid) during motion. The hanging payload was free to oscillate in the radial and tangential direction of the jib. This was achieved by assuming that a tower crane is a 5-DOF robot (5 joints) whereas 3 were the actuated joints and two the unactuated joints.

The proposed framework and controller were validated in two scenarios: transporting and hoisting the payload from a blind spot (the part of a site which is not visible to a crane operator). In Gazebo, a construction site was designed with four under-construction buildings, an array of LiDAR sensors and a tower crane with payload. 3 LiDAR sensors were placed around the tower crane mast and 1 in front t of the construction site. Then, their point cloud data was fused to acquire the occupancy grid, as shown in Figure 4(a). MoveIt! (motion planner) can consider the occupancy grid as the obstruction in the space and plan the path accordingly in order to avoid any collision between the payload, tower crane and construction site during motion.



Figure 4. (a) Planning scene in RViz. (b) Simulated World in Gazebo

6.1 Transporting the Payload from Blindspot

Figure 5 shows the jib, trolley and hook trajectory tracking with input forces, input torque and swing angles. MoveIt computed the collision-free path, and then the trajectory was estimated using a quintic spline (5th order spline). Quintic spline is differentiable till the fifth order, thus minimising discontinuities in the derivative of acceleration (jerk). This induces lesser oscillation in hook and load at the start of motion. Figure 5(a) and 5(c) show that the trolley and jib (blue line) follows the trajectory (orange line) while inducing smaller radial and tangential swings, as shown in Figure 5(g) and 5(h). Figure 5(h) shows the radial swing during trajectory tracking and after 50s (see Figure 5(a)) when the desired trajectory of the trolley requires higher velocities, resulting in higher radial swings. However, the controller reduced the swings while bringing the trolley to the desired state.

6.2 Hoisting the Payload from Blindspot

Figure 6 shows the jib, trolley and hook tracking of trajectory with input forces, torque and swing angles during the second scenario. MoveIt computed the collision-free path; then, the trajectory was estimated using a quintic spline (5th-order spline). Figure 6(a), (c) and (e) show that the trolley, jib and hook (blue line) follow the trajectory (orange line) while inducing smaller radial and tangential swings, as shown in Figure 6(g) and (h). In this scenario, the desired distance of the trolley was relatively shorter than the jib and hook; thus, minute radial swings can be observed from Figure 6(g). The jib's desired state and current state are relatively closer in this scenario than the previous one. Thus, the jib controller inputs the highest allowable torque to the jib at the start and end of the planned motion in order to efficiently track the planned trajectory. Still, the tangential oscillation was considerably lower. As the planned trajectory is differentiable to the fifth order, the accelerations were continuous, resulting in smooth transitions of states from rest to motion and vice versa. The video demonstration of the proposed methods can be


Figure 5. Transporting Trajectory Tracking. In graphs (a), (c) and (e), the dotted red line shows the planned (desired) trajectory and the blue lines in all graphs show the current state of the TC and input quantities. ε_x , ε_y and ε_l are the difference between the desired and current state of trolly, jib and hook.

on https://github.com/muddassir93/ISARC2024_ Demo/raw/main/Demo.mp4.

7 Discussion

This study aimed to assess the potential of implementing the framework for robotic applications in the context of tower cranes (TCs) and construction sites. While no direct comparison was made with other TC automation techniques, the research emphasises leveraging ROS's path/trajectory planning algorithms and programming frameworks for TC automation. The successful performance in path planning and trajectory execution showcased in the video supports the viability of our proposed framework.

The assumption of non-stretchable cable and rigid tower cranes is ideal, but these elements could experience stretch or deformation. This assumption simplifies the mathematical modelling and analysis of the system. The proposed



Figure 6. Hoisting Trajectory Tracking. In graphs (a), (c) and (e), the dotted red line shows the planned (desired) trajectory and the blue lines in all graphs show the current state of the TC and input quantities. ε_x , ε_y and ε_l are the difference between the desired and current state of trolly, jib and hook.

framework remains valid for several reasons. The cable stretch and tower crane deformation are relatively smaller compared to the scale of the TC structure and can often be neglected. Control strategies and algorithms are robust and adaptive, enabling them to compensate for these minor deviations. Future research will incorporate cable stretch and tower crane deformation into the model. This can be achieved through advanced mathematical models or sensor-based feedback control strategies. This will enhance the accuracy and reliability of the proposed framework.

Gazebo does not directly support flexible or non-rigid structures. The flexibility of TC's structure and rope can be represented as a link of multiple prismatic (linear) and revolute joints, allowing for near-to-real dynamic behaviour but significantly increasing the computation load on physics simulation. Future research will focus on quantifying or estimating the degree of error in both approaches. Regarding payload orientation, using double cables for the hook (reeving) and spreaders for longer payloads naturally dampens twisting oscillations, leading to the assumption that the payload orientation follows the jib's orientation.

This framework is designed to be modular and scalable, allowing for easy customisation and adaptation to different crane models. The control algorithms and strategies can be adjusted based on specific crane characteristics, operational requirements, and dynamic models. However, scaling and customising the system may pose challenges, requiring extensive testing, validation, and additional sensors or hardware components. To tackle these challenges, we propose comprehensive testing and validation processes, compatibility with various sensors and hardware components, and developing advanced control strategies and algorithms to enhance applicability across the construction industry.

While the Robot Operating System (ROS) is widely used, alternative robotics frameworks such as PyRobot, Orca, Yet Another Robot Platform (YARP), Mobile Robot Programming Toolkit (MRPT), Robotics Library (RL), and Dartsim/Dart also offer valuable tools. However, these alternatives often specialise in specific areas or need more active development and community adaptation. For instance, PyRobot is an excellent choice for research or education, while Orca, despite its suitability for embedded systems, has not been developed in recent years. YARP primarily provides robust communication channels among various robot components. MRPT is tailored towards mobile robots, and RL exclusively offers a C++ API. These specialisations and limitations should be considered when selecting a robotics framework for a particular application.

8 Conclusion

This paper presents a framework for planning the collision-free path, jerk-free trajectory and control of the jib, trolley and hoisting mechanism to follow a planned path using sensory input based on ROS and MoveIt. The developed framework is validated on a simulated realscaled tower crane in Gazebo. Unlike several other pathplanning and controlling approaches, we aim to validate our methods on a real-scaled tower crane, which poses different challenges. The developed framework is scaleable to a real system as far as the desired data types are provided to ROS framework, for example, occupancy grid of surroundings and all TC states and their first derivatives in real-time. In the future, we will scale our framework to a scaled-down version of a tower crane while incorporating more realistic parameters like joint frictions, TC body deformation and cable flexibility.

References

- [1] C. Huang, C.K. Wong, and C.M. Tam. Optimization of tower crane and material supply locations in a high-rise building site by mixed-integer linear programming. *Automation in Construction*, 20(5):571–580, August 2011. doi:10.1016/j.autcon.2010.11.023.
- [2] C. M. Tam and Thomas K. L. Tong. GA-ANN model for optimizing the locations of tower crane and supply points for high-rise public housing construction. *Construction Management and Economics*, 21(3):257–266, April 2003. doi:10.1080/0144619032000049665.
- Yehiel Rosenfeld and Aviad Shapira. Automation of existing tower cranes: economic and technological feasibility. *Automation in Construction*, 7(4):285– 298, May 1998. doi:10.1016/S0926-5805(98)00049-1.
- [4] Songbo Hu, Yihai Fang, and Yu Bai. Automation and optimization in crane lift planning: A critical review. *Advanced Engineering Informatics*, 49:101346, August 2021. doi:10.1016/j.aei.2021.101346.
- [5] Menghua Zhang and Xingjian Jing. Adaptive Neural Network Tracking Control for Double-Pendulum Tower Crane Systems With Nonideal Inputs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(4):2514–2530, April 2022. doi:10.1109/TSMC.2020.3048722.
- [6] G. Rigatos, M. Abbaszadeh, and J. Pomares. Nonlinear optimal control for the 4-DOF underactuated robotic tower crane. *Autonomous Intelligent Systems*, 2(1):21, August 2022. doi:10.1007/s43684-022-00040-4.
- [7] Šandor Ileš, Jadranko Matuško, and Mircea Lazar. Piece-wise ellipsoidal set-based model predictive control of linear parameter varying systems with application to a tower crane. *Asian Journal of Control*, 23(3):1324–1339, May 2021. doi:10.1002/asjc.2515.
- [8] Lobna T Aboserre and Ayman A El-Badawy. Robust integral sliding mode control of tower cranes. *Journal of Vibration and Control*, 27(9-10):1171–1183, May 2021. doi:10.1177/1077546320938183.
- [9] Zheng Tian, Lili Yu, Huimin Ouyang, and Guangming Zhang. Transportation and swing reduction for double-pendulum tower cranes using partial enhanced-coupling nonlinear controller with initial saturation. *ISA Transactions*, 112:122–136, June 2021. doi:10.1016/j.isatra.2020.11.028.

- [10] Tong Yang, Ning Sun, He Chen, and Yongchun Fang. Neural Network-Based Adaptive Antiswing Control of an Underactuated Ship-Mounted Crane With Roll Motions and Input Dead Zones. *IEEE Transactions on Neural Networks and Learning Systems*, 31(3):901–914, March 2020. doi:10.1109/TNNLS.2019.2910580.
- [11] Pham Van Trieu, Hoang Manh Cuong, Hoang Quoc Dong, Nguyen Huu Tuan, and Le Anh Tuan. Adaptive fractional-order fast terminal sliding mode with fault-tolerant control for underactuated mechanical systems: Application to tower cranes. *Automation in Construction*, 123:103533, March 2021. doi:10.1016/j.autcon.2020.103533.
- [12] Yiming Wu, Ning Sun, He Chen, and Yongchun Fang. Adaptive Output Feedback Control for 5-DOF Varying-Cable-Length Tower Cranes With Cargo Mass Estimation. *IEEE Transactions on Industrial Informatics*, 17(4):2453–2464, April 2021. doi:10.1109/TII.2020.3006179.
- [13] Huimin Ouyang, Zheng Tian, Lili Yu, and Guangming Zhang. Adaptive tracking controller design for double-pendulum tower cranes. *Mechanism and Machine Theory*, 153:103980, November 2020. doi:10.1016/j.mechmachtheory.2020.103980.
- [14] Anh Tuan Le and Soon-Geul Lee. 3D cooperative control of tower cranes using robust adaptive techniques. *Journal of the Franklin Institute*, 354(18):8333–8357, December 2017. doi:10.1016/j.jfranklin.2017.10.026.
- [15] Menghua Zhang, Yongfeng Zhang, Bing Ji, Changhui Ma, and Xingong Cheng. Modeling and energy-based sway reduction control for tower crane systems with double-pendulum and spherical-pendulum effects. *Measurement and Control*, 53(1-2):141–150, January 2020. doi:10.1177/0020294019877492.
- [16] Huai-Tao Shi, Jian-Qi Huang, Xiaotian Bai, Xiang Huang, and Jie Sun. Nonlinear Anti-swing Control of Underactuated Tower Crane Based on Improved Energy Function. *International Journal of Control, Automation and Systems*, 19(12):3967–3982, December 2021. doi:10.1007/s12555-020-0292-1.
- [17] Dongfeng He, Junping He, Zhibin Li, Bo Huang, and Hongliang Cheng. An anti-collision system for tower cranes considering lifting heavy objects. *Journal of Physics: Conference Series*, 2365 (1):012019, November 2022. doi:10.1088/1742-6596/2365/1/012019.

- [18] Malak Al Hattab, Emile Zankoul, and Farook R. Hamzeh. Near-Real-Time Optimization of Overlapping Tower Crane Operations: A Model and Case Study. *Journal of Computing in Civil Engineering*, 31(4):05017001, July 2017. doi:10.1061/(ASCE)CP.1943-5487.0000666.
- [19] Jiahao Tian, Sang Luo, Xinming Wang, Jing Hu, and Jun Yin. Crane Lifting Optimization and Construction Monitoring in Steel Bridge Construction Project Based on BIM and UAV. Advances in Civil Engineering, 2021:1–15, July 2021. doi:10.1155/2021/5512229.
- [20] Souravik Dutta, Yiyu Cai, Lihui Huang, and Jianmin Zheng. Automatic re-planning of lifting paths for robotized tower cranes in dynamic BIM environments. *Automation in Construction*, 110:102998, February 2020. doi:10.1016/j.autcon.2019.102998.
- [21] Wenyu Zhang, Hong Zhang, and Lu Yu. Collaborative Planning for Stacking and Installation of Prefabricated Building Components Regarding Crane-Collision Avoidance. *Journal of Construction Engineering and Management*, 149(6):04023029, June 2023. doi:10.1061/JCEMD4.COENG-12955.
- [22] Leon C. Price, Jingdao Chen, Jisoo Park, and Yong K. Cho. Multisensor-driven real-time crane monitoring system for blind lift operations: Lessons learned from a case study. *Automation in Construction*, 124:103552, April 2021. doi:10.1016/j.autcon.2021.103552.
- [23] Shih-Chung Jessy Kang. Physics based model for simulating the dynamics of tower cranes. 2004. URL https://api.semanticscholar. org/CorpusID:210718033.
- [24] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566), volume 3, pages 2149–2154, Sendai, Japan, 2004. IEEE. ISBN 9780780384637. doi:10.1109/IROS.2004.1389727.
- [25] Hanafy M Omar and Ali H Nayfeh. Gain Scheduling Feedback Control of Tower Cranes with Friction Compensation. *Journal of Vibration and Control*, 10(2):269–289, February 2004. doi:10.1177/1077546304035610.
- [26] Sachin Chitta, Ioan Sucan, and Steve Cousins. Moveit! [ros topics]. *IEEE Robotics & Automation Magazine*, 19(1):18–19, 2012.

Development of Robotics for Building Exterior Inspection: A Literature Review

Tianxi Chen¹, Mi Pan¹, Thomas Linner² and Honghao Zhong¹

¹Department of Civil and Environmental Engineering, University of Macau, Macau, China ²Department of Civil Engineering, OTH Regensburg, Regensburg, Germany <u>chen.tianxi@connect.um.edu.mo</u>, <u>mipan@um.edu.mo</u>, <u>thomas.linner@oth-regensburg.de</u> <u>zhong.honghao@connect.um.edu.mo</u>

Abstract -

The aging of buildings is a global concern, with potential risks to human safety and property. Building inspection and maintenance are crucial for ensuring structural integrity and safety. However, traditional manual methods are time-consuming and pose safety risks, especially for exterior inspection at height. Robotics offer a promising alternative to enhance building inspection efficiency and costeffectiveness, but still in the early development stage. This paper aims to review and analyze the state-ofthe-art design and development of robotics for building exterior inspection, referring to the literature published in the last two decades. Firstly, the review classifies different types of robots for building exterior inspection in terms of locomotion and adhesion modes, and discusses the capability of robots from navigation, obstacle surmounting, wallto-wall/floor transition, curved wall climbing, grasping, barrier avoidance, and self-protection. Secondly, the paper examines the applicability of robots to various building materials for inspections and summarizes the most typical applications (i.e. glass curtain walls, tile walls, and concrete walls). Thirdly, the paper discusses the typical data collection and analysis methods for building exterior inspection using robots. The paper also explored potential enhancements for robotic inspection through the integration of building information modeling, augmented reality/virtual reality, and the involvement of human-in-the-loop. Finally, the paper summarizes the typical application of robotics in building exterior inspection regarding robot types, inspection applications, data collection and analysis methods, discusses the challenges, and outlines the future directions.

Keywords -

Wall inspection; Building exterior inspection; Robotics; Locomotion; Adhesion; Non-destructive testing.

1 Introduction

The phenomenon of building aging presents a significant global concern, as it entails the gradual deterioration of both interior and exterior walls, leading to potentially severe consequences for human safety and property [1]. Compared to interior walls, exterior walls are more exposed to environmental factors and external forces with issues such as deterioration or delamination. Thus, the demand for robotic inspection targeting exterior walls is relatively stronger than for interior ones. Compared to interior wall inspection, the demolition and reconstruction of building exteriors require significant financial and labor resources. Traditional approaches for inspecting building walls typically involve manual visual exams or hammering tests [2]. The inspection tasks are performed by human operators who ride on gondolas suspended in mid-air and utilize various handheld facilities together with telescopes and cameras for the inspection [3]. However, these methods present safety concerns, particularly when working at elevated heights. In recent decades, there has been a growing trend towards the adoption of robots for building inspection. In particular, robotic inspection offers several advantages, including enhanced efficiency and consistency in inspection results and presenting a cost-effective alternative to manual inspection methods [4]. Human operators can remotely operate robots from a secure location and obtain the data and images from the facilities mounted on the robots [5]. However, robotics for building exterior inspection is still an emerging area with limited real-world applications. How different types of robots could be effectively developed and applied to inspect building exterior walls to ensure building safety and functionality remains unexplored.

This paper aims to review and analyze the state-ofthe-art design and development of robotics for building exterior inspection, referring to the literature published in the last two decades. The objectives were to i) explore the different types of robots that are used in exterior wall

inspection of buildings and the specific inspection applications conducted by robots, ii) examine the methods for robots to inspect and detect building exterior defects, and iii) identify challenges and future directions of robotics for exterior wall inspection of buildings.

2 Methodology

The study was conducted in three major stages. Firstly, a comprehensive literature review was conducted to collect information on the utilization of robotics in inspecting exterior walls in buildings, focusing on their specific applications. Web of Science (WoS) and Scopus

databases were utilized to search for relevant papers. The keywords (robot* OR aerial OR drone OR uav OR "unmanned aerial vehicle") AND (wall OR building) AND (inspection OR maintenance) were used. Secondly, data screening and supplementation were conducted on the identified papers. Only relevant publications regarding robots for building exterior wall inspection were kept. Additional articles identified by crossreferencing were supplemented. Thirdly, content analysis was performed to extract relevant information from the included papers regarding the types of robots and their capability, inspection applications, inspection methods, and challenges regarding robots for building exterior inspection.



- (2) Three-module wheeled robot
- ③ Hybrid system with a crawler robot and a microdrone
- (4) Double-chamber wheeled robot
- (6) Multi-legged robot with twist-based crouching
- ⑦ Double-propellers wall inspection robot
- (8) Multi-legged robot with suckers

Figure 1. Exemplars of robots utilized in exterior wall inspection.

Robots for Building Exterior Inspection 3

This section focuses on the classification of robots and their capability in practical scenarios.

3.1 **Classification of Robots**

The classification of robots for building exterior inspection considers two key aspects: locomotion and adhesion mechanisms. Locomotion refers to the capability of the robot to move, while adhesion refers to its adaptability to building surfaces and materials. Figure

1 shows some examples of inspection robots developed in various types.

3.1.1 **Modes of Locomotion**

The locomotion systems allow the efficient and autonomous mobility of robots during the inspection of exterior walls. The selection of suitable locomotion depends on the structural and built environment. The primary types of locomotion for robots applied for building exterior inspection can be classified into wheeled, tracked, legged, cable-driven parallel, aerial, and hybrid.

Wheeled robots, inspired by automobiles, offer fast speed and efficient movement. However, the presence of large gaps and inconsistencies on wall surfaces often prevents proper contact between the wheels and the building surface [6].

Tracked robots, inspired by tanks, utilize continuous treads to achieve basic movement and improve climbing stability by ensuring contact with the wall surface. However, this locomotion method may lead to increased frictional resistance and higher energy consumption [7].

Legged robots, including bipedal (two-feet), quadrupedal (four-feet), and hexapod (six-feet) locomotion systems, offer superior adaptability for wall surface inspection [8]. However, the utilization of multiple legs increases the complexity of locomotion and adhesion force control.

Cable-driven parallel robots, utilizing cables as actuators, offer a mechanism for vertical and inclined movement control in high-rise building wall inspection [9]. However, their reliance on secure anchoring points and complex setup processes distinguishes them from other wall inspection robots.

Aerial robots, such as drones and unmanned aerial vehicles (UAVs), offer straightforward control mechanisms. However, aerial systems encounter complexity in maintaining a consistent distance from walls during inspection, as challenges from external factors like wind and rain arise [10].

Hybrid robots, integrating multiple locomotion mechanisms, offer enhanced capabilities and increased mobility for rapid motion on wall surfaces, obstaclespanning, and seamless mode switching [11]. Ogusu et al. [12] designed a robot system where a tracked robot moves to the front of a concrete wall, while a drone takes off and flies vertically against the wall, utilizing a camera to capture images. However, the integration of diverse locomotion modes and functionalities in hybrid robots often poses challenges to control mechanisms [13].

3.1.2 Adhesion Methods

Another distinguished feature of robots for building exterior inspection lies in the technologies for adhering to building surfaces, mostly applied for climbing-type inspection robots [6]. The adhesion methods can be classified as active or passive, depending on whether an external energy supply is needed to support the robot. A more general classification is based on the nature of adhesion forces required: pneumatic or air pressure, magnetic, biomimetic, and electrostatic [14].

Pneumatic adhesion methods, including vacuum or suction cups and negative pressure, are commonly used for climbing flat surfaces [14]. Vacuum adhesion, employing suction cups and vacuum pumps, offers simplicity and effective wall steps but can be problematic on rough or grooved surfaces [2]. Negative pressure adhesion, utilizing impellers or eddy currents, provides secure attachment but may face potential suction chamber leakage on vertical walls [15].

Magnetic adhesion is commonly used for inspection robots on ferrous surfaces, employing either permanent magnets or electromagnets. It allows for rapid locomotion and eliminates the need for power application but faces challenges in maintaining stable magnetic power on non-smooth surfaces and is limited in suitability for non-steel structures [16].

Biomimetic adhesion methods, inspired by climbing animals like cockroaches and cicadas, utilize robust dry adhesives based on Van der Waals forces [4]. These methods allow robots to attach to various surface materials and maintain attachment without power consumption. However, they have limitations in terms of payload capacity.

Electrostatic adhesion utilizes compliant electrode patterns to generate strong electrostatic forces between the robot and the wall surface, allowing for exceptional adhesion pressure on uncharged surfaces [17]. While this method offers minimal power consumption, and noisefree operation, it may face challenges of unstable adhesion and reduced reliability in the presence of external disturbances like wind and heavy rain.

3.2 Capability of Robots

The capability of robots for building exterior inspection considers seven aspects: navigation, obstacle surmounting, wall-to-wall/floor transition, curved wall climbing, grasping, barrier avoidance, and selfprotection. Most of these abilities are particularly related to climbing-type robots.

3.2.1 Navigation Ability

Navigation ability is critical for robots to perform building exterior inspection, which includes localization and mapping, and path planning.

Localization refers to the estimation of robot position and orientation, while mapping denotes the creation of a digital representation of the robot environment, both are fundamental processes for robot navigation [18]. Robots utilize sensors such as RGB-D cameras, stereo cameras and Light Detection and Ranging (LiDAR) to gather the environment data for localization and mapping [7]. Simultaneous localization and mapping (SLAM) is widely used to enable the robot to create a map of its environment while simultaneously determining its position within the map [19].

Path planning determines the optimal path or trajectory for the robot to navigate from its current location to a desired goal location while avoiding obstacles [20]. Path planning for robots to conduct building exterior inspection often considers factors such as finding local optimal paths for defect detection to minimize the energy consumption of the robot and navigating the robot safely from building exterior and mid-air obstacles [20].

3.2.2 Obstacle Surmounting Ability

The obstacles on the exterior wall mainly include grooves and strips. For climbing robots, having a high ability to overcome obstacles is essential during the design stage of robotic gait planning. For example, Bian et al. [4] developed a robot with a gear transmission system inspired by cicadas and geckos, incorporating a stable gait design for obstacle surmounting.

3.2.3 Wall-to-Wall/Floor Transition Ability

The ability of wall-to-wall/floor transition involves smoothly navigating and changing position between vertical and horizontal surfaces, which is a key feature for mobility in inspection robots. Guan et al. [2] designed a modular biped robot that can smoothly transit between walls by adapting climbing gaits and utilizing suction modules for reliable attachment to different surfaces.

3.2.4 Curved Wall Climbing Ability

The ability of robots to climb curved walls can enhance the efficiency of inspections by enabling them to detect complex geometries and reach areas with limited human accessibility. Saito et al. [8] developed a flexiblelegged robot with a sucker mechanism that allowed it to climb unknown curved walls.

3.2.5 Barrier Avoidance Ability

Autonomous barrier avoidance for robots in complex outdoor scenes should be considered, especially for unknown environments to prevent crashes or falls. For example, Chang et al. [19] developed a robust fuzzy logic controller for building inspection robots, enabling autonomous navigation and barrier avoidance to achieve precise wall-following behavior.

3.2.6 Grasping Ability

The setup of grippers is essential as the interacting grasping media between the robot and the object. The design of grippers should overcome the challenge of grasping exterior walls, considering their shapes, materials, and roughness [7]. Previous researchers have explored various grasping methods. For example, Xu et al. [21] designed a cross-structured gripper for climbing robot consisting of multiple claws and hooks that can slide along the wall surface to locate attachable uplifts.

3.2.7 Self-Protection Ability

Protective devices for robots are crucial to ensure their safety and stability during operation on the building exterior by protecting against overturning during climbing, falls, and external objects. For example, Altaf et al. [14] designed an inspection robot with protective devices, including a support frame, EVA shell, airbag, and established a mechanical model to analyze the forces involved during a fall.

4 Inspection Applications

The applications of robots for building exterior inspection are mostly considered for maintaining glass curtain walls, tile walls, and concrete walls.

4.1 Glass Curtain Wall Inspection

The inspection of glass curtain walls has gained increasing attention, driven by concerns surrounding the safety risks associated with accidental glass breakage and the subsequent hazards of falling glass. Scholars proposed a specific design of end-effectors for glass curtain wall inspections. Liang et al. [6] introduced a robotic system that utilizes vibration response signals obtained by knocking the glass with a mechanical arm.

4.2 Tile Wall Inspection

Robotic techniques also play a crucial role in the inspection of tile wall delamination, which occurs between the tiles and substrates due to aging and improper installation. Pan et al. [23] proposed an efficient method for inspecting the status of tile walls using forward-looking infrared technology and camera on UAVs. The combination of hammering tests and cameras is also commonly employed to identify tile delamination caused by bonding degradation and thin tile layers [9].

4.3 Concrete Wall Inspection

Concrete used for wall surfaces can deteriorate over time, resulting in issues such as peeling, lifting, and cracking. Traditional visual inspection methods for concrete surfaces primarily rely on human workers. Crack detection and inspection in walls benefit from advanced robotic techniques. For example, Wang et al. [24] developed an automatic detection of building surface cracks using UAV, demonstrating high accuracy and potential for practical application.

5 Inspection Methods

This section focuses on the inspection methods in terms of data collection and data analysis for robots to conduct building exterior inspection and defect detection.

5.1 Data Collection

Data collection methods primarily utilize nondestructive testing (NDT) techniques to gather

information without causing damage to the building. Essentially, robots offer the platform to equip different NDT techniques for data collection, which could reduce the impact of human factors and achieve more efficient, accurate, cost-effective, and damage-avoided inspection of buildings. Typical NDT methods for building exteriors are as follows: 1) visual inspection utilizes cameras to examine defects that are widely employed in concrete wall crack detection [25]; 2) infrared thermography uses infrared thermal imagers to detect temperature variations for defect identification, which is commonly employed for wall moisture and air leakage detection [26]; 3) laser scanning emits laser beams to capture three-dimensional (3D) point cloud for defect detection, which suits for 3D crack dimension detection and localization [27]; 4) impact acoustic inspection analyzes sound waves generated by impacting the object for defect evaluation, and is mostly employed for tile wall debonding and integrity detection [9].

5.2 Data Analysis

Based on the collected data, a number of studies focus on the development of new algorithms to process and analyze different types of inspection data, like RGB and thermal images [26, 28, 29], point cloud data [27], and acoustic signals [30].

Image-based analysis has received the most attention. Early research works focus on simple image processing, such as greyscale for removing image color information and histogram equalization for adjusting image pixel value distribution [20]. Recently, deep learning techniques, especially convolutional neural networks (CNNs) have been successfully developed in the field of computer vision, and are widely applied for object detection and image segmentation for building exterior inspection [7]. For example, Hu et al. [7] proposed deep convolutional neural network (DCNN) with reduced parameters and low latency for feature extraction in the crack detection for an inspection robot; Woo et al. [28] utilized the fast and accurate YOLOv5 model for UAVbased crack detection system for concrete wall inspection; Wang et al. [24] proposed a UAV-based approach integrating ResNet50 and YOLOv8 for accurate detection of concrete cracks on building exteriors. Some studies combine traditional machine learning and deep learning to achieve more accurate inspection results. For example, Chaiyasarn et al. [31] proposed a crack detection system that combined support vector machines with CNN, achieving a detection accuracy of approximately 86% in validation images.

Point cloud-based analysis extracts relevant inspection information from 3D data in depicting object geometry for defect detection, which heavily relies on the quality of the point cloud data. Some relevant studies combined 3D point clouds with 2D images and employed image processing techniques for defect detection [7]. For example, Yuan et al. [27] developed a novel inspection robot with deep stereo vision for 3D concrete damage detection and quantification, which employed Mask R-CNN for the captured RGB imaging data for crack segmentation and localization and conducted spatial mapping from 2D plane to 3D space coordinates for damage quantification.

Acoustic signal-based analysis utilizes the acoustic signals generated by tapping the wall surface. Since the collected signals are generally not perfect with background noise and missing signals, different data processing techniques have been used to address the noisy environment and data quality issues. For example, Nishimura et al. [30] proposed a novel method to clarify the features of hammering sound and propeller noise and conducted acoustic analysis of recorded hammering sounds collected by a multi-copter type mobile robot for defect detection.

6 Integration with Other Advanced Technologies

This section discusses the integration of inspection robots with other advanced technologies such as building information modeling (BIM), augmented reality/virtual reality (AR/VR), and human-in-the-loop (HILP).

6.1 BIM

For managing inspection data and planning inspection paths, BIM can be integrated with robots to provide rich geometric and semantic information. Tan et al. [20] proposed BIM-based inspection area extraction, optimized UAV flight path using A star and genetic algorithms, and real-time crack identification through edge computing and a zoom camera.

6.2 AR/VR

AR serves as a valuable tool for visualizing the inspection process, enabling efficient remote inspections by overlaying models onto captured images [32]. The use of AR technology in wall inspection robots allows for virtual walkthroughs and real-time visualization. Nishimura et al. [30] developed a robot that utilizes an AR marker for localization, allowing for precise identification of the hammering position on a structure.

VR can simulate the real environment of buildings, enabling operators to identify issues and take appropriate actions in a virtual setting. It reduces the need for robotics, lowers risks and costs, and provides more practice opportunities. Albeaino et al. [33] developed a VR-based flight simulator for training drone operators in inspecting building exteriors, focusing on flying near targets, maintaining stable hovering, and collecting data.

6.3 HITL

HITL generally refers to interactive simulation systems that enable direct human intervention during the operation of robots or artificial intelligence models [3]. It combines human intelligence with machine or robot capabilities for risk-free and improved system performance, which is preferable for hazardous building exterior inspection tasks. For example, Saleem et al. [34] investigated human eye gaze patterns during a façade damage inspection using eye tracking which should facilitate information-sharing and decision-making for collaborative human-robot teams for building inspection.

Table 2. Summary of typical robots for building exterior inspection.

Example	Robot type	Inspection application	Data collection	Data analysis
A multi-chamber adhesive	Wheeled with	Glass curtain wall	Impact acoustic	-
climbing robot [6]	pneumatic	defects		
Alicia ³ [35]	Wheeled with	Concrete wall defects	-	-
	pneumatic			
CROMSCI [15]	Wheeled with	Concrete wall defects	Visual, cover meter,	-
	pneumatic		impulse radar	
A four-wheel drive robot [27]	Wheeled	Concrete wall crack	Visual, laser scanning	Mask R-CNN, statistical
		assessment		outlier removal filter, KNNs
An autonomous mobile ground	Tracked	Concrete wall crack	Visual, laser scanning	DCNN, random sample
robot [7]	~	defects		consensus
A wall-climbing robot inspired	Legged with	Stone and glass curtain	-	-
by cicadas and geckos [4]	biomimetic	wall defects		
A	adhesion	XX7 1 1.4 ⁻¹ 11		
A semi-autonomous multi-	Legged with	wooden and tile wall	-	-
An alastra adhasiya wall	I accord with	Class surfairs wall		
alimbing robot [17]	electrostatic	defects	-	-
	Cable driven	Tile wall bonding	Impact acoustic	ANN
wieboi [9]	Cable-unven	integrity	impact acoustic	Ainin
A UAV-based system [24]	UAV	Concrete crack defects	Visual	ResNet50, YOLOv8
Quasar [™] [26]	UAV	General moisture, air	Infrared thermography	Gaussian low-pass filter
		leakage	017	I.
A UAV-based system [25]	UAV	General crack defects	Visual	CNN
DJI MJ200 [22]	UAV	Concrete wall crack	Visual	CNN, CycleGAN
		defects		
DJI Phantom4 RTK [28]	UAV	Concrete wall crack	Visual	YOLOv5
		defects		
ABECIS [36]	UAV	General crack defects	Visual	Xception
DJI Phantom4 [31]	UAV	Masonry wall crack	Visual	CNN, support vector machine
		defects		
Microdrone-equipped mobile	UAV+tracked	Concrete wall crack	Visual	AWS crack detection AI
crawler robot [12]	*****	defects	•	system
A multi-copter mobile robot [30]	UAV+wheeled	Concrete wall defects	Impact acoustic	Short-time Fourier transform

7 Discussions

Based on the review and analysis, a summary of typical robots for building exterior inspection is provided in Table 1, covering robot types, inspection applications, and methods for data collection and data analysis. The findings provide useful insights for research and development in robotics for building exterior inspection, and are elaborated as below.

Firstly, prior studies have demonstrated two major streams of research regarding robotics for building exterior inspection. The first stream focused on developing inspection robot prototypes and emphasized the mechanical mechanisms of the robots, but many did not cover specific data collection and analysis methods for inspection [4]. The second stream mainly employed UAV as the robotic platform and focused on specific data collection and analysis methods for inspection [24]. The future direction for improvement could be the integration of the two streams to develop comprehensive robotic solutions for building exterior inspection. These solutions could combine the high flexibility and adaptability of robot mechanics with advanced data collection and analysis methods for building exterior inspection.

Secondly, most of the research focused on the inspection application of concrete crack detection as concrete cracks are typically visible and possess distinct features that can be effectively captured using imaging techniques and analyzed with computer vision algorithms. However, many buildings, especially high-rise buildings, feature a variety of wall materials rather than being limited to a single wall type and still lack comprehensive and robust data collection and analysis methods. Besides, glass curtain walls and tile walls are more challenging to achieve timely and accurate defect detection [6, 9]. More efforts are required to develop adaptable and generic solutions for inspection robots integrating different data collection methods that are not limited to specific types of wall materials or defects.

Thirdly, only limited studies focused on the development of end effectors of the inspection robot to enable the multi-functional capability of the robot for multiple inspection tasks [7]. In this regard, the integration of repair or renovation functions with the inspection robots should be considered to guide the selection of robot body type and the design of different inspection and repair end-effectors as well as control mechanisms.

8 Conclusions

This paper reviews and analyzes the state-of-the-art design and development of robots for building exterior inspection. Specifically, these inspection robots are classified based on locomotion and adhesion modes, and their capabilities are discussed from navigation, obstacle surmount, wall-to-wall/floor transition, curved wall climbing, grasping, reflective and transparent barrier avoidance, and self-protection. The review further summarizes the most typical building exterior inspection applications for robotics considering the wall materials, namely glass curtain, tile, and concrete, and discusses the detailed data collection and data analysis methods for inspection. Furthermore, the paper explored the potential integration of BIM, AR/VR, and HITL for improving robotic building exterior inspection. Blending the review findings, useful insights are provided regarding robot types, inspection applications, data collection and analysis methods for building exterior inspection. Future research is needed toward comprehensive and tailored solutions for robotic building exterior inspection, adaptable and generic inspection robots that suit different wall materials and defect types, and multi-functional end effectors for inspection robots and integration with building repair capability.

Acknowledgement

We acknowledge funding support from University of Macau (File no. SRG2023-00006-FST).

References

- Pan, M., Linner, T., Pan, W., Cheng, H., and Bock, T. Structuring the context for construction robot development through integrated scenario approach. *Automation in construction*, 114, 103174, 2020.
- [2] Guan, Y., Zhu, H., Wu, W., Zhou, X., Jiang, L., Cai,

C., Zhang, L., and Zhang, H. A modular biped wallclimbing robot with high mobility and manipulating function. *IEEE/ASME Transactions on Mechatronics*, 18(6):1787–1798, 2013.

- [3] Zheng, Z. J., Pan, M., and Pan, W. Virtual Prototyping-Based Path Planning of Unmanned Aerial Vehicles for Building Exterior Inspection. In ISARC Proceedings of the International Symposium on Automation and Robotics in Construction, pages 16-23, Kitakyushu, Japan, 2020.
- [4] Bian, S., Xu, F., Wei, Y., and Kong, D. A novel type of wall-climbing robot with a gear transmission system arm and adhere mechanism inspired by Cicada and gecko. *Applied Sciences*, 11(9):4137, 2021.
- [5] Bock, T., and Linner T. Robot-Oriented Design: Design and Management Tools for the Deployment of Automation and Robotics in Construction, pages 156 - 230. Cambridge University Press, Cambridge, 2015.
- [6] Liang, R., Altaf, M., Ahmad, E., Liu, R., and Wang, K. A low-cost, light-weight climbing robot for inspection of class curtains. *International Journal* of Advanced Robotic Systems, 11(7):106, 2014.
- [7] Hu, K., Chen, Z., Kang, H., and Tang, Y. 3D vision technologies for a self-developed structural external crack damage recognition robot. *Automation in Construction*, 159, 105262, 2024.
- [8] Saito, A., Nagayama, K., Ito, K., Oomichi, T., Ashizawa, S., and Matsuno, F. Semi-autonomous multi-legged robot with suckers to climb a wall. *Journal of Robotics and Mechatronics*, 30(1):24-32, 2018.
- [9] Luk, B. L., Liu, K. P., Jiang, Z. D., and Tong, F. Robotic impact-acoustics system for tile-wall bonding integrity inspection. *Mechatronics*, 19(8):1251-1260, 2009.
- [10] González-deSantos, L. M., Martínez-Sánchez, J., González-Jorge, H., and Arias, P. Active UAV payload based on horizontal propellers for contact inspections tasks. *Measurement*, 165, 108106, 2020.
- [11] Sukvichai, K., Maolanon, P., and Songkrasin, K. Design of a double-propellers wall-climbing robot. In 2017 IEEE International Conference on Robotics and Biomimetics, pages 239-245, Macau, Macao, 2017.
- [12] Ogusu, Y., Tomita, K., and Kamimura, A. Microdrone-equipped mobile crawler robot system, DIR-3, for high-step climbing and high-place inspection. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1261-1267, Macau, Macao, 2020.
- [13] Tokura, Y., Toba, K., and Takada, Y. Practical applications of HORNET to inspect walls of structures. *Journal of Robotics and Mechatronics*,

28(3):320-327, 2016.

- [14] Altaf, M., Ahmad, E., Xu, Y., Liu, R., Li, Y., and Na, H. Design of a climbing robot platform with protection device. *International Journal of Advanced Robotic Systems*, 14(4):1–14, 2017.
- [15] Hillenbrand, C., Schmidt, D., and Berns, K. CROMSCI: development of a climbing robot with negative pressure adhesion for inspections. *Industrial Robot: An International Journal*, 35(3):228-237, 2008.
- [16] Phlernjai, M., and Ratsamee, P. Multi-Legged Inspection Robot with Twist-Based Crouching and Fine Adjustment Mechanism. *Journal of Robotics* and Mechatronics, 34(3): 588-598, 2022.
- [17] Chen, R. A gecko-inspired electroadhesive wallclimbing robot. *IEEE Potentials*, 34(2):15-19, 2015.
- [18] Tan, Y., Li, G., Cai, R., Ma, J., and Wang, M. Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. *Automation in Construction*, 139, 104284, 2022.
- [19] Chang, S., Siu, M. F. F., and Li, H. Development of a Fuzzy Logic Controller for Autonomous Navigation of Building Inspection Robots in Unknown Environments. *Journal of Computing in Civil Engineering*, 37(4):04023014, 2023.
- [20] Tan, Y., Yi, W., Chen, P., and Zou, Y. An adaptive crack inspection method for building surface based on BIM, UAV and edge computing. *Automation in Construction*, 157, 105161, 2024.
- [21] Xu, F., Wang, B., Shen, J., Hu, J., and Jiang, G. Design and realization of the claw gripper system of a climbing robot. *Journal of Intelligent & Robotic Systems*, 89:301-317, 2018.
- [22] Munawar, H. S., Ullah, F., Heravi, A., Thaheem, M. J., and Maqsoom, A. Inspecting buildings using drones and computer vision: A machine learning approach to detect cracks and damages. *Drones*, 6(1):5, 2021.
- [23] Pan, N. H., Tsai, C. H., Chen, K. Y., and Sung, J. Enhancement of external wall decoration material for the building in safety inspection method. *Journal of Civil Engineering and Management*, 26(3):216-226, 2020.
- [24] Wang, J., Wang, P., Qu, L., Pei, Z., and Ueda, T. Automatic detection of building surface cracks using UAV and deep learning - combined approach. *Structural Concrete*, 2024.
- [25] Chen, K., Reichard, G., Xu, X., and Akanmu, A. Automated crack segmentation in close-range building façade inspection images using deep learning techniques. *Journal of Building Engineering*, 43, 102913, 2021.
- [26] Gil-Docampo, M., Sanz, J. O., Guerrero, I. C., and Cabanas, M. F. UAS IR-Thermograms Processing

and Photogrammetry of Thermal Images for the Inspection of Building Envelopes. *Applied Sciences*, 13(6):3948, 2023.

- [27] Yuan, C., Xiong, B., Li, X., Sang, X., and Kong, Q. A novel intelligent inspection robot with deep stereo vision for three-dimensional concrete damage detection and quantification. *Structural Health Monitoring*, 21(3):788-802, 2022.
- [28] Woo, H. J., Hong, W. H., Oh, J., and Baek, S. C. Defining Structural Cracks in Exterior Walls of Concrete Buildings Using an Unmanned Aerial Vehicle. *Drones*, 7(3):149, 2023.
- [29] De Filippo, M., Asadiabadi, S., Kuang, J. S., Mishra, D. K., and Sun, H. AI-powered inspections of facades in reinforced concrete buildings. *HKIE Trans*, 30(1):1-14, 2023.
- [30] Nishimura, Y., Takahashi, S., Mochiyama, H., and Yamaguchi, T. Automated hammering inspection system with multi-copter type mobile robot for concrete structures. *IEEE Robotics and Automation Letters*, 7(4):9993-10000, 2022.
- [31] Chaiyasarn, K., Khan, W., Ali, L., Sharma, M., Brackenbury, D., and DeJong, M. Crack detection in masonry structures using convolutional neural networks and support vector machines. In *ISARC Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 1-8, Berlin, Germany, 2018.
- [32] Park, J., Chang, S., Lee, H., and Cho, Y. K. Inspection data exchange and visualization for building maintenance using AR-enabled BIM. In ISARC Proceedings of the International Symposium on Automation and Robotics in Construction, pages 483-490, Bogotá, Colombia, 2022.
- [33] Albeaino, G., Eiris, R., Gheisari, M., and Issa, R. R. DroneSim: A VR-based flight training simulator for drone-mediated building inspections. *Construction Innovation*, 22(4):831-848, 2022.
- [34] Saleem, M. R., Mayne, R., and Napolitano, R. Analysis of gaze patterns during facade inspection to understand inspector sense-making processes. *Scientific reports*, 13(1):2929, 2023.
- [35] Longo, D., and Muscato, G. The Alicia/sup 3/climbing robot: A three-module robot for automatic wall inspection. *IEEE Robotics & Automation Magazine*, 13(1):42-50, 2006.
- [36] Ko, P., Prieto, S. A., and de Soto, B. G. ABECIS: An automated building exterior crack inspection system using UAVs, open-source deep learning and photogrammetry. In *ISARC Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 637-644, Dubai, UAE, 2021.

Cable-Driven Parallel Robot (CDPR) for Panelized Envelope Retrofits: Feasible Workspace Analysis

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

Buildings and Transportation Science Division, Oak Ridge National Laboratory, United States of America * liuy5@ornl.gov zhangr2@ornl.gov hayesnw@ornl.gov hunde@ornl.gov maldonadopbp@ornl.gov

Abstract -

Recent decades have seen remarkable progress in the field of robotic-assisted construction. Cable-driven parallel robots (CDPRs) emerge as promising tools for automating construction processes, due to their advantageous features such as scalability, reconfigurability, compact design, and high payload-to-weight ratio. This paper uses a simple static model to determine the feasibility of a CDPR for overclad panel installation in building envelope retrofits. Given that the building facade needs to be a subset of the CDPR's wrench-feasible workspace, we focus on the sensitivity of the workspace concerning various cable arrangements and CDPR frame sizes (e.g., height and width extensions). Our analysis indicates that no cable arrangement satisfies the requirement of complete facade coverage and avoids cable-topanel collisions. Thus, frame extension is needed to enhance coverage. However, in densely populated areas where width extension is limited by space constraints, height extension alone is insufficient to guarantee full facade coverage. This paper pioneers the investigation of CDPRs for panelized envelope retrofits, showcasing their advantages and limitations and paving the way for further research and development.

Keywords -

CDPR, construction robotics, panelized envelope, envelope retrofits

1 Introduction

Buildings account for more than 35% of the total carbon dioxide emissions in the United States [1]. About 52% of the existing residential buildings were built before the implementation of the 1980 energy codes [2]. Consequently, these buildings are more likely to have inadequate thermal insulation and air barriers, or none altogether [3]. The absence of adequate insulation and airtightness in these structures leads to high energy losses through the building envelope that account for 70% of carbon emissions



Figure 1. Conceptual view of a cable-driven parallel robot for retrofitting a residential building using overclad panels in a densely populated urban area.

generated locally by households [4]. Therefore, bringing these outdated structures in line with current energy codes can reduce energy usage and greenhouse gas emissions in the building sector. Overclad panel envelope retrofits using prefabricated panels effectively minimize waste and improve energy efficiency [5]. Traditional approaches for installing overclad panels in multi-story buildings typically involve panels being carried by laborers or hoisted with a crane and placed at their final location by hand [6]. These practices incur installation errors, safety concerns for construction workers, low efficiency, and complications in densely populated urban areas due to limited space and disruptions to adjacent structures and pedestrians. Robotic construction provides an alternative for faster, more accurate, and safer installation processes.

The unpredictable nature of typical construction sites and the need to minimize disruptions to surrounding areas require an adaptable and compact robotic system. Cable-Driven Parallel Robots (CDPRs) [7–9] are a type of parallel mechanism where several flexible cables suspend the end-effector. As depicted in Figure 1, CDPRs offer a compact, small-footprint, portable, and easily reconfigurable design, making them ideal for overclad panel installation in densely populated areas with minimal disruption to surroundings. Although the effectiveness of CDPRs at reducing installation time, errors, and costs has been demonstrated with robotic curtain wall installation [10], to the best of our knowledge, this is the first paper on the theoretical optimization of CDPR's size that broadens

^{*}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).

their applicability in construction within the constraints of crowded urban environments.

While CDPRs present advantages for panel installation, they face specific challenges and limitations. Unlike traditional rigid-link parallel robots, CDPRs uniquely face the challenge of having cables that can only be driven by positive tension. This characteristic has spurred extensive research in design [11], analysis [12], control [13], and path planning [14]. One particular challenge for automated overclad panel installation is workspace coverage. Ideally, the height and width of the CDPR frame should not exceed the facade dimensions, especially for densely populated areas as pictured in Figure 1. The frame size constraints, coupled with cable force limitations, result in the system being unable to bring the panel to the building corners. This simulation study investigates all possible CDPR designs to determine how much facade coverage the wrench-feasible workspace can achieve. The different models were defined based on their cable configurations. For the most promising models, a frame extension analysis was conducted for both planar and special CDPRs to identify strategies for maximizing coverage.

In the following sections, we begin by reviewing panelized envelope retrofits and CDPRs in construction (Section 2). Next, we discuss the details of CDPRs that can be used for overclad panel installation (Section 3) and wrench-feasible workspace (Section 4). Finally, Section 5 concludes the paper and outlines our future work.

2 Related Work

In this section, we will first talk about the background associated with panelized envelope retrofits, then we will review relevant research on CDPRs in construction.

2.1 Panelized envelope retrofits

With the growing focus on energy efficiency and sustainability, panelized envelope retrofits [5] have emerged as an essential solution to improve the energy efficiency of older buildings. This retrofit practice uses prefabricated panels for overclad installation over the existing exterior walls and roof. Panelized envelope systems are designed to provide a more efficient and controlled construction process, as the panels are produced under controlled conditions, reducing the potential for on-site errors and improving construction speed. In addition, this method minimizes disruptions to occupants since the overclad panels can be installed with building occupants in place [15].

Traditional approaches for panel installation in multistory buildings, such as scaffolding, suspended platforms, forklift telehandlers, and stationary or mobile cranes, come with significant drawbacks. They often prove costly, timeconsuming, prone to inaccuracies, and require a large footprint for equipment. For example, scaffolding and suspended platforms, while worker-friendly, are slow and subject to human error, restricting panel size to what can be manually carried. Cranes offer an alternative for large panels, reducing installation time compared to manual methods. However, they remain expensive and relatively slow, with each precast concrete panel taking 15 to 60 minutes to install [16]. Additionally, their large footprint poses challenges in densely populated areas. While addressing a different application, Iturralde et al. [10] investigated the work time of using CDPR to install a curtain wall module, reducing installation by 0.51 hours with promising accuracy and repeatability. This achievement holds significant potential to reduce labor costs and increase construction efficiency. With the potential to significantly improve installation speed, reduce costs, and enhance safety, CDPRs are attracting attention for panel installation.

2.2 CDPR in Construction

Recently, CDPRs have been explored in various applications and fields [10, 17-24]. Due to their desirable characteristics, such as scalability, reconfigurability, compact design, and high payload-to-weight ratio, they are widely investigated in construction applications such as bricklaying [19-21], 3D printing [22], and solar power plant assembly [23]. For building envelope applications, in particular, Izard et al. [24] explored the use of a CDPR for inspecting building facades. However, the robot was not designed for tight tolerances or heavy payload required for retrofits. In addition, Iturralde et al. [10] designed and implemented a CDPR for modular curtain wall installation in real-world application. They investigated whether CDPR could install curtain wall modules with sufficient accuracy and shorten manual installation time. To our knowledge, no prior research has explored the application of CDPRs for automated overclad panel envelope retrofits.

3 CDPR for Panelized Envelopes

In this section, we will discuss the specific requirements for CDPR panel installation. We will then analyze various methods to achieve these requirements.

3.1 Panelized Envelopes Requirements

Three main requirements are essential for a successful robotic overclad panel envelope retrofit system. Firstly, the CDPR should control the panel's 6 Degrees of Freedom (DOF), including translations and rotations around all three axes. This capability is essential to accommodate non-flat and non-perpendicular walls encountered in realworld residential buildings. Secondly, the CDPR should have enough maneuverability to install panels at any location on the facade (including near corners) without posi-



Figure 2. Geometric definitions for a kinematic model of a general CDPR design.

tioning cables between the panel and the facade. Thirdly, the CDPR should have a simple, compact design and a small footprint given that the available installation space may be limited to the building dimensions and the space in front of the building (e.g., sidewalk). Compactness is particularly critical in densely populated urban areas where adjacent buildings or narrow sidewalks impose limitations on the frame size. Simplicity of the frame allows the robot to be installed quickly without the use of a crane.

The CDPR in Figure 1 fulfills the first and third requirements. However, achieving the second requirement requires further exploration. To that end, this paper will study various cable configurations (Section 4.1) and will consider extending the frame size (Section 4.2) in order to improve the facade coverage for panel installation.

3.2 CDPR Kinematic Model

In general, a CDPR contains three parts: a fixed frame described in a world coordinate system \mathcal{A} , a mobile platform (e.g., the panel to be installed) described in a local coordinate system \mathcal{B} , and *n* cables connecting the platform to the frame. The geometric description of the CDPR is shown in Figure 2, which defines the following vectors:

- The constant vectors a_{i∈{1,...,n}} denote the proximal anchor points A_i in the frame with respect to A.
- The constant vectors b_{i∈{1,...,n}}, denote the distal anchor points B_i in the platform with respect to B.
- The platform pose (\mathbf{r}, R) is defined by the vector \mathbf{r} , which is the location of the platform's center of mass with respect to \mathcal{A} , and the rotation matrix $R \in SO_3$, which represents the orientation of the mobile platform's frame of reference \mathcal{B} with respect to \mathcal{A} .
- The vectors l_{i∈{1,...,n}} represent ideal cables and can be calculated as l_i = a_i − r − Rb_i with respect to A.
- The unit vector along the cable $\mathbf{u}_i = \mathbf{l}_i / ||\mathbf{l}_i||_2$.

• The cable forces $\mathbf{f}_i = f_i \mathbf{u}_i$, where $f_i \ge 0$ are the tensile forces action on each cable $i \in \{1, ..., n\}$.

A simple kinematic model can be constructed by solving the force and torque equilibrium equations as follows:

$$\sum_{i=1}^{n} \mathbf{f}_{i} + \mathbf{f}_{p} = 0 \quad \text{and} \quad \sum_{i=1}^{n} R\mathbf{b}_{i} \times \mathbf{f}_{i} + \tau_{p} = 0.$$
(1)

Here, \mathbf{f}_{p} and τ_{p} are the external forces and torques applied to the platform. For our particular application, note that $\mathbf{f}_{p} = m\mathbf{g}$ and $\tau_{p} = 0$, where *m* is the mass of the overclad panel. Rewriting Eqn. (1) into matrix form we obtain:

$$\underbrace{\begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_n \\ R\mathbf{b}_1 \times \mathbf{u}_1 & \dots & R\mathbf{b}_n \times \mathbf{u}_n \end{bmatrix}}_{A^{\mathrm{T}}} \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}}_{\mathbf{f}} + \underbrace{\begin{bmatrix} \mathbf{f}_p \\ \tau_p \end{bmatrix}}_{\mathbf{w}_p} = 0. \quad (2)$$

This can be written in a compact matrix-vector form as $A^{T}\mathbf{f} + \mathbf{w}_{p} = 0$, where the A^{T} is the transpose of the Jacobian matrix and referred to as the *structure matrix*, \mathbf{f} is the vector of tensile forces, and \mathbf{w}_{p} is the total wrench applied.

3.3 Wrench-Feasible Workspace

The pose (\mathbf{r}, R) is called *wrench-feasible* for a given wrench \mathbf{w}_p if there exists a vector **f** that satisfies:

$$A^{\mathrm{T}}(\mathbf{r}, R)\mathbf{f} + \mathbf{w}_{\mathrm{p}} = 0 \quad \text{s.t.} \quad 0 \le f_{\mathrm{min}} \le \mathbf{f} \le f_{\mathrm{max}} \quad (3)$$

where f_{\min} and f_{\max} denote the lower and upper bounds for the feasible force range of the cables. Considering the limited rotation angles required for overclad panel installation, our analysis assumes that rotation angles are set to zero. Therefore, the wrench-feasible workspace in our study can be defined as:

$$\mathcal{W} = \left\{ \mathbf{r} \mid \text{Eqn. (3) holds with } \mathbf{w}_{p} = \begin{bmatrix} m\mathbf{g}^{T} & 0 \end{bmatrix}^{T} \right\}.$$
(4)

In general, Eqn. (3) leads to an underdetermined system with either zero or infinitely many solutions. To estimate W, a set of poses **r** was selected from a discretized grid, and feasibility was determined using linear programming.

4 Analysis of *W* for different CDPRs

For ease of visualization and understanding, we will first analyze the wrench-feasible workspace coverage using a planar CDPR. Figure 3 shows the dimensions for a ¹/₃ scaled model of a 3-story residential building currently used at ORNL for building envelope retrofit experiments. The planar CDPR shown in the left plot of Figure 3 corresponds to a 2-translation 1-rotation (2T1R) model with



Figure 3. Left: planar CDPR for square panel installation where the frame matches the facade dimensions (blue rectangle). Complete coverage is achieved if the centroid (purple dot) can be moved within the yellow-shaded region. Right: W (green area), the red area indicates lack of full coverage.

n = 4 cables. The CDPR frame (blue solid line) has the same dimensions as the facade and the square panel (black solid line) can be installed anywhere within the facade. The yellow-shaded region describes the area where the panel's center of mass r (purple dot) needs to be maneuvered to fully cover the facade with the panels. Due to inherent limitations imposed by the CDPR's cable arrangement and force restrictions, certain regions of the facade may be difficult to access by most of the configurations. We categorize these regions into three types: corners, top or bottom margins, and lateral bands. The right plot of Figure 3 shows W in green color for $f_{\min} = 132.3$ N, $f_{\text{max}} = 1200 \text{ N}$, and m = 13.5 kg. The red colored area shows locations where \mathbf{r} is required to reach, but there is no feasible solution for Eqn. (3). For this particular CDPR cable arrangement, the top left and right corners, the top and bottom margins, and the side bands of the facade are not reachable, therefore the model in Figure 3 does not provide full facade coverage.

The next sections explore two potential solutions to achieve full facade coverage: adjusting cable configurations and extending the frame size. While this paper ignores cable-to-cable and cable-to-panel collisions, these potential issues can be detected, mitigated, or eliminated in real-world applications through various approaches, such as optimizing cable routing design, applying advanced path planning algorithms, and optimizing anchor positions [7, 25-27]. The following assumptions have been made: (1) all cables are massless, (2) cables don't sag and behave as straight line segments, and (3) the center of mass of the panel coincides with its centroid.

4.1 Analysis of Cable Configurations

The first method to increase the workspace coverage involves exploring alternative cable configurations. For the

CDDD

	5 I. F	W IOI	most j	JIOHIIS	ang pi	anar	UPRS.	
Model#	21	13	15	19	11	12	22	4
$P_{\mathcal{W}}$	1	.88	.83	.83	.79	.78	.76	.73
Tab	le 2. I	Desire	d spati	al wor	kspac	e defi	nition.	
				x		у	z	_
Prism start corner (m) Prism end corner (m)			0.38 1.44	$\begin{array}{ccc} 1 & 0. \\ 7 & 2. \end{array}$	381 743	0.254	-	

planar CDPR with 4 cables, the total number of possible cable configurations is 4! = 24. Figures 4 and 5 illustrate all possible cable configurations and their corresponding wrench-feasible workspaces. Let P_W denote the wrenchfeasible workspace coverage of the CDPR, calculated as the area of W divided by the required coverage, i.e., yellow area = green + red areas in Figure 3. Table 1 shows P_{W} for the most promising CDPR models. Note that only model 21 achieves $P_W = 1$, satisfying the full coverage requirement. By observing all possible cable configurations we concluded that, to maximize P_W , it is advantageous to connect the proximal anchor to the distal anchor such that the angle between the cable and the closest frame edge is maximized. This is especially useful when the panel needs to reach the corners of the facade. For instance, if the CDPR in model 4 needs to reach an upper corner, the upper cables will form a small angle with respect to the upper frame edge. If the upper cables are close to a horizontal position, their vertical force component will be too small to compensate for gravity. To accommodate for this, the tension on the upper cables needs to increase significantly, saturating the upper limit f_{max} and resulting in an infeasible pose. Therefore, in the planar case, the best cable configuration that achieves $P_W = 1$ is the following: $\{A_1 \to B_3, A_2 \to B_4, A_3 \to B_1, A_4 \to B_2\}.$

Extending our analysis to the spatial case, we consider configurations with 8 cables, 8 distal anchors, and 8 proximal anchors. The parameters employed are identical to those used in the planar case, including a frame depth of 1.219 m and a platform depth of 0.254 m. Note that in real-world panel installation using CDPRs, the dimension of the platform may not be the same as the panel itself. Designing a thicker platform that can carry multiple panels offers a promising approach to reduce the loading time, which can be a significant bottleneck in the installation process. For the spatial analysis, both the CDPR frame and the region where the panel centroid is to be manipulated take the form of rectangular prisms. Table 2 summarizes the dimensions of the desired workspace. Here, the x and y axes are the same as in Figure 3, while the z axis is determined following the right-hand rule. Similar to the previous case, we assumed $R = I_{3\times 3}$ for calculating W.

The spatial CDPR case yields 8! = 40,320 possible cable configurations. However, 63% of them (25,500) lead to



Figure 4. Visualization of all possible cable configurations for planar CDPRs where the frame corners are attached to the panel corners. The model numbers are given for reference.



Figure 5. Wrench-feasible workspace \mathcal{W} (represented in green) for the corresponding planar CDPRs when exploring all possible cable configurations. The red area represents a lack of full facade coverage.



Figure 6. Histogram of $P_W > 0.5$ for cable-to-panel collision-free spatial CDPR configurations.



Figure 7. Left: Spatial CDPR model with the largest workspace. We only show cables between the facade and the platform for better visualization. Right: *W*.

cable-to-panel collisions and a substantial portion results in small workspace coverage. Figure 6 shows a histogram depicting the distribution of P_W for all collision-free configurations with $P_W > 0.5$. This analysis reveals that only 82 configurations, out of the original 40,320, achieve a $P_{W} > 0.5$. Due to the retrofitting requirements, positioning cables between the panel and the facade is undesirable. For example, Figure 7 shows the spatial CDPR model with cable configuration $\{A_1 \rightarrow B_3, A_2 \rightarrow B_4, A_3 \rightarrow$ $B_2, A_4 \rightarrow B_2, A_5 \rightarrow B_7, A_6 \rightarrow B_8, A_7 \rightarrow B_5, A_8 \rightarrow B_6 \}$ that achieves $P_W = 0.95$. However, the configuration presents an obstruction caused by all the cables between the facade and the platform, hindering panel attachment to the facade. Therefore, we focused our analysis on four specific models that have a large P_{W} but do not have cables obstructing the panel installation. Such models are depicted in the first row of Figure 8. From Figures 6 and 8, we can see that achieving 100% coverage of the entire facade is not feasible. Therefore, we should consider an alternative solution: CDPR frame extension.

4.2 Frame Extension

Extending the frame size presents another viable solution for increasing the wrench-feasible workspace. Theoretically, the frame can be extended along both width and height, but each approach presents certain limitations. Height extension can only be applied on top of the building and can be impacted by wind loads. Width extension is applied symmetrically to both the left and right sides and is often constrained by the presence of adjacent buildings. Several factors, including local regulations, infrastructure differences, and community dynamics, can influence the successful implementation of the frame extension. Setting aside local regulations, in densely populated areas, the depth of the frame shouldn't exceed the sidewalk length. For width extension implementation, it is crucial to ensure that the extension does not interfere with adjacent buildings or have any adverse effects on them. Additionally, height extension considerations should be mindful of potential high wind effects. To address safety concerns, a truss-like structure is recommended for supporting the upper cantilevers. Furthermore, it is recommended to shield the frame to prevent pedestrians and workers from harm in the event of cable snapping. Considering the densely populated urban environment depicted in Figure 1, height extension presents a more favorable option due to the limitations inherent in width extension.

To achieve $P_W = 1$, different configurations may require different extensions. In the planar case, Model 4 requires height and width extensions, while Model 12 only needs width extension, as illustrated in Figure 9. In the spatial case, Model S1, as shown in Figure 8, fails to reach all corners and leaves a significant gap at the top of the desired workspace. Similar to Model 4 in the planar case, increasing P_{W} for Model S1 requires height and width expansions. However, as shown in the third row of Figure 8, achieving full coverage needs substantial extensions on both dimensions, potentially impractical for real-world scenarios. Models S2 and S3 overcome the limitation of not reaching the top of the facade. However, their W do not cover the side bands, which demands a width extension to augment P_W . However, the coverage depicted in plots C(2) and C(3) indicate that even with a 25% width extension, P_W remains below 1. Model S4 successfully covers the side bands but misses the top and bottom margins. While a height extension can address the top margin coverage, the bottom margin remains inaccessible for this model. When restricted to a single cable configuration, Model S4 is one of the most promising options for maximizing P_W . Based on our discussion so far, there is currently no single spatial model capable of covering the entire facade area with practical frame extensions. In cases where multiple cable configurations are permitted, a combination of Models S3 and S4 proves effective in covering nearly all areas within the required retrofitting space, except for the four corner areas. The following section will explore potential solutions to address this issue. In addition, depending on the configuration, width or height extensions may not always be beneficial for increasing $P_{\mathcal{W}}$. For instance, models S1 and S3 from Figure 8 showcase scenarios where height extension decreases P_W .



Figure 8. Top: Selected cable configurations. Middle: W without frame extension, the black rectangle indicates the panel centroid area needed for a complete retrofit. Bottom: P_W obtained through frame extension.



Figure 9. Planar CDPR P_W comparison for models 4 and 12 with height and width extensions.

5 Conclusion and Future Work

This paper studied the use of CDPRs for automated overclad panel installation. To fulfill envelope retrofit requirements, we explored strategies for maximizing the wrenchfeasible workspace. Different cable configurations lead to different wrench-feasible workspace coverage. In general, both width and height extensions improved the coverage for a given configuration. However, if width extension is not feasible, implementing height extension can still improve coverage. This study highlights the importance of considering different cable configurations and frame extension to optimize coverage in different scenarios.

Future research will continue to explore strategies to enhance coverage. Designing reconfigurable anchors offers a potential solution to achieve full coverage. For instance, the combination of Models S3 and S4 can cover nearly all the required retrofitting area, except for the four corners. To address this limitation, one can attach the distal anchor points to a panel carrier instead of directly attaching them to the panel. The panels can be attached off-center on the carrier, with an offset towards the respective corners, thereby compensating for the infeasible space by the original models. Additionally, approaches like optimizing panel and carrier dimensions or designing advanced path planning algorithms by strategically sagging specific cables offer the potential for further increasing workspace coverage. We also intend to expand our research to incorporate a real-world model, aiming to offer a more comprehensive understanding of the entire system. This expansion includes dynamic modeling and control algorithms to evaluate aspects like the system's robustness, autonomy, and applicability to different construction processes, including considerations of economic viability and long-term durability.

Acknowledgements

This research was supported by the DOE Office of Energy Efficiency and Renewable Energy (EERE), Building Technologies Office, under the guidance of Sven Mumme, and Integration Center (BTRIC), a DOE-EERE User Facility at Oak Ridge National Laboratory.

References

- [1] U.S. Energy Information Administration. Monthly energy review-november 2023. Technical report, US Energy Information Administration, 2023.
- [2] The American Council for an Energy-Efficient Economy. Energy efficiency impact report. Technical report, 2022.
- [3] Chrissi A Antonopoulos, Cheryn E Metzger, et al. Wall upgrades for residential deep energy retrofits: A literature review. 2019.
- [4] Office of Energy Efficiency & Renewable Energy. Affordable home energy shot. On-line: https://www.energy. gov/eere/affordable-home-energy-shot#:~: text=Through%20this%20Energy%20Earthshot%2C% 20DOE, to%20achieving%20this%20bold%20goal., Accessed: 03/14/2024.
- [5] Mikael Salonvaara, Antonio Aldykiewicz Jr, et al. Cost assessment of building envelope retrofits. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2020.
- [6] Christian Mascheck. Serielles sanieren: So funktioniert der "energiesprong". On-line: https://cradle-mag. de/artikel/serielles-sanieren-energiesprong. html, Accessed: 12/15/2023.
- [7] Andreas Pott and Tobias Bruckmann. Cable-driven parallel robots. Springer, 2013.
- [8] Zhaokun Zhang, Zhufeng Shao, et al. State-of-the-art on theories and applications of cable-driven parallel robots. Frontiers of Mechanical Engineering, 17(3), 2022.
- [9] N. G. Dagalakis, J. S. Albus, et al. Stiffness Study of a Parallel Link Robot Crane for Shipbuilding Applications. Journal of Offshore Mechanics and Arctic Engineering, 111(3):183-193, 08 1989. ISSN 0892-7219.
- [10] Kepa Iturralde, Malte Feucht, et al. Cable-driven parallel robot for curtain wall module installation. Automation in Construction, 138:104235, 2022.
- [11] Jean-Baptiste Izard, Alexandre Dubor, et al. On the improvements of a cable-driven parallel robot for achieving additive manufacturing for construction. In Cable-Driven Parallel Robots: Proceedings of the Third International Conference on Cable-Driven Parallel Robots, pages 353-363. Springer, 2018.
- [12] Zane Zake, Stéphane Caro, et al. Stability analysis of pose-based visual servoing control of cable-driven parallel robots. In Cable-Driven Parallel Robots: Proceedings of the 4th International Conference on Cable-Driven Parallel Robots 4, pages 73-84. Springer, 2019.
- [13] João Cavalcanti Santos, Ahmed Chemori, and Marc Gouttefarde. Model predictive control of large-dimension cabledriven parallel robots. In Cable-Driven Parallel Robots: Proceedings of the 4th International Conference on Cable-Driven Parallel Robots 4, pages 221-232. Springer, 2019.
- [14] Ricard Bordalba, Josep M Porta, and Lluís Ros. Randomized kinodynamic planning for cable-suspended parallel robots. In Cable-Driven Parallel Robots: Proceedings of the Third International Conference on Cable-Driven Parallel Robots, pages 195-206. Springer, 2018.

- and used resources at the Building Technologies Research [15] Robert P Sroufe, Craig E Stevenson, et al. The building envelope holds the key. The Power of Existing Buildings: Save Money, Improve Health, and Reduce Environmental Impacts, pages 85-105, 2019.
 - [16] Nolan Hayes and Diana Hun. Expediting precast concrete installation. On-line: https: //lsc-pagepro.mydigitalpublication.com/ publication/?i=806162&article_id=4666276& view=articleBrowser, Accessed: 12/18/2023.
 - Andreas Pott, Hendrick Mütherich, et al. Ipanema: a family [17] of cable-driven parallel robots for industrial applications. Cable-Driven Parallel Robots, pages 119-134, 2013.
 - [18] Rendong Nan. Five hundred meter aperture spherical radio telescope (fast). Science in China Series G, 49(2), 2006.
 - [19] Yulong Wu, Hung Hon Cheng, et al. Cu-brick cabledriven robot for automated construction of complex brick structures: From simulation to hardware realisation. In 2018 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIM-PAR), pages 166–173. IEEE, 2018.
 - [20] Ilija Vukorep. Autonomous big-scale additive manufacturing using cable-driven robots. In Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC), pages 254-259, Taipei, Taiwan, July 2017. Tribun EU, s.r.o., Brno. ISBN 978-80-263-1371-7.
 - [21] Tobias Bruckmann, Christopher Reichert, et al. Concept studies of automated construction using cable-driven parallel robots. In Clément Gosselin, Philippe Cardou, Tobias Bruckmann, and Andreas Pott, editors, Cable-Driven Parallel Robots, pages 364-375, Cham, 2018. Springer International Publishing. ISBN 978-3-319-61431-1.
 - [22] Paul Bosscher, Robert L. Williams, et al. Cable-suspended robotic contour crafting system. Automation in Construc-tion, 17(1):45–55, 2007. ISSN 0926-5805.
 - [23] Andreas Pott, Christian Meyer, and Alexander Verl. Largescale assembly of solar power plants with parallel cable robots. In ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics), pages 1-6, 2010.
 - [24] Jean-Baptiste Izard, Marc Gouttefarde, et al. Integration of a Parallel Cable-Driven Robot on an Existing Building Façade, pages 149-164. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-31988-4.
 - [25] Vincentius A Handojo, Adlina T Syamlan, et al. Cable driven parallel robot with big interference-free workspace. In Mechanism and Machine Science: Select Proceedings of Asian MMS 2018, pages 43-56. Springer, 2020.
 - [26] Jeong-Hyeon Bak, Sung Wook Hwang, et al. Collision-free path planning of cable-driven parallel robots in cluttered environments. Intelligent Service Robotics, 12:243-253, 2019.
 - [27] Martin J-D Otis, Simon Perreault, et al. Determination and management of cable interferences between two 6-dof foot platforms in a cable-driven locomotion interface. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 39(3):528-544, 2009.

Development of a Prototype Data-system Integration Platform for As-Built Inspection

F. Matsushita¹, K. Miyaoka², B. Miyazaki³, M. Matsubara⁴ and K. Ozawa⁵

^{1,2,5} Institute of Engineering Innovation, School of Engineering, The University of Tokyo, Japan ³Mitsubishi Research Institute, Inc., Japan ⁴AMDlab, Inc., Japan

<u>matsushita@i-con.t.u-tokyo.ac.jp, miyaoka@i-con.t.u-tokyo.ac.jp, bumiyaza@mri.co.jp, masamiki.matsubara@amd-lab.com, ozawa@i-con.t.u-tokyo.ac.jp</u>

Abstract -

This study focuses on developing a data-system integration platform tailored for the construction It employs application programming phase. interfaces (APIs) and a range of sensor devices to enhance construction management practices. The platform is crafted to cater to various requirements while maintaining scalability, utilizing a microservice architecture to facilitate seamless integration with external systems. A prototype of the platform was developed and subsequently assessed for its practicality in as-built inspections, receiving positive evaluations for its effectiveness and efficiency. Additionally, the platform showed its adaptability by successfully integrating with external services for comprehensive analysis and data collection. Future endeavors will focus on customizing the platform for diverse applications and collaborating with stakeholders to foster its practical implementation in construction projects.

Keywords -

APIs; As-built inspection; Integration platform

1 Introduction

The rapid advancement of sensor technology in recent years has greatly facilitated the collection of crucial data for construction management from worksites. This development, coupled with the progression of applications designed for data analysis and processing, suggests a trend toward the tailored combination and usage of these devices and applications to meet the specific needs of various construction projects. application programming interfaces (APIs) are a key method for combining these technological tools. However, challenges arise when system integration occurs in a one-to-one relationship between API providers (e.g., device developers) and users (e.g., application developers), as depicted in Figure 1 (above). To address this issue, the integration platform illustrated in Figure 1 (below) is proposed. This study focuses on developing a prototype data-system integration platform for the construction phase, aiming to enable the seamless integration of devices and applications.

A critical aspect of developing a data system integration platform is defining requirement specifications based on user needs. The objective of this study is to specify the system requirements for a datasystem integration platform targeting the construction phase, develop a prototype, and assess its effectiveness through practical application.



Figure 1. Comparison of device and application collaboration with and without an integration platform (above: without platform, below: with platform).

2 What is data-system integration platform

The evolution of integration platforms is a global

phenomenon. Notable examples include DATA-EX [1] and the Connector Architecture for Decentralized Data Exchange (CADDE) [2] in Japan, and GAIA-X [3] and the International Data Spaces Association (IDSA) [4] internationally. Industry-specific integration platforms are emerging in sectors such as agriculture [5], smart cities [6], and the energy sector [7].

Moreover, the proliferation of cloud computing technologies has led to diverse service offerings, including infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) [8]. Gartner, a leading American consulting firm, defines integration platform as a service (iPaaS) [6] as a comprehensive suite of cloud services that facilitates the development, execution, and governance of integration flows. These services enable the connection of a variety of on-premises and cloud-based processes, services, applications, and data within or across organizations [9].

This research and development project focuses on a data system integration platform tailored for the construction phase, categorizing it as an iPaaS. Regarding with construction phase, several platforms have been proposed and are being implemented like Infrakit [10] in Finland and Landlog [11] in Japan. Infrakit is one of integration platforms for collaboration among stakeholders in construction phase. Using this feature, real-time quality control reporting in construction projects is being attempted [12]. Landlog is one of the integration platforms for utilizing construction management information from heavy machinery. This platform is developed by 4 companies including Komatsu [13] and provide various services for not only construction but also construction planning. Compared to these cases, this platform is designed as a versatile integration tool that allows for the flexible combination and utilization of devices and applications in accordance with on-site requirements.

3 Specifications for a data-system integration platform

3.1 Stakeholders of the platform

Table 1 delineates the attributes and specific examples of potential users. These users span a wide range, including clients, main contractors, subcontractors, and manufacturers, all of whom play integral roles in the construction process. Additionally, IT vendors, startups, and various manufacturers are identified as key stakeholders for the integration of existing applications and devices. To cater to the varied demands of these users, developers of new services, notably startups and construction firms engaged in technological innovations, are also considered vital stakeholders.

3.2 Characteristics of the construction phase and issues to be resolved

Table 2 outlines the unique characteristics and challenges of the construction phase. A notable feature of this phase is the variability of the supply chain across different sites. The inconsistency of the supply chain poses a challenge, as it is impractical for all suppliers to adopt a specific commercial system. Additionally, determining the essential functionalities within a single organization is complex, given that main contractors, subcontractors, leasing companies, manufacturers, and others must jointly consider and implement necessary features. To overcome these challenges, the system must possess extensibility, the capability to enhance functionalities based on each company's specific needs.

Table 1 User attributes and representative examples

Attributes	Specific examples
Users	clients, main contractors, sub- contractors, and manufacturers
Application developers	IT vendors and startups
Device developers	survey instrument manufacturers, heavy machinery manufacturers and IoT device manufacturers.
New service developers	startups and contractors

Table 2 Key characteristics and challenges in the construction phase

Characteristics	Issues
Supply chain varies by site	Issue 1: Difficult to use certain commercial systems Issue 2: Difficult to determine the functions required by one organization
Single item production and construction methods vary depending on site	Issue 3: A wide variety of required functions Issue 4: The developers can be diverse
Different sensing is required for each site	Issue 5: Necessary of data collection methods from devices

3.3 System architecture

To achieve extensibility, it is crucial to modularize functions and develop a system that allows for effortless replacement and API integration. To meet this criterion, the data-system integration platform utilizes a microservice architecture. Microservices are a distributed architecture comprising individual service components. In the context of the construction phase, as indicated as issue 3 and 4 in Tables 2, the required functionalities and their developers can vary significantly. It is impractical to incorporate all components as functions of the data system integration platform. Therefore, it becomes necessary to implement microservices that can effectively collaborate with external systems.

In considering the integration with external systems, the system configuration of the data-system integration platform is depicted in Figure 2. This diagram categorizes the system into three distinct layers: the platform layer, the analysis and processing service layer, and the construction management information collection service layer.

The analysis and processing service layer is designed to facilitate collaboration between web and desktop applications. In this layer, functionalities are represented as individual components. Meanwhile, the construction management information collection service layer is tailored for the interaction between survey instruments and internet of things (IoT) devices, with each device and instrument being depicted as a component.

The platform layer consists of the data system integration platform and an existing, previously developed platform. Within this layer, the data system integration platform is equipped with essential functions on a component basis.



Figure 2. System configuration diagram of the data-system integration platform

4 **Prototype development**

4.1 Use case configuration

The prototype of the data system integration platform, targeting the construction stage, has been specifically developed for as-built inspections. This use case involves a range of users, including construction companies and their clients. To ensure the reliability of construction management information, the platform envisages API integration with a blockchain-based system [15]. Figure 3 illustrates the system usage flow when conducting asbuilt inspections of ICT earthwork and new Austrian tunneling method (NATM) construction.



Figure 3. System usage workflow for as-built inspections

4.2 System requirements definition

In the process of defining system requirements, functionalities have been established to adequately address two key aspects: the capacity to perform as-built inspections and the ability to maintain system extensibility. Initially, for the development of a prototype capable of as-built inspections, functionalities such as visualization, analysis, and required APIs at each stage of the system usage flow were determined.

Table 3 enumerates the specific items targeted for visualization and analysis in the prototype, along with their respective formats. The visualization function encompasses essential features for verifying construction management, design, and as-built data. Additionally, for analysis purposes, features critical for determining the scope of the as-built inspection and for the evaluation of the as-built itself are included.

The focus of this study also encompasses the management of large-scale point-cloud data, which is essential for construction management information. Consequently, analysis functions such as trimming and noise removal of point clouds are necessary. The process of discrepancy calculation for as-built evaluations, involving comparisons between point-cloud data and 3D design information, varies depending on the construction type. For the trial, as information and communications technology (ICT) earthwork was the focal point, a specialized as-built earthwork evaluation function was developed. The evaluation parameters adhere to the standards set by Japan's Ministry of Land, Infrastructure, Transport, and Tourism.

Table 3 Visualization and analysis functions with corresponding file formats

Function	File formats
Visualization	
3D design	IFC, LandXML, obj, stl
information	
2D design	dxf
information	
Point cloud	las, csv, txt, xyz
As-built	CSV
Alignment	CSV
<u>Analysis</u>	
Trimming for point	las, csv, txt, xyz
cloud	
Noise removal for	las, csv, txt, xyz
point cloud	
As-built evaluation	Input: las, csv, txt, xyz
for ICT earthwork	Output: csv, json

To maintain system extensibility, APIs were integrated to enable interaction with external services. Table 4 outlines the service layers anticipated for integration and implementation within these APIs. In the integration with the analysis and processing service layer, APIs have been developed not only for uploading alignment and as-built information but also for retrieving various types of data necessary for analysis and processing. This allows external applications to extract data from the data-system integration platform and store their analysis results. In terms of integration with the construction management information collection service layer, APIs facilitate uploading of point clouds from survey instruments. Furthermore, if various data types include reference point coordinates, there is an API to update these coordinates. Given that diverse types of information are uploaded over the network, there exists a potential risk of upload failures. To address this, a file upload completion notification API has been implemented. This API can be accessed by an application or survey instrument upon successful data upload, providing a confirmation of the upload's completion.

Table 4 Service layers anticipated for integration and associated APIs

	Implemented APIs
-	API for uploading alignment
	information
-	API for uploading as-built
	information
-	API for file upload completion
	notification
-	API for getting alignment
	information file list
-	API for getting point cloud file
	list
-	API for getting design
	information list
-	API for downloading file
-	API for uploading large-scale
	point cloud upload
-	API for updating reference
	points of various files
-	API for file upload completion API

5 Trial and evaluation

5.1 Setting trial contents

The trial for this data-system integration platform was assessed from two main perspectives: "usefulness for the user" and "satisfaction of system characteristics". Consequently, two distinct trials were established: "Trial 1: Evaluation of Usefulness" and "Trial 2: Verification of System Characteristics", each aligning with the respective perspectives. These trials were conducted as part of the activities of a sub-working group (sub-WG) under a committee formed by the Japan Federation of Construction Contractors in 2022. The sub-WG saw participation from over 20 companies representing a diverse range of industry sectors, including main contractors, subcontractors, startups, IT vendors, manufacturers, and consultants. Trial 1 primarily involved main contractors and subcontractors, while startups and manufacturers were the focus of Trial 2.

5.2 Trial 1: Evaluation of usefulness

For the evaluation of usefulness, a questionnaire survey was administered following the testing of the prototype. This trial included ten contractors. Test data were specifically prepared to ensure that the evaluation results were not influenced by variances in construction management and 3D design information. Table 5 presents a summary of these trials.

The trial encompassed contractor's the implementation within the system usage flow (depicted in blue in Figure 3), while the client's implementation scope (indicated in purple in Figure 3) was executed by the study's authors. Figure 4 depicts an example of the results obtained by adhering to the system usage flow, illustrating as-built information in a color-coded format to highlight discrepancies between the design and actual construction. It is important to note that for the analysis of as-built evaluations, the use of external services was intentionally avoided to ensure uniform conditions among the participating companies. Instead, the analysis function integrated into the prototype was utilized.

	Content
-	ICT earthwork
-	File format: txt
-	Amount of data: 2,856 MB
	(2.9 GB)
-	File format: LandXML
-	Amount of data: 5 MB
-	Average value
-	maximum and minimum
	value for discrepancy
	between design and actual
-	Evaluation area,
-	Number of evaluation point
	data
-	Number of rejected points
	- - - - - -

Table 5 Overview of the trial process



Figure 4. Illustrative case from the trial.

The evaluation of the trials was aligned with the standards outlined in ISO/IEC 25010 [16]. This framework defines the quality characteristics of software as evaluation metrics. The questionnaire items, derived from these standards, are compiled in Table 6. Notably, while operability and willingness to use are not specified as quality characteristics by the ISO, they were included as crucial aspects for assessing usefulness. Additionally, inquiries were made regarding the practicality and operability of each function, with the specifics of these functions detailed in Table 7.

A questionnaire was developed, employing a fivepoint scale for responses. A score of 5 represented the highest rating, while 1 was the lowest. The results, depicted in Figure 5, revealed that all users rated the effectiveness and efficiency as 5 and 4, respectively. This indicates a certain level of effectiveness and efficiency achieved for the users. In terms of practicality and operability, over 60% of users awarded high ratings for all items, with the exception of the analysis operability. Free responses concerning the operability of the analysis suggested improvements, such as the ability to specify the scope for as-built inspections using coordinate values. The practicality and operability ratings for each function, when compared to commercially available software, were likely influenced by factors such as the prototype's limited feature implementation, resulting in relatively lower ratings.

Table 6 Questionnaire items	assessing	software	quality
characte	eristics		

quality		Ouestions
characteristics		
Effectiveness	-	Is it possible to achieve as-built
		inspection works by utilization
		of the system?
Efficiency	-	Does the use of the system
		make as-built inspection works
		more efficient?
Practicality	-	Do you think the features of
		visualization, analysis, and as-
		built information visualization
		are useful in as-built works?
Operability	-	Are the features of
		visualization, analysis, and as-
		built information visualization
		easy to operate?
Willingness	-	If the system is fully
to use		operational, would you like to
		implement it in your company's
		operations?

Table 7 Detailed descriptions of each system function.

Functions			De	tails		
Visualization -		Various visualization functions such as 3D design information				
		and p	oint cio		1	
Analysis	-	Irimi	ning of	point clo	bud	
	-	Selec	tion of s	cope for	as-built	
		inspection				
	-	analy	tical fur	octions for	or as-bui	lt
		evalu	ation			
Visualization	-	The v	isualiza	tion fund	ction of t	he
of as-built		analy	sis resul	ts (visua	lization	of
information		heat r	naps)			
	-	The v	isualiza	tion fund	ction of a	18-
		built	evaluati	on infor	nation	
Effectiveness • 0		2	4	6	8	1
Efficiency		1	1			2
Practicality						-
Operability						
Practicality		_	-			2
Operability for analysis						
Practicality for as-built						2
Operability		-		-	-	
Willingness	No ans	wer F	valuation 1	- Evaluation	2	
	Evalu	ation 3	Evaluation 4	Evaluatio	n 5	

Figure 5. Questionnaire results analysis.

5.3 Trial 2: Verification of system characteristics

The system characteristic of extensibility was a crucial aspect to be verified in this study. To this end, 9 companies, including startups and IT vendors, participated in the verification process. The focus of this verification was on the integration capabilities of the data-system integration platform, specifically regarding the components of the analysis and processing service layer and the construction management information collection service layer.

The external services integrated for this verification included a desktop application with an as-built evaluation function, laser scanners, and mobile devices capable of collecting point cloud data as shown in table 8. Integration with these services was facilitated mainly using the API provided by the data-system integration platform.

A case study was conducted to demonstrate the practical application of this system in the as-built lining concrete evaluation for the NATM. The point clouds required for this evaluation were of shotcrete and lining concrete, with the objective to assess the inner width, height, and lining thickness by analyzing these point clouds. ENZAN's Application and iXs's device is utilized for this trial of NATM.

Туре	Developers and suppliers
Application	
As-built analysis for	ENZAN [17]
NATM	
As-built analysis for	FUKUICOMPUTER,
ICT earthwork	Inc. [18]
Structure from	SkymatiX, Inc. [19]
Motion/ Multi-View-	
Stereo (SfM/ MVS)	
analysis enfine	
Device	
Terrestrial laser	iXs [20] [21]
scanner	
Terrestrial laser	Nikon/Trimble CO.,
scanner (Using	LTD. [23]
Trimble Connect	
[22])	
Lidar installed on	OPTiM Corporation [24]
mobile	
Unmanned aircraft	Skydio [25]
vehicle (UAV)	
Laser scanner	Fujita Corporation [26]
equipped on heavy	Leica Geosystems [27]
machinery	

The integration with the external analysis and processing service was accomplished through the available API, and the outcomes of the as-built evaluation for NATM are illustrated in Figure 6. The results show that the essential evaluation information could be effectively outputted, enabling a trouble-free asbuilt evaluation. The visualization, storage, and conversion functions, previously employed in ICT earthworks, were repurposed here. For visualization, additional line-drawing functions indicating the inner width and height were implemented, while the analysis function's as-built evaluation program was integrated with an external service. This trial successfully demonstrated the platform's capacity for functional expansion. Furthermore, the trial confirmed the feasibility of integrating the platform with various survey instruments and mobile device applications.



Figure 6. As-built evaluation example for the new Austrian tunneling method (NATM).

6 Conclusion

The aim of this study was to define the essential requirements for a data-system integration platform tailored to the construction phase. In pursuit of this objective, a prototype was developed, and its practicality assessed through various trials. Upon analyzing user needs, microservices were selected for their capacity to fulfill the system's extensibility requirement, facilitating integration with external services via APIs. This approach led to the development of a prototype, with trials focusing on ICT earthwork and NATM construction as practical use cases. The trials demonstrated that the system was highly effective and efficient in meeting user needs. The verification of system characteristics confirmed that extensibility, a key system attribute, was successfully achieved. In future studies, it will be essential to explore the platform's adaptability to a broader range of use cases. In addition, it is needed a comparative case study to highlight the advantages of the proposed platform over existing platforms. Furthermore, ongoing discussions with stakeholders involved in the platform's development and

operation will be crucial for its readiness for broader practical implementation in construction projects.

7 Acknowledgment

This research was undertaken as part of the initiatives of School of Engineering, at the University of Tokyo, and the Sub-Working Group of the Committee formed by the Japan Construction Industry Federation. We received substantial support from numerous stakeholders for field trials and various other activities. We wish to extend our heartfelt gratitude for their invaluable contributions and assistance. In addition, in designing the system, we received advice from Mr. Tsujioka of Gartner Japan and Dr. Morita of the School of Engineering, at the University of Tokyo. We would like to express our gratitude to both of them as well.

References

- [1] DATA-EX. On-line: https://data-societyalliance.org/data-ex/, Accessed: 21/12/2023.
- [2] CADDE. On-line: https://sip-cyber-x.jp/overview/, Accessed: 21/12/2023.
- [3] GAIA-X. On-line: https://www.datainfrastructure.eu/GAIAX/Navigation/EN/Home/ho me.html, Accessed: 21/12/2023.
- [4] IDSA. On-line: https://internationaldataspaces.org/, Accessed: 21/12/2023.
- [5] Cestari R., Sebastien D. and Ernesto E. IPaaS in agriculture 4.0: an industrial case. *IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2020.
- [6] Cirillo F., Solmaz G., Berz E., Bauer M., Cheng B. and Kovacs E. A standard-based open source IoT platform: FIWARE. *IEEE Internet of Things Magazine*, pages 12-18, 2019.
- [7] Genge B., Beres A. and Haller P. A survey on cloud-based software platforms to implement secure smart grids. 49th International Universities Power Engineering Conference (UPEC) IEEE, pages 1-6, 2014.
- [8] Saraswat M. and Tripathi R. Cloud computing: Analysis of top 5 CSPs in SaaS, PaaS and IaaS platforms, 9th International Conference System Modeling and Advancement in Research Trends (SMART) IEEE, pages 300-305), 2020.
- [9] Gartner. On-line: https://www.datainfrastructure.eu/GAIAX/Navigation/EN/Home/ho me.html, Accessed: 21/12/2023.
- [10] Infrakit. On-line: https://www.infrakit.com , Accessed: 21/12/2023.
- [11] Landlog. On-line: https://www.landlog.jp/, Accessed: 21/12/2023.

- [12] Kivimäki T and Heikkilä R. Infra BIM based realtime quality control of infrastructure construction projects. *Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 32., 2015.
- [13] Komatsu. On-line: https://www.komatsu.jp/en, Accessed: 21/12/2023.
- [14] Saisho T. Innovation in Japanese Manufacturing: The Case of KOMATSU. *Kokushikan University Library and Information Commons*, 2023.
- [15] Matsushita F. and Ozawa K. Proof-of-concept for a reliable common data environment utilizing blockchain and smart contracts for supply-chain of public civil works, *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 214-221, Chennai, India, 2023.
- [16] ISO/IEC25010. On-line: https://www.iso.org/standard/35733.html, Accessed: 21/12/2023.
- [17] ENZAN. On-line: https://www.enzank.com/eng,html, Accessed: 21/12/2023.
- [18] FUKUICOMPUTER Inc. On-line: https://www.fujita.com/, Accessed: 21/12/2023.
- [19] SkymatiX. On-line: https://skymatix.co.jp/smx_media/?p=3842, Accessed: 21/12/2023.
- [20] iXs. On-line: https://www.ixs.co.jp/, Accessed: 21/12/2023.
- [21] Yamasaki F. and Maeda K. High-Precision Point-Cloud Automatic Combination Utilizing the Retroreflective Target Marker, 4th Symposium for Promotion of i-Construction, pages 57-69, 2022.
- [22] Trimble Connect. On-line: https://connect.trimble.com/, Accessed: 21/12/2023.
- [23] Nikon-Trimble Co., LTD., On-line: https://www.nikontrimble.co.jp/english/index.html, Accessed: 21/12/2023.
- [24] OPTiM Corporation. On-line: https://www.optim.com/, Accessed: 21/12/2023.
- [25] Skydio. On-line: https://apidocs.skydio.com/reference/introduction, Accessed: 21/12/2023.
- [26] Fujita Corporation. On-line: https://www.fujita.com/, Accessed: 21/12/2023.
- [27] Leica Geosystems. On-line: https://leicageosystems.com/en-gb/, Accessed: 21/12/2023.

Automated Defect Inspection in Building Construction with Multi-Sensor Fusion and Deep Learning

Juhyeon Kim^a, Jeehoon Kim^a, Yulin Lian^a, and Hyoungkwan Kim^a

^aDepartment of Civil and Environmental Engineering, Yonsei University, Korea E-mail: <u>kah5125@yonsei.ac.kr</u>, john101010@yonsei.ac.kr, yulin.lian@yonsei.ac.kr, and <u>hyoungkwan@yonsei.ac.kr</u>

Abstract –

The occurrence of defects during the building construction process significantly impacts housing quality. One such defect, the distortion of a building's framework, affects both sustainability and aesthetics. This study presents an automated technique for inspecting framework distortion in building construction by measuring the angles between walls. The proposed method employs a portable data acquisition system that allows for dynamic data collection. The system's accuracy is enhanced through calibration based on terrestrial laser scanning (TLS) data as a reference. Point cloud data are registered to form a map of the interior space, leveraging a deep learning algorithm to visualize framework distortions. When tested in an apartment construction environment, the method reduces data acquisition time compared to the TLS-based approach, while maintaining precision with an average angular error of 0.28 degrees. This study demonstrates a costeffective and accurate solution for defect inspection in the construction industry.

Keywords -

Defect Inspection; Mobile Data Acquisition System; Multi-Sensor Calibration; Point Cloud Registration; Defect Visualization

1 Introduction

Defects occurring during the building construction process profoundly affect housing quality. Critical defects encompass structural flaws, substandard finishing work, inadequate electrical and plumbing systems, insufficient moisture and waterproofing measures, and faulty mechanical systems [1]. Addressing these defects is vital for ensuring the building's safety and longevity. Among these, distortion of the building's framework is a critical structural defect. It directly impacts the building's long-term durability and its aesthetic integrity.

Recent studies have leveraged 2D and 3D computer vision technologies along with deep learning algorithms

to detect building defects more effectively. Wang et al. [2] developed a photogrammetry-based pipeline for 3D reconstruction of buildings, enabling automated identification of building surface defects such as moulds and cracks on reconstructed 3D scenes. Guo et al. [3] proposed a rule-based deep learning approach for detecting façade defects, including cracks, delamination, peeling, and spalling. Chow et al. [4] presented an automated system for detecting cracks and spalling in buildings using mobile data collection, deep learning, and scene reconstruction. Tan et al. [5] developed a method for integrating crack data from unmanned aerial vehicle images into building information models, thereby improving the inspection of high-rise building facades. In addition, several studies have been conducted to monitor various defects in buildings, such as leakage and heat loss [6, 7, 8]. Despite the importance of monitoring framework distortion, research in this area remains scarce. Moreover, there is a lack of technologies that can inspect framework distortion accurately and efficiently for onsite application. Terrestrial Laser Scanning (TLS) is one of the accurate methods for quality assurance and control in construction [9], and is applicable for inspecting framework distortion. However, while this method is highly precise, it comes with considerable constraints in terms of cost and time.

To address the above issue, this paper proposes a novel approach for accurately and cost-effectively measuring framework distortion, as shown in Figure 1. We calibrate multiple Azure Kinect DK depth cameras using TLS data as a reference, and then acquire point cloud data. For visualization purposes, the point cloud data are registered to form a map. The angles between walls are analyzed from each point cloud data, and these values are visualized on the map. A deep learning-based registration algorithm is used for the calibration of the sensors and the formation of the map.



Figure 1. Overview of the proposed method.

2 Methodology

2.1 Deep Learning-Based Sensor Calibration and Data Acquisition

A backpack platform was developed for data acquisition. As shown in Figure 2(a), the platform was equipped with multiple Azure Kinect sensors and batteries for powering the sensors and a laptop. Four Azure Kinect sensors were mounted to broaden the field of view for data acquisition. As shown in Figure 2(b), an operator can carry the platform to collect data while moving around. The point cloud data were saved in the MKV file format, and later PCD files were extracted from the MKV file. During data acquisition, we ensured that the starting and ending positions were identical.



Figure 2. A backpack platform for data acquisition; (a) hardware configuration, and (b) operational demonstration of the backpack platform in field.

For the extrinsic calibration of multiple Azure Kinect sensors, TLS data were used as the ground truth. As shown in Figure 3, point cloud data from each of the four different Azure Kinect sensors were registered with the TLS data. We utilized deep global registration (DGR) [10], a deep learning-based registration algorithm, for aligning Azure Kinect data with TLS data. Through registration, four transformation matrices were obtained, which reveal the relative positional relationships among the Azure Kinect data.

The TLS data in Figure 3 were used solely for calibration purposes; the data were obtained in a laboratory setting. In this study, additional TLS data were acquired to serve as ground truth for calculating the performance of the proposed defect inspection method (Figure 6(b)). The TLS data are different from those in Figure 3 and were acquired from an actual apartment construction site.

Once the calibration of multiple Azure Kinect sensors was completed using TLS data, there was no need to repeat the calibration process. The relative transformations among the sensor coordinate systems were calculated through calibration, and these calculated values remain valid as long as the relative positions of the



Data acquired with multiple Azure Kinect sensors

Figure 3. Calibration of multiple Azure Kinect sensors using TLS data and DGR.

2.2 Point Cloud Registration for Visualization

The point cloud data, obtained from the calibrated multiple Azure Kinect sensors, were employed to inspect the distortion of the building's framework (detailed in Section 2.3). A point cloud map was formed through registration, which was later used to visualize the results of the angle measurement between walls. Point cloud data extracted from the MKV file were sequentially registered using DGR. As shown in Figure 4, once all the point cloud data were registered, a map of the entire interior space could be generated.

In the initial map, as shown in Figure 4(a), misalignment occurred between the point cloud data sets; this was due to the failure to recognize that the data acquisition starting and ending positions were identical. Therefore, we used DGR to register two point cloud data obtained at the start and end points of data acquisition, thereby calculating the degree of discrepancy (Figure 4(c)). The amount of discrepancy was propagated across all point cloud data between the two data. Ultimately, a complete map without mismatch between the point cloud data was generated, as shown in Figure 4(b). This map was used solely for visualizing defect information and was not utilized in defect analysis.



Figure 4. Point cloud registration and misalignment resolution for visualization using DGR; (a) initial map exhibiting misalignment, (b) refined map with aligned point clouds, and (c) calculation of the discrepancy between start and end point cloud data.

2.3 Defect Inspection

Each Azure Kinect data underwent a defect inspection process. Figure 5 shows the procedure for calculating angles between walls. Using random sample consensus (RANSAC), planes were segmented from the raw point cloud data. Subsequently, only the vertical walls were extracted from these planes, and the angles between them were calculated. The angles were visualized on the map generated in Section 2.2. During this process, angles located in close proximity on the map were merged and represented by their average value.





3 Experiments and Results

3.1 Datasets

Point cloud data for one unit of the apartment were acquired at an apartment construction site. The effectiveness of the proposed method was validated through a performance comparison with the TLS-based method. Figure 6(a) and Figure 6(b) show data acquisition using the backpack platform developed in this study and data acquisition using TLS, respectively. As shown in Table 1, a total of five scans were conducted using TLS, taking about 25 minutes excluding the time to move the sensor. With the proposed method, data were continuously acquired for about 3 minutes and saved in the MKV file format. From the acquired MKV file, 191 PCD files were extracted.



(a) (b) Figure 5. Data acquisition for comparative experiment; data acquisition with (a) backpack platform, and (b) TLS.

Table 1. Comparison of data acquisition methods using TLS and multiple Azure Kinect sensors: analysis of scan frequency, time, and cost.

Method	Number of scans	Total acquisition time (mm:ss)	Sensor prices (USD, in thousands)
TLS	5	24:40 (excluding sensor relocation time)	38.3
Four Azure Kinect sensors	1	2:34	2.5

3.2 Implementation Details

When acquiring Azure Kinect data, all sensors were synchronized and connected in a daisy-chain configuration; this configuration refers to the sequential interconnection of pairs of sensors [11]. For registration, we utilized the pre-trained DGR algorithm, which had been trained on the 3DMatch dataset [12]. For calculating the angles between walls, four planes were extracted from each point cloud data using RANSAC. When executing the RANSAC algorithm, the maximum distance for a point to be classified as an inlier was set to 2 cm and the number of points randomly sampled for plane estimation was set to 3. Data acquisition was conducted on a laptop equipped with an Intel Core i7-10750H CPU and an RTX 2060 GPU. Data processing was performed on the Ubuntu 16.04 operating system with an Intel Xeon Gold 6240M CPU processor and an RTX 3080 GPU based on the Python programming language.

3.3 Experimental Results

3.3.1 Sensor Calibration

We conducted comparative experiments to validate the proposed calibration method for multiple Azure Kinect sensors. Figure 7(a) shows the results of a typical calibration method using an AprilTag marker and the iterative closest point (ICP) algorithm. Figure 7(b) shows the results of calibration using the proposed method with TLS data and DGR. As shown in the figure, when the sensors were calibrated in a typical way, a misalignment occurred between the Azure Kinect data. The suggested calibration technique addressed this issue, thereby enhancing the data quality.

When performing calibration using the proposed method, it was possible to combine point cloud data with precision comparable to TLS data, and it eliminated the need for labor-intensive processes like AprilTag marker detection. Figure 8 shows examples of point cloud data from multiple Azure Kinect sensors calibrated using TLS data and DGR.



Figure 6. Comparative analysis of Azure Kinect sensor calibration methods; (a) calibration with an AprilTag marker and ICP, and (b) calibration using TLS data and DGR.



Figure 7. Examples of point clouds from multiple Azure Kinect sensors calibrated using TLS data and DGR.

3.3.2 Point Cloud Registration for Visualization

All Azure Kinect data were combined using DGR to form a map, which was then utilized for visualizing the results of defect inspection. Figure 9 shows the registered point cloud map generated after adjusting the discrepancy between the point cloud data sets. The figure demonstrates that the created map was precise enough to visually comprehend the building's internal elements adequately.

3.3.3 Defect Inspection

The angles between walls measured from each Azure Kinect data were visualized on the registered point cloud map (Figure 10). As shown in Table 2, angles between a total of 10 pairs of wall surfaces were measured. To evaluate the accuracy of the proposed method, the same defect inspection process was applied to TLS data. Figure 11(a) and Figure 11(b) show examples of wall angle measurement using the TLS-based approach and the proposed method, respectively. The proposed method demonstrated an average angular measurement error of 0.28 degrees when compared against the TLS-based approach.



Figure 8. Results of point cloud registration for visualization using DGR.



Figure 9. Visualization of calculated angles between walls in a registered point cloud map.



(a) (b)
Figure 10. Examples of wall angle measurements from point cloud data; angle measurement using (a) TLS data, and (b) Azure Kinect data.

Table 2. Comparative analysis of wall angle measurements from TLS and Azure Kinect data.

		Azure	
Wall	TLS angle	Kinect	Error
pair	(degrees)	angle	(degrees)
		(degrees)	
1	90.92	90.34	0.58
2	89.89	89.59	0.30
3	89.78	89.90	0.12
4	90.44	89.98	0.46
5	90.01	89.75	0.27
6	89.65	89.76	0.12
7	89.81	90.05	0.24
8	90.05	89.73	0.33
9	91.02	91.01	0.01
10	89.64	90.00	0.36

3.4 Discussion

The experimental results demonstrate the potential of the developed backpack platform and defect inspection technique. However, there is still room for improvement in the proposed method through future studies.

First, there is a need for preprocessing point cloud data in the defect inspection process. Noise may occur when acquiring data from a distance using Azure Kinect sensors. Removing such noise based on the data acquisition distance can enhance defect detection performance. Statistical outlier removal and radius outlier removal are common methods used to remove noise from point cloud data. However, these methods do not effectively remove scattered noise points that are acquired from distant ranges in Azure Kinect data. Therefore, using density-based clustering for noise removal can be an effective solution.

Second, the data acquisition platform can be improved to increase its on-site applicability. To achieve this, several strategies can be employed: using a Mini PC instead of a laptop, utilizing efficient batteries to reduce weight, implementing real-time data processing for defect inspection, redesigning the backpack to reduce worker fatigue, and providing visual guides to for data acquisition.

Third, in addition to analyzing angles between walls, a wider array of defects should be addressed in future studies. Sagging in ceilings and floors is also critical defect information, and such defects can be detected using the proposed method. In addition to structural defects, surface anomalies such as cracks, voids, and spalling on the structure's surface can also be detected. These surface defects can be inspected using not only point cloud data but also by applying vision-based methods that utilize images.

Fourth, the developed technology should be validated in more field applications to increase its robustness. The proposed method was validated on a single type of apartment construction site. There is a need to apply the method to a wider variety of building construction sites and address the various challenges that arise in the process. Such diverse real-world implementations will aid in enhancing the applicability of the proposed method.

Addressing these four key improvement areas could significantly enhance the efficacy of the proposed method, making it a highly valuable tool in the field of defect inspection and building analysis.

4 Conclusion

This study proposed a novel pipeline for inspecting framework distortion in building construction employing multiple Azure Kinect sensors. By calibrating the sensors against TLS data and implementing a deep learning algorithm for registration, the system created a comprehensive 3D map of the building's interior. The angles between walls analyzed from the Azure Kinect data were visualized on the 3D map. This approach not only significantly reduced the time required for data acquisition but also maintained a high level of accuracy. If the proposed method is further developed, it could revolutionize the way building construction projects are managed by ensuring effective defect inspection and enhanced safety.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. 2018R1A6A1A08025348) and the "National R&D Project for Smart Construction Technology (No.23SMIP-A156488-04)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation. This research was conducted with the support of i-thetto Co., Ltd., and was funded by DL E&C Co., Ltd.

References

- Ministry of Land, Infrastructure and Transport. Investigation of Defects in Apartment Buildings, Estimation of Repair Costs, and Standards for Determining Defects. On-line: <u>http://www.molit.go.kr/USR/policyData/m_34681/</u> <u>dtl.jsp?id=4503</u>, Accessed: 14/12/2023.
- [2] Wang, T., Gan, V. J., Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning. *Automation in Construction*, 149: 104810, 2023.
- [3] Guo, J., Wang, Q., Li, Y., Evaluation-oriented façade defects detection using rule-based deep learning method. *Automation in Construction*, 131: 103910, 2021.
- [4] Chow, J. K., Liu, K. F., Tan, P. S., Su, Z., Wu, J., Li, Z., Wang, Y. H., Automated defect inspection of concrete structures. *Automation in Construction*, 132: 103959, 2021.
- [5] Tan, Y., Li, G., Cai, R., Ma, J., Wang, M., Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. *Automation in Construction*, 139: 104284, 2022.
- [6] Lee, K., Lee, S., Kim, H. Y., Bounding-box object augmentation with random transformations for automated defect detection in residential building façades. *Automation in Construction*, 135: 104138, 2022.
- [7] Chen, J., Lu, W., Lou, J., Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. *Computer-Aided Civil and Infrastructure Engineering*, 38(8): 1079-1098, 2023.
- [8] Puliti, M., Montaggioli, G., Sabato, A., Automated subsurface defects' detection using point cloud reconstruction from infrared images. *Automation in Construction*, 129: 103829, 2021.
- [9] Tang, X., Wang, M., Wang, Q., Guo, J., Zhang, J., Benefits of terrestrial laser scanning for construction QA/QC: a time and cost analysis. *Journal of Management in Engineering*, 38(2): 05022001, 2022.
- [10] Choy, C., Dong, W., Koltun, V. Deep global registration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2514-2523, Seattle, Washington, United States, 2020.
- [11] Microsoft. Synchronize multiple Azure Kinect DK devices. On-line: <u>https://learn.microsoft.com/en-us/azure/kinect-dk/multi-camera-sync</u>, Accessed: 14/12/2023.
- [12] Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T. 3DMatch: Learning local

geometric descriptors from RGB-D reconstructions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1802-1811, Honolulu, Hawaii, 2017.

Towards Efficient Construction Monitoring: An Empirical Study on Action Recognition Models

Sudheer Kumar Nanduri¹ and Venkata Santosh Kumar Delhi¹

¹Indian Institute of Technology Bombay, Mumbai, India <u>n.sudheer@iitb.ac.in</u>, venkatad@iitb.ac.in

Abstract -

Monitoring fatigue is challenging under computer-vision-based action recognition due to the changes in motion patterns caused by fatigue. Particularly in the construction scenario, the motion patterns are unique per trade and longer than daily life actions, causing challenging scenarios. This paper aims to understand the patterns that can guide the selection of optimal clip durations for aggregating motion features specific to each task. We compare the performance of three action recognition models (I3D, MViT, and VideoMAE) on different construction tasks (excavation, masonry, plastering, etc.) at varying clip lengths. We evaluate the models based on frame-wise accuracy, sequence predictability error, and normalized evaluation duration. Our results show that the transformer-based models outperform the convolutional neural network-based models. The model trained directly over videos performs better than those trained on images. Also, the clip duration affects the model performance differently depending on the task type. Neither the 3s context window from the Atomic Visual Actions (AVA) dataset nor the 10s context window from the Kinetics-400 dataset is suitable for construction tasks. Instead, we suggest a variable clip duration between 5s and 7s, which is preferable depending on the tasks and model architecture. Our work provides insights for developing a dynamic and context-aware duration selection system for action recognition in construction.

Keywords -

Action Recognition; Construction Activities; Clip Durations

1 Introduction

Worker fatigue is a much-studied problem in construction, considering the adverse effects on productivity, safety, and health. Prior attempts to automate fatigue detection utilized computer vision (CV) or on-body sensors for collecting necessary data. Sensors are limited by the contextual information they can collect. For example, an IMU sensor can collect the motion of a specific body part to which it is connected. Computer vision is a better fit for field application because it can collect information from the worker and surroundings simultaneously. In CV, current action recognition approaches analyze patterns in features aggregated from a set of frames. For measuring the duration of a specific action, existing works run action recognition on fixed-length input clips in sequence and append the results.

In prior works for developing fatigue monitoring, work-rest status [1] is set manually for biomechanical evaluation of joint movements. In a pragmatic approach, manual identification and biomechanical evaluation reduce their applicability to real-time monitoring. Utilizing the changes in the movement patterns is a better approach for automating part of these tasks.

While conventional understanding associates muscular fatigue with a decline in performance, the literature suggests that performance is maintained with changes in movement patterns under fatigue. Depending on the variable selected, movement variability may increase or decrease under fatigue [2] [3]. Muscle groups behave differently under fatigue [4]. Fatigue diminishes the force-producing capacity and the ability for smooth and controlled action. This aspect can be utilized for fatigue monitoring and skipping biomechanical analysis. Humans cannot detect movement variability due to cognitive limitations, so computer vision is the best fit. However, breaking down the action into small clips will not be sufficient for fatigue monitoring as the models will lose the context and motion patterns they can use.

One solution is to adopt a dynamic context-aware approach in selecting the clip durations for aggregating motion features specific to each task. The dynamic selection will improve the detection performance while reducing the resource usage for recognition models. Context awareness will also be helpful for safety monitoring and improve the interpretability of action recognition models. In developing a dynamic and context-aware selection system, this work focuses on the first step of understanding the patterns that can guide the selection of clip durations.

2 Literature Review

Though empirical data collection is preferred, two significant works have already provided such data in construction. The first video-based dataset on construction activities [5] is available with 11 action classes. Bricklaying and plastering are manually identified as 9.8 s of mean clip length, with a variance of 3.6s for bricklaying and 5.2 seconds for plastering. An average clip length of 6.8s is observed in the case of all 11 activities, with a variance of 2.7 s. In a later dataset [6], the average activity lengths for bricklaying with subactivities ranged between 27.33 and 33.30 frames. Considering a 25 FPS video or 30 FPS video, which are the standard practices, these range to only one second of video. The plastering application has 50 frames, which comes close to 1.5 to 2 seconds of video. This difference in clip duration is a significant concern. The subsequent models developed using the datasets are expected to learn correctly from the motion patterns available within the dataset. This assumption makes it difficult to transfer the model trained on the first dataset to the second dataset. Thus, it is preferable to study the performance of different models on standard datasets to identify suitable selection patterns. Developing a new dataset needs to consider this aspect to reduce potential bias.

Recent approaches to action recognition utilize three seconds as a fixed clip length following the Atomic Visual Actions (AVA) dataset [7] standards. The AVA dataset focuses on the action recognition of a single person in a frame with a context window of one and a half seconds before and after the frame. The smaller context window enables fine-scale annotation and improves the action boundary precision. The dataset contains 430 15-minute video clips, 1.58 million class labels, and 80 classes. The dataset is built from movies without actual construction-related actions.

Another important dataset is the Kinetics-400 [8], which has 400 action classes. It has 306,245 clips sourced from YouTube videos, mostly from amateur videographers. Thus, it also provides variety in how the action is performed, along with the clothing, pose, and other parameters. Each clip lasts around ten seconds, providing a context window of 5 seconds before and after the keyframe. Some of the action classes in the dataset, like 'laying bricks', 'plastering', 'welding', and 'bending metal', can be utilized for evaluating construction action videos. For this reason, in the current work, we utilized the models with pretraining using the Kinetics-400 dataset.

CV-based action recognition models utilize four major algorithmic approaches - Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Transformers, and Graph Neural Networks (GNN). Transformer-based models provide the most accurate results and can be considered the latest upgrade over RNNs. GNNs need human joint key points to create graph nodes for analysis. However, they have yet to be proven to perform better than the Transformers, which utilize frame features similar to CNNs.

Two-stream inflated 3D ConvNet (I3D) is a CNNbased model proposed along with the Kinetics-400 dataset [9]. It performs well and is considered a baseline for models and datasets developed afterward. Multiscale Vision Transformers (MViT v2) is a Transformer-based image classification model [10] extended for video classification. Video Masked Autoencoders with dual masking (VideoMAE v2) are also transformer-based models but are built with a specific focus on video data for all types of tasks on videos [11]. Both transformer models utilize 3D convolutions to convert the video frames into patches for training and testing. After conversion, both models use attention networks to identify patterns from the data.

3 Methodology

3.1 Data Collection

The study focuses on evaluating CV models on standard construction processes. Processes selected include excavation, scaffolding, formwork, reinforcement, concreting, masonry, and plastering. The videos are carefully chosen from YouTube, with actions relevant to the processes. The selection of the videos considered video clarity, showing critical activities in the operations without cropping and focusing on the workers doing the actions. Activities within the processes are identified concerning construction literature while ensuring a representative dataset capturing realistic scenarios. A total of 14 videos, ranging from 00:15 to 26:50 minutes (average 08:11 min), are utilized for the current study. Most videos have a 30 FPS frame rate and 1280x720 frame dimensions. Other valuable details are presented in Table 1 below, and the example frames for each task category are presented in Appendix 1 for reference.

The videos are then annotated for the ground truth labels. However, the ground truth labels are a subset of action classes from the Kinetics-400 dataset. This subset comprises actions that can be observed in construction sites. For example, Kinetics-400 does not have any formwork-related class label. In the video of formwork, the annotated labels include classes like 'moving furniture', which has the keyword 'moving' relevant to the action context.

However, several frames only fall under some of the subsets of the action labels. These frames are marked with a new class named 'Background'. Frames that show transitions, empty land, and other problems are marked in this class.
File Name	Task Category	FPS	Total Frames
CONC_2	Concreting	30	30425
EXCV_2	Excavation	29.94	450
EXCV_6	Excavation	30	3709
EXCV_7	Excavation	30	16382
FMWK_2	Formwork	29.97	17069
FMWK_4	Formwork	30	12651
MASO_2	Masonry	25	4065
MASO_3	Masonry	30	48304
MASO_4	Masonry	29.97	24581
PLAS_1	Plastering	30	18151
PLAS_2	Plastering	30	4350
RFMT_2	Reinforcement	29.97	8364
RFMT_3	Reinforcement	29.97	2131
SCFL_1	Scaffolding	30	15010

Table 1 Frame Count & Task Category of Dataset Used

3.2 Model Selection

The three action recognition models mentioned in the literature section (i3D, MViT v2, VideoMAE v2) are utilized for comparative analysis. All three models are utilized from the same open-source toolbox, MMAction2, based on PyTorch, to standardize the comparison. Transformer-based models typically have more parameter count and can perform better in general. However, a comparison with the CNN model will help us evaluate the strengths and weaknesses when applying the models to construction scenarios. The I3D model is trained directly on the Kinetics-400 dataset. The MViT model is primarily an image detection model, and hence, it is pre-trained on the ImageNet dataset before training on the Kinetics-400 dataset. The VideoMAE v2 model architecture makes it difficult to train with small datasets. Hence, it is trained with larger hybrid datasets, and the classification head is trained for the Kinetics-400 dataset. Other model details are presented in Table 2 below. Thus, comparing the performance of models also helps us compare the Kinetics-400 dataset with the construction scenario.

Table 2 Model Details

Model	Sampling protocol	FLOPs	Params
I3D	10 clips x 3 crop	43.5G	28.0M
MViT	5 clips x 1 crop	225G	51.2M
MAE	5 clips x 3 crops	180G	87M

A high-performance computer with 2xIntel-Xeon G-6348 CPU and 4x64 GB RAM is used. The available GPUs are not utilized as the work focuses only on model evaluation, and no training is involved.

3.3 Evaluation Parameters

The two parameters mentioned in Table 2 – floating point operations (FLOPs) and parameters (Params)indicate model performance. Lower FLOPs and Higher Params are the best combinations for CV models. However, a few other considerations also come into play while evaluating the model throughput. The two most important considerations are the model architecture and input variations.

The current work evaluates three aspects - the models' performance on different construction tasks at varying clip lengths. Models and Tasks are detailed in the earlier subsections. Clip length is the final variable discussed in the present subsection.

When a video is chunked into multiple clips, two other parameters that can be useful are the gap duration between two subsequent clips and the overlap duration of the first clip over the second clip. In general, a gap between clips increases the speed at the cost of accuracy, and overlap increases the accuracy by providing additional context at the cost of reduced speed. However, the actual performance might differ due to the model and input variations.

For the current work, clip lengths of 1s, 3s, 5s, 7s, 9s are utilized to cover various temporal scales. Overlaps and gaps are not mixed; when the overlap is present, a gap is not considered, and vice versa. Overlaps chosen are 0s, 2s, and 4s, provided they are always less than the clip duration. For a 1s clip, overlap cannot be 2s as it is the same as a 3s clip length and takes in more features than expected. The gaps chosen are 0s, 1s, 10s. In general practice, gaps are provided such that the frame rate is only 1 Hz, that is, a gap of 1s. However, larger gaps can be considered for the action of longer durations, typically observable in construction sites. A 10s gap is chosen to verify whether a large gap will be helpful. With the given conditions, 22 combinations are formed for durations.

In prediction, the gap durations will have no outputs. This approach improves model performance by reusing the last frame results for all frames within the gap duration. Overlap durations only provide the context for current frame prediction, so there is no effect on the outputs for each frame.

3.4 Evaluation Criteria

Evaluation is based on the accuracy of models without any fine-tuning or transfer learning to avoid any biases from additional training. Doing so will also help maintain classification consistency, even when the specific class labels are absent in the pre-trained dataset. Hence, the dataset prepared is utilized for model output evaluation.

Three evaluations are made on the model's overall predictions and task-wise predictions - Frame-wise evaluation, Sequence-wise evaluation, and Time duration evaluation. Frame-wise evaluation captures the model's qualitative performance per frame. Sequence-wise evaluation captures the model's sensitivity to the motion pattern changes surrounding the frames. Finally, the time duration evaluation captures the model's quantitative performance per frame. By comparing the results of these three criteria, analysis is carried out to compare construction tasks and extract useful patterns.

Frame-wise evaluation matches the current frame prediction with manual annotations. Accuracy is the ratio of total positive to total positive and negative classifications for a given action. The models are bound to provide noisy predictions for frames annotated as 'Background'. So, the frame prediction is skipped in evaluation for accuracy, but the duration is considered for measuring model prediction performance. Although a multi-class confusion matrix can be utilized, we use the simple metric given the choice of models and a small dataset.

The methodology followed for the sequence-wise predictions is as follows. The previous frame (A) and current frame (B) predictions together form the consecutive clip for evaluation. The confusion matrix is built based on whether the sequence is correctly predicted, as shown in the table below.

Table 3 Confusion Matrix for Sequence-Wise Evaluation

Predicted	True Sequence				
Sequence	A-B	A-A			
A-B	TP	FP			
A-A	FN	TN			

The prediction should correctly capture the change in actions for good sensitivity. Hence, the change in action class is marked as positive, and no change is marked as negative. If consecutive clips have different actions, but the same classifications are provided, a 'False Negative' is considered, and when a different classification is provided, a 'True Positive' is considered. Suppose consecutive clips have the same actions, but a different classification is provided for the second frame. In that case, a 'False Positive' is considered, and a 'True Negative' is considered if the same classification is provided.

In general, accuracy and precision metrics are evaluated from the confusion matrix. An issue with these usual metrics occurs when comparing the sequential predictions. The change in the action may be detected at a very different frame than annotations. This can occur due to changes in the frames, which are invisible to the human eye or missed easily during annotations. Hence, a different metric is developed for the specific case using the same elements of the confusion matrix.

$$PC = (TP + FP)/(TN + FN)$$
(1)

$$AC = (TP + FN)/(TN + FP)$$
(2)

$$SPE = (PC / AC) - 1 \tag{3}$$

Where,

TP = True Positive TN = True Negative FP = False Positive FN = False Negative PC = Predicted Changes AC = Actual Changes SPE = Sequence Predictability Error

In cases where the actual changes can be zero, that is, no changes in the actions in the video, the denominator will be considered as 1 to overcome the division by zero error. Also, the predicted changes can be far more than actual changes. The sequence predictability error must be close to zero.

Finally, the time duration evaluation compares the durations for inference of all the models over the clip duration combinations specified earlier. The time taken per inference is captured and averaged for each model, video, and task. As we compare video clip size variations, considering the number of evaluations made within each variation will provide a better metric for model performance evaluation. The average duration per video is divided by the total number of evaluations made within the video to evaluate the performance.

$$T_{NE} = 100 * T_{AE} / N_E$$
 (4)

Where,

 T_{NE} = Normalized Evaluation Duration, in seconds T_{AE} = Average Evaluation Duration, in seconds N_E = Total number of evaluations



Figure 1 Annotation and Evaluation Methodology

The annotation and evaluation methodology is depicted in 'Figure 1' above, except the duration evaluation.

4 Results

4.1 Frame-wise Evaluation

The overall model performances are unexpectedly low. The best-performing model is MAE with 56.42%, followed by MViT model with 49.27%, and I3D model with 43.13%. The clip overlap and gap durations show slight improvements but do not form any meaningful patterns in frame-wise accuracy. Models offer good performance for tasks with action classes available in the Kinetics-400 dataset. Across models, the average accuracy stays in a similar ratio, as shown in 'Figure 2' below.



Figure 2 Average Accuracy Percentage Per Task Type

The need for more accuracy indicates disagreement between manual annotations and model predictions for action classes unavailable in the pre-trained dataset. Utilizing the results from the top three tasks – Excavation, Masonry, and Plastering, the frame-wise accuracies are reported in 'Figure 3' below.



Figure 3 Model-wise Average Accuracy Percentage Per Clip Duration

The I3D model shows a considerable increase in performance with an increase in the input clip duration. After 5-second clip length, a plateauing of I3D performance and a decrement of MViT performance can be observed. The decrement in the MAE model might be related to the fact that it is trained on datasets beyond the K-400 dataset, owing to its architectural needs.

4.2 Sequence-wise Evaluation

Models might capture more action transitions than annotated ones because they can see more details than humans. However, the duration, overlap, and gap combinations also affect the predictive capabilities due to the sampling strategies for testing. A 10-clip x 3-crop strategy takes ten clips from the given video, crops three different zones within each clip, and utilizes the information for prediction. Thus, a longer video duration, an overlap, and a gap between clips will all provide different features.

The Sequence Predictability Error (SPE) of models for different durations is presented from 'Figure 4' to 'Figure 6' below. The negative predictability shows that the models predict less than the actual, and positive values indicate that the models predict more. Being closer to zero is preferred, as the models are expected to perform best in correctly identifying sequences.







Figure 5 Gap Duration-wise Sequence Predictability Error of Models



Figure 6 Overlap Duration-wise Sequence Predictability Error of Models

As the duration increases, the predictability of the models – I3D and MViT- improves. However, the MAE model over-predicts the number of changes in the video. A slight gap in the durations improves the predictability, but too much will throw the models far away. Also, a slight overlap improves the predictability.

4.3 Time Duration Evaluation

The models are compared with different clip lengths and durations and presented in 'Figure 7' below.



Figure 7 Average Evaluation Duration Per Model at Different Clip Lengths

The I3D, a CNN-based model, shows a linear increase in the evaluation duration with increasing clip lengths. Combined with the need for more accuracy beyond 5 seconds, using large clip lengths for CNN-based models is not valuable. The Transformer-based models do not show any linear increase and are stable across the clip lengths.

Although evaluation duration increases with time, that is the case when there is no overlap. Additional overlaps of 2 and 4 seconds did not show the same incremental behavior in performance time. One exception is the 9-second limit for the video clips. The I3D model took longer in any duration-overlap combination while evaluating the clips of length 9 seconds. But beyond the 9 seconds, the evaluation duration reduces. There is no clear explanation for this behavior.



Figure 8 Gap Duration-wise Normalized Evaluation Durations of Models

From 'Figure 8' above, having a large gap between clips increased the normalized evaluation duration across all models. Technically, the models consider each clip a separate video and only predict the action within the clip context. But, the behavior here suggests that the models utilize the previous videos as context.

4.4 Task-wise Evaluation

Breaking down the model performance task-wise, 'Figure 9' below shows the different performance of models for the tasks.



Figure 9 Average Accuracy Percentage per Task Type for Models

The need for correct labels for other tasks limits our

ability to evaluate model performance systematically. However, MViT performs well for concreting and scaffolding, while MAE performs well for formwork.

Task-wise evaluations are focused only on the bestperforming tasks. The models show different behavior under different clip durations as presented in Table 4.

Table 4 Average Accuracy of Models for Best Performing Task Types under Different Clip Lengths

Task Type	Model	Clip Duration (in seconds)							
		1	3	5	7	9			
	I3D	66	68	74	72	76			
Excavation	MViT	92	92	94	95	93			
	MAE	100	95	94	94	95			
Masonry	I3D	77	74	76	75	77			
	MViT	77	77	76	75	71			
	MAE	95	96	96	95	96			
Plastering	I3D	58	71	75	80	74			
	MViT	73	76	77	76	78			
	MAE	90	89	87	84	89			

Table 5 Average Accuracy of Models for Best Performing Task Types for Overlap-Gap Combinations

Task Type	Model	Overlap - Gap Combinations							
		0-0	0-1	0-10	2-0	4-0			
	I3D	72	70	78	73	64			
Excavation	MViT	93	93	92	95	94			
	MAE	95	96	94	95	97			
	I3D	76	76	75	77	77			
Masonry	MViT	74	74	79	74	73			
	MAE	95	96	95	96	96			
	I3D	67	75	69	75	81			
Plastering	MViT	76	76	76	77	78			
	MAE	87	88	89	86	88			

The rounded-off average accuracies for overlap-gap duration combinations is presented in Table 5 above. For the I3D model, increasing the overlap duration increases the model performance for plastering tasks but decreases accuracy for the excavation task. A 2-sec overlap in the I3D model improves the performance when the clip durations are below 9 seconds.

Comparing the task-wise sequence predictability of the models, it was observed that the plastering task has too many change predictions than actual. The results are presented in 'Figure 10' below. Across the tasks, none of the models can predict the number of changeovers sufficiently. Although performance for formwork tasks seems promising, the high background percentage might also lead to this erroneous evaluation.



Figure 10 Task-Type wise Sequence Predictability Error of Models

Separating the task-wise performance, each model shows a different pattern for the duration. Results from the comparison are presented in 'Figure 11' below. The excavation task takes most of the time for evaluation, followed by reinforcement and plastering.



Figure 11 Normalized Evaluation Duration per Task Type for Models

5 Discussion

Summarizing the results across models, duration combinations, and task types, a few patterns are found useful. Frame-wise accuracy suggests that the Transformer models are the best-performing models. The MAE model, which is trained on videos directly, performs better than MViT, which is trained on images and extended to videos. For I3D, a CNN-based model, the clip duration affects the performance by providing a larger context window. However, a 5-second limit occurs across models trained on images. The context window of 5 seconds seems sufficient for our use cases.

For the I3D model, the excavation task with a context window of 5 seconds and the plastering task with a context window of 7 seconds perform best. However, the exact durations are the worst performing for the MAE model. The MAE model is best performing for clips of short durations.

For excavation and masonry, providing a gap between clips increased the accuracy, but extra overlap increased the accuracy for plastering. This suggests that the motion patterns within the former two actions are highly repeated, whereas, for the latter, a considerable difference occurs.

In the sequence predictability, higher clip durations increase the predictability of the I3D and MViT models but reduce the MAE model. A slight gap or overlap in the durations improves the predictability.

For most tasks, all the models predict less than actual sequence changes. Only for the plastering task, the models predict more than the actual. This might be a oneoff case and needs further investigation. The excavation and reinforcement tasks have higher evaluation durations than the average model performances.

In the concreting activity videos, the actions are the concreting of a floor slab and a road. The most repetitive actions are dumping the concrete and evening (smoothing) the surface. The annotations given are 'unloading the truck' and 'sweeping the floor', considering the closeness of these labels to the actions. However, most model detections classified the surface evening as 'digging'. The motion patterns relevant to these two classes need to be differentiated. Similarly, most formwork tasks are annotated under 'moving furniture' since they involve moving and fixing the components. The best-performing model, MAE, detects some of these actions as 'building shed' and 'bending metal'.

There is no pre-training involved in the study, and the annotations are mapped to nearest action class of the Kinetics-400 dataset. Consequently, the accuracy results in task-wise results are not useful for concreting, formwork, scaffolding and reinforcement works. Since the focus of the current work is on temporal precision rather than frame level accuracies, some useful interpretations can be derived from the task-wise results of these actions also. The sequential predictability error identifies how well the model can detect the action switching from one to the next. Even in the mapped action classes, the actual action is irrelevant, and only the change of action is important. From the task-level results, a negative SPE is seen in most cases suggesting that the models predict less switches than that can be detected by the manual annotators. Also, the normalized evaluation duration results hint at a possible correlation with the video clip lengths. The smallest video clip of excavation has only 450 frames and the results of excavation point to a very high duration for evaluation. However, this is contrary to the general expectation that a smaller clip can be evaluated faster. There is no relation found between the accuracy and duration because the accuracies are very less for the reinforcement work yet the action class took higher duration for evaluation. For the action classes existing in the dataset and are directly related to construction activities like masonry, plastering and excavation, results across all evaluations are useful.

These observations lead to the development of a classification system for tasks and models. Excavation tasks can be predicted better with 5-second context windows but need too much time for prediction. Plastering tasks can be predicted best with 7-second context windows, but the number of switches detected can be far higher than actual. Masonry tasks can be predicted with 5-second context windows without any drawback. Overall, additional overlap or gap increases the prediction performance on speed and accuracy. However, models trained on video datasets directly perform the best. If the sequence changes are not a particular concern, then using the video data-trained models with the least clip durations will improve the model performance dramatically. However, if sequence predictability is the primary concern, the transformerbased model trained on the image dataset will work best with a clip duration between 5 and 7 seconds.

Overall, this work suggests that neither the 3-sec context window from the AVA dataset nor the 10-sec context window from the Kinetics-400 dataset is suitable for construction tasks. Instead, it is a variable that needs to be carefully evaluated and considered for better performance.

The study offers useful insights for the construction organizations adopting automated visual surveillance for applications like fatigue and safety monitoring. One of the important components of such systems is the worker action recognition model. Action recognition is a resource-intensive task, requiring the model to classify a fixed set of frames from the input videos into relevant actions. Changing the fixed set of frames to a variable set according to the task is a valuable optimization. Consider an analogy: Suppose you watch a live stream of your favorite action-oriented sport. Usually, there is a minimum of 30-second delay in live streaming the sport, depending on the technology. Now, imagine a computer watching the live match at the location, informing you whether there is any useful action in progress. If it can

analyze the match faster and more accurately, your cost and time will also be saved. In the game, two people can be doing different actions simultaneously, one running towards the opponents and another standing in some corner for quite some time. We too, focus on the running person rather than the standing person. By enabling the computer to focus on the running person, we can save the cost and time of operating the computer while increasing accuracy. Our proposed approach is one such method to achieve this. By remembering how humans move for different tasks, the computers can speed up their performance, by using smaller clip lengths for important actions and vice versa. Using technical terminology, the proposed method improves the speed of action recognition by aligning the clip durations with taskspecific motion patterns. This strategic optimization reduces the data analysis costs while increasing the speed of evaluations for the organizations. Also, some action recognition models depend on upstream models like object detection. When these upstream models cause erroneous detections or missed predictions, our methodology can improve the action recognition as the systems can still make correct predictions due to the varying clip lengths according to the task.

6 Limitations and Future Work

Compared to other related works on action recognition for construction activities, this study stands out in its comprehensive evaluation of different models and clip durations. While most previous studies have focused on a single model or a fixed clip duration, this study provides a more nuanced understanding of how the model performance varies with the clip duration and task type.

However, like all studies, this one also has its limitations. The study is based on a limited number of construction tasks and a specific dataset on which the selected models are pre-trained. The generalizability of the findings to other tasks or datasets remains to be tested. The reliance on YouTube videos for data collection may only partially capture the complexity and diversity of construction activities in real-world scenarios. Future studies could benefit from using a dataset collected directly from construction sites. The study does not discuss potential challenges such as sensitivity to noise, occlusion, camera motion, or lighting conditions, which are common in construction sites. It is expected that the videos collected from YouTube will include these issues in the dataset.

The need for correct labels for all the tasks limits the applicability of the pre-trained models directly. Yet, the work proceeds to utilize the model, and as a result, the respective estimates for tasks like reinforcement could have yielded more useful clip duration estimates. However, these observations proved helpful in studying the model performance for these tasks. The higher evaluation time for reinforcement tasks and higher accuracy of the MViT model for concreting and scaffolding compared to the MAE model are useful predictions for further work.

Lastly, the sequence predictability error metric measures the total number of predicted sequence changes. It does not capture the sensitivity of correctly identifying the change precisely at the annotated frame. This is an approximation, considering that the models might see the frames in videos differently than humans and capture more details for their analysis. Future studies can also modify the metric and measure the duration between annotated and predicted change for more sensitivity.

There are several promising directions for extending this work. One possibility is to incorporate more construction tasks into the study to validate further and refine the findings. Another possibility is to create a dataset of the same set of actions with varying execution time and use the dataset for evaluation following the methodology discussed in the current work. If the performance of models depends on execution time also, then the context-awareness of models can add another dimension of variance. For example, a dataset of masonry work is created with action lengths between 3 and 9 seconds. The context-aware systems can identify the mean and variance from this dataset. In a future application, the system can consider this variance while varying the clip lengths. Another direction is to develop a new dataset with more accurate and diverse annotations, which could help to improve the model's performance and robustness. Fine-tuning or transferring the models to the construction domain could also be explored to exploit the domain-specific knowledge and data. Finally, other datasets (like AVA), modalities (using skeletal frame, optical flow), architectures for action recognition, and tasks (recognition, segmentation, localization) could be investigated. Evaluation for different mechanization levels (manual, tools, equipment, machinery) can be done for more detailed analysis. Additionally, the application of these models to other tasks, such as safety analysis or productivity assessment, could be explored.

7 Conclusion

This paper compares three action recognition models for construction activities: I3D, MViT, and VideoMAE. The models are evaluated on a YouTube video dataset covering seven standard construction processes. The paper analyzes the effect of clip length, overlap, and gap on the model performance, using frame-wise accuracy, sequence predictability error, and normalized evaluation duration as the criteria. The results show that the transformer-based models outperform the CNN-based model in accuracy but have different sensitivity to the clip duration and motion patterns. The paper also suggests that the optimal clip length for construction action recognition is between 5 and 7 seconds, depending on the task and the model. The paper contributes to understanding the strengths and weaknesses of different action recognition models for construction scenarios. It provides insights for developing a dynamic and context-aware selection system for clip durations.

References

- Y. Yu, H. Li, X. Yang, L. Kong, X. Luo, and A. Y. L. Wong, "An automatic and non-invasive physical fatigue assessment method for construction workers," *Autom. Constr.*, vol. 103, pp. 1–12, Jul. 2019, doi: 10.1016/j.autcon.2019.02.020.
- [2] N. Cortes, J. Onate, and S. Morrison, "Differential effects of fatigue on movement variability," *Gait Posture*, vol. 39, no. 3, pp. 888–893, Mar. 2014, doi: 10.1016/J.GAITPOST.2013.11.020.
- [3] K. L. Mudie, A. Gupta, S. Green, and P. J. Clothier, "Adaptation of lower limb movement patterns when maintaining performance in muscle fatigue," *Hum. Mov. Sci.*, vol. 48, pp. 28–36, Aug. 2016, doi: 10.1016/J.HUMOV.2016.04.003.
- [4] J. C. Cowley and D. H. Gates, "Proximal and distal muscle fatigue differentially affect movement coordination," *PLoS One*, vol. 12, no. 2, p. e0172835, Feb. 2017, doi: 10.1371/JOURNAL.PONE.0172835.
- [5] J. Yang, Z. Shi, and Z. Wu, "Vision-based action recognition of construction workers using dense trajectories," *Adv. Eng. Informatics*, vol. 30, no. 3, pp. 327–336, Aug. 2016, doi: 10.1016/j.aei.2016.04.009.
- [6] D. Roberts, W. Torres Calderon, S. Tang, and M. Golparvar-Fard, "Vision-Based Construction Worker Activity Analysis Informed by Body Posture," *J. Comput. Civ. Eng.*, vol. 34, no. 4, p. 04020017, Jul. 2020, doi: 10.1061/(ASCE)CP.1943-5487.0000898.
- [7] C. Gu *et al.*, "AVA: A Video Dataset of Spatiotemporally Localized Atomic Visual Actions," May 2017, Accessed: Dec. 15, 2023. [Online]. Available: http://arxiv.org/abs/1705.08421.
- [8] W. Kay *et al.*, "The Kinetics Human Action Video Dataset," May 2017, Accessed: Dec. 15, 2023.
 [Online]. Available: http://arxiv.org/abs/1705.06950.
- [9] J. Carreira and A. Zisserman, "Quo Vadis, action recognition? A new model and the kinetics dataset," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR* 2017, 2017, vol. 2017-Janua, pp. 4724–4733, doi:

10.1109/CVPR.2017.502.

- [10] Y. Li et al., "MViTv2: Improved Multiscale Vision Transformers for Classification and Detection," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022, vol. 2022-June, pp. 4794–4804, doi: 10.1109/CVPR52688.2022.00476.
- [11] L. Wang *et al.*, "VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking," Mar. 2023, Accessed: Dec. 16, 2023. [Online]. Available: http://arxiv.org/abs/2303.16727.

Appendix 1: Sample Task Frames



Preliminary Study on Enhancing Detection of Concrete Bridge Surface Spalling by Infrared Thermography

Ching-Wei Chen¹, Hui-Ping Tserng¹, En-Wei Chiang¹

¹Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. D09521018@ntu.edu.tw, <u>hptserng@ntu.edu.tw</u>, r10521701@ntu.edu.tw

Abstract

Regular inspections, effective management, and timely maintenance are critical issues to ensure bridge safety and quality. Currently, visual inspection remains the predominant method employed worldwide for bridge inspection. However, visual inspection heavily relies on the training, experience, and subjective judgment of inspectors, leading to inconsistent assessments. When applying deep learning techniques to assist in identifying bridge crack formations, challenges persist. Some images may not clearly display the crack's location. Infrared thermography, with its non-contact, non-destructive properties, effectively detects surface delamination in concrete bridges. However, most research employs higher-spec infrared thermography, which comes with higher instrument costs and less economic viability. Hence, this study aims to investigate the feasibility of using lower-spec infrared thermography to detect surface delamination in concrete bridges as well as analyze the potential of using lower-spec infrared thermography results to assist AI image recognition of bridge surface defects.

Keywords -

Infrared Thermography; Bridge Inspection; Concrete Defects; Passive Infrared Thermography technology; AI Image Recognition

1 Introduction

Regular inspections, effective management, and timely maintenance are critical issues to ensure bridge safety and quality. Currently, visual inspection remains the predominant method employed worldwide for bridge inspection. However, visual inspection heavily relies on the training, experience, and subjective judgment of inspectors, leading to inconsistent assessments.

As for the literature review related to this research objectives, Yahui Liu et al. (2019)[1] propose a deep hierarchical convolutional neural network (CNN) to detect the concrete crack without using thermal imaging cameras. Chia-Chi Cheng et al. (2008) [2] use infrared thermography associated with elastic waves to detect concrete structure defects in the lab. Using this elastic waves in the lab maybe not suitable for this research utilizing to the bridge. This study utilizes the test method modified from ASTM [3] to detect cracks in concrete using infrared thermography. Rocha and Povoas (2017)[4] just review the state of the art using infrared thermography to inspect the concrete bridges. However, the infrared instruments used in these literature [3,4] are too heavy and large to be employed on unmanned aerial vehicles (UAV). For next further research objectives, it will be utilized a lighter and smaller infrared instrument to employ the defect detection with UAV. Nevertheless, there is still a significant challenge to overcome AI image recognition. Previous literature has highlighted the effectiveness of thermal imaging cameras in detecting concrete bridge surface spalling. Currently, this study first aims to investigate the feasibility of employing lowstandard infrared thermal imaging cameras (lighter and smaller) to detect concrete bridge surface spalling and support AI image recognition technology in bridge inspections. In the further research, this lighter and smaller infrared instrument can be easily to be employed to the defect detection with UAV.

Based on laboratory research and on-site inspections of bridges, the FLIR E5 infrared thermal imaging camera has demonstrated the recommended measurement distance of 1 to 2 meters. While it may not be suitable for measuring shallow-depth and small-area defects, it excels in effectively detecting various shapes of spalling and supporting AI image recognition. Consequently, employing low-standard infrared thermal imaging cameras for the detection of concrete bridge surface spalling and integration with AI image recognition technology in bridge inspections appears to be a feasible approach.

2 Methodology

2.1 Research Procedure

This research first comprehends the background of bridge inspection and infrared inspection in Taiwan. It

describes the research motivations, objectives, scope, and limitations, followed by a literature review that gathers relevant domestic and international studies along with pertinent theories. After the literature review, laboratory experiments and on-site bridge inspections will be conducted. In the laboratory experiments, dimensions of test specimens and defect sizes will be determined based on multiple domestic and international studies. Test specimens will be created according to the actual composition of Taiwanese concrete bridges (designed strength of 350 kgf/cm²). The entire process of specimen creation, including template assembly, material preparation, mixing, defect model creation, grouting, and final curing, will be executed independently. During the first phase of experiments, equipment suitability (FLIR E5, shown in Figure 1 and 2)) for measurement dimensions and distances will be determined based on the analysis results. The second phase will primarily explore whether defects of different shapes can be detected similarly. The third phase will involve redesigning and creating new specimens, conducting thermal imaging analysis to assist Deepcrack AI image recognition, and drawing final conclusions. In the bridge inspections, identical measurement and analysis methods will be employed to derive conclusive results.



Figure 1. FLIR tools/tools+ software screenshot



Figure 2. FLIR tools/tools+ Software editing model

2.2 Bridge Crack Inspection Experiments

2.2.1 First-Phase of Experiment

The concrete specimens in this study will be fabricated to match the actual design strength of Taiwanese concrete bridges, which is 350 kgf/cm^2 . The dimensions of the concrete specimens will be $60 \times 70 \times 10$ cm³, with the defective portions made using perlite

boards whose sizes vary according to the experimental phases. The concrete mix will have a water-cement ratio of 0.55, requiring 18.9 kg of cement, 11.1119 kg of water, 29.8918 kg of mainland sand, and 41.2113 kg of coarse aggregates.

Passive infrared thermography will be utilized in this phase of experiments. The prepared specimens will be laid flat in an area devoid of shadows and exposed to direct sunlight. When capturing images, the concrete specimens will be positioned upright, and the FLIR E5 infrared camera will conduct two sets of captures every half hour. Each set will encompass five measurement distances (1m, 2m, 3m, 4m, 5m). Image capture sessions are scheduled from 11:00 a.m. to 2:00 p.m. and from 7:00 p.m. to 9:30 p.m. A distance meter, UNI-T LM80, will ensure the infrared camera is perpendicular to the specimen, and an infrared thermometer, TECPEL-DIT300B, will measure the surface temperature of the specimen.

2.2.2 Second and Third-Phase Experiments

In this phase of the experiment, passive infrared thermography will be employed, and the fabricated specimens will be laid flat in an area without shadows and exposed directly to sunlight. During the capture process, the operator will stand on an A-frame ladder and utilize the FLIR E5 infrared camera to capture images in sets every half hour. Each set will involve two measurement distances (1m, 2m), and the image capture sessions will occur from 3:00 p.m. to 5:00 p.m. A distance meter, UNI-T LM80, will ensure the infrared camera is perpendicular to the specimen, and an infrared thermometer, TECPEL-DIT300B, will measure the surface temperature of the specimen.

2.3 Data Collections

The temperature data acquisition through thermal imaging will be performed using FLIR tools to obtain the average temperature of the defect (T_i) and the average temperature around the defect (T_S). The steps are as follows:

- Import the desired image into FLIR tools (using defect number 1 for demonstration purposes). Double-click to open the image and enter the operating interface. Switch the image mode to digital thermal camera (visible light image).
- (2) In the visible light image within the digital thermal camera mode, use the temperature measurement function to select the area of the defect for measurement. This action will yield the average temperature of the defect. For demonstration purposes using defect number 1 from the first set of specimens (refer to Figure 3), the average temperature (T_1) of defect number 1 (Bx1) is measured to be 42.5°C.



Figure 3. Example of selecting defect temperature value in the visible light image

 The selected area on the defect measures 40×40 (Figure 4) in pixels, representing 1600 temperature values within the selected area. When comparing the 40×40 (pixels) area to the actual defect size of 10×10 (cm²), there exists a 4:1 ratio. Therefore, in this study, the perimeter of the defect is extrapolated to 3 centimetres from the defect distribution (Figure 5), meaning the real size is 16×16 (cm²). FLIR tools utilize a size of 64×64 (pixels). In this example, after the selection process, the overall average temperature (T_A) within the 64×64 (pixels) area is measured at 41.9°C (marked as Bx2 in Figure 5).



Figure 4. Diagram illustrating defect selection in digital thermal camera mage mode



Figure 5. Diagram illustrating selection in infrared imaging mode

2. However, the overall average temperature value (T_A) within the selected 64×64 (pixels) area in Step 3 includes the temperature of the defect itself. To obtain the defect's surrounding average temperature (T_S) required for this study, this temperature should be deducted. Therefore, by employing the following formula:

$$T_{s} = \frac{T_{A} \times A_{t} - \sum_{i=1}^{n} T_{i} \times A_{i}}{A_{t} - \sum_{i=1}^{n} A_{i}}$$
(1)

- T_s : Defect surrounding average temperature
- T_A : Overall average temperature
- T_i : Defect average temperature
- A_t : Total area
- A_i : Defect area
- *i* : Defect numbe

The obtained defect surrounding average temperature (T_s) is 41.52°C. According to the defect assessment method adopted by ASTM-D4788 in this study, to identify concrete defects in thermal imaging, the temperature of the defect must differ by at least 0.5°C from the adjacent area. In the demonstrated example, the temperature difference (Δ_T) is calculated by subtracting the average temperature of defect number $1(T_1)$ from the defect surrounding average temperature (T_s) , resulting in 0.98°C. As this value is greater than or equal to 0.5°C, it indicates that defect number 1 in this example has been successfully detected as a defect.

$$\Delta_T = T_1 - T_s = 0.98 \ge 0.5 \,(^{\circ}\text{C}) \tag{1}$$

3 Research Results

In both the first and second phases of the experiment, the research results will be presented using line graphs (as in Figure 6) and pie charts (as in Figure 7). The third stage will involve the use of line graphs and bar charts for discussion purposes. For on-site bridge inspections, discussions will be presented in tabular form. The following is a summary of the three stages of the experiment.





The total number of measurements for a distance of 1m	30						
Defect number	1	2	3	4	5	6	total
Detectable number	5	4	4	5	5	5	28
ratio	17%	13%	13%	17%	17%	17%	93%

Detectable proportion of different defect shapes when measuring distance 1m



Figure 7. Pie chart depicting the detection rates of different defect shapes at a distance of 1 meter.

3.1 Result of first-phase experiment

- 1. The relative area of the defect's depth significantly affects detection outcomes, with sizes smaller than 1 centimeter in depth proving less detectable by the FLIR E5 infrared thermal imager utilized in this study.
- 2. Defects with a depth of 3.5 centimetres but an area smaller than 1 square centimetre are challenging to detect with the equipment employed in this research.
- 3. The recommended measurement distance for the equipment used in this study is between 1 to 2 meters, allowing for the detection of a broader range of sizes. The average measurement error is also smaller, consistently below 5%.
- 4. The detection rate in this stage of experiments was notably low, attributed to insufficient sunlight exposure. Adjustments in measurement timings during the second stage aim to enhance the detection rate.

3.2 Result of Second-phase experiment

- 1. For the thermal imager FLIR E5 utilized in this study, the detectability at a measurement distance of 1 meter is 93%, and at 2 meters, it is 90%. This indicates a substantial increase in detection rates during the 3:00 p.m. to 5:00 p.m. period compared to the first stage's imaging sessions from 11:00 a.m. to 2:00 p.m.
- 2. The FLIR E5 thermal imager employed in this research can detect not only rectangular defects but also various other shapes, aligning with practical applications in bridge inspection.

3.3 Result of Third-phase experiment

1. Infrared thermal imaging results indicated a recognition rate of 38% for the fourth set of specimens (pure cracks) and 88% for the fifth set

(cracks with localized delamination). This suggests that the presence of localized delamination around cracks contributes to enhancing passive infrared thermal imaging detection.

- 2. In Deepcrack AI image recognition, the detection rate for the fourth set of specimens (cracks) stood at 100%, surpassing the detection rate of 88% for the fifth set (cracks with localized delamination). This indicates that localized delamination might potentially lower the detection rate in Deepcrack AI image recognition.
- 3. During false defect recognition in Deepcrack AI image recognition, the misjudgement rates for the fourth set of specimens (pure cracks) and the fifth set (cracks with localized delamination) were 63% and 50%, respectively. However, in infrared thermal imaging, the misjudgement rates for both were 0%.

Based on the above, it's evident that infrared thermal imaging effectively assists AI image recognition in defect detection. Defects with very shallow depths or small areas may not be detectable by the FLIR E5 infrared thermal imager used in this study. However, given that practical applications often prioritize identifying larger or deeper defects that pose higher risks, the method used in this study is suitable for bridge inspection practices.

4 Conclusion

- 1. Based on the experimental results, this study recommends using the FLIR E5 infrared thermal imager for imaging at an optimal measurement distance of 1 to 2 meters. This range effectively detects larger, deeper, and variously shaped defects posing higher risks.
- 2. Laboratory research and on-site bridge inspections indicate that the FLIR E5 infrared thermal imager is less suitable for measuring defects with shallower depths or smaller areas. However, defects with shallower depths or smaller areas do not significantly impact bridge safety.
- 3. The detection rate significantly improves during imaging sessions between 3:00 p.m. and 5:00 p.m. compared to those from 11:00 a.m. to 2:00 p.m., suggesting that using passive infrared thermal imaging for defect detection is advisable when the subject has adequate exposure to heat.
- 4. The study observed that elongated and small-area defects resembling cracks are less detectable by the FLIR E5 infrared thermal imager used in this research. However, the presence of localized delamination around cracks enhances the effectiveness of passive infrared thermal imaging.
- 5. In the third stage of experiments, it was found that

the presence of localized delamination around cracks improves the detection rate of passive infrared thermal imaging but lowers the detection rate of AI image recognition. Additionally, AI image recognition exhibits a significantly higher false defect identification rate compared to infrared thermal imaging.

The contribution of this work is to confirm that smaller and lighter infrared devices can enhance the accuracy of detecting concrete cracks. These smaller and lighter devices can be easily installed on unmanned aerial vehicles in the future research to conduct real-time bridge inspection work.

References

- [1] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, Li Li (2019) .Deepcrack : A deep hierarchical feature learning architecture for crack segmentation.
- [2] Chia-Chi Cheng, Tao-Ming Cheng, Chih Hung Chiang, (2008). Defect detection of concrete structures using both infrared thermography and elastic waves.
- [3] D4788 ASTM (1997): Standard Test Method for Detecting Delaminations in Bridge Decks Using Infrared Thermography.
- [4] J. H. A. Rocha, Y. V. Povoas (2017). Infrared thermography as a non-destructive test. for the inspection of reinforced concrete bridges : A review of the state of the art.

Variable Speed Limits Control for Smart Work Zone with Connected Vehicles

Shuming Du¹, Saiedeh Razavi¹, and Harith Absulsattar¹

¹Department of Civil Engineering, McMaster University, Canada dus5@mcmaster.ca, razavi@mcmaster.ca, abdulh30@mcmaster.ca

Abstract

Freeway work zones with lane closures can cause traffic disruption and congestion with increased travel time, safety risks, and emissions. Variable speed limit (VSL) control has been widely studied to mitigate the negative impacts of work zone lane closures. To ensure effective VSL control, accurate detection of traffic conditions is important. However, stationary sensors from distant upstream or merging areas of a work zone can only provide location-based detections, which may not be sufficient for effective VSL control. Therefore, this study proposes a VSL control system for a smart work zone with a weighted traffic density estimation algorithm using data from stationary sensors and connected vehicles (CV). With location-based traffic detections from distant upstream and merging areas of a smart work zone, data from CVs in locations between the stationary sensors are used to obtain the weighted density. This weighted density is then fed to a feedback VSL controller, which can dynamically define the appropriate speed limits to reduce the negative impacts of the work zone closures. The proposed VSL controller was implemented and evaluated under a calibrated freeway work zone environment using the traffic simulator SUMO. The results show that the system can provide accurate traffic density estimation and effectively improve traffic mobility, safety, and environmental sustainability near the work zone area.

Keywords -

Variable Speed Limits Control; Stationary Sensors; Connected Vehicle; Weighted Traffic Density; Smart Work Zone

1 Introduction

Freeway work zones play an important role in maintaining and upgrading freeways. However, closing one or more lanes in a work zone area can cause bottlenecks and traffic disruption near work zone areas. This disruption can easily lead to congestion, increased travel time, safety risks to motorists and work zone workers, emissions, and fuel consumption. Therefore, developing an effective strategy and mitigating the negative impacts of work zones is essential.

Variable Speed Limit (VSL) control is one of the Intelligent Transportation Systems that has been studied to improve traffic conditions near work zone areas. By generating dynamic speed limits, VSL control can regulate the travel speeds of vehicles approaching a work zone to alleviate the impacts of capacity loss due to lane closures. Different VSL controllers have been developed using control strategies such as rule-based [1], feedbackbased [2], and reinforcement learning [3] approaches. These VSL controllers rely on accurate traffic condition detections to ensure the control effectiveness.

Stationary sensors can detect traffic conditions at fixed locations. Many VSL controllers have used traffic measurements by fixed sensors near work zone areas to achieve control objectives [4][5]. In practice, fixed sensors may be present at distant locations upstream of a work zone and may not provide accurate traffic measurements in work zone areas. However, sensors at a smart work zone's merging area can detect traffic conditions immediately upstream of a work zone. The density variations between locations of distant upstream and merging areas of a work zone can hardly be estimated using stationary sensors unless dense stationary sensors are installed, which is cost-prohibitive.

Connected vehicles (CVs) can provide real-time traffic state observations along their travel path. When a CV travels through work zone areas, the vehicle can become a probe sensor, sending the location and travel speed to fill the gap between the stationary sensors. Many approaches have used massive data from connected vehicles to estimate traffic states [6] [7]. However, work zones particularly short-term work zones exist for a limited period of time, which makes it challenging to collect a large amount of historical CV data for traffic state estimation.

This paper proposes a VSL control system with a weighted traffic density estimation algorithm for a smart work zone using stationary sensors and CV data. CV data is used to bridge the gap that exists between the data from sensors located at distant upstream and the ones in merging areas of a smart work zone.

The rest of this paper is organized as follows: studies on VSL control and density estimation are reviewed first. Then, the development of the weighted density approach and VSL controller are discussed in the Methodology section. Later, the system evaluation and results are discussed. Finally, the conclusion and future work are presented in the conclusion section.

2 Background and Literature Review

A hypothetical freeway smart work zone scenario in which one lane is closed is shown in Figure 1. Due to the capacity loss, traffic congestion occurs at the merging area of the work zone with a queue propagating upstream. The work zone is a smart zone with two sets of stationary sensors. Traffic Sensor (TS) 1 is a stationary sensor that is present upstream of the work zone, and TS 2 is the smart work zone stationary sensor at the merging area of the work zone. CVs capable of transmitting speed and location, along with conventional vehicles with no connectivity travel through the work zone area from free flow to congested and then recover to free flow once they pass the bottleneck area.



Figure 1. Layout of a Hypothetical Freeway Work Zone with Stationary Sensors and CVs

The traffic queue forms first at the merging area and then propagates upstream. Since it takes time for the queue to reach the location of TS 1, the traffic states detected by TS 2 could be heavily congested, while the traffic states from TS 1 could be free flow. The discrepancies in traffic detections from stationary sensors due to different sensor locations could impact the effectiveness of VSL Control.

VSL control has been studied to mitigate congestion near bottleneck areas with lane closures. The control strategies can be broadly categorized into rule-based [1], feedback-based [2], and reinforcement learning-based [3] controllers. Rule-based VSL controllers, such as the fuzzy-logic controller [8], post speed limits using predefined rules and thresholds of traffic flow, density, and speed. The control objective, such as the critical traffic density [2], is used by feedback-based VSL controllers to obtain speed limits. Reinforcement learn-based VSL controllers leverage learning algorithms and traffic measurements to train a model and generate dynamic speed limits to reduce travel time [3]. All these VSL controllers require accurate traffic state measurements. However, the types of traffic sensors, e.g. stationary sensors and probe vehicles, and the locations of sensors are commonly not investigated in such studies. The sensor factors could impact the accuracy of traffic state estimation near work zone areas and, consequently, the effectiveness of VSL control.

Dense stationary installation sensor could approximate the continuous measurements in space to provide higher accuracy in traffic state estimation. However, the high installation and maintenance costs are prohibitive for smart work zone systems. Instead of dense sensor installation, mathematical models, such as observers, are developed to provide traffic estimations using nearby sensors when sensor detections are missing [9][10]. For such models, the traffic fundamental diagram is calibrated to minimize the discrepancies between model-generated and real sensor data so that model-generated traffic state estimations can be used by applications such as VSL control [11]. However, traffic congestion can easily occur near work zone areas due to capacity loss, which leads to frequent traffic state transitions from free flow to congestion when a queue starts forming or from congestion to free flow when a queue starts resolving. This traffic state transition would make it difficult for mathematical approaches to estimate traffic state accurately between the locations of stationary sensors.

CVs travelling through work zone areas can be considered as probe sensors to estimate traffic transition between the locations of stationary sensors. Relying on the conservation law of traffic flow, the traffic state is estimated using the spacing and speed of CVs [12]. A deep convolutional neural network approach is studied to estimate traffic speed and congestion qualitatively and quantitatively using a large amount of historical probe vehicle data from different congestion scenarios [13]. [7] proposes a macroscopic model to estimate the traffic states of that segment based on the CV market penetration rate. These approaches use probe vehicles with either advanced onboard sensors like radar to track the headway between vehicles or a large amount of historical CV data to estimate traffic states. However, advanced onboard sensors may not be always available on CVs and lack of a large amount of historical CV data could occur due to the limited time periods of works zones, which would make it challenging to apply these probe sensors approaches and accurately estimate traffic states near work zone areas.

In this study, a VSL control system with a weighted density estimation algorithm was developed. Considering stationary sensors and CVs, weighted density is calculated and fed into a VSL controller. Specifically, sensors from distant upstream and merging areas of a smart work zone are used to obtain locationbased traffic densities. Then, the weight of two locationbased densities is derived using CV vehicles between the two sensor locations based on kinematic wave theory. The system proposed in this study mainly offers the following contributions:

- It investigates whether sensor locations have any impacts on VSL control. The effectiveness of VSL control is evaluated using fixed sensors at distant upstream and merging areas of a work zone.
- It provides weighted traffic density estimations for VSL control. The weight is calculated using CVs based on kinematic wave theory.
- It produces good VSL control performance, despite the impacts of stationary sensors locations at distant upstream and merging areas of a work zone. The system evaluation shows consistent improvement in mobility, safety, and sustainability.

3 Methodology

The effectiveness of VSL control relies on the accuracy of traffic state estimation using traffic sensors near work zone areas. A VSL control system with a weighted density algorithm is illustrated in Figure 2. The weighted traffic density is obtained using stationary sensors and CVs to approximate traffic density at a work zone area. By using this weighted density, a VSL controller is designed to regulate traffic speeds. The framework of the proposed system is discussed, followed by traffic density estimation and VSL control in the subsequent subsections.



Figure 2. VSL Control System Framework

A feedback-based VSL controller is developed with critical traffic density as the control objective. The VSL controller processes the critical density and weighted density error to calculate speed limit. These speed limits regulate vehicle travel speeds and influence traffic conditions in the vicinity of the work zone area. This study assumes a regulatory VSL system with the full driver compliance.

The weighted density component of the framework

effectively approximates the traffic density within the work zone area. The baseline traffic density is detected from sensors at the distant upstream and the merging area and then is adjusted by weights that are calculated using the speed and location of CVs. The weighted density is obtained by applying the weight to traffic density from stationary sensors.

Comparing the error between the critical density and weighted density, VSL control will generate dynamic speed limits to minimize the error. This proactive approach ensures that the traffic flow can sustain the maximum work zone capacity, thereby enhancing traffic conditions without inducing a capacity drop [14]. The capacity drop phenomenon occurs when a queue forms at the bottleneck area leading to increased capacity loss [15]. Details of the estimation of the weighted traffic density and VSL controller are presented in the subsequent sections.

3.1 Traffic Density Estimation

Traffic sensors at a work zone's distant upstream and merging area can detect location-based traffic density. However, the two sensor locations may have a density discrepancy, as illustrated in Figure 1. A weighted density is calculated using CVs based on kinematic wave theory to address this discrepancy.

The traffic state transition is shown in the triangular fundamental diagram in Figure 3.



Figure 3. Traffic States from Fundamental Diagram

The triangular fundamental diagram has the free flow branch with densities smaller than the density ρ_r and the congested branch with densities larger than the density ρ_r .States *a*, *d*, and *c* lie on the free flow branch while States *b* and *b'* lie on the congested branch. State *a* represents high traffic demand, which is higher than the work zone capacity *b'* at the congested status. Because the demand is higher than the road capacity, congestion occurs at the merging area and leads to the capacity drop phenomenon. Traffic flow further drops to State *b*. State *c* represents low traffic demand, lower than the work zone capacity so that the queue forms due to congestion can be resolved. State *d* represents the same work zone capacity at the free flow status. v_b and v_f represent travel speeds at the State *b* and *a* respectively. State *d* and *c* share the same free-flow speed v_f with State *a*. Traffic density ρ_a under free flow status and ρ_b under congested status for State a and b are also illustrated in Figure 3. The VSL control objective is to maintain traffic flow at the work zone capacity State *d* without a capacity drop. The work zone critical density ρ_{cr} as the VSL control goal corresponding to State *d* is shown in the figure as well, which will be discussed in detail under the VSL Control section below.

Figure 4 illustrates the traffic state transition from the space-time diagram under the kinematic wave. The bottleneck location is bold in blue. Two traffic sensors, as shown as grey blocks in Figure 4, are located in two areas, one close to the bottleneck at the merging area and the other at the distant upstream of the bottleneck. When the high traffic demand reaches the bottleneck, a queue forms at the bottleneck area and transfers the traffic state from *a* to *b*. This queue propagates upstream at the shockwave speed ω_{ab} . When the low traffic demand *c* appears and meets at the queue, this queue starts resolving and propagating downstream towards the work zone at shock wave speed ω_{cb} . Once the queue is resolved, the traffic state *c* will prevail along the work zone area.



Figure 4. Traffic States from Space-time Diagram

From Figure 4, traffic states at two traffic sensors are different until the queue propagates upstream and reaches the distance upstream of the work zone. Consequently, the traffic densities $\rho_{ma}(k)$ and $\rho_{mb}(k)$ from measurements of sensors located distant upstream and merging area of the work zone respectively, likely face discrepancies before the queue reaches the sensor location distant stream. Therefore, a weighted density ρ_w is calculated using Equation (1) below.

$$\rho_w(k) = (1 - \alpha(k))\rho_{ma}(k) + \alpha(k)\rho_{mb}(k) \qquad (1)$$

The discrete time step is represented by k. The weight $\alpha(k)$ is applied to traffic density $\rho_{ma}(k)$ and $\rho_{mb}(k)$ detected by traffic sensors at time step k to calculate the weighted density $\rho_w(k)$.

To obtain the weight $\alpha(k)$, the speed and location transmitted by CVs are utilized. In Figure 4, at time step

k, the length of the queue from the bottleneck is l_1 and the total distance between two traffic sensors is l_2 . Because of the queue propagating upstream, at the congested area within the queue l_1 , vehicles travel at the speed v_b while vehicles outside the queue but between two traffic sensors l_2 travel at the speed v_f . Therefore, the ratio of l_1 and l_2 is used to calculate the weight as

$$\alpha(k) = 0.5 + 0.5 \, l_1(k) / l_2(k) \tag{2}$$

From Equation (2), the weight $\alpha(k)$ equals 0.5 when the queue length is 0, meaning the same weight is applied to traffic densities detected by sensors at distant upstream and merging areas of a work zone when there is no queue or congestion. Only traffic density from the merging area is used when the queue reaches the sensor located distant upstream. Therefore, more weights are assigned to the sensor at the merging area when the queue becomes longer.

To determine the location of a queue $l_1(k)$, the speed and location of CVs are used. We assume the speed of CVs is similar to conventional vehicles, as studied in [7]. Then the length $l_1(k)$ can be approximated in Equation (3) as the distance between a CV, which is at the furthest upstream of the bottleneck travelling below a speed threshold v_{tr} , and the location of the bottleneck. $l_{i,CV}$ represents the distance between a CV and the bottleneck. This CV satisfies the requirement in Equation (4) where the CV $l_{i,CV}(k)$ is not larger than l_2 and speed $v_{i,CV}(k)$ is not higher than v_{tr} .

$$\begin{cases} l_1(k) = \max(l_{i,CV}) \\ l_{i,CV}(k) = distance \ between \ a \ CV \ and \\ bottleneck \end{cases}$$
(3)

$$\begin{cases} l_{i,CV}(k) \leq l_2 \\ v_{i,CV} \leq v_{tr} \end{cases}$$
(4)

The speed threshold v_{tr} is obtained using Equation (5).

$$v_{tr} = \frac{v_f + v_b}{2} \tag{5}$$

The free flow speed v_f and speed v_b at state b can be obtained from the fundamental diagram in Figure 3. It should be noted that there may be cases where no CVs travel below the speed threshold v_{tr} or no CVs are available between the two traffic sensors located at distant upstream and merging areas due to a low market penetration rate. In those cases, the same weights are assigned to traffic densities detected by the two stationary sensors at different locations.

3.2 Variable Speed Limit Control

3.2.1 Controller

A feedback VSL controller [16] can generate dynamic speed limits by controlling traffic density at a work zone area towards the control objective, which is the critical density ρ_{cr} as shown in Figure 3. When the traffic state can be maintained at State *d* in Figure 3, the flow rate can be preserved at the maximum work zone capacity without causing a capacity drop.

The VSLs can be calculated using Equations (6) and (7) below.

$$\beta(k) = \beta(k-1) + K(\rho_{cr} - \rho_w(k)) \tag{6}$$

$$v_{vsl}(k) = \beta(k)v_{vsl}(k-1) \tag{7}$$

The factor $\beta(k)$ is calculated using the factor $\beta(k - 1)$ from the last time step and the error between the weighted density $\rho_w(k)$ and the critical density ρ_{cr} . The parameter K is the gain applied to the density error. By applying the factor $\beta(k)$ to the speed limit $v_{vsl}(k - 1)$ from the last time step, the current speed limit is obtained.

The setup of this feedback VSL control system consists of a VSL control zone, acceleration zone and work zone. VSLs are posted at the VSL control zone to regulate traffic speed while vehicles passing the VSL control zone into the acceleration zone can accelerate to free flow speed before entering the work zone. Readers can refer to [13] for more details of this VSL control zone and accelerate zone setup.

3.2.2 Speed Limit Constraints

Constraints are considered when VSLs are applied in practice. The following constraints are included in the VSL control system:

- Speed limits are within the range of the minimum v_{min} and the maximum v_{max} Speed limit.
- Speed limits are discrete values rounded to the close integer speed limits at the incremental speed Δv .
- Speed limits between two consecutive time steps are within the maximum incremental speed. Δv_{max} .
- Speed limits control time interval is Δt , so frequent speed limit changes can be avoided.

4 Experiment and Results

4.1 Experimental Setup

The proposed system was evaluated using a calibrated

freeway work zone segment on SR99 northbound in California. The freeway segment has a maximum speed limit of 105 km/h (65 mi/h). A work zone with a closure of one of the three lanes was observed starting from State Postmile (PM) 19.7 on May 3, 2018, as shown in green in Figure 5. Traffic congestion occurred in the presence of the work zone. A 2.2 km freeway segment upstream of the work zone and a 650 m two-lane work zone was built in the simulation environment under the traffic simulator SUMO. The traffic network was calibrated using travel speed and flow from loop detectors in California Freeway Database at work zone areas under the static speed limit [17].

Two loop detectors were added in the simulated freeway segment to study the impacts of stationary sensors and CVs. One set is located 500 m upstream of the work zone [18] and acts as the stationary sensor at a distant upstream of the work zone. The other set was added 50 m upstream of the work zone and acted as the stationary sensor at the merging area. Both loop detectors can detect traffic density at respective locations every 15 s. The market penetration rate of CVs was set as 10% to transmit speeds and locations as probe sensors. The length of the VSL control zone and acceleration zone were set as 850 m and 550 m, respectively. The lengths of the VSL control zone and acceleration zone could range from 500 m to 1 km, as studied in [14]. Further studies on more selections of sensor locations near work zones could be assessed in future work.



Figure 5. Freeway Work Zone on SR99 Northbound in California

The fundamental diagram was calibrated with free flow speed v_f , critical density ρ_{cr} , the threshold speed v_{tr} , as 105 km/h (65 mi/h), 50 veh/km, and 65 km/h. The gain K was selected as 0.01 via extensive tests. The maximum speed limit v_{max} , minimum speed limit v_{min} , incremental speed Δv , and the maximum incremental speed Δv_{max} and control time interval Δt were selected as 105 km/h (65 mi/h), 24 km/h (15 m/h), 8 km/h (5 mi/h), 8 km/h (5 mi/h) and 60 s.

4.2 **Results Analysis**

To demonstrate the effectiveness of the proposed

VSL control system, four scenarios are considered in the simulation: 1) no VSL control, in which the traffic conditions were collected as the baseline; 2) VSL control with only the stationary sensor at distant upstream of the work zone; 3) VSL control with only stationary sensors at the merging area of the work zone and 4) VSL control with stationary sensors at distant upstream and merging area of the work zone and CVs. The calibrated freeway work zone segment was first built using the traffic simulator SUMO. An interface TraCI was used to communicate with SUMO to retrieve simulated vehicle and network data and update speed limits [19]. Traffic demand for the simulation was generated as follows. The simulation was run for 600 s as the warm-up period. Simulation data from this warm-up period was discarded. Then, the simulation was run for 5400 s. The low traffic demand of 3000 veh/h was generated for the first 1200 s, and gradually increased to 3700 veh/h at 1800 s and maintained at this high demand for 1500 s. Following the high demand, traffic demand gradually reduced to 2500 veh/h at 3900 s and dropped to 2000 veh/h until the end of the simulation.

The traffic density within the acceleration zone upstream of the work zone under the four scenarios can be found in Figure 6.

The black dash line refers to the traffic density collected using the traffic simulator, which is used as the density reference. The detected density refers to the density outputs from the stationary sensors while the weighted density is derived from outputs of stationary sensors using Equation (1).





Figure 6. Traffic Density Comparison under (a) no control (b) VSL control with stationary sensors at distant upstream (c) VSL control with stationary sensors at merging area (d) VSL control with both stationary sensors and CVs

Figure 6 shows that the density under no control increases to around 150 veh/km when traffic demand is high and reduces to around 20 veh/km when the low traffic demand appears at the end of the simulation.

Despite VSL control being in effect under scenarios (b) and (c), traffic density has reached above the critical density and caused some control failure when traffic demand is high. This control failure is mainly due to impacts of sensor locations. Under scenario (b), because stationary sensors are located at distant upstream of the work zone, sensors cannot detect congestion in a timely manner before 1800 s when a queue has already formed at the work zone at around 1500 s. This congestion detection delay causes a delay for VSL control, and consequently, the queue is not effectively resolved. Under scenario (c), although traffic congestion is detected once the queue forms at the work zone area, it overestimates traffic density between 1500 s and 2200 s, which shows that the detected density is significantly higher than the density reference. This overestimation restricts traffic heavily and causes more traffic to enter the work zone, and congestion cannot be effectively resolved. Under scenario (d), the weighted density can provide relatively more accurate density estimation for VSL control. Accordingly, VSL control is more effective under scenario (d) with density remains around the critical density.

The density estimation error analysis of scenarios (b), (c), and (d) can also be found in Table 1. RMSE of Traffic Density. The results show that scenario (d) has the lowest RMSE (root mean square error) of density.

Table 1. RMSE of Traffic Density

Scenario	Density (veh/km)
(b)	23.1
(c)	15.0
(d)	8.2

The speed profile upstream of the work zone, the work zone, and downstream of the work zone under all four scenarios can be found in Figure 7. The time-space diagram of Figure 7 is color-coded by travel speeds. The speed color-coded bars on the right of the sub-figures use different colors representing different travel speeds from 0 (dark red) to 110 km/h (dark blue). Different vehicle travel speeds are shown on the time-space diagram of Figure 7. Traffic congestion occurs under no control scenario and the queue propagates towards upstream first. Then, when the traffic demand reduces, the queue propagates downstream until it is fully resolved. Under scenario (b), the queue forms only at the beginning of the congestion around 1800 s due to the delay of traffic density, while under scenario (c), the queue forms when higher traffic demand reaches the work zone due to traffic density overestimation.





Figure 7. Speed Profile under (a) no control (b) VSL control with stationary sensors at distant upstream (c) VSL control with stationary sensors at merging area (d) VSL control with both stationary sensors and CVs

Scenario (d) shows limited congestion occurs in the work zone area, and vehicles can remain at high travel speed once passing the control zone.

The performance of the VSL control system in terms of travel time, NOx, CO2 emissions, fuel consumption and time-to-collision risks is also assessed for the four scenarios as shown in Table 2. The travel time, NOx, CO2 emissions, and fuel consumption were obtained using the built-in libraries with default models [19]. The time-tocollision risks were calculated using the minimum 1.5 s as the safe distance [13].

Table 2. VSL Control Performance

Scenario	Travel Time (min)	NOx (kg)	CO2 (t)	Fuel (l)	Time-to- Collision (%)
(a)	3.0	1.0	2.6	882.8	11.5
(b)	2.7	0.9	2.4	775.8	2.5
(c)	3.2	1.0	2.6	821.1	2.5
(d)	2.6	0.9	2.4	767.6	0.7

Scenario (d) shows the consistent improvement in travel time, NOx, CO2 emissions, fuel consumption and time-to-collision risks. However, scenario (c) causes longer travel time compared to the no-control scenario, while the lower fuel consumption and time-to-collision risks than no control scenario are shown. Scenario (b) also demonstrates consistent improvements, which are not as significant as the scenario (d).

5 Conclusions

A VSL control system for a smart work zone with CVs is proposed in this study to address the discrepancies in traffic detections from different stationary sensors. Traffic densities from stationary sensors at the distant upstream and merging area of the work zone are weighted by applying kinematic wave theory and using speed and location transmitted from CVs. The weighted density approximates traffic density at work zone areas with lower estimation error. While the discrepancies caused by the locations of stationary sensors negatively impact the VSL control performance, the weighted density ensures the effectiveness of VSL control. The results show the proposed VSL control system can consistently improve traffic mobility, safety, and sustainability near work zone areas.

References

- [1] Allaby P, Hellinga B, and Bullock M. Variable Speed Limits: Safety and Operational Impacts of a Candidate Control Strategy for Freeway Applications. *IEEE Transactions on Intelligent Transportation Systems*, 8(4):671-680, 2007.
- [2] Iordanidou GR, Papamichail I, Roncoli C, Papageorgiou M. Feedback-Based Integrated Motorway Traffic Flow Control With Delay Balancing. *IEEE Transactions on Intelligent Transportation Systems*, 18(9):2319-2329., 2017.
- [3] Li Z, Liu P, Xu C, Duan H, Wang W. Reinforcement Learning-Based Variable Speed Limit Control Strategy to Reduce Traffic Congestion at Freeway Recurrent Bottlenecks. *IEEE Transactions on Intelligent Transportation* Systems, 18(11):3204-3217., 2017.
- [4] Pu Z, Li Z, Jiang Y, Wang Y. Full Bayesian Before-After Analysis of Safety Effects of Variable Speed Limit System. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1-13, 2020.
- [5] Yang X, Lu Y (Carl), Lin Y. Optimal Variable Speed Limit Control System for Freeway Work Zone Operations. *Journal of Computing in Civil Engineering*, 31(1):04016044, 2017.
- [6] Al Mallah R, Quintero A, Farooq B. Cooperative Evaluation of the Cause of Urban Traffic Congestion via Connected Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 21(1):59-67, 2020.
- [7] Bekiaris-Liberis N, Roncoli C, Papageorgiou M.

Highway Traffic State Estimation With Mixed Connected and Conventional Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 17(12):3484-3497, 2016.

- [8] Li D, Ranjitkar P. A Fuzzy Logic-based Variable Speed Limit Controller. *Journal of Advanced Transportation*, 49(8):913-927, 2015.
- [9] Du S, Razavi S. Fault-tolerant Control of Variable Speed Limits for Freeway Work Zone Using Likelihood Estimation. *Advanced Engineering Informatics*, 45:101133, 2020.
- [10] Du S, Razavi S. Fault-Tolerant Control of Variable Speed Limits for Freeway Work Zone With Recurrent Sensor Faults. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):10815-10826, 2022.
- [11] Bramich DM, Mónica Menéndez, Lukas Ambühl. Fitting Empirical Fundamental Diagrams of Road Traffic: A Comprehensive Review and Comparison of Models Using an Extensive Data Set. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):14104-14127, 2022.
- [12] Lee H, Lee J, Chung Y. Traffic Density Estimation Using Vehicle Sensor Data. *Journal of Intelligent Transportation Systems*, 26(6):1-15, 2021.
- [13] Rempe F, Franeck P, Bogenberger K. On the Estimation of Traffic Speeds with Deep Convolutional Neural Networks Given Probe Data. *Transportation Research Part C: Emerging Technologies*, 134:103448, 2022.
- [14] Du S, Razavi S. Variable Speed Limit for Freeway Work Zone with Capacity Drop Using Discrete-Time Sliding Mode Control. *Journal of Computing in Civil Engineering*, 33(2), 2019.
- [15] Yuan K, Knoop VL, Hoogendoorn SP. Capacity Drop: Relationship Between Speed in Congestion and the Queue Discharge Rate. *Transportation Research Record*, 2491(1):72-80, 2015.
- [16] Grumert EF, Tapani A, Ma X. Characteristics of Variable Speed Limit Systems. *European Transport Research Review*, 10(2), 2018.
- [17] C. D. of Transportation. Caltrans Performance Measurement System (PeMS). On-line: http://pems.dot.ca.gov/, Accessed: 10/12/2023.
- [18] Carlson R. C., Papamichail I. and Papageorgiou M. Local Feedback-based Mainstream Traffic Flow Control on Motorways Using Variable Speed Limits. *IEEE Transaction on Intelligent Transportation Systems*. 12(4): 1261–1276, 2011.
- [19] Krajzewicz D., Erdmann J., Behrisch M., and Bieker L. Recent Development and Applications of SUMO – Simulation of Urban MObility. International Journal on Advances in Systems and Measurements, 5(3/4), 128-138, 2012.

VL-Con: Vision-Language Dataset for Deep Learning-based Construction Monitoring Applications

Shun-Hsiang Hsu¹, Junryu Fu² and Mani Golparvar-Fard³

¹PhD student of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, USA ²MSc of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, USA ³Professor of Civil Eng., Computer Science, and Technology Entrepreneurship, University of Illinois Urbana-Champaign, USA

hsus2@illinois.edu, junryuf2@illinois.edu, mgolpar@illinois.edu

Abstract -

Recently, vision-language research has gained significant interest by successfully connecting visual concepts to natural language, advancing computer vision-based construction monitoring using a wide variety of text queries. While vision language models demonstrate high capability, performance degradation can be expected when adapting the model to the ever-changing construction scenarios. In contrast to the source image-text pairs, it is more challenging to cover the multitude of potentially involved objects and their naming conventions for construction activities. To bridge the domain gap, this study aims to collect construction-specific imagetext pairs of building elements and related site work based on the ASTM Uniformat II. The image-text pairs of 641 activities in Uniformat are retrieved from the LAION-5B dataset based on the image and text embeddings using CLIP with two different prompts. Then, the collected images are labeled at the image level to conclude the requirements of the vision-language datasets for further development. Based on the results, a vision-language dataset, VL-Con, consisting of image-text pairs for construction monitoring applications is proposed with the aid of a construction semantic predictor and prompt engineering. The proposed VL-CoN dataset can be accessed through https://github.com/huhuman/VL-Con.

Keywords -

Vision-Language Dataset; Construction Monitoring; Foundation Model

1 Introduction

Vision tasks such as scene understanding and object recognition have been widely studied for automation in construction monitoring [1]. A significant amount of reality capture data is collected to generate actionable insights for construction monitoring [2]. Supervised learning models, such as Mask R-CNN [3] and SegFormer [4], have been predominantly adopted in the past decade [5], necessitating large-scale annotated image datasets with hardcoded indexes of the target objects. The results can be utilized to compare with BIM or 4D BIM for progress monitoring. However, for recognition at a finer level (e.g., image segmentation), labeling images is extremely expensive, and such large-scale datasets still remain unavailable in the industry [6]. Without details in project schedules in today's Virtual Design and Construction (VDC) and Project Control practices, the lack of formal definitions around what needs to be segmented in photos causes the gap between different construction monitoring applications.

Applications based on supervised learning methods are limited to pre-defined classes and require extensive post-processing to derive actionable insights. Integrating knowledge from different models or enhancing their capability to broaden the recognition scope proves to be a challenging task. While the AEC/O research community has been focusing on these application-driven challenges, the AI community has worked on developing foundation models, such as MAE [7] and GPT-4 [8], respectively for vision and language understanding. In this research line, emphasis has been placed on self-supervised techniques, allowing models to learn from large-scale data without any annotation. Well-trained foundational models can be seamlessly adapted to generate fine-grained segmentation masks for various objects [9] and to create a chat agent capable of handling diverse requests [10]. The robustness and generalization of learned knowledge enable the applications for a wide range of downstream tasks.

Since large language models (LLMs) have achieved remarkable success in language understanding through selfsupervised learning from the billion-scale training corpus, the robustness of text embeddings has turbocharged the development of open-ended vision systems by incorporating arbitrary text queries. Vision-language models have gained increasing interest in enhancing vision understanding by aligning image features with text embeddings in the latent space. Radford et al. [11] proposed the visionlanguage model, CLIP, using linear projection to map the outputs of the image and text encoders through contrastive pre-training. CLIP under natural language supervision has shown promising zero-shot transfer performance on various public image datasets. Peng et al. [12] used CLIP features to segment 3D points according to their similarities to text queries, enabling open-vocabulary scene understanding. Tsai et al. [13] fine-tuned the CLIP model to translate construction safety-related semantics in images to textual information for site inspectors. These studies have demonstrated a more applicable user interface that facilitates the direct extraction of actionable insights, potentially advancing computer vision-based applications for construction monitoring.

Despite the recent breakthrough of the vision-language models, their performance on the construction scenes remains untapped, leaving room for potential improvements in the current vision system. Considering that most visionlanguage models are trained with a wide variety of imagetext pairs instead of specific domain knowledge, the models may not contain enough construction knowledge to be adapted to downstream tasks for construction monitoring. To address the issue, this paper leverages the CLIP model to collect images according to the text embeddings of the building construction activities in the Uniformat standard. The Uniformat standard encompasses eight main activities related to building construction. The clip-retrieval [14] API is employed to retrieve the top ten search results from the LAION-5B dataset [15]. To provide the model with additional textual clues regarding construction semantics, a second round of image collection is conducted using the prompt template "A photo of {activity} in construction" to observe any noticeable improvements. Consequently, about 6200 images of 641 categories are collected in each round. Each retrieved image is reviewed whether it is correct or is within the construction context. Zero-shot performance of CLIP on the construction image classification dataset [16] is presented.

The results show that even using the prompt template instead of contextless activity names, the CLIP model is still not good enough at retrieving correct images for the target activity as well as zero-shot classification. Improving the prompts in this way only ensures the retrieval of construction-related content and not for the correct category. To further improve the dataset, strategies aiming for enhancing text prompts and visual construction semantics to obtain more accurate images are proposed. Similar to [11], where different context prompts are ensembled to enhance the zero-shot performance, three different prompt templates are used to increase the opportunities of retrieving the correct images of the work activities. Additionally, a construction semantic classifier is developed to remove non-construction images by estimating the score of how an image is construction-related. Both of the two strategies expedite the image collection of the proposed VL-Con dataset and help ensure the quality of the collected images.

Based on the enhanced image retrieval through CLIP API, the images of different activities are collected and reviewed, and another round of manually image collection on web is conducted to finalize our dataset. Only 142 of the 641 activities have additional 5 images from the manual collection because some activities are considered ambiguous (e.g., specialty and supplementary activities) or cover too broad range of definitions (e.g., high-level activities). Note that the reason could also be why the previous collections fail to find proper images to some activities. By comparing the collected images of different activities at different round through large visionlanguage model and our manual collection, the gap in image features is significant and highlights the need of more construction-specific image-text pairs to improve in-domain knowledge. To adapt the large-scale visionlanguage models to construction monitoring applications, such datasets containing construction knowledge would be required. As a preliminary and pioneer work, the proposed VL-Con dataset is publicly available through https://github.com/huhuman/VL-Con.

2 Related work

2.1 Computer vision in construction monitoring

In construction progress monitoring, reality capture data were collected to be compared against BIM or 4D BIM for tracking element changes and confirming project schedules. Pal et al. [17] estimated the completion percentage of building construction progress in combination of site images, reconstructed point clouds, and BIMs. These efforts have been constrained by low LOD across various model disciplines of BIM and a lack of details in project schedules. Jung et al. [18] proposed a transformer model to ensure the consistency among project schedule data by aligning them with Uniformat classifications. Núñez-Morales et al. [19] generated synthetic images from high-LOD BIMs to help develop supervised learning models recognize different under-construction elements as the well-annotated datasets satisfying a certain quantity and quality to train the models are not available. Despite transfer learning from another pre-trained supervised model with large-scale datasets, Lin et al. [20] presented that the dataset bias may prevent the model from learning construction-specific contents, and the transferlearning Faster R-CNN even had poorer performance than the one trained from scratch on bridge defects.

Inspired by LLMs, vision foundation models using selfsupervised learning techniques have shown promising results and become popular alternatives. He et al. [7] proposed masked autoencoders (MAE) following the idea of masked language modeling to pre-train the large vision model with a wide variety of image data without labels. Kirillov et al. [9] proposed the large-scale dataset of 1B



Figure 1. Overview of the image collection and labeling

Label	А	В	С	D	Е	F	G	Total
		"{activity}"						
Correct	18.6%	48.8%	62.6%	28.4%	21.7%	33.0%	43.2%	37.7%
Incorrect but related	30.4%	18.8%	23.1%	10.2%	15.1%	11.7%	14.3%	16.5%
Incorrect	51.0%	32.4%	14.3%	61.4%	63.2%	55.3%	42.5%	45.8%
Total	494	738	854	1677	503	528	1352	6146
		1	'{activ	ity} in	constr	uction'	I	
Correct	26.0%	50.1%	53.4%	26.1%	21.5%	25.5%	44.1%	36.4%
Incorrect but related	54.7%	34.3%	40.1%	29.0%	24.2%	34.5%	26.0%	32.6%
Incorrect	19.3%	15.7%	6.5%	44.9%	54.2%	40.0%	29.9%	31.1%
Total	494	738	854	1677	503	528	1352	6212

Table 1. The label distributions of the two collected image sets

masks and 11M images to develop the segment anything model (SAM) using MAE pre-trained vision transformer as the image encoder for class-agnostic segmentation. Taking advantage of the robustness of such vision foundation models, Ahmadi et al. [21] combined SAM with U-Net to enhance crack detection in concrete. Ge et al. [22] fine-tuned SAM for crack segmentation to improve crossdataset generalization.

However, a number of challenges still hinder the computer vision-based applications for construction monitoring, including (1) lack of available ground truth segmentation for relevant physical assets in reality capture datasets and (2) lack of formal definitions around what needs to be segmented in pictures in the first place. Vision-language models that take the advantage of robust language understanding have broadened vision understanding to alleviate the limitations. The extracted image features reflecting the context relationship can be more easily adapted to various construction scenes.

2.2 Vision-language in construction monitoring

Before vision-language foundation models emerge, research has been focused on interpreting construction images in the form of natural language. For example, image captioning that can directly generate actionable insights for construction monitoring has been studied in the past decade. Without the robust text embeddings from foundation models, creating new and meaningful textual labels or captions of various construction scenes is the core hindrance. Xiao et al. [23] proposed the image captioning dataset for common construction machines and their activities. Liu et al. [24] proposed the image captioning dataset of five construction activities with the details of worker actions and safety gears. Zhai et al. [25] created the image captioning dataset for perceiving unsafe behavior of workers in construction.

The reviewed image captioning methods mostly adopted the encoder-decoder architecture to perform image-to-text translation, where CNN models were used as the image encoder, and RNN models were used as the text decoder. Bang and Kim [26] extracted features of object regions from drone images through Faster R-CNN as the image encoder and decode the features using LSTM to produce dense captioning. Wang et al. [27] used Mask R-CNN as the image encoder and LSTM with the attention layer as the text decoder for construction works, including masonry, reinforcement steel bar tying, and tiling. The adopted single-modal models were only trained with their proposed datasets to connect the representations across vision and language. As the studies focused on specific scopes and scenarios, the learned knowledge of their fully supervised models was limited to the adopted datasets.

The limitations of the encoder-decoder architecture made the applications difficult to be scaled. In contrast, Radford et al. [11] proposed the dual-encoder model, CLIP, to first jointly train text and vision encoders with numerous image-text pairs of a wide range of cases. By bridging multi-modal understanding through natural language supervision, the vision-language model was capable of handling various scenarios with more robust image embeddings. A text decoder can be specifically trained for a downstream task that needs text generation [28]. As being a promising alternative, the feasibility analysis of the vision-language foundation model for construction monitoring applications is needed to explore and validate model's understanding of construction contexts.

3 Vision-Language Understanding of Construction Context

3.1 Data collection and labeling

This paper retrieves the corresponding images using the Uniformat work item as the text query from the LAION dataset through their clip-retrieval [14] API (see Figure 1). The default parameter values are adopted to search and rank the images, including aesthetic scoring. Additionally, the keyword, "in construction", is prepended to the original names as text queries to collect the images in a second run. This is expected to provide more semantic clues of construction and help improve the performance because some of the names are not exclusive in the construction industry, and the model does not specifically learn to recognize them.

After that, this paper manually review every image and classify them into three groups: (1) **correct** - the image represents the corresponding activity, (2) **incorrect but related** - the image does not indicate the corresponding activity but contains construction semantics, (3) **incorrect** - the image is not related to any construction activi-

ties. Figure 1 illustrates the examples of the defined three groups.

Table 1 presents the summary of the dataset over eight different main activities. As presented, only about onethird of the images are correctly retrieved for the given construction activities. Despite the increase in the ratio of the class of incorrect but related as shown in Figure 2 when using the prompt "{activity} in construction", the overall accuracy is not significantly improved. As a result, construction-specific image-text pairs are needed to enhance the construction knowledge for developing a more robust foundation model in the construction domain.



Figure 2. Image retrieval results with and without *"in construction"*

3.2 Zero-shot performance on construction images

The zero-shot performance of the CLIP model on construction images is evaluated by the BCS dataset [16], which contains about 212,000 photos of buildings and construction sites for classification. To evaluate the understanding of construction contexts, 104,484 images of 54 categories for construction sites in the BCS dataset are employed to perform the zero-shot classification. The image numbers of different classes are *Bridge* (6752), *Site fence* (4980), *Wood floor* (4808), *Ordinary Door* (4568), etc. As stated in [11], the same prompt for zero-shot transfer to existing image classification datasets, "A photo of a {label}.", is used to wrap the inputs instead of using contextless class names.

Cheng et al. [16] has achieved the top-1 accuracy of up to 94.7% on the dataset by a fully supervised model while zero-shot CLIP underperforms by over 35%. The CLIP model only achieves the top-1 accuracy of 59.39% and top-3 accuracy of 81.41%. Figure 3 presents the zero-shot accuracy distribution over different categories. Among all the categories, the model has the highest accuracy of 99.40% for *Site vest* and the lowest accuracy of 0% for *U-steel*. The *U-steel* images are mostly misclassified into other steel-related classes, such as *Sheet steel* and *Angle steel*. The situation infers that the lack of construction-

specific knowledge limits the model to only recognize general contents. The professional terms with only minor differences significantly confuse the model. The evidence can be found as the top-3 accuracy is increased by over 20%.

4 VL-Con: Vision-Language Dataset for Construction Monitoring

4.1 Requirements of vision-language datasets

Based on the CLIP's understanding of construction context, construction-specific vision-language datasets are required to enhance the construction knowledge. Through manually inspecting the collected images, Figure 4 presents that potential causes of the poor performance on image retrieval. The requirements of the visionlanguage dataset preparation reflecting the issues for construction monitoring are summarized as follows:

1. Ambiguous description/name of work activity

Although construction knowledge is required to understand the semantics of the activities, some of their names are too ambiguous to be easily interpreted even by people with construction backgrounds. The names become more abstract when the corresponding activities are at a higher level because they need to cover various children's items. To keep the simplicity, only a few words are used to define the whole scope, leading to failures in capturing the semantics behind the words without any additional context. For example, in the adopted Uniformat system, Floor Construction (B1010), Roof Construction (B1020), and Stairs (B1080) are all sub-items of Superstructure (B10). Figure 4(a) demonstrates that the CLIP model fails to retrieve correct images of the superstructure but is capable of recognizing other common words like floor and roof. Consequently, prompt engineering is needed to include more context for CLIP to retrieve more correct images.

2. Missing photos of work activity

Though the examples of image retrieval show correct semantics, more construction-related and progressdetailed images are expected. Most of the activities suffer from missing photos in the training data because the construction activity or corresponding assemblies contain numerous components and steps. The issue limits the model to return only the photos of finished states or irrelevant content. Figure 4(b) shows that this is especially true for any categories associated with "supplementary" or "specialties". For example, *Exterior Wall Specialties* (B2090) include below-grade egress, window wells, and any kind of finished product tangent to the exterior wall [29]. The exhibited semantics overlapped with other categories such as *Exterior Fabrications* (B2080.70) leading to failures in differentiating between each other, whereas exterior fabrications are more about column covers or decorative finishing directly applied onto the wall. This phenomenon can also be attributed to the ambiguous nature of the activity names. Without specific images that demonstrate the difference between such similar sub-categories, even engineers could be confused with the definitions.

3. Searching preference for construction needs

In the adopted CLIP image retrieval, the image quality can be determined by the aesthetic predictor, ensuring the retrieved images are closer to what users are expecting. A photo of a document copy or procedure diagram may get a low score and be ranked behind because such an image contains less vision information, meaning little object information is included. For example, Figure 4(c) shows the image retrieval results of heat generation under different aesthetic score thresholds and average weights. As searching preference could significantly impact image retrieval, the construction-specific preference can be developed to help retrieve more correct images of our interest to build the vision-language dataset.

4.2 Prompt engineering and construction semantic classification

To improve image retrieval of the Uniformat activities, three different prompts with more details are used: (1) "A photo of {activity}, a type of building construction activity", (2) "A photo of {full-activities-hierarchy}", and (3) "A photo of {activity}, revit". The first and second one is to replace the previous "in construction" keyword with more specific definitions of the activities. The final one is taking advantage of the exclusive word in the construction industry to force the retrieved images to be related to construction while the images are mostly about virtual scenes. The enhanced prompts can help collect more potentially correct image-text pairs as presented in our final proposed VL-CoN dataset.

For the semantic predictor, the original image retrieval already employed the aesthetic predictor to ensure the images contain more useful information and significant object appearance instead of diagrams and flow charts. This paper trained a ResNet-18 model for binary construction semantic classification with the previously labeled dataset of Uniformat categories, followed by the prompt "{activity} in construction". After removing duplicate images, nearly 5000 images were then separated into an 8 to 2 ratio for training and validation sets. The



Figure 3. The zero-shot top-1 and top-3 accuracy of different classes



Figure 4. The examples for the requirements of vision-language datasets

images were all resized to 224 by 224, and the training set was augmented with random crop and random horizontal flip. The model after 25 epochs achieves the accuracy of 73.63% on the validation set with a wide range of images associated with various categories in the Uniformat. Figure 5 demonstrates the model predictions and scores of the construction semantics of the images. One noticeable feature of this very simple trained model is to identify diagram-like images or "unrealistic images" and filter more realistic scenes such as those of construction sites. With the two proposed strategies, another three rounds of image collection are conducted, and the retrieved images are firstly filtered by the proposed construction semantic predictor. After that, manual inspection is still required to finally complete the vision-language dataset for enhancing construction knowledge of foundation models. Note that because some of the activities may not exist in the LAION-5B dataset, the image retrieval possibly fail to find any correct images of those activities. In that case, another image sources will be needed in the future to acquire the corresponding photos of the activi ties. To ensure certain quantity and quality of the dataset, the dataset additionally include 5 images for each of 142 activities manually collected through the web. The final VL-CoN dataset are publicly available and scalable, allowing other researchers to add more image-text pairs to any of Uniformat categories.



Figure 5. Construction semantic classification

5 Conclusions

This paper conducted preliminary analysis of visionlanguage understanding of CLIP to construction scenes with regards to Uniformat. The images from the LAION-5B dataset were retrieved using the CLIP model to assess its applicability of understanding construction scenes. Upon detailed review of the retrieved images, the limitations of the existing vision language model were identified: appropriate prompt to maximize the likely result, insufficient images that precisely describe all the activities in Uniformat, and inability to contextualize construction scenes from images. To address these limitations, two strategies, prompt engineering and a semantic classifier of construction scenes, were proposed to complete the visionlanguage dataset for construction monitoring. Another manual image collection is also conducted to further enhance the dataset. The final VL-Con dataset was released to provide the basis for further method development and model training benchmark. As a pioneer work for adapting vision-language models to construction monitoring applications, the dataset can be scaled by adding more images for any of the activities, boosting the construction-specific foundation models.

References

- Shuai Tang, Dominic Roberts, and Mani Golparvar-Fard. Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction*, 120:103356, 2020.
- [2] Youngjib Ham, Kevin K Han, Jacob J Lin, and Mani Golparvar-Fard. Visual monitoring of civil infrastructure systems via camera-equipped unmanned aerial vehicles (uavs): a review of related works. *Visualization in Engineering*, 4(1):1–8, 2016.

- [3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017.
- [4] 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.
- [5] Shun-Hsiang Hsu, Ting-Wei Chang, and Chia-Ming Chang. Impacts of label quality on performance of steel fatigue crack recognition using deep learningbased image segmentation. *Smart Structures and Systems*, 29(1):207, 2022.
- [6] Yeji Hong, Somin Park, Hongjo Kim, and Hyoungkwan Kim. Synthetic data generation using building information models. *Automation in Construction*, 130:103871, 2021.
- [7] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings* of the IEEE/CVF conference on computer vision and pattern recognition, pages 16000–16009, 2022.
- [8] OpenAI et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2021.
- [9] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. arXiv preprint arXiv:2304.02643, 2023.
- [10] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, et al. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, page 100017, 2023.
- [11] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.
- [12] Songyou Peng, Kyle Genova, Chiyu Jiang, Andrea Tagliasacchi, Marc Pollefeys, Thomas Funkhouser, et al. Openscene: 3d scene understanding with open vocabularies. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 815–824, 2023.

- [13] Wei Lun Tsai, Jacob J Lin, and Shang-Hsien Hsieh. Generating construction safety observations via clipbased image-language embedding. In *European Conference on Computer Vision*, pages 366–381. Springer, 2022. doi:10.1007/978-3-031-25082-8_24.
- [14] Romain Beaumont. Clip retrieval: Easily compute clip embeddings and build a clip retrieval system with them. https://github.com/rom1504/ clip-retrieval, 2022.
- [15] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35:25278–25294, 2022.
- [16] Xuanhao Cheng, Mingming Jia, and Jian He. A largescale dataset of buildings and construction sites. *Computer-Aided Civil and Infrastructure Engineering*, 2023.
- [17] Aritra Pal, Jacob J Lin, Shang-Hsien Hsieh, and Mani Golparvar-Fard. Activity-level construction progress monitoring through semantic segmentation of 3dinformed orthographic images. *Automation in Construction*, 157:105157, 2024.
- [18] Yoonhwa Jung, Julia Hockenmaier, and Mani Golparvar-Fard. Transformer language model for mapping construction schedule activities to uniformat categories. *Automation in Construction*, 157: 105183, 2024.
- [19] Juan D Núñez-Morales, SHUN-HSIANG HSU, Amir Ibrahim, and Mani Golparvar-Fard. Realityenhanced synthetic image training dataset for computer vision construction monitoring. *Proceedings of International Structural Engineering and Construction*, 10(1):CON–29, 2023.
- [20] Jacob J Lin, Amir Ibrahim, Shubham Sarwade, and Mani Golparvar-Fard. Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3d mapping, defect detection, analysis, and reporting. *Journal of Computing in Civil Engineering*, 35(2):04020064, 2021.
- [21] Mohsen Ahmadi, Ahmad Gholizadeh Lonbar, Abbas Sharifi, Ali Tarlani Beris, Mohammadsadegh Nouri, and Amir Sharifzadeh Javidi. Application of segment anything model for civil infrastructure defect assessment. arXiv preprint arXiv:2304.12600, 2023.

- [22] Kang Ge, Chen Wang, and Yutao Guo. Fine-tune vision foundation model for crack segmentation in civil infrastructures. *arXiv preprint arXiv:2312.04233*, 2023.
- [23] Bo Xiao, Yiheng Wang, and Shih-Chung Kang. Deep learning image captioning in construction management: A feasibility study. *Journal of Construction Engineering and Management*, 148, 7 2022. ISSN 0733-9364. doi:10.1061/(asce)co.1943-7862.0002297.
- [24] Huan Liu, Guangbin Wang, Ting Huang, Ping He, Martin Skitmore, and Xiaochun Luo. Manifesting construction activity scenes via image captioning. *Automation in Construction*, 119:103334, 2020. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2020.103334.
- [25] Peichen Zhai, Junjie Wang, and Lite Zhang. Extracting worker unsafe behaviors from construction images using image captioning with deep learning-based attention mecha-Journal of Construction Engineering nism. and Management, 149(2):04022164, 2023. doi:https://doi.org/10.1061/JCEMD4.COENG-12096.
- [26] Seongdeok Bang and Hyoungkwan Kim. Contextbased information generation for managing uavacquired data using image captioning. *Automation in Construction*, 112, 4 2020. ISSN 09265805. doi:10.1016/j.autcon.2020.103116.
- [27] Yiheng Wang, Bo Xiao, Ahmed Bouferguene, Mohamed Al-Hussein, and Heng Li. Visionbased method for semantic information extraction in construction by integrating deep learning object detection and image captioning. Advanced Engineering Informatics, 53:101699, 2022. doi:10.1016/j.aei.2022.101699.
- [28] Chenliang Li, Haiyang Xu, Junfeng Tian, Wei Wang, Ming Yan, Bin Bi, Jiabo Ye, Hehong Chen, Guohai Xu, Zheng Cao, et al. mplug: Effective and efficient vision-language learning by cross-modal skip-connections. arXiv preprint arXiv:2205.12005, 2022.
- [29] Jim Bedrick, Will Ikerd, and Jan Reinhardt. Level of development (lod) specification. https://bimforum.org/resource/ lod-level-of-development-lod-specification/, 2022.

Acoustic Signal-Based Excavator Fault Detection Using Deep Learning Method

Yuhan Zhou¹, Foo Zhen Lin¹, Koh Yong Zhi¹, and Justin K.W. Yeoh¹

¹Department of Civil and Environmental Engineering, National University of Singapore, Singapore.

Abstract

As machinery assumes a critical role in modern construction, particularly in Singapore's development initiatives, maintaining excavators becomes paramount. Despite the prevalence of faults within this equipment, the scarcity of skilled mechanics compounds the challenge of timely diagnoses and maintenance. Leveraging deep learning methodologies, this research endeavors to analyze audio signals from excavators, aiming to identify distinctive patterns indicative of faults. Unlike existing studies primarily relying on vibration signals, this research focuses on audio signals for excavator fault prediction. Challenges involving ambient noise in construction sites and limitations in dataset size and imbalance compel the need for robust machine learning models capable of accurate fault diagnoses. The proposed methodology involves dataset collection, audio signal processing, feature extraction, and neural network training to differentiate normal operation from faulty conditions. This study delves into the application of machine learning and signal processing techniques to discern excavator conditions, aiming to classify their operational state as either faulty or operational. With an achieved 89.33% accuracy, 94.74% precision, and 85.71% recall, the method demonstrates promising performance. This research offers the potential to fortify excavator maintenance practices, potentially mitigating the impact of faults on construction productivity and costs.

Keywords -

Excavator Maintenance; Deep Learning; Fault Detection; Audio Signal Processing

1 Introduction

Throughout history, both equipment and labor have constituted essential components within the built environment [1]. The utilization of manual labor in construction has perennially remained a cornerstone. However, with technological advancements, the reliance on manual labor diminishes, giving way to a greater dependence on cutting-edge machinery at construction sites. These machines facilitate construction processes, fostering heightened productivity and long-term cost savings. Unlike humans, susceptible to fatigue, machinery operates tirelessly, accomplishing tasks equivalent to multiple human capacities without succumbing to exhaustion. Nevertheless, machines are susceptible to distinct forms of fatigue, manifesting as engine overheating or operational stress due to inadequate maintenance, leading to component faults.

This study aims to investigate excavators, extensively used across Singapore, particularly as the government endeavors to intensify subterranean development and coastal land reclamation. A prevalent challenge arises from the difficulty in instilling a sense of ownership and care for excavators. Primarily, excavator operators do not possess ownership, potentially impeding their understanding of how malfunctioning machinery impacts project outcomes, thus neglecting proper maintenance. Further compounded by demanding project schedules, these excavators endure heavy utilization, increasing susceptibility to faults necessitating servicing and upkeep. Unfortunately, only a select few excavator brands demonstrate prolonged durability with minimal maintenance costs.

Optimal excavator maintenance, beyond routine servicing, necessitates pre-empting potential faults. Minor issues such as hydraulic oil or radiator water leakage, or even loose bucket attachments, have the propensity to escalate into significant problems, potentially damaging the excavator if not promptly addressed. However, amidst a scarcity of foreign domestic workers and the industry's shift toward technologically advanced machines, the availability of experienced mechanics capable of accurately diagnosing and rectifying excavator faults diminishes. This makes manual diagnosis of excavator faults more difficult and troublesome.

To address this challenge, an effective approach involves deploying advanced machine learning techniques to autonomously discern machine activities or construction equipment by analyzing their distinctive sound patterns [1]. The progression of machine learning methodologies has facilitated fault detection successfully across diverse machinery, encompassing construction excavators. Typically, sensor fusion strategies, such as employing microphone arrays to capture machinegenerated sounds, are utilized to gather audio signals. Even a single microphone holds the capability to capture sound from multiple machines.

While machine learning models leveraging sound signals have demonstrated success in domains like rotary machines [2] and combustion engines [3], their application specifically to excavators remains to be explored. Prior studies on excavator fault prediction [4], primarily relied on vibration signals as the primary input data. Contrary to vibration signals, audio signals are more intuitive and simpler to collect. However, the prevalent ambient noise within construction sites significantly undermines sound detection accuracy, complicating fault detection via sound signals. Therefore, establishing a dataset acquired in relatively quiet environments, encompassing both normal operating conditions and faults, becomes imperative as a standard resource for fault diagnosis.

Moreover, machine learning techniques rely on extensive, well-balanced datasets to construct highly accurate models. However, practical scenarios often present challenges in collecting such vast datasets. Instead, smaller and imbalanced datasets tend to be more common. Regrettably, limited research explores excavator fault detection efficacy using such data configurations. Hence, it becomes imperative to devise robust machine learning methodologies capable of effectively diagnosing excavator conditions, even when working with small and imbalanced datasets.

The primary objective of this research is to classify excavators as either faulty or operational through the utilization of signal processing methodologies in tandem with deep learning techniques. The process involves collecting faulty diagnosis datasets, denoising audio files, extracting audio features, and training a neural network to distinguish between normal operation and faulty conditions in excavators. Subsequently, the neural network's classification efficacy will be assessed using a validation dataset, employing pertinent evaluation metrics. An underlying assumption integral to this research involves acknowledging that the model's predictive capacity may not extend to anticipating excavator faults in their nascent or imminent stages, as the recordings of faults are already captured in their final state.

2 Related Work

2.1 Machine Fault Detection

The utilization of diverse sensor signals for machine fault detection has witnessed significant advancements in recent years. Leading researchers, such as Janssen & Arteaga (2020), have adeptly leveraged machinelearning methodologies to identify faults in rotary and structural machinery [5]. Signals play a crucial role in discerning activity and vibrations in large-scale mining and material handling equipment, where effective vibration management is paramount to mitigate the risk of potential mine collapses during ongoing excavation and movement. Wieckowski et al. (2020) utilized the Fast Fourier Transform (FFT) to convert vibration waveforms from the time domain to the frequency domain, subsequently devising a vibration control algorithm [6].

MEMS (Micro-Electro-Mechanical Systems)-based vibration sensors have gained widespread adoption various applications, showcasing across their effectiveness in accurately recognizing activity across a spectrum of equipment types [7]. However, the practical challenges associated with directly attaching these sensors to equipment, especially for smaller machinery or construction tools like concrete cutting saws and small concrete mixers, cannot be overlooked [8]. Additionally, the typical deployment of vibration sensors for individual machines presents limitations [9], particularly in scenarios where multiple machines operate concurrently. In contrast, an audio-based system offers a solution by obviating the need to attach a microphone to each machine, instead relying on strategically positioned microphones throughout the site.

Therefore, audio signals have emerged as valuable data resources for activity recognition and fault detection. Typically, this methodology adheres to a standard framework: (1) recording audio data using a single microphone, (2) utilizing FFT and Inverse Fast Fourier Transform (IFFT) to denoise the audio and extract pertinent features within specific time frames by transitioning between time and frequency domains, and (3) training a machine-learning algorithm using these extracted features [1]. Furthermore, pioneering techniques, such as employing a mobile microphone for spatial information in machine condition monitoring, have been explored [5]. The utilization of an array of multiple microphones has also proven effective in predicting the position of a mass on a vibrating plate, emulating structural flaws or engine imbalances. Its data processing involves employing peak-finding methods and three-dimensional imaging techniques. Collectively, these research findings underscore the efficacy of leveraging microphone signals as a valuable tool for fault detection and condition monitoring across a spectrum of diverse applications.

2.2 Sound-based Classification Using Neural Network

Acoustic sensors offer distinct advantages over other sensors due to their affordability and ease of placement, rendering them highly practical for event classification [10]. The collection of ambient sounds through simple microphones or sensors has paved the way for advanced signal-processing models. Consequently, there has been a surge in research focused on classifying construction work, machine types, and detecting faults using soundbased methodologies, driven by their cost-effectiveness and widespread applicability.

There have been studies in the literature focus on sound-based Construction Site Monitoring (CSM), aiming primarily at identifying activities or classifying brands and models of working machines. Multiple methodologies have been developed, often using Machine Learning (ML) approaches [11], [12]. Some commonly employed methods include Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), Artificial Neural Networks (ANNs), Hidden Markov Models (HMMs), and Gaussian Mixture Models (GMMs) [13].

Recently, deep learning methods using Neural Network become much more popular with their promising results. The most used model in the field of DL is the Convolutional Neural Network (CNN). Maccagno et al. employed a CNN-based model to classify different types and brands of machines on construction sites, obtaining 97.08% accuracy [14]. Similarly, Scarpiniti et al. presented a deep recurrent neural network (DRNN) model to classify five classes of multiple vehicles and tools using sound signals, reporting 97% accuracy [15]. While the mentioned studies exhibit promising advancements in classifying machine types and brands based on sound, an area that remains unexplored is the utilization of audio signals for fault detection. Despite the successes in identifying construction equipment and activities through sound analysis, there is a lack of research focused on leveraging these signals specifically for detecting faults or malfunctions within construction machinery such as excavators.

2.3 Imbalanced Data Processing

Conventional machine-learning techniques commonly rely on training models using balanced datasets, ensuring an equitable number of data samples for each class. Such datasets facilitate unbiased learning and streamline model performance evaluation. Many fault detection studies leverage extensive datasets encompassing several thousand samples for each class. However, datasets collected often exhibit imbalance, particularly when certain classes are rare, leading to limited samples representing these rare occurrences. Imbalanced datasets markedly affect model quality, resulting in poorer predictive performances for the minority class and potential overfitting to training samples, constraining the model's ability to generalize during testing [16]. This limitation is crucial, especially when the accuracy of predicting the minority class holds more significance than that of the majority class, particularly in scenarios where the cost of false negative predictions for the minority class is substantial [17].

Hence, various methods have emerged to address imbalanced datasets, primarily within the realm of datalevel methods aiming to enhance the performance of standard machine learning algorithms. Resampling, a prevalent data-level technique, aims to rectify class distribution imbalances.

Random over-sampling is a widely used technique where, for instance, Hensman and Masko (2015) applied this approach to an image dataset by randomly duplicating minority class samples until reaching a balanced representation [18]. Results indicated that random over-sampling notably enhanced model performance, akin to that of a balanced dataset. An advanced technique, Synthetic Minority Over-sampling Technique (SMOTE), generates synthetic samples derived from the minority class data to alleviate overfitting issues typically associated with regular oversampling [19]. While commonly used, SMOTE's efficacy has raised concerns. Sharma et al. (2018) indicated that in highly imbalanced classes (e.g., ratios around 1:1000), SMOTE-generated samples might negatively impact model performance by incorporating features overlapping with the majority class [20]. Elreedy and Atiya (2019) supported this observation, noting that due to SMOTE's nature, synthetic samples for very small or high-dimensional classes become less representative [21]. Nonetheless, despite its drawbacks, SMOTE improves classification, albeit not to the extent of a balanced dataset.



Figure 1. Overview of the proposed method

3 Method

Figure 1 presents an outline of the proposed method of diagnosing excavator conditions. Generally, the methodology comprises four key steps: a) data collection; b) data preprocessing; c) feature extraction; and d) fault detection. Detailed explanations of each of these steps follow in the subsequent sections.

3.1 Data Collection

Audio data from excavators was gathered using a single microphone to capture the noise emitted during their operation. Microphone placement differed based on whether the excavator was idle or in operation. For idle instances, the microphone was positioned in close proximity to the fault, while during operation, it was placed at a safe distance of 1 to 2 meters from the excavator. Given that most microphones possess 360° sensitivity [9], the utilization of a single microphone suffices for data collection in this study. However, for more complex on-site implementation scenarios, employing an array of consistently positioned microphones may represent a more effective strategy.

The study encompassed examinations of both "Good" and "Bad" conditioned excavators. "Good" noises represent well-conditioned excavator sounds without apparent faults, while "Bad" noises represent sounds indicative of faults. These faults encompass various issues such as hydraulic leakage in the arm, boom hose bursting, low engine power, control valve spool leaks, or visual indicators like smoke emanating from the excavator. Some faults manifest audibly, such as hydraulic leaks or hose bursts, while others are visually detectable, like smoke emissions.

3.2 Data Preprocessing

To enable the subsequent machine learning methodology, it was imperative to initially label the gathered dataset. Before audio recording, skilled mechanics conducted diagnostic assessments on the involved excavators. Their expertise and valuable insights into the excavators' conditions were pivotal. Based on their evaluations, each file was meticulously labeled either as "0" denoting normal operational status or "1" indicating a faulty condition in the excavator. Furthermore, the reasons behind the faults were meticulously documented during this process. Apart from data labeling, the imbalanced dataset is another issue. Neural network effectiveness is hindered by small or imbalanced datasets, which are common in machine learning due to challenges in acquiring large, representative samples. Addressing this, an enhanced oversampling technique, SMOTE, was employed on minority classes. SMOTE generates synthetic samples by interpolating minority class samples with their nearest neighbors of the same class, creating similar yet distinct synthetic samples. This mitigates overfitting risks, enhancing the model's generalization capability [19].

3.3 Feature Extraction

Mel Frequency Cepstral Coefficient (MFCC) methodology was employed for audio signal feature extraction. Unlike the widely used Fast Fourier Transform, MFCC is designed to mirror human hearing capabilities. Human auditory perception of frequency increases logarithmically, meaning what's perceived as a uniform frequency increase to humans isn't linear. Additionally, human hearing is more adept at distinguishing lower frequencies compared to higher ones. MFCC efficiently captures unique features from recorded audio signals by employing the Mel scale, where lower-frequency coefficients possess greater spacing, while higher-frequency coefficients have narrower spacing, mimicking human hearing properties [22]. The Mel scale ensures that sounds equidistant on the scale are perceived as equally spaced by humans [23]. Leveraging MFCC, the model can learn crucial sound features indicative of faults.

3.4 Fault Detection

For fault detection in excavators, a binary classification approach - normal operation versus faulty - was undertaken. Utilizing MFCC spectrograms, specific spectrogram values at each point served as inputs. A proposed CNN-based neural network, depicted in Figure 2, was employed. Each convolutional layer utilized ReLU as the activation function. Max-pooling layers downsampled convolved features to save processing time and reduce image size. At each step, maximum values within each window were pooled into an output matrix. The model utilized binary cross-entropy as the loss function and the Adam optimizer. To optimize the performance of the proposed neural network, hyperparameter tuning was utilized.



Figure 2. Proposed CNN structure

4 Validation

The dataset comprises 305 audio files of varied durations, spanning 1 to 88 seconds, recorded using a single microphone positioned according to the operational status of each excavator. Included in this dataset are recordings from diverse excavator brands such as Caterpillar (CAT), Sumitomo, Kobelco, Hitachi, Yanmar, Volvo, and Doosan. The excavators ranged in size from 5-tonne mini excavators to 38-tonne large excavators. Among the 305 files, 187 are classified as "good," indicating excavators in optimal working condition, while the remaining 118 files are categorized as "bad," representing excavators experiencing faults, which ranged from singular to multiple faults concurrently. These faults encompass various issues such as hydraulic leakages in excavator arms, burst boom hoses, low engine power, control valve spool leakages, smoke emissions, and others. While some faults, like hydraulic leakages, emit distinct sounds, others are only identifiable through visual cues like alarm lights or smoke emissions. Figure 3 shows some samples of the collected audio data, including excavators in normal operation and excavators with different types of faults.



Figure 3. Samples of the recorded excavator operations

After preprocessing, the original imbalanced dataset has been expanded to 374 samples, with 187 "good" noises and 187 "bad" noises. To validate the proposed method, the dataset is divided into two subsets: an 80% training dataset and a 20% test dataset. Evaluation of the proposed network's performance uses three key metrics: accuracy (see Equation (1)), precision (see Equation (2)), and recall (see Equation (3)). Accuracy denotes the ratio of correctly predicted observations (True Positives) to the total number of observations. Precision isolates the actual positive instances from the predicted positive dataset. Recall computes the count of true positives among all identified positives, including true positives and false negatives. Notably, in this study, recall holds more significance than the other metrics, as a false negative could significantly impact the excavator's lifespan if faults remain unidentified and unresolved.

$$ccuracy = \frac{TP + TN}{Total \ Predictions} \tag{1}$$

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

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

**TP*: true positive; *FP*: false positive; *TN*: true negative; *FN*: false negative.

a



Figure 4. A sample of the MFCC spectrogram

Following the application of MFCC, the resulting MFCC spectrogram serves as the input for the proposed neural network. Illustrated in Figure 4 is a sample of the generated MFCC spectrogram. The X-axis delineates

time, while the Y-axis represents the distinct MFCC coefficients. Additionally, it depicts frequencies transformed into Mel scale values, a logarithmic representation of signal frequencies. Each plotted point in the diagram signifies a specific MFCC at a precise moment in time. This spectrogram offers a visual portrayal of sound in the Mel scale, presenting an alternative to the frequency domain.

5 Results and Discussion

The findings presented in Table 1 demonstrate the model's good performance across both the training and test datasets. The training accuracy achieved 98.32%, while the test accuracy maintained a robust 89.33%. In terms of precision, the training and test sets displayed values of 97.35% and 85.71%, respectively, indicating the model's ability to minimize false positives, ensuring accurate positive predictions. Moreover, examining recall metrics reveals significant effectiveness. The training set exhibited a recall of 99.32%, while the test set maintained 94.74%. These results suggest the model's proficiency in capturing the majority of positive instances without missing many, highlighting its consistent performance on unseen test data.

The model's ability to generalize without overfitting is evident, showing consistent predictive power across various datasets. Figure 5 further illustrates the evolution of these metrics across epochs, depicting a rapid decrease in training and validation loss from the initial epochs, eventually converging to minimal values with a slight gap between the two losses. This presents an efficient learning ability of the proposed network.

The implementation of the proposed automatic fault detection method carries profound implications for machinery productivity and cost-effectiveness. By prioritizing the minimization of false-negative results, especially in the context of fault detection within excavators, this method significantly mitigates the risk associated with undetected faults. Such oversights can lead to operational disruptions, downtime, and potential safety hazards, all of which can incur substantial costs and impair machinery productivity.

By ensuring the timely identification and resolution of faults, the proposed method helps avert catastrophic failures that could result in extensive financial losses and pose risks to end-users. This proactive approach not only safeguards equipment integrity but also enhances operational efficiency by minimizing unplanned downtime and maintenance efforts. Moreover, the reliability, effectiveness, and accuracy demonstrated by this model underscore its capacity to optimize machinery performance and reduce operational expenses. By providing early and precise fault detection capabilities, this method enables maintenance interventions to be strategically planned, thereby minimizing costly repairs and maximizing equipment uptime.



Figure 5. Results of the training and test process of the proposed CNN
In essence, the successful implementation of the proposed automatic fault detection method not only bolsters machinery reliability and safety but also translates into tangible benefits in terms of enhanced productivity and cost-effectiveness. Its robust performance positions it as a valuable asset for fault detection in excavators, offering significant potential for improving overall operational efficiency and profitability.

Table 1. Results of the proposed CNN

Dataset	Accuracy	Recall	Precision
Training	98.32%	99.32%	97.35%
Test	89.33%	94.74%	85.71%

6 Conclusion

This paper presents a CNN-based network designed for excavator fault detection using audio signals. The sound of excavator operations is captured through a single microphone, and expert mechanics assist in diagnosing the excavators' statuses. Both normally functioning excavators and those with faults are recorded, with subsequent data labeling based on the mechanics' assessments. To address dataset imbalance, the SMOTE oversampling method is employed to balance the dataset. The extraction of frequency features from audio signals is accomplished through the generation of MFCC spectrograms. These spectrograms serve as input for the proposed CNN-based network. The evaluation of the network's performance employs accuracy, precision, and recall metrics, demonstrating good generalization and robustness in excavator condition recognition.

This study contributes to leveraging machine learning applications to improve excavator serviceability. The COVID-19 pandemic highlighted Singapore's heavy reliance on foreign manual labor, resulting in significant downtime and project slowdowns. There is a growing consensus that a paradigm shift in site operations is necessary, especially considering the migration of talent to other sectors, potentially reducing expertise in this industry. Implementing machine learning could alleviate dependency on senior mechanics for fault identification. It could empower on-site engineers or workers to recognize fault-related sounds promptly, enabling immediate servicing to prevent excavator breakdowns.

A limitation of this study lies in the range of its dataset. As previously mentioned, audio data capturing excavator operations was acquired using a single microphone within a relatively quiet environment with minimal ambient noise. There is merit in extending the dataset to encompass a diverse range of real-world conditions, particularly those characterized by higher levels of background noise, to further validate the proposed method. Apart from gathering data from realworld conditions, acquiring a comprehensive range of normal operation sounds holds significant value in enriching the dataset. This augmentation can enhance the rationality of fault diagnosis and provide additional support for anomaly detection, which presents itself as a potential avenue for future research exploration. Additionally, the data labeling process relied on expert input, which can be time-consuming and labor-intensive, particularly as the dataset expands. This process could be streamlined by implementing automated or semiautomated labeling procedures.

In terms of future research directions, a more granular classification of faults into major and minor categories could yield substantial insights. Moreover, an in-depth analysis of the network's discriminatory capabilities across varied fault scenarios is warranted. Additionally, expanding efforts to correlate specific faults with their respective excavator brands, leveraging an augmented dataset, could enhance the network's proficiency in accurately identifying excavator-specific faults. The inclusion of datasets spanning a broader spectrum of excavator models holds promise for enriching research outcomes.

References

- B. Sherafat *et al.*, "Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review," *Journal of Construction Engineering and Management*, vol. 146, no. 6, p. 03120002, Jun. 2020, doi: 10.1061/(ASCE)CO.1943-7862.0001843.
- [2] O. Janssens *et al.*, "Convolutional Neural Network Based Fault Detection for Rotating Machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331– 345, Sep. 2016, doi: 10.1016/j.jsv.2016.05.027.
- [3] L. Barelli, G. Bidini, C. Buratti, and R. Mariani, "Diagnosis of internal combustion engine through vibration and acoustic pressure non-intrusive measurements," *Applied Thermal Engineering*, vol. 29, no. 8, pp. 1707–1713, Jun. 2009, doi: 10.1016/j.applthermaleng.2008.07.025.
- [4] Q. Zhou, G. Chen, W. Jiang, K. Li, and K. Li, "Automatically Detecting Excavator Anomalies Based on Machine Learning," *Symmetry*, vol. 11, no. 8, Art. no. 8, Aug. 2019, doi: 10.3390/sym11080957.
- [5] L. A. L. Janssen and I. Lopez Arteaga, "Data processing and augmentation of acoustic array signals for fault detection with machine learning," *Journal of Sound and Vibration*, vol. 483, p. 115483, Sep. 2020, doi: 10.1016/j.jsv.2020.115483.
- [6] J. Więckowski, W. Rafajłowicz, P. Moczko, and E. Rafajłowicz, "Data from vibration measurement in a bucket wheel excavator operator's cabin with the aim of vibrations damping," *Data in Brief*, vol. 35,

p. 106836, Apr. 2021, doi: 10.1016/j.dib.2021.106836.

- [7] B. Sherafat, A. Rashidi, Y.-C. Lee, and C. R. Ahn, "A Hybrid Kinematic-Acoustic System for Automated Activity Detection of Construction Equipment," *Sensors*, vol. 19, no. 19, Art. no. 19, Jan. 2019, doi: 10.3390/s19194286.
- [8] B. Sherafat, A. Rashidi, and S. Asgari, "Soundbased multiple-equipment activity recognition using convolutional neural networks," *Automation in Construction*, vol. 135, p. 104104, Mar. 2022, doi: 10.1016/j.autcon.2021.104104.
- [9] C.-F. Cheng, A. Rashidi, M. A. Davenport, and D. V. Anderson, "Activity analysis of construction equipment using audio signals and support vector machines," *Automation in Construction*, vol. 81, pp. 240–253, Sep. 2017, doi: 10.1016/j.autcon.2017.06.005.
- [10] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, "Acoustic Scene Classification: Classifying environments from the sounds they produce," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 16–34, May 2015, doi: 10.1109/MSP.2014.2326181.
- [11] S. Ahmad *et al.*, "Environmental sound classification using optimum allocation sampling based empirical mode decomposition," *Physica A: Statistical Mechanics and its Applications*, vol. 537, p. 122613, Jan. 2020, doi: 10.1016/j.physa.2019.122613.
- [12] S. Scardapane, M. Scarpiniti, M. Bucciarelli, F. Colone, M. V. Mansueto, and R. Parisi, "Microphone array based classification for security monitoring in unstructured environments," *AEU International Journal of Electronics and Communications*, vol. 69, no. 11, pp. 1715–1723, Nov. 2015, doi: 10.1016/j.aeue.2015.08.007.
- [13] R. V. Sharan and T. J. Moir, "An overview of applications and advancements in automatic sound recognition," *Neurocomputing*, vol. 200, pp. 22–34, Aug. 2016, doi: 10.1016/j.neucom.2016.03.020.
- [14] A. Maccagno, A. Mastropietro, U. Mazziotta, M. Scarpiniti, Y.-C. Lee, and A. Uncini, "A CNN Approach for Audio Classification in Construction Sites," in *Progresses in Artificial Intelligence and Neural Systems*, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds., in Smart Innovation, Systems and Technologies. , Singapore: Springer, 2021, pp. 371–381. doi: 10.1007/978-981-15-5093-5_33.
- [15] M. Scarpiniti, D. Comminiello, A. Uncini, and Y.-C. Lee, "Deep Recurrent Neural Networks for Audio Classification in Construction Sites," in 2020 28th European Signal Processing Conference

(EUSIPCO), Jan. 2021, pp. 810–814. doi: 10.23919/Eusipco47968.2020.9287802.

- [16] Z. Li, K. Kamnitsas, and B. Glocker, "Analyzing Overfitting Under Class Imbalance in Neural Networks for Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1065–1077, Mar. 2021, doi: 10.1109/TMI.2020.3046692.
- [17] S. Vucetic and Z. Obradovic, "Classification on Data with Biased Class Distribution," in *Machine Learning: ECML 2001*, L. De Raedt and P. Flach, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2001, pp. 527–538. doi: 10.1007/3-540-44795-4 45.
- [18] D. Masko and P. Hensman, "The Impact of Imbalanced Training Data for Convolutional Neural Networks," 2015. Accessed: Dec. 15, 2023.
 [Online]. Available: https://www.semanticscholar.org/paper/The-Impact-of-Imbalanced-Training-Data-for-Neural-Masko-Hensman/62e81797fff75603a3d7c7759e6efac4fd2 b6b31
- [19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Oversampling Technique," *jair*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [20] S. Sharma, C. Bellinger, B. Krawczyk, O. Zaiane, and N. Japkowicz, "Synthetic Oversampling with the Majority Class: A New Perspective on Handling Extreme Imbalance," in 2018 IEEE International Conference on Data Mining (ICDM), Nov. 2018, pp. 447–456. doi: 10.1109/ICDM.2018.00060.
- [21] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Information Sciences*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.
- [22] A. Meghanani, A. C. S., and A. G. Ramakrishnan, "An Exploration of Log-Mel Spectrogram and MFCC Features for Alzheimer's Dementia Recognition from Spontaneous Speech," in 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China: IEEE, Jan. 2021, pp. 670– 677. doi: 10.1109/SLT48900.2021.9383491.
- [23] A. Mahmood and U. Köse, "Speech recognition based on convolutional neural networks and MFCC algorithm," *Adv. Artif. Intell. Res.*, vol. 1, no. 1, Art. no. 1, Jan. 2021.

UAV Thermography for Building Energy Audit: Comparing Image Acquisition Strategies

Cheng Zhang^a, Yang Zou^a, Johannes Dimyadi^{b,c} and Vicente Gonzalez^d

^aDepartment of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand ^bSchool of Computer Science, University of Auckland, Auckland 1010, New Zealand ^cCodify Asset Solutions Limited (CAS), Auckland 1010, New Zealand ^dDepartment of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada

czha626@aucklanduni.ac.nz, yang.zou@auckland.ac.nz, jdimyadi@cas.net.nz, yagonzal@ualberta.ca

Abstract -

Maintaining energy performance of ageing building stock has become a global priority towards a more environmentally sustainable future in the age of climate change. Unmanned Aerial Vehicle (UAV) equipped with thermal camera is increasingly used for building inspection. However, scant attention has been paid to the satisfactory image acquisition strategy for a rigorous building energy audit. To narrow this research gap, this paper undertakes a comparative analysis of the impact of three primary features of UAV thermal image collection strategies on accurate thermal anomaly evaluation. Firstly, a laboratory-scale building featuring artificial thermal anomalies was developed and constructed. Then, UAV thermal images were collected under various inspection conditions, examining features including (1) temperature difference between interior and exterior environments, (2) ground sampling distance (GSD), and (3) UAV oblique angle. GSD herein refers to the GSD of the collected UAV thermal images. The collected thermal images underwent a comprehensive analysis and comparison to understand the influence of the three features. The results suggest that: (1) thermal images collected with highest interior and exterior temperature difference are recommended for identifying all potential thermal anomalies; (2) a GSD of within 5 mm is recommended to ensure the visibility of thermal damages in the collected thermal images; (3) a multi-scale thermal image collection strategy is recommended for an efficient and accurate evaluation of thermal anomalies, especially minor defects; (4) a UAV oblique angle within 30° is recommended to ensure a high image contrast between damaged and undamaged areas on the inspected façade.

Keywords -

Building energy audit; Unmanned aerial vehicle (UAV); Aerial thermography; Image acquisition strategy

1 Introduction

Reducing energy consumption and greenhouse gas emission has become a global priority in the age of climate change. The built environment currently accounts for around 30% of global final energy consumption and over 33% of global energy- and process-related CO_2 emission [1]. Maintaining the energy performance of ageing buildings is crucial towards an environmentally sensitive and sustainable future. In recent years, considerable research efforts have been made to understand the building façade energy performance and conduct timely maintenance.

Thermal anomalies in building façades, such as heat loss and insulation failure, are the primary reasons for poor energy performance [2]. One commonly used method to diagnose these thermal anomalies is infrared thermography (IRT), which captures thermal images representing surface temperatures [3]. IRT has been widely used for building energy audits, considering its non-invasive nature [4]. Thermal anomalies generally feature significant variations in the surface temperatures, which are easily identifiable in thermal images [5]. However, traditional terrestrial or hand-held IRT requires inspectors to physically visit the site, which is timeconsuming, labour-intensive, and subject to safety risks, especially when it comes to hard-to-reach areas, such as high-rise buildings.

With the rapid evolution and use of Unmanned Aerial Vehicle (UAV) technology, aerial thermography has become a safe, cost-efficient, and fast solution for collecting close-range thermal images of building façades [6]. Thermal anomalies, such as thermal bridges, heat loss, and insulation failure, can be automatically extracted from UAV-collected thermal images through digital image processing [7] and deep learning [8] algorithms. Additionally, a 3D thermal model can be generated to facilitate a comprehensive understanding of the building energy performance [9].

A key factor in achieving successful UAV thermography for building energy audits is the implementation of a suitable image acquisition strategy, i.e., UAV flight path planning [10]. Two primary factors should be considered during the path planning, i.e., UAV oblique angle and ground sample distance (GSD) [11]. Theoretically, collecting close-range thermal images with a perpendicular view to the inspected surface is recommended [12]. However, meeting these conditions during real-world UAV-enabled building inspections can be challenging since: (1) UAV needs to maintain a safe distance to the inspected façades to avoid collisions; (2) as the GSD decreases, the inspection distance decreases and the inspection duration increases significantly, posing constraints on completing the inspection within the UAV's flight time; and (3) surrounding obstacles, such as trees and power lines, may obstruct areas that would allow the UAV to achieve a perpendicular view. Therefore, it is crucial to understand the influence of different UAV inspection configurations on inspecting thermal anomalies.

Current practices have examined the influence of UAV oblique angle in thermal image-based 3D reconstruction [13] and surface temperature measurement [14], while its impact on detecting thermal anomalies stays unrevealed. Pan et al. [15] obtained similar thermal anomalies detection accuracy from images with an inspection distance of 5 m and 10 m. However, this observation was based on relatively large thermal anomalies with a dimension larger than 1.5 m. Additionally, Mayer et al. [11] investigated the influence of different UAV settings (e.g., flight speed, oblique angle and inspection distance) on the quality of the collected thermal images. However, their analysis focused only on the overall quality (e.g., image resolution, coverage, and contrast) of thermal images with a large field of view, and thus failing to provide insights in the optimal UAV configurations for close building thermal inspection and anomaly detection.

In addition to the UAV configurations, the environmental conditions also play a vital role [16]. Particularly, a minimum temperature difference of 10 °C between interior and exterior environments is recommended for thermal building inspection [12]. Although this temperature difference can be achieved by conducting inspections during night or early morning with heating system switched on [16], this time restraint severely impedes the widespread use of UAV-enabled building energy audits. Therefore, it is vital to explore the feasibility of identifying thermal anomalies from images collected with a relatively low difference in interior and exterior temperatures.

To better understand the influence of different factors on accurate thermal anomaly inspection, it is crucial to collect thermal images with known thermal anomalies under various conditions. However, collecting these thermal images during real-world building inspections presents challenges. To address this issue, this study designed and developed a laboratory-scale building with artificial thermal anomaly (e.g., sub-surface insulation failure and surface deterioration). Thermal images were further obtained under different UAV oblique angles, GSD, and interior and exterior temperature differences. The collected thermal images were subsequently analysed to compare the influence of different image acquisition strategies.

The structure of this paper is organised as follows: Section 2 presents the design and production of the laboratory-scale building model. Section 3 describes the laboratory experiment. Results and discussions are illustrated in Section 4. Finally, Section 5 summarises the findings with a conclusion.

2 Laboratory-scale building

To support the comparison of different image acquisition strategies for UAV-enabled building thermography, a laboratory-scale building was developed, as shown in Figure 1. Timber buildings have garnered global popularity due to their unique attributes, such as eco-friendliness, widespread availability, and relative ease of handling [17,18]. As such, a timber building was selected as the subject for this laboratory investigation. The structure decided upon was a singlepanel timber-frame wall element, composed of three layers: a timber frame, a single-sided sheathing board, and thermal insulation [19,20]. This research primarily focused on thermal anomalies in an individual wall, thus the dimension and shape of the building are considered less significant and are designed in reference to [21]. A timber frame with a dimension of $600 \text{ } mm \times 600 \text{ } mm \times$ 600 mm was first assembled, and further, the plywood panels with a thickness of 7 mm were affixed to the timber frame to form the façade exterior, as demonstrated in Figure 1(a) and (b). Additionally, 25 mm thick polystyrene ThermaSlab, with a thermal resistance rating of 0.66 $m^2 K/W$, was applied to the façades and roof for insulation purposes.

It should be noted that conventional timber-based walls used in construction industry often incorporate a building wrap, also known as membrane, which serves primarily to weatherproof the structure by blocking rainwater while allowing for vapour transfer to mitigate condensation [22]. However, its impact on the building energy performance is minimal compared to the thermal insulation layer, particularly given that the laboratory investigation takes place indoors, absent of any rain or vapour. Therefore, to simplify the construction process, the building wrap was excluded.

Two types of thermal anomalies on building façades, i.e., surface deterioration [23] and sub-surface insulation failure [21], were added in the laboratory-scale building. Figure 1(c) shows the surface deterioration with various dimensions, achieved by accurately cutting the surface panel using a laser cutting machine. Additionally, insulation failures, such as missing and thin insulations, were achieved by manually cutting the ThermaSlab, as shown in Figure 1(d). Particularly, for the thin insulation, the thickness of the insulation was reduced to 10 mm. Detailed dimensions of the artificial thermal anomalies are summarised in Table 1. It should be noted that although the laboratory-building was downscaled, the anomalies were designed to mirror real-world dimensions for accurate representation.



Figure 1. Laboratory-scale building with artificial anomaly: (a) exterior; (b) interior; (c) surface deterioration; (d) sub-surface insulation failure.

Thermal anomaly	Dimension
Surface deterioration	150 mm imes 100 mm
	150 mm imes 50 mm
	150 mm imes 20 mm
	$150 \ mm imes 10 \ mm$
	150 mm imes 5 mm
Missing insulation	$200 \ mm imes 100 \ mm$
Thin insulation	$200 mm \times 100 mm$

3 Laboratory experiment

Laboratory experiments were further conducted using the developed building model. Figure 2(a) provides an overview of the laboratory experiment. The UAV adopted in this study is DJI Mavic 2 Enterprise Dual, equipped with a dual camera system for capturing thermal and RGB images simultaneously. The RGB camera has an image resolution of 4056×3040 , and the thermal camera has a resolution of 640×480 . It should be noted that the thermal camera was calibrated and registered with the RGB camera in our previous work [9].

During the experiments, a fan heater was utilised to warm up the building interior, as shown in Figure 2(b). A wireless temperature sensor was used to monitor the interior temperature. By combining with the exterior temperature (approximately 20 °C) measured by another sensor, the temperature difference can be measured. Additionally, AprilTag [24] was attached to the building façade to calibrate the collected images and measure their corresponding GSD. The experiment mainly includes three cases, each related to a different inspection feature, i.e., interior and exterior temperature difference, GSD and UAV oblique angle.





Case 1 aimed to compare the influence of various interior and exterior temperature differences. The UAV maintained a distance of 2 m (corresponding to a GSD of around 3 mm) and a perpendicular view to the inspected façade. Then, the heater was turned on to gradually warm up the building interior. Thermal images were collected

as the temperature difference increased.

Case 2 focused on studying the influence of different GSD. After the building interior was warmed up and the temperature was stabilised, thermal images were collected with a perpendicular view to the inspected façade and at varying distances ranging from 1 m to 9 m to achieve a GSD ranging from 1.5 mm to 13 mm.

Case 3 examined the influence of different UAV oblique angles, as shown in Figure 2(c). The experiment was conducted at a distance of 2 m (corresponding to a GSD of around 3 mm) and with a stable interior and exterior temperature difference. The maximum UAV oblique angle was set at 50°, considering that thermal images collected with a larger oblique angle have a massive field of view and are rarely used for actual building inspection.

4 Results and discussion

This section demonstrates the laboratory experiment results and discusses findings regarding the image acquisition strategy for UAV-enabled building energy audit.

4.1 Temperature difference

Figure 3 shows the thermal images collected to identify insulation anomalies under different temperature differences, ranging from 1.6 °C to 19.5 °C. It can be observed that the presence of missing insulation becomes noticeable at a temperature difference of 3.8 °C, while thin insulation requires a higher temperature difference of approximately 8.1 °C. Particularly, the colour of the thin insulation closely resembles the undamaged façade area surrounding the missing insulation, even at higher temperature difference of over 14.5 °C. This similarity may be attributed to a small gap between the façade panel

and insulation layer, allowing heat transfer from the missing insulation area to the surrounding area. This effect is particularly prominent at high temperature differences exceeding $14.5 \,^{\circ}$ C, where the missing insulation areas appear red and a small surrounding area stands out with a deep yellow colour, indicating a higher temperature. Additionally, as the temperature difference increases, the missing insulation areas becomes more distinguishable. Thus, it is recommended to capture thermal images with a higher temperature difference for identifying and quantifying missing insulation.

Thermal images collected for identifying surface deteriorations under different temperature differences are displayed in Figure 4. Surface deteriorations, even small defects, start to appear at low temperature differences, such as 3.7 °C. As the temperature difference increases, surface deterioration becomes more noticeable. Similar to the missing insulation, the surrounding area of the surface deteriorations also exhibits a different colour compared to other undamaged façade areas. Particularly, with a significant temperature difference exceeding 11.9 °C, surface deterioration with relatively large such as over 20 mm, becomes dimensions, distinguishable (i.e., red colour) compared to the surrounding areas (i.e., deep yellow colour). This means that a higher temperature difference allows accurately identifying and quantifying these surface deteriorations. However, smaller surface deteriorations, such as 10 mm and 5 mm defects, stay merged with the surrounding areas, making it challenging to accurately extract and segment them, even at an extremely high temperature difference of 22.6 °C. Therefore, although small deteriorations can be visible in thermal images, accurately quantifying these deteriorations poses a challenge. Collecting thermal images with a smaller GSD may help with identifying minor anomalies, which will be discussed in the following section.



Temperature difference increasing

Figure 3. Thermal images of insulation anomalies under different temperature differences: GSD of around 3 mm and UAV oblique angle of approximately 0°.

3.7°C	5.6°C	7.7°C	9.6°C	11.9℃
			ET.	EEL
		in the second second		
4				
22.6°C	20.5°C	18.6°C	16.4°C	14.4°C
22.6°C 5 mm 50 mm 10 mm 100 mm	20.5°C	18.6°C	16.4°C	14.4°C

Temperature difference increasing

Figure 4. Thermal images of surface deterioration under different temperature differences: GSD of around 3.3 mm and UAV oblique angle of approximately 0°.

GSD increasing



Figure 5. Thermal images of insulation anomalies under different GSD: temperature difference of around 18 °C and UAV oblique angle of approximately 0°.



Figure 6. Thermal images of surface deterioration under different GSD: temperature difference of around 29 °C and UAV oblique angle of approximately 0°.

To summarise, an interior and exterior temperature difference exceeding 8 °C is required to detect various

thermal anomalies, including minor defects, such as thin insulation. As the temperature difference increases,

thermal anomalies become more noticeable in the collected UAV thermal images. Therefore, it is advisable to conduct building inspections with the highest temperature difference to achieve maximum thermal image contrast between damaged and intact areas.

4.2 GSD

To understand the influence of GSD in the UAVenabled building thermography, thermal images were collected with a GSD ranging from 1.5 mm to 16 mm, as shown in Figure 5 and Figure 6. The missing insulation remains visible as a distinct red colour, even with a GSD of 1.5 mm, as shown in Figure 5. However, the thin insulation starts to blend with the undamaged areas when the GSD reaches 6.6 mm. Similarly, small surface deteriorations of 5 mm, 10 mm and 20 mm become invisible with a GSD of 6.6 mm. Additionally, even large surface deteriorations of 50 mm and 100 mm disappeared from the thermal image taken with a GSD of 10.2 mm. This may be due to the thin insulation and surface deteriorations exhibiting a smaller temperature difference compared to the undamaged areas, as they are relatively minor thermal damages compared to the missing insulation. Based on these observations, a GSD of within 5 mm is recommended to ensure that all thermal damages are visible in the collected thermal images.

However, a smaller GSD may be needed to accurately measure and quantify these thermal damages, especially with regard to minor thermal anomalies. For example, surface deterioration with a width of 10 mm is distinguishable as red colour in the thermal image collected with a GSD of 1.6 mm, while starting merged with the surrounding areas (appearing as deep yellow colour) with a GSD of 3.2 mm, as shown in Figure 6.

Therefore, a multi-scale thermal image collection strategy is recommended considering the limited flight time of most UAV (around 30 mins). Specifically, thermal images may first be collected with a GSD less than 5 mm distance to identify all possible areas with thermal damages on building façades. Then, close-range thermal images with a smaller GSD should be taken around these identified areas to ensure accurate detection and measurement of the thermal damages. Considering that most commercial UAV can not fly too close to the inspected surface due to safety concerns, it is recommended to develop customised UAV (e.g., wallclimbing UAV [25]), that are specifically designed for collecting close-range images.

4.3 UAV oblique angle

Figure 7 and Figure 8 display thermal images collected from various UAV oblique angles, ranging from 0° to 50° . Although thermal anomalies stay visible

in the thermal images, even with a large oblique angle of 50°, the contrast between the damaged and the undamaged façade areas decreases as the oblique angle increases. Particularly, when the oblique angle reaches 30°, the missing insulation starts to blend with the surrounding areas, as shown in Figure 7. This could be attributed to the fact that a large oblique angle may capture reflections from other sources, thereby reducing the ability to accurately capture the emissivity of the target surfaces [14]. Additionally, as the UAV oblique angle increases, the field of view of the thermal images also increases significantly, covering unwanted backgrounds. These backgrounds can also reduce the image contrast of the target areas. Therefore, a UAV oblique angle within 30° is recommended when collecting thermal images for building energy audits.

UAV oblique angle increasing



Figure 7. Thermal images of insulation anomalies under different UAV oblique angles: temperature difference of around 16 °C and GSD of approximately 3.3 mm.



UAV oblique angle increasing

Figure 8. Thermal images of surface deterioration under different UAV oblique angles: temperature difference of around 28 °C and GSD of approximately 3.3 mm.

5 Conclusion

This paper compared three features of image acquisition strategies of UAV thermography for building energy audit. A laboratory-scale building with artificial thermal anomalies, such as sub-surface insulation failure and surface deterioration, was first developed. Furthermore, thermal inspections were conducted under various conditions, such as different UAV oblique angles, GSD, and interior and exterior temperature differences. The results indicate that:

- (1) An interior and exterior temperature difference exceeding 8 °C is necessary to identify potential thermal anomalies from UAV thermal images. Conducting building inspections with the highest temperature difference is advisable to achieve maximum thermal image contrast between damaged and intact areas.
- (2) A GSD of within 5 *mm*, is required to ensure the visibility of thermal anomalies in the collected thermal images. A multi-scale thermal image collection strategy is recommended to ensure efficient and accurate measurement of various thermal anomalies, especially minor defects.
- (3) An UAV oblique angle of within 30° is recommended to be maintained during the thermal

building inspection.

The comparison presented in this paper has some limitations and room for improvement. As an ongoing research project, future work will involve quantitatively evaluating thermal damage detection from the collected thermal images and investigating the interferences between the three features. Additionally, given the relatively straightforward nature of the constructed laboratory-scale building, forthcoming investigations will involve field studies on real-world buildings, which may contain more complex anomalies, to validate the identified image acquisition strategies.

Acknowledgements

The authors appreciate the assistance provided by two undergraduate students, Yuvraj Behal and Mohammed Samdani of the University of Auckland, in the development of the laboratory-scale building.

References

- [1] C. Delmastro, O. Chen, Buildings. The International Energy Agency (IEA) (2023). Retrieved 28 February 2024 from https://www.iea.org/reports/buildings.
- [2] M.H. Shariq, B.R. Hughes, Revolutionising building inspection techniques to meet largescale energy demands: A review of the state-ofthe-art, Renewable and Sustainable Energy Reviews 130 (2020), pp. 109979, <u>https://doi.org/10.1016/j.rser.2020.109979</u>.
- [3] M. Martin, A. Chong, F. Biljecki, C. Miller, Infrared thermography in the built environment: A multi-scale review, Renewable and Sustainable Energy Reviews 165 (2022), pp. 112540,

https://doi.org/10.1016/j.rser.2022.112540.

- [4] A. Kirimtat, O. Krejcar, A review of infrared thermography for the investigation of building envelopes: Advances and prospects, Energy and Buildings 176 (2018), pp. 390-406, <u>https://doi.org/10.1016/j.enbuild.2018.07.052</u>.
- [5] A. Ramón, A. Adán, F. Javier Castilla, Thermal point clouds of buildings: A review, Energy and Buildings 274 (2022), pp. 112425, <u>https://doi.org/10.1016/j.enbuild.2022.112425</u>.
- [6] T. Rakha, A. Gorodetsky, Review of Unmanned Aerial System (UAS) applications in the built environment: Towards automated building inspection procedures using drones, Automation in Construction 93 (2018), pp. 252-264,

https://doi.org/10.1016/j.autcon.2018.05.002.

- [7] N. Bayomi, S. Nagpal, T. Rakha, J.E. Fernandez, Building envelope modeling calibration using aerial thermography, Energy and Buildings 233 (2021), pp. 110648, <u>https://doi.org/10.1016/j.enbuild.2020.110648</u>.
- [8] Z. Mayer, J. Kahn, Y. Hou, M. Götz, R. Volk, F. Schultmann, Deep learning approaches to building rooftop thermal bridge detection from aerial images, Automation in Construction 146 (2023), pp. 104690, https://doi.org/10.1016/j.autcon.2022.104690.
- [9] C. Zhang, Y. Zou, J. Dimyadi, R. Chang, Thermal-textured BIM Generation for Building Energy Audit with UAV Image Fusion and Histogram-based Enhancement, Energy and Buildings 301 (2023), pp. 113710, https://doi.org/10.1016/j.enbuild.2023.113710.
- [10] M. Mahmoodzadeh, V. Gretka, P. Mukhopadhyaya, Challenges and opportunities in quantitative aerial thermography of building envelopes, Journal of Building Engineering 69 (2023), pp. 106214, https://doi.org/10.1016/j.jobe.2023.106214.
- [11] Z. Mayer, A. Epperlein, E. Vollmer, R. Volk, F. Schultmann, Investigating the Quality of UAV-Based Images for the Thermographic Analysis of Buildings, Remote Sensing 15 (2) (2023), pp. 301, <u>https://doi.org/10.3390/rs15020301</u>.
- [12] International Organization for Standardization (ISO), Thermal insulation-Qualitative detection of thermal irregularities in building envelopes-Infrared method, Standard ISO 6781, 1983, <u>https://www.iso.org/standard/13277.html</u>.
- [13] S. Dabetwar, N.N. Kulkarni, M. Angelosanti, C. Niezrecki, A. Sabato, Sensitivity analysis of unmanned aerial vehicle-borne 3D point cloud reconstruction from infrared images, Journal of Building Engineering 58 (2022), pp. 105070, <u>https://doi.org/10.1016/j.jobe.2022.105070</u>.
- [14] J. Ortiz-Sanz, M. Gil-Docampo, M. Arza-García, I. Cañas-Guerrero, IR thermography from UAVs to monitor thermal anomalies in the envelopes of traditional wine cellars: Field test, Remote Sensing 11 (12) (2019), pp. 1424, https://doi.org/10.3390/rs11121424.
- [15] N.H. Pan, C.H. Tsai, K.Y. Chen, J. Sung, Enhancement of external wall decoration material for the building in safety inspection method, Journal of Civil Engineering and Management 26 (3) (2020), pp. 216-226, <u>https://doi.org/10.3846/jcem.2020.11925</u>.
- [16] E. Lucchi, Applications of the infrared thermography in the energy audit of buildings: A review, Renewable and Sustainable Energy

Reviews 82 (2018), pp. 3077-3090, https://doi.org/10.1016/j.rser.2017.10.031.

- [17] Z. Chen, M. Popovski, C. Ni, A novel floorisolated re-centering system for prefabricated modular mass timber construction – Concept development and preliminary evaluation, Engineering Structures 222 (2020), pp. 111168, <u>https://doi.org/10.1016/j.engstruct.2020.11116</u> <u>8</u>.
- [18] M. Caniato, A. Marzi, S. Monteiro da Silva, A. Gasparella, A review of the thermal and acoustic properties of materials for timber building construction, Journal of Building Engineering 43 (2021), pp. 103066, <u>https://doi.org/10.1016/j.jobe.2021.103066</u>.
- [19] V.Ž. Leskovar, M. Premrov, Influence of glazing size on energy efficiency of timberframe buildings, Construction and Building Materials 30 (2012), pp. 92-99, <u>https://doi.org/10.1016/j.conbuildmat.2011.11.</u> 020.
- [20] K. Vogrinec, M. Premrov, E. Kozem Šilih, Simplified modelling of timber-framed walls under lateral loads, Engineering Structures 111 (2016), pp. 275-284, <u>https://doi.org/10.1016/j.engstruct.2015.12.029</u>.
- [21] M. Puliti, G. Montaggioli, A. Sabato, Automated subsurface defects' detection using point cloud reconstruction from infrared images, Automation in Construction 129 (2021), pp. 103829,

https://doi.org/10.1016/j.autcon.2021.103829.

[22] K. Orlowski, Automated manufacturing for timber-based panelised wall systems, Automation in Construction 109 (2020), pp. 102988,

https://doi.org/10.1016/j.autcon.2019.102988.

- [23] E. Bauer, E. Pavón, E. Barreira, E. Kraus De Castro, Analysis of building facade defects using infrared thermography: Laboratory studies, Journal of Building Engineering 6 (2016), pp. 93-104, https://doi.org/10.1016/j.jobe.2016.02.012.
- [24] E. Olson, AprilTag: A robust and flexible visual fiducial system, 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 3400-3407, https://doi.org/10.1109/ICRA.2011.5979561.
- [25] S. Jiang, J. Zhang, Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system, Computer-Aided Civil and Infrastructure Engineering 35 (6) (2020), pp. 549-564, <u>https://doi.org/10.1111/mice.12519</u>.

NeRF-Con : Neural Radiance Fields for Automated Construction Progress Monitoring

Yuntae Jeon¹, Almo Senja Kulinan¹, Dai Quoc Tran², Minsoo Park³ and Seunghee Park⁴

¹Department of Global Smart City, Sungkyunkwan University, Suwon, South Korea

²Global Engineering Institute for Ultimate Society, Sungkyunkwan University, Suwon, South Korea

³Sungkyun AI Research Institute, Sungkyunkwan University, Suwon, South Korea

⁴School of Civil, Architectural Engineering and Landscape Architecture, Sungkyunkwan University, South Korea

jyt0131@g.skku.edu, almosenja14@skku.edu, daitran@skku.edu, pms5343@skku.edu, shparkpc@skku.edu

Abstract -

The monitoring of construction progress is crucial for ensuring project timelines, budget adherence, and quality control. Traditional methods often involve manual inspection, which is labor-intensive and prone to human error. We introduce NeRF-Con, an innovative approach utilizing Neural Radiance Fields (NeRF) to automate the process of construction progress monitoring. NeRF-Con can infer images that render the construction site with a level of quality comparable to reality by utilizing NeRF, which synthesizes novel views of complex scenes from a sparse set of images. Additionally, by employing a segmentation model, NeRF-Con can compare these rendered images with BIM to evaluate the progress of the work. This capability is achieved by training the model using handheld smartphone-captured video. This paper details a method for applying NeRF in real construction sites with data collection, pre-processing, and progress evaluation. In assessing the model's performance, comparisons are made with data from mobile-LiDAR, stand-LiDAR, and BIM. With this research, we suggest potential future studies in applying NeRF models to construction progress monitoring systems.

Keywords -

NeRF; 3D Computer Vision; Deep Learning; Segmentation; Construction Progress Monitoring

1 Introduction

In the field of construction, progress monitoring stands as a essential work ensuring timely and cost-effective project delivery. The advent of advanced AI and deep learning technologies has initiated a new era of innovation in this domain, enabling automated progress monitoring with remarkable accuracy and efficiency. In recent years, AI advancements utilizing computer vision, such as object detection and instance segmentation for construction object recognition, have been increasingly adopted, transforming traditional monitoring techniques with automated, data-driven approaches. Among these advancements, Neural Radiance Fields (NeRF) [1] have emerged as a innovative approach in the field of 3D data processing and visualization. This study introduces NeRF as a deep learning model that excels in synthesizing photo-realistic images by considering light and material properties, rendering images on novel views in construction sites or built environments that closely replicate real-life visuals. The integration of NeRF into construction progress monitoring marks a significant advancement, providing a method that not only improves visual comprehension but also greatly contributes to the automation and precision of tracking construction progress.

For automating construction progress monitoring, the integration of vision sensors and deep learning methods has drastically changed traditional approaches. Beginning with the use of traditional image processing skills like filtering, edge and corner detection to analyze site images [2], the approach has evolved to incorporate deep learning for object detection [3] and segmentation [4]. This advancement significantly improves the accuracy of construction progress assessments from 2D sensors by enabling more precise comparisons of site images with designs derived from Building Information Modeling (BIM).

Furthermore, the progression in construction monitoring has greatly benefited from the adoption of 3D scanning technologies like LiDAR [5], which have revolutionized the field by enabling comprehensive three-dimensional site captures. These methods allow for detailed and precise comparisons between ongoing construction and BIM designs. Advancements such as real-time 3D point cloud mapping with Simultaneous Localization and Mapping (SLAM) [6], further enhance geometry analysis in construction environments. Combining these cutting-edge 3D scanning techniques with AI and deep learning significantly improves the accuracy and efficiency of construction monitoring, setting a new standard in the industry.

While previous studies in automated construction progress monitoring have significantly utilized 2D and 3D sensing technologies for gathering building or construction site data, they commonly entail transforming scanned data into a 2D image with orthogonal view [4, 7, 8]. However, challenges remain, such as: 1) Achieving efficiency and quality in rendered parallel 2D images. The creation of parallel 2D images from RGB cameras is a detailed, rulebased process requiring manual refinement. Moreover, while LiDAR or SLAM methods often lack the realistic appearance of actual images, resulting in lower quality renderings. 2) The cost and user-convenience of data acquisition. Methods such as SLAM, which utilize robotic or drone sensing, necessitate predefined operational paths. Compared to manual, hand-held capture, these methods are operationally more complex and constrained by environmental factors like limited pathways or airspace, reducing their feasibility in diverse construction environments.

To address the issues of existing 2D and 3D sensor-based methods in automated progress monitoring, we propose an approach utilizing NeRF. This approach utilizes deep learning to achieve a degree of realism in spatial rendering that significantly exceeds the capabilities of traditional methods. A key advantage of our methodology is the use of smartphone-captured video as input. Furthermore, our approach is not limited to rendering the site in 3D; it also generates orthogonal views, which can be directly compared with BIM for accurate construction process monitoring. Our NeRF-based method's ability to generate both realistic 3D renderings and orthogonal views establishes it as a versatile and effective solution for construction progress monitoring. We further enrich our research by testing and comparing various NeRF models-vanilla NeRF [1], Instant-NGP [9], and Nerfacto [10]. Utilizing the Nerfstudio [10] platform, we efficiently train and visualize our models. Our research includes the collection and analysis of data from two different indoor scenes and one outdoor scene, all derived from actual built environments. Our main contributions are:

- We utilize the concept of neural radiance fields (NeRF) to comprehend the 3D spatial information of construction sites and render images from novel views that closely resemble the actual environment.
- We demonstrate the use of a common smartphone camera, easily handheld and maneuvered, to capture videos in a user-friendly and uncomplicated manner. These videos are then used as the input for NeRF model training.
- We evaluate and apply the NeRF model in various built environments, including indoor and outdoor settings, specifically for the purpose of automated construction progress monitoring.

2 Background

2.1 Automated progress monitoring

Computer vision technology has increasingly been applied in automated construction monitoring. Initial approaches involved image processing techniques like edge detection and deep learning-based object segmentation to compare material edges with as-designed BIM [2, 3, 4]. The focus then shifted to LiDAR-based 3D scanning [5], providing detailed site comparisons with BIM, typically evaluated using Root Mean Square Error (RMSE). Advancements continued with SLAM [6], using moving robots capable of capturing diverse scenes, thereby enhancing segmentation and detection for more accurate progress tracking against as-designed BIM. Recently, Pal et al. [8] employed vanilla NeRF [1] to generate orthographic views of under-construction elements, performing semantic segmentation to monitor construction progress in comparison with BIM designs. In this paper, we utilize various NeRF models such as vanilla NeRF, Instant-NGP [9], and Nerfacto [10].

2.2 Neural Radiance Field (NeRF)

Neural Radiance Fields [1], or NeRF, represent a novel approach in the field of 3D scene reconstruction from 2D images. Traditionally, rendering realistic 3D objects involved the use of expensive 3D scanners or photogrammetry that transform images into voxel, point cloud, or mesh forms, NeRF introduces a novel approach in novel view synthesis. In recent, research in the field of 3D computer vision is largely centered around the use of NeRF. First, Vanilla NeRF [1], as the foundational model, utilizes an Multi-Layer Perceptron (MLP) with 8 linear layers, offering a distinctive approach to 3D scene representation. It processes 3D coordinates through positional encoding to enrich the input data, thereby enhancing the details captured in the scene. This architecture extracts density outputs and integrates ray viewing directions, allowing the final RGB output to dynamically reflect how the appearance of objects changes with the viewer's perspective. Building on this, Instant-NGP [9] innovates by encoding coordinates with HashMap and Linear Interpolation to significantly reduce computational load and accelerate training. This approach efficiently creates feature vectors from selected coordinates and auxiliary values, streamlining the MLP processing. Finally, Nerfacto [10] builds upon previous NeRF advancements by combining several techniques for real data capture of static scenes. It integrates camera pose refinement, per-image appearance conditioning, proposal sampling, scene contraction, and hash encoding.



Figure 1. Overview of NeRF-Con pipeline for construction progress monitoring

3 Method

Fig. 1 shows our NeRF-based method for automated construction progress monitoring, starting with data collection. This model is then trained to accurately render photo-realistic 2D images of the site from novel perspectives, aiming for realistic visualizations of the 3D space. The final comparison stage involves aligning NeRF-generated images, potentially orthographic views, with BIM designs, employing instance segmentation for precise progress assessment. This method, leveraging NeRF's strengths in 3D spatial representation and image synthesis, offers a novel, accurate, and efficient approach to quantifying construction progress.



Figure 2. Visualization of the actual move path for capturing a video (above) and the Nerfacto model application on the Nerfstudio platform (below)

3.1 Data collection

For video capture in construction progress monitoring, various methods are employed, such as smartphones, drones, and robots, which can be hand-held or integrated into automated systems. We predominantly utilize the iPhone 15 Pro, chosen for its high accessibility and efficiency, and employ COLMAP [11], a structure-frommotion (SfM) technique, to extract camera poses from image sequences. This approach, as depicted in Fig. 2, involves using SfM, a photogrammetric method, to estimate three-dimensional structures from two-dimensional images. The process identifies key features across images and uses their relative motion to infer depth and structure, with a focus on the epipolar line, which indicates the trajectory of a point in one image across another, based on camera movement. COLMAP processes video frames to generate accurate 3D coordinates and view directions for the camera, constructing a 3D point cloud of the site and determining the camera's position and orientation for each frame. This method maintains the practicality and convenience of data collection, ensuring regular monitoring feasibility across various environments without specialized equipment. The detailed process guarantees a precise representation of the construction site, facilitating high-precision training of the NeRF model, aligning with our goals for efficient and comprehensive construction progress monitoring.

3.2 NeRF application

The fundamental concept of NeRF [1] involves sampling points in a 3D space along rays that emanate from



Figure 3. The architecture of the Nerfacto model

the camera's viewpoint. These sampled points are used to estimate both the color and the density at every location through the scene, which are then composited into a final image via volumetric rendering techniques. The input to a NeRF model typically includes the 3D coordinates of sample points, the direction of the viewing rays, and any appearance features that characterize the scene's properties, such as lighting or texture. The output is the rendered 2D image that approximates the real-world scene from the camera's perspective. Fig. 3 shows Nerfacto [10] model's pipeline for NeRF application at construction sites by creating photo-realistic images from captured videos.

3.2.1 Piecewise Sampler

The rendering pipeline begins with the Piecewise Sampler, selecting sample points along camera rays to evaluate the scene. It distributes half of the samples uniformly within a distance of 1 (unit distance) from the camera, ensuring a detailed sampling of nearby areas. The remaining samples are placed with increasing step sizes, effectively scaling the sampling frustums and allowing for a broader coverage that includes distant objects without compromising the sampling density for closer areas. This two-part approach can be expressed as:

$$d_i = \begin{cases} \frac{i}{N/2} & \text{if } i \le \frac{N}{2}, \\ f(i) & \text{if } i > \frac{N}{2}, \end{cases}$$
(1)

where *i* is the index value of the samples, d_i is the distance from the camera, *N* is the total sample count, and f(i), a monotonically increasing function, adjusts the samples based on conical frustum.

3.2.2 Proposal Sampler

After the initial sampling phase, the Proposal Sampler, utilizing two density functions, refines the sample locations. Its primary goal is to maximize sampling around surface boundaries, which are crucial for the scene's visual accuracy. These density functions, constituting the Density Field, guide the sampling process. Each density function in the Proposal Sampler is an MLP that receives 3D coordinates as input and is combined with hash encoding. This structure is designed to provide a coarse representation of density, which is crucial for efficient sampling. The density function can be expressed as:

$$\rho(\mathbf{x}) = \Theta_{\text{density}}(\phi(\mathbf{x})) \tag{2}$$

where **x** is a spatial coordinate, ϕ is a hash encoder [9], Θ_{density} is MLP for density and $\rho(\mathbf{x})$ is the estimated density at that location. The hash encoding transforms the 3D coordinates into a suitable format for the MLP, enabling it to compute the density. The two density fields in the Proposal Sampler work together to concentrate sample points around significant areas like surface boundaries. The design of these density fields focuses on capturing only a coarse representation of scene density. This approach is sufficient for guiding the sampling process, ensuring that the model concentrates computational resources on the most important aspects of the scene without being burdened by the intricacies of high-frequency details.

3.2.3 Nerfacto Field

The Nerfacto Field is an integral component of the rendering pipeline that takes as input the 3D coordinates \mathbf{x} , the view direction \mathbf{d} , and the appearance features \mathbf{f} , and outputs both the color *C* and the density $\rho(\mathbf{x})$ at the given spatial location. For extracting the RGB color, the Nerfacto Field employs a neural network function which can be expressed as:

$$C(\mathbf{x}, \mathbf{d}) = \Theta_{\text{RGB}}(\phi(\mathbf{x}), SH(\mathbf{d}), \mathbf{f})$$
(3)

where Θ_{RGB} is MLP for density, *SH* is the spherical harmonic encoding of the view direction, and the appearance features **f** capture the variations in scene appearance such as lighting and material properties. The density $\rho(\mathbf{x})$ is inferred using the same equation to Eq. 2.

3.2.4 Volumetric Render

The last stage in the pipeline is the Volumetric Render, which integrates the density and color information along the rays to form the final rendered image. This integration can be described by the following equation:

$$\text{RGB}_{\text{final}} = \int \rho(\mathbf{x}) \cdot C(\mathbf{x}, \mathbf{d}) \, d\mathbf{x}, \tag{4}$$

where the integration is performed along the ray path, accumulating the product of density and color to yield the final pixel color value. The rendered color RGB_{final} is then compared to the ground truth image's RGB values, using the L-2 distance as a loss function during the training process. This loss function quantifies the difference between the rendered image and the actual image, guiding the optimization of the network parameters to minimize these discrepancies.

We delve into the application of NeRF for creating 2D novel images. These orthogonal projected images offer a distinctive view of construction sites. Utilizing the NeRF model, we efficiently segment building elements through the semantic-segment-anything [12]. This segmented output is then compared with 2D plane images derived from BIM model, which similarly employ orthogonal projections. By evaluating the segmented outcomes from the NeRF model against those from BIM, we are able to not only gauge construction progress with great precision but also visually confirm the consistency with the original design. This approach provides a layered insight into project development, facilitating a thorough comparison between what was planned and what is being constructed.

4 **Experiments**

4.1 Dataset

Our study involved three experiments - two indoors and one outdoors - using an iPhone 15 Pro for data collection. In the first indoor experiment, a 90-second site video was captured for NeRF model training, complemented by a 120-second mobile LiDAR (iPhone 15 Pro) scan and a 210-second FARO LiDAR scan, the latter offering higher accuracy but at a significantly higher cost (50x expensive) than mobile LiDAR. The scanning time differences between mobile LiDAR and FARO LiDAR are due to their operational designs. Mobile LiDAR, a handheld device, necessitates manual navigation for comprehensive site imaging, conversely, FARO LiDAR, a stationary system, automates image capture from all directions.

The second indoor experiment used a 120-second video capture and the last outdoor experiment used a 20-second video. Notably, in each experiment, we downsampled the video frames to one-third of the total frames for both training and testing, dividing the data in a 0.9 to 0.1 training-to-

testing ratio. This methodology created a diverse dataset, integrating various technologies for a comprehensive assessment of our NeRF-based monitoring system.

4.2 Implementation details & Metrics

We utilize the Nerfstudio [10] platform for train and visualization, and our experiments involved three NeRF models – Nerfacto, instantNGP, and vanilla NeRF – to compare their performance. Common settings across these models included 200k iterations and 4096 for train/test number of rays per batch. For Nerfacto and instantNGP, the optimizer used was Adam with a learning rate of 0.01. In contrast, vanilla NeRF utilized the RAdam optimizer, featuring a lower learning rate of 0.0005. These models were trained on an NVIDIA RTX4090 GPU, using PyTorch version 2.0.1 and CUDA 11.8, ensuring efficient computation and model optimization.

To evaluate the performance of our models, we employed three key metrics: PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index Measure), and LPIPS (Learned Perceptual Image Patch Similarity) [13].

- **PSNR:** Expressed in the logarithmic decibel scale, with values typically ranging from 20 to 30. Higher values indicate better image quality.
- **SSIM:** Values range between -1 and 1. A value of 1 indicates perfect similarity between the test image and the reference image. SSIM assesses visual impacts based on luminance, contrast, and structure, aligning more with human visual perception than PSNR.
- LPIPS: Scores typically range from 0 to 1, where a lower score indicates greater perceptual similarity between compared images. Unlike PSNR and SSIM, LPIPS leverages deep learning models to better approximate human visual perception.

4.3 Results

In our research, we conducted a comparison using two different LiDAR sensors with NeRF-based approaches. One of the LiDAR sensors is a mobile LiDAR embedded in the iPhone 15 Pro, utilizing Pix4Dcatch for analysis. The other is a stationary Faro LiDAR, known for its exceptional precision and high cost. In contrast, for our NeRF-based approaches, we used a smartphone or drone equipped with only a RGB camera. Thus, we experimented with three different NeRF models' rendering image quality (Tab. 1) related to creating the parallel view images and performed an additional comparison between NeRF, stable LiDAR, and BIM with SAM [12] (Fig. 5). We also visualized the infeasible result from the mobile LiDAR (Fig. 6).



Corridor

Under construction



Figure 4. Qualitative comparison of three NeRF models on three different scenes

1157



Figure 5. Visualization of semantic segmented results on 2D orthogonal images from three different sources

Method	PSNR ↑	SSIM ↑	LPIPS ↓						
exp#1 : smart	city room (i	indoor)							
NeRF Instant-NGP Nerfacto	10.8 30.7 31.0	0.68 0.91 0.92	0.64 0.20 0.19						
exp#2 : corrid	exp#2 : corridor (indoor)								
NeRF Instant-NGP Nerfacto	14.7 24.8 24.4	0.63 0.74 0.73	0.71 0.35 0.37						
exp#3 : under construction (outdoor)									
NeRF Instant-NGP Nerfacto	11.4 16.8 19.1	0.43 0.56 0.57	0.93 0.38 0.22						

Table 1. Quantitative comparison of three NeRF models on three different scenes

We trained three distinct NeRF models - vanilla NeRF. Instant-NGP, and Nerfacto - in diverse environments: a smartcity room (indoor), a corridor (indoor), and an under construction site (outdoor). The results, detailed in Tab. 1, exhibit a notable trend as the spatial scale increases from a confined room to a more expansive corridor and then to an open outdoor space, there's a discernible decrease in model accuracy, as reflected by metrics such as PSNR, SSIM, and LPIPS. This pattern suggests that the complexity and size of the environment negatively impact the rendering quality of these models. In particular, the outdoor scene (exp#3) highlighted the strengths of the Nerfacto model. It achieved a PSNR of 19.1 and an LPIPS of 0.22, surpassing the Instant-NGP model, which managed a PSNR of 16.8 and an LPIPS of 0.38. Furthermore, the qualitative visual results in Fig. 4 corroborate this finding, showing that in the construction site scene of exp#3, Nerfacto outperforms Instant-NGP, providing relatively superior visual quality.



Figure 6. Visualization of semantic segmented results on 2D orthogonal images from mobile LiDAR

Fig. 5 presents a comparison of parallel view images from different sources. The NeRF image was obtained from the Nerfacto model in exp#1, alongside 2D views acquired from stand-LiDAR and BIM models. These images were further processed using the SAM model. This comparison highlights the practical utility of applying segmentation to NeRF-generated images, considering the higher cost and lower usability of stable LiDAR sensor. We also tested scanned images from mobile LiDAR shown in Fig. 6, but the resolution of images acquired from mobile-LiDAR are inferior compared to stand LiDAR or NeRF. Therefore, we can use smartphones to easily capture visual information at construction sites and then utilize NeRF and SAM models to visualize the level of progress.

5 Conclusion

In conclusion, this study has successfully demonstrated the efficacy of Neural Radiance Fields (NeRF) in automating construction progress monitoring, marking a significant leap over traditional methods. By leveraging NeRF-Con, we have shown that it is possible to generate photorealistic, 3D rendered images from simple smartphonecaptured videos, offering a more efficient, accurate, and cost-effective solution compared to existing 2D and 3D sensor-based methods. The application of NeRF in various environments - small room, corridor, construction site - proves the robustness and versatility. The integration with segmentation models to compare these renders with BIM designs, ensuring more precise and automated monitoring of construction progress. In conclusion, our proposed methods, NeRF-based rendering and SAM-based comparison with BIM, can enable more efficient project planning and facilitate communication among construction site stakeholders.

In future work, we aim to address two main challenges: the decline in NeRF model's rendering accuracy with increased spatial scale, especially outdoors, and the current reliance on only qualitative SAM result images for progress monitoring. Our focus will be on optimizing NeRF's application for large outdoor sites and developing quantitative assessment methods, such as completion percentages, to enhance automated progress monitoring.

Acknowledgment

This research was supported by a grant [2022-MOIS38-002 (RS-2022-ND630021)] from the Ministry of Interior and Safety (MOIS)'s project, a grant from the Korean Government (MSIT) to the NRF [RS-2023-00250166] and Korea Ministry of Land, Infrastructure and Transport(MOLIT) as Innovative Talent Education Program for Smart City.

References

- Srinivasan P. P. Tancik M. Barron J. T. Ramamoorthi R. Ng R. Mildenhall, B. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. doi:https://doi.org/10.1145/3503250.
- [2] Kim B. Kim H. Kim, C. 4d cad model updating using image processing-based construction progress monitoring. *Automation in Construction*, 35:44–52, 2013. doi:https://doi.org/10.1016/j.autcon.2013.03.005.
- [3] Bienvenido-Huertas D. Carretero-Ayuso M. J. Della Torre S. Marín-García, D. Deep learning model for automated detection of efflorescence and its

possible treatment in images of brick facades. *Automation in Construction*, 145:104658, 2023. doi:https://doi.org/10.1016/j.autcon.2022.104658.

- [4] Lee-S. Ying, H. Q. A mask r-cnn based approach to automatically construct as-is ifc bim objects from digital images. In *ISARC*, pages 764–771, 2019. doi:https://doi.org/10.22260/ISARC2019/0103.
- [5] Turkan Y. Puri, N. Bridge construction progress monitoring using lidar and 4d design models. *Automation in Construction*, 109:102961, 2020. doi:https://doi.org/10.1016/j.autcon.2019.102961.
- [6] Chen J.-Cho Y. K. Kim, P. Slam-driven robotic mapping and registration of 3d point clouds. *Automation in Construction*, 89:38–48, 2018. doi:https://doi.org/10.1016/j.autcon.2018.01.009.
- [7] Bosché F.-Lu C. X.-Wilson L. Li, J. Occlusionfree orthophoto generation for building roofs using uav photogrammetric reconstruction and digital twin data. In *ISARC*, pages 371–378, 2023. doi:https://doi.org/10.22260/ISARC2023/0051.
- [8] Lin J. J. Hsieh S. H. Golparvar-Fard M. Pal, A. Activity-level construction progress monitoring through semantic segmentation of 3d-informed orthographic images. *Automation in Construction*, 157:105157, 2024. doi:https://doi.org/10.1016/j.autcon.2023.105157.
- [9] Evans A. Schied C. Keller A. Müller, T. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (ToG), 41(4):1–15, 2022. doi:https://doi.org/10.1145/3528223.3530127.
- [10] Weber E. Ng E. Li R. Yi-B. Wang T. ... Kanazawa A. Tancik, M. Nerfstudio: A modular framework for neural radiance field development. In ACM SIGGRAPH, pages 1–12, 2023. doi:https://doi.org/10.1145/3588432.3591516.
- [11] Frahm J. M. Schonberger, J. L. Structurefrom-motion revisited. In *CVPR*, 2016. doi:https://doi.org/10.1109/CVPR.2016.445.
- [12] Arfeto B. E. Zhang C. Shin H. Tariq, S. Segment anything meets semantic communication. arXiv preprint, 2023. doi:https://doi.org/10.48550/arXiv.2304.02643.
- [13] Isola P. Efros A. A. Shechtman E.-Wang O. Zhang, R. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. doi:https://doi.org/10.1109/CVPR.2018.00068.

Identification of Factors and Metrics to Compare Vision Based Data Acquisition Devices

Aakar Garg¹, Megha S Pradeep^{1, 2}, and Koshy Varghese¹

¹Department of Civil Engineering, Indian Institute of Technology Madras, India ²Department of Infrastructure Engineering, The University of Melbourne, Australia gargaakar8101998@gmail.com, meghaspradeep64@gmail.com, koshy@iitm.ac.in

Abstract –

Construction projects rely on several technologies for progress monitoring of projects. Among these (CV) technologies, Computer Vision based technologies are gaining popularity as they enable direct acquisition of physical site data. There are numerous devices available using vision-based technologies. Several studies have attempted to compare these technologies to identify appropriateness to meet the project requirements. However, there is no structured framework to compare and select a CV-based data acquisition device based on the requirements of a project. To develop a framework, it is critical to identify the factors and associated metrics that enable a systematic device comparison. Through a systematic review of literature of comparative studies on CV-CPM technologies this work identifies several factors and defines the metrics that form the basis for a structured framework. An approach to forming the framework based on these factors is also proposed.

Keywords -

Progress monitoring technologies; Computer Vision; Data acquisition; Comparative framework; Literature review

1 Introduction

Effective progress monitoring is crucial during a construction project's life cycle to control cost and time overruns. Further, prompt and accurate progress updates from a site avoids stakeholder disputes and related complexities by eliminating unexpected circumstances.

Data acquisition is a crucial step in progress monitoring process, which contributes for accurate project control. Project control data is increasingly being obtained through automated data acquisition technologies. Among these, Computer Vision (CV) based technologies are gaining significance as they have the potential to capture the physical state of a site [1,2].

There are several devices in the present-day market

for acquiring as-built status based on CV-based data inputs. These devices range from hand-held portable – low-resolution ones to tripod-mounted high-resolution ones. Correspondingly, the outputs of these devices can be used for different levels of progress monitoring. These levels could vary from basic visualization to detailed quantification of as-built components. Four levels of progress monitoring have been defined by earlier studies [1]. Identifying a suitable device suitable for the required level of progress monitoring specified for a project is an important requirement.

Existing papers have focused on the comparison of specific technologies, [3] or devices [4]. For a robust implementation, first, there is a need to systematically structure this comparison of the data acquisition devices and, secondly, create a framework to select the suitable device given an intended level of progress monitoring for a project. Hence, this paper aims to:

- 1. Review the existing studies on the comparison of various devices and technologies.
- 2. To identify the factors that enable a systematic comparison of automated data acquisition technologies for Computer Vision based Construction Progress Monitoring (CV-CPM).

The paper is structured as follows. In Section 2, the review methodology is initially discussed and then a table listing the various comparative studies and factors considered for technology comparison is presented and discussed. In Section 3, the metrics for each of the factors are defined and an approach to develop a structured framework is outlined. Discussion on the work is presented in Section 4 and followed by summary and future work in Section 5.

2 Review of Literature

2.1 Methodology

The reference literature for the review was collected from the Scopus database using a keyword search-based method followed by snowballing technique. Out of 312 results from the Scopus database and 24 papers from the snowballing technique, a total of 42 papers were identified through the PRISMA methodology [5], and an exhaustive review with analysis was performed. In this review, the papers that focuses on CV-CPM and specific comparison of data acquisition devices were included. The chronological distribution of the selected papers varies from 2011 to 2023, with majority concentrated in the years 2021, 2022 and 2023.

The search attributes used in the review with the keywords used and search scope are as shown in Table 1. The relevant articles for the construction domain were filtered after reading the abstracts. The filtered articles were considered for meta-analysis.

Table 1. Search attributes

Search attributes	Values used in the search
Databases	Scopus
Language	English
Duration	2012-2023
Туре	Journal and conference articles
Keywords	Construction, Automated
	progress monitoring

2.2 CV-based data acquisition devices

As mentioned in Section 1, CV-CPM is an emerging field focusing on information retrieval through visual inputs. These inputs include digital images, videos, thermal images, as-built point clouds (PCs), panoramas, and photospheres.

Major CV-based techniques include fixed surveillance, photogrammetry, videogrammetry, range or depth imaging and 3D laser scanning, each with their own advantages and limitations [1,2]. In all these technologies, the acquired data as image frames or point clouds are retrieved in multiple file formats compatible with corresponding processing software.

These technologies are combined with suitable mounting options including Unmanned Aerial Vehicles and Unmanned Ground Vehicles to enable progress monitoring for construction projects. There are numerous popular devices of varying combinations of the abovementioned technologies and mounting options being used in the industry for efficient progress monitoring.

To develop a framework for comparing CV-based device for varied requirements of progress monitoring, the first step is to perform a detailed literature review of existing specific comparative studies.

As presented in Table 2, eighteen studies are reviewed in detail to identify the devices being compared along with the technology categorization of these devices. The devices that are included consists of 3D laser scanners (Terrestrial- TLS and Mobile- MLS), iPhone or iPad LiDAR sensors, digital cameras, depth camera, etc. It is to be noted in the reviewed studies that some of the researchers identified a third reference in their study as a benchmark [3,4,6-15], whereas others evaluated a particular device in comparison to another device, keeping the latter as a reference [16-19]. In a few studies, both the above cases are evaluated [20,21]. In all three cases, the benchmark or reference is mentioned in the table. The table is further organized based on the technologies compared, grouping the similar technology comparisons together.

Most of the studies focused on comparing devices working on the same CV-based technology [3,6–8]. However, in studies where cross-technology comparisons were done, the images captured using depth or digital cameras were subjected to photogrammetric reconstructions where the input data is converted to asbuilt point clouds [12–14]. Further, these point clouds were compared to the directly obtained point clouds from the laser scanners or LiDAR sensors.

Various quantitative and qualitative factors, based on which the comparison is performed in the studies, are also stated in Table 2 along with the methodology adopted for comparison in the studies. It is interesting to note that a significant portion of the studies focuses only on quantitative factors [4,6,8,13,16–21] with a comparatively lesser studies taking into account both the factors [3,7,9–12,14,15].

The results from these studies are not included in the table, as the focus of the work is to identify and document the factors that were used for comparison and define suitable metrics than can be used to quantify the factors.

Both the quantitative and qualitative factors, listed in the table, must be suitably quantified based on their context. This will form the basis for a structured comparison framework. However, the scope of this article is limited to the characterization of quantitative factors, as provided in Section 3.

3 Metrics for quantitative factors

The major quantitative factors identified earlier are summarized in Table 3. These factors include resolution; accuracy; time; surface coverage; cost; system storage and device moving speed. The corresponding metric that can be used for quantification of these factors are also mentioned in the table along with the description to quantify them.

It can be noted that different studies use one or more factors to compare the devices along with different terminology for the same methods. All studies examined accuracy for comparison, with the majority also addressing resolution and quantifying them using various metrics, as depicted in Table 3. Accuracy ensures the reliability of data capture, while resolution determines the level of detail and clarity in the output. Given their

\mathbf{s}
es
· <u>5</u>
qe
Ψ
C
õ
Lis.
0a.
Ξ
S
ĩ
\mathbf{f}_{0}
R
.ie
ē
q I
ē
ai.
G
Ц.
c,
le
ab
H

		. Data A	cquisition				Comparison	
Nar	I"L ne	Jevice Technology	Z ^{au} De	vice Technology	Benchmark/ Reference	Quantitative Factors	Qualitative Factors	Methodology
Parac las scar	cosm ter nner	Handheld mobile laser scanner	Faro Focus 3D S120	Terrestrial laser scanner	Manually collected field measurements	Accuracy, time, cost	Workflow, quality of scans	 Accuracy is compared using variance calculation Comparison of total time
DotP-DPI- DPI- GeoS ZE sca	roduct 8 and 8 LAM 8 LAM 8 LAM 8 LAM 8 DI 1 mner	Handheld mobile laser scanner	Faro Focus3D X330, Leica Nova MS50 MultiStation, Leica ScanStation C5	Terrestrial laser scanner	Standard tape measuring tool	Accuracy, scanning time, post-processing time	r	• Error in distances is calculated
L	eica K360	Terrestrial laser scanner	RIEGL-VZ- 400i	Terrestrial laser scanner	Topographic equipment (total station and laser distance measurer)	Accuracy, points density, distance & orientation deviation	·	 Cloud-to-cloud (C2C) distance analysis using CloudCompare software Number of points in 5 × 5 cm²
iPa Lil (A _J	d Pro DAR pp 1)	Handheld mobile laser scanner	iPad Pro LiDAR (Three other Apps)	Handheld mobile laser scanner	Two TLSs (Faro Focus 3D and the Leica HDS 7)	Accuracy, number of points, scanning times	Visual quality	 C2C distance analysis using CloudCompare software Total number of points within the object
iPa Lil Sca	d Pro DAR - 1 - 3D nner)	Handheld mobile laser scanner	iPad Pro LiDAR (App 2 - RTAB-Map)	Handheld mobile laser scanner	Leica Disto D810 ranger (distance measurement), GeoSLAM ZEB Horizon scanner (MLS), Leica RTC360 (TLS)	Accuracy, point density, point spacing	·	 Distance analysis between points and their best-fitting plane C2C distance analysis using CloudCompare software Scan-to-BIM Model and compare dimensions to the reference dimensions
iPho Pro J (A	one 12 LiDAR pp 1)	Handheld mobile laser scanner	iPhone 12 Pro LiDAR (two other apps)	Handheld mobile laser scanner	Terrestrial laser scanner (Faro Focus3D X330)	Accuracy	Visual quality, influence of lighting and object materials	 C2C distance analysis using CloudCompare software Error in distances is calculated

		Data A	Acquisition				Compariso	5
Papers	1 st D	evice	2 nd Dev	rice	Benchmark/	Quantitative	Qualitative	Mathadalam
	Name	Technology	Name	Technology	Reference	Factors	Factors	INTELLIOUOIOGY
[10] (2017)	Google Tango tablet	Mobile range camera (Infrared- based scanning)	ZED camera	Mobile range camera (passive stereo vision)	Leica Laser Disto (for distance measurement)	Preparation time, scanning time, post- processing time, accuracy	Rescanning effect, ease of use, influence of lighting, and object materials	• The deviation in dimensions of scan values and actual values are compared and the average percentage error is calculated
[11] (2015)	Kinect v1	Mobile range camera (Infrared- based scanning)	Kinect v2	Mobile range camera (Time-of- flight sensor)	NextEngine 2020i Desktop laser scanner, laser meter for distance	Accuracy	Influence of sunlight	C2C distance analysisError in distances is calculated
[12] (2011)	Nikon D-80 camera	Structure from Motion (SfM) for unordered photos	Leica ScanStation 2	Terrestrial laser scanner	Standard tape measuring tool	Accuracy, number of points, cost, storage space on computer	Applicability, automation, training need, extra effort required	• PCs are converted to 3D CAD objects. The percentage error in ratios for each dimension (x, y, z) was used for comparison.
[13] (2013)	e Canon Vixia HF S100 camera & Point Grey Flea-2 camera	Photogram- metric and videogram- metric recon- struction	Leica ScanStation C10	Terrestrial laser scanner	Total station (SOKKIA 30R)	Accuracy, completeness, number of points	I	 Error between a point from the surface of the reference model Dividing the surfaces of the reference model into small regions and the existence of points is checked Number of points per square meter
[14] (2022)	DSLM- type camera	Photogram- metric re- construction	iPad Pro LiDAR (Scaniverse App)	Handheld mobile laser scanner	Laser scanner, CAD model (measuring with a laser range finder and tape), set of calibration balls	Accuracy, number of points	Quality (discontinuity , deformation, displacement, duplication)	 C2C distance analysis using CloudCompare software Total number of points within the object
[15] (2021)	Canon EOS 6D DSLR camera	Photogram- metric re- construction	iPad Pro LiDAR (SiteScape & EveryPoint App)	Handheld mobile laser scanner	Terrestrial laser scanner (Faro Focus X330)	Accuracy, completeness cost	ease of use, Influence of lighting	 Multiscale Model to Model Cloud Comparison (M3C2) analysis Mesh to cloud distance analysis

	Mathodalaav		2C distance analysis histance analysis between the hints and their best-fitting plane	ror between the object parameters ith respect to the reference model 2C distance analysis	istance between the points and eir best-fitting plane 2C distance analysis stal number of acquired points istance between two consecutive ints	2C distance analysis	istance between a point from the rface of the reference model 2C distance analysis	istance analysis between the pints and their best-fitting plane istance between Apple PC and the osest point into the TLS PC istances between the PC and the sest surface in the 3D BIM rthogonally projecting the points the corresponding surface to mstruct a 2D alpha-shape
uo			9 A C	• • C2 wi		•	• • C2	
Comparise	Qualitative	Factors	ı	I	ı	ı	I	ı
	Quantitative	Factors	Accuracy	Accuracy	Accuracy, number of points, storage space on the system	Accuracy, number of points	Accuracy	Local precision, global correctness, surface coverage
	Benchmark/	Reference	2 nd device	2 nd device	2 nd device	2 nd device	Quadrics, 2 nd device	2 nd device and manual 3D BIM model
	Device	Technology	Terrestrial laser scanner	Terrestrial laser scanner	Terrestrial laser scanner	Terrestrial laser scanner	Terrestrial laser scanner	Terrestrial laser scanner
ition	2 nd]	Name	Faro LS 880	Riegl VZ-400	Faro Focus3D X330	FARO Photon 120/20	Konica- Minolta VI-9i	Faro Focus3D X330
Data Acquis	Jevice	Technology	Mobile range camera (Infrared-based scanning)	Mobile range camera (Infrared-based scanning)	Handheld mobile laser scanner	Photogram- metric re- construction	Photogram- metric re- construction	Handheld mobile laser scanner
	1 st I	Name	Kinect Sensor	Kinect Sensor	iPad Pro LiDAR (SiteScape App)	iPhone 3GS	Canon EOS 300D, SONY DSC-HX20V, Smartphone SONY Z1	iPad Pro LiDAR (3D Scanner app)
	Papers		[16] (2012)	[17] (2014)	[18] (2021)	[19] (2014)	[20] (2014)	[21] (2022)

fundamental importance, they emerge as pivotal factors in device selection. Factors like surface coverage, cost, system storage, and device moving speed are explored only by a few studies for comparison.

Factors	Papers		
	3D density of points	Number of points per unit volume / Total number of	[7,9,12,14,
		points in the target object	18,19]
Resolution	Number of points / 2D density of points	Number of points per unit square / Total number of points in the target area	[6,8,9,13,18]
Resolution	Average percentage completeness	Dividing the surfaces of the reference model into small regions and the existence of points is checked	[13,15]
	Consecutive point distance / Point spacing	Consecutive pointDistance between two consecutive pointslistance / Point spacing(Absolute/average)	
	BIM/mesh to cloud Average distance between one PC and the closest		515.013
	distance analysis	surface in 3D BIM/mesh	[15,21]
	C2C distance analysis	C2C distances are determined by calculating the mean of all Euclidean distances between the nearest neighboring points of two-point clouds	[6–9,11,14, 16–21]
Accuracy	Local precision / roughness	Distance analysis between the points and their best- fitting plane	[8,16,18,21]
	M3C2 analysis	Multiscale Model to Model Cloud Comparison	[15]
	3D BIM distance analysis	Building the BIM model from the PC (Scan-to-BIM) and comparing its dimensions with reference dimension	[8]
	Error in distance measurement	Percentage/absolute/average error in measurement of distances as compared to the reference	[3,4,9–12]
	Average error / Average percentage error	Distance between a point from the surface of the reference model where this point is supposed to be located / error between the object parameters with respect to the reference model	[10,13,17,20]
	Total time per setup	-	[3,12,13]
Time	Preparation time		[3,10]
Time	Scanning time	-	[3,4,7,10]
	Post-processing time	-	[3,4,10,13]
Surface coverage	-	Points are orthogonally projected on the corresponding surface to construct a 2D shape	[21]
Cost	-	Costs of actually purchasing or renting the devices (May also add man-hour cost as per time)	[3,12,13]
System storage	File size on the system	-	[12,18]
Device moving speed	Range of the speed	The speed of moving the device at which errors are least	[10]

Table 3. Q	uantitative	factors	and	corresponding	metrics
· · ·				1 0	

3.1 Device Comparison

To compare devices, the factors identified in Table 3 can be weighted based on the project requirements such as project type and complexity, required level of progress monitoring, and level of details to be captured. Using these, the weighted average of factors can be calculated for a particular device, and this can be used to compare devices and select the appropriate option for the project.

The methodology for device comparison and selection is shown in Figure 1. Table 4 illustrates an example for the comparison approach, with a few factors as an example for selecting between two devices. The weights can be determined using the Analytic Hierarchy Process (AHP). However the weights presented in the table are hypothetical, and the specifics of weight estimation are not addressed in this paper, but are a part of ongoing research. The third and fourth columns of the table denote High, Medium, or Low values for each device, with corresponding numerical values of 3, 2, and 1 respectively. For factors where lower values are preferable, such as time, the numerical values are inverted to appropriately represent High, Medium, or Low. Finally, a weighted average of factors is computed for each device and compared, leading to the conclusion that Device 1 should be selected in the given example.

Factor	Factor Weight	Device 1	Device 2	Weighted value Device 1	Weighted value Device 2
Resolution	0.2	3 (High)	2 (Med)	0.6	0.4
Accuracy	0.5	2 (Med)	1 (Low)	1.0	0.5
Time	0.3	1 (High)	3 (Low)	0.3	0.9
Weighted Average				1.9	1.8

Table 4. Example of the comparison approach for device selection



Figure 1. Methodology for device selection

4 Discussion

In the reviewed literature, a notable gap exists as the range of the devices are not evaluated Range of the device is one of crucial quantitative factors that should be considered based on field study and site conditions.

Most of the literature reviewed in this paper has conducted the field experiments to compare the devices on a site that is available based on convenience. These sites vary in monitoring requirements and physical conditions. As a result, developing a standardized benchmark for the devices is not possible. There is a need to develop a standardized testbed that would allow for the systematic comparison of devices under controlled physical conditions, including factors such as lighting and different types of construction. This will ensure a more comprehensive and reliable evaluation, of data acquisition devices.

The qualitative factors are discussed by lesser studies as compared to the quantitative factor. However, several significant qualitative factors have been identified, including ease of use, influence of lighting, influence of object materials, visual quality, training need, and more. These factors play crucial roles in assessing the overall performance and suitability of devices. While this paper has addressed quantitative factors, more work is required to characterize the qualitative factors. Using both the quantitative and qualitative factors, a holistic framework for CV-based device comparison to meet progress monitoring requirements of a project can be developed.

5 Summary and further work

This paper provides a systematic review of

comparative studies on CV-based data acquisition technologies and devices from the relevant publications to understand the state-of-the-art in this domain.

Based on the comparative studies reviewed, key quantitative factors and the measurement metrics are identified. However, qualitative factors should also be included in developing a systematic device comparison and selection framework.

It is proposed that these factors can be weighted based on project requirements and scores for devices being considered for a project can be computed. These scores can assist in selecting the appropriate device.

Ongoing work is focused on developing a decision support system using this holistic framework. As a part of this framework, models for comparing performance of a mobile hand-held device with a terrestrial laser scanner is also being developed through a field-based study.

References

- Reja, V. K., Varghese, K., and Ha, Q. P., "Computer Vision-Based Construction Progress Monitoring," *Automation in Construction*. Volume 138. https://doi.org/10.1016/j.autcon.2022.104245
- [2] Reja, V. K., Pradeep, M. S., and Varghese, K., "A Systematic Classification and Evaluation of Automated Progress Monitoring Technologies in Construction," Vols. 2022-July, 2022, pp. 120–127. https://doi.org/10.22260/isarc2022/0019
- Blinn, N., and Issa, R. R. A., "Comparison of Traditional Laser Scanning and Mobile Lidar Technology for AECO Applications," *Congress on Computing in Civil Engineering, Proceedings*, Vols. 2017-June, 2017, pp. 113–121. https://doi.org/10.1061/9780784480830.015
- [4] Sepasgozar, S. M. E., Forsythe, P., and Shirowzhan, S., "Evaluation of Terrestrial and Mobile Scanner Technologies for Part-Built Information Modeling,' Journal of Construction Engineering and 12, Management, Vol. 144, No. 2018. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001574
- [5] Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., Mcguinness, L. A., Stewart, L. A.,

Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Mckenzie, J. E., "PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews," *BMJ*, Vol. 372, 2021. https://doi.org/10.1136/BMJ.N160

- [6] Moyano, J., Justo-Estebaranz, Á., Nieto-Julián, J. E., Barrera, A. O., and Fernández-Alconchel, M., "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, Vol. 62, 2022. https://doi.org/10.1016/j.jobe.2022.105190
- [7] Vacca, G., "3D Survey with Apple LiDAR Sensor— Test and Assessment for Architectural and Cultural Heritage," *Heritage*, Vol. 6, No. 2, 2023, pp. 1476– 1501. https://doi.org/10.3390/heritage6020080
- [8] Teo, T. A., and Yang, C. C., "Evaluating the Accuracy and Quality of an IPad Pro's Built-in Lidar for 3D Indoor Mapping," *Developments in the Built Environment*, Vol. 14, 2023. https://doi.org/10.1016/j.dibe.2023.100169
- [9] Teppati Losè, L., Spreafico, A., Chiabrando, F., and Giulio Tonolo, F., "Apple LiDAR Sensor for 3D Surveying: Tests and Results in the Cultural Heritage Domain," *Remote Sensing 2022, Vol. 14, Page 4157*, Vol. 14, No. 17, 2022, p. 4157. https://doi.org/10.3390/RS14174157
- [10] Senthilvel, M., Soman, R. K., and Varghese, K., "Comparison of Handheld Devices for 3D Reconstruction in Construction," 2017. https://doi.org/10.22260/ISARC2017/0097
- [11] Zennaro, S., Munaro, M., Milani, S., Zanuttigh, P., Bernardi, A., Ghidoni, S., and Menegatti, E., "Performance Evaluation of the 1st and 2nd Generation Kinect for Multimedia Applications," *Proceedings - IEEE International Conference on Multimedia and Expo*, Vols. 2015-August, 2015. https://doi.org/10.1109/ICME.2015.7177380
- [12] Golparvar-Fard, M., Bohn, J., Teizer, J., Savarese, S., and Peña-Mora, F., "Evaluation of Image-Based Modeling and Laser Scanning Accuracy for Emerging Automated Performance Monitoring Techniques," *Automation in Construction*, Vol. 20, No. 8, 2011, pp. 1143–1155. https://doi.org/10.1016/j.autcon.2011.04.016
- [13] Dai, F., Rashidi, A., Brilakis, I., and Vela, P., "Comparison of Image-Based and Time-of-Flight-Based Technologies for Three-Dimensional Reconstruction of Infrastructure," *Journal of Construction Engineering and Management*, Vol. 139, No. 1, 2013, pp. 69–79. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000565
- [14] Łabędź, P., Skabek, K., Ozimek, P., Rola, D., Ozimek, A., and Ostrowska, K., "Accuracy

Verification of Surface Models of Architectural Objects from the IPad LiDAR in the Context of Photogrammetry Methods," *Sensors*, Vol. 22, No. 21, 2022. https://doi.org/10.3390/s22218504

- [15] Murtiyoso, A., Grussenmeyer, P., Landes, T., and Macher, H., "FIRST ASSESSMENTS INTO THE USE OF COMMERCIAL-GRADE SOLID STATE LIDAR FOR LOW COST HERITAGE DOCUMENTATION," *The International Archives* of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vols. XLIII-B2-2021, 2021, pp. 599–604. https://doi.org/10.5194/isprsarchives-XLIII-B2-2021-599-2021
- [16] Khoshelham, K., and Elberink, S. O., "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications," *Sensors*, Vol. 12, No. 2, 2012, pp. 1437–1454. https://doi.org/10.3390/s120201437
- [17] Hämmerle, M., Höfle, B., Fuchs, J., Schröder-Ritzrau, A., Vollweiler, N., and Frank, N., "Comparison of Kinect and Terrestrial LiDAR Capturing Natural Karst Cave 3-D Objects," *IEEE Geoscience and Remote Sensing Letters*, Vol. 11, No. 11, 2014, pp. 1896–1900. https://doi.org/10.1109/LGRS.2014.2313599
- [18] Spreafico, A., Chiabrando, F., Teppati Losè, L., and Giulio Tonolo, F., "THE IPAD PRO BUILT-IN LIDAR SENSOR: 3D RAPID MAPPING TESTS AND QUALITY ASSESSMENT," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vols. XLIII-B1-2021, Nos. B1-2021, 2021, pp. 63– 69. https://doi.org/10.5194/ISPRS-ARCHIVES-XLIII-B1-2021-63-2021
- [19] Sirmacek, B., and Lindenbergh, R., "Accuracy Assessment of Building Point Clouds Automatically Generated from Iphone Images," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vols. XL–5, 2014, pp. 547–552. https://doi.org/10.5194/isprsarchives-XL-5-547-2014
- [20] Skabek, K., and Tomaka, A., "Comparison of Photgrammetric Techniques for Surface Reconstruction from Images to Reconstruction from Laser Scanning," *Theoretical and Applied Informatics ISSN 1896-5334*, Vol. 26, Nos. 3, 4, 2014, pp. 159–176.
- [21] Díaz-Vilariño, L., Tran, H., Frías, E., Balado, J., and Khoshelham, K., "3D MAPPING OF INDOOR AND OUTDOOR ENVIRONMENTS USING APPLE SMART DEVICES," *The International* Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vols. XLIII-B4-2022, Nos. B4-2022, 2022, pp. 303–308. https://doi.org/10.5194/ISPRS-ARCHIVES-XLIII-B4-2022-303-2022

Robotic IoT-Enabled 1D Line Scanner Integration for 3D Point Cloud Data Processing

Emre Ergin¹, Sven Stumm¹ and Sigrid Brell-Cokcan¹

¹The Chair of Individualized Production in Architecture, RWTH Aachen University, Germany E-mail: <u>office@ip.rwth-aachen.de</u>

Abstract -

The integration of measurement systems and sensors is critical to advancing construction automation processes and aligning digital models with physical production realities. However, incorporating field measurement and inspection tools into digital fabrication is challenging due to the dynamic nature of construction sites. While point cloud sensors have been widely used in field robotics, there is still potential for more accurate, flexible measurement systems for component assembly. This paper presents progress in integrating an automated, precise measurement system into a robotic fabrication process to facilitate adaptive process planning and control, focusing on the prefabrication stage of structural steel assembly and node prefabrication. The implemented scanning system, a typically static 1D laser line profiler, is attached to an industrial robotic arm. Unlike traditional full system integration approaches, we adopt an extensible distributed network approach, that considers the sensor and robot as separate modular systems. By leveraging automated calibration and synchronization between robot and sensor, we achieve a 3D-capable approach. Our work describes an overarching, streamlined system that covers the entire process from calibration over registration up to mesh reconstruction, that proposes a modular integration through a network communications middleware. This will allow faster component replacement and system retooling without the need for typical hard wiring or embedded computation, allowing better integration or extension for potential multi-sensor data fusion. In addition, we propose an edge detection system for potential automatic adaptive weld path generation.

Keywords -

Line scanning, Robotics, Internet of Things, IoT, 3D Point Cloud, Sensor Calibration

1 Introduction

The Architecture, Engineering, and Construction (AEC) industry continues to evolve at a rapid pace, marked by the integration of innovative materials, methods, and new technologies. The increasing requirements of sustainability, high performance functionality, and complexity of structural forms is driving an ongoing digitalization in the industry. As new techniques emerge, they become an integral part of pushing the limits and reducing the constraints of manufacturability. This introduces new physical and digital systems that need to be integrated with low barriers to utilization, especially in manufacturing and construction phases where digital fabrication is becoming a more essential part of production. With the integration of digital processes comes the need for more data regarding processes, to both ensure quality, that is in need of precision due to the dominance of manual inspection in the industry, as well as monitor progress for optimization of production workflow. Monitoring of construction processes has been recognized as a critical factor in minimizing uncertainties [1].

Robotics have emerged as a transformative force in inspection and monitoring of prefabrication processes. Following research shows the potential in quality assurance of construction components through 3D laser scanning, such as [2], where non-contact measurement technique utilizing terrestrial laser scanning (TLS) is utilized, [3], proposes a comprehensive method integrating building information modelling (BIM) and 3D laser scanning technology, to assess the dimensions and quality of precast concrete panels, and [4] introduces a refined metric for evaluating point cloud quality in automated construction progress monitoring using the Scan-vs-BIM method. Most research in the area of construction application delves into the quality assurance by not explicitly detailing the 3D data gathering procedure.

For the visual inspection sensor integration to kinematic systems, two primary configurations are commonly utilized; eye-in-hand and eye-to-hand [5]. Each configuration offers distinct advantages. Eye-inhand approach, which is utilized for the scope of this paper, ensures accuracy at tool center point (TCP) scale but offers limited environment sight, while eye-to-hand provides panoramic visibility albeit with slightly reduced precision at manipulator TCP.[5]

For the application of high-precision industrial inspections, line scanning sensors (1D, linear sensors, laser profilers) are widely implemented, where a single line of measurement points is collected at each exposure [6]. In order to integrate such measurement systems with other type of tools and systems, they need to be calibrated context-aware. A dynamic measurement system, where the laser profiler is integrated with a 6-DoF robotic arm, as opposed to the movement of a linear conveyor belt in the factory setting, requires precise identification of real-world coordinates for line scan image data, referred to as geo-referencing or mapping [7].

Achieving this requires accurate calibration of a sensor's intrinsic characteristics and extrinsic parameters. This is crucial considering that slight inaccuracies in calibration results in major deviations of scanned measurements based on changes due to orientation and distance to the object that is scanned. A method for compensation of measurement inaccuracies is the registration of line measurements against the already measured object. Within a distributed approach, the Internet-of-Things (IoT) aspect of implementation, the synchronization between robotic sensor positioning and the measurement is another important aspect that needs to be considered.

While using an IoT enabled or rather a distributed networked approach through an Industry 4.0 communication layer creates this additional challenge of data synchronization it also has a number of advantages such as:

- Easy retrofit for existing systems
- Easy expansion and replacement of sensor systems
- Capabilities for multi-sensor data fusion and mapping of data from multiple sources
- Easy upgrade of data processing systems
- Independence of data acquisition, processing, and visualization
- Data exchange for sensor-based process adaptation

This paper investigates the streamlined calibration of a 1D laser profiler and its integration with robot kinematics via IoT for precise 3D scanning. Our approach includes a modular system with IoT communication for synchronization, covering the entire process from calibration over registration up to mesh reconstruction. We further propose an edge detection method for adaptive weld path generation based on existing geometry.

2 Methodology and Evaluation

This section delves into the steps of implementing and setting up the required environment, starting from physical assets, tool calibration, IoT adaptation and finalizes at software development documentation a well as the results.

The test environment consists of KUKA KR30-3 F 6axis robot arm equipped with Keyence LJ-X8400 Line Profiler measurement head, that is connected to the Keyence Raw Data Output Controller LJ-X8000A with dedicated cabling. The controller connection to a data processing computer is achieved through an Ethernet port. This project software is developed in Python environment, with the implementation of mainly Keyence LJ-X8000A Communication Library, Open3D, NumPy, Paho-mqtt, and SciPy-Spatial libraries amongst other peripheral packages.

The industry 4.0-compliant communication layer and IoT infrastructure is established through the use of Cloud Remote Control (CRC), developed by Robots in Architecture Research [8]. The CRC framework enables distributed registration, state, and command communication of various assets and facilitates easy integration of automation components to simplify the addition of new devices and replacement of devices in automated processes.

This setup ensures seamless data integration, enabling robots, devices, and sensors as networkconnected things to communicate status and control commands to a centralized or multiple distributed control units in a local intranet. It also enables secure, gatewaycontrolled remote access to production environments and cloud computing. These potential features of remote access have not been realized in the approach described, and the benefits of Intranet-of-Things communication are more apparent.

The reference geometry to be scanned is a to-be-weld steel node connection, consisting of a perpendicular SHS 40x40x3 mm square tube and RO 42.5x2.5 mm round tube at 45° angle, on the upper surface of a HEA 200 steel beam, point weld for initial alignment as depicted in Figure 1.



Figure 1. Used reference object; steel node section.

2.1 Setup Environment

In the following section, the main components of the setup and the deployment process are described.

2.1.1 Tool Calibration

A Keyence LJ-X8400 Line Profiler measurement head is attached to the robot flange as shown in Figure 2. The device has a reference distance of 380 mm with measurement range of ± 60 mm in Z-axis and 210mm of line length at the reference distance¹. Since the laser pointer is enclosed in the profiler itself, the reference distance and the middle of the line is chosen to be calibrated as TCP (where X is 105mm and Z 0mm). Unlike the traditional 4-point calibration method, where a sharp, physical calibration reference point is utilized in the robot work environment, the laser calibration is done by referring to a reflective marker and luminance values at given spot. The reason behind this is the occlusion between reflective mirror and laser projection at detecting sharp objects, as previewed in Figure 3.

¹ keyence.com/products/measure/laser-2d/lj-x8000 last visited 21/11/2023.



Figure 2. Keyence LJ-X8400 Line Profiler Head attachment to the flange of KUKA KR30-3 F Robot arm



Figure 3. Preview of measurement occlusion and calibration

The proposed methodology of referring to a single reflective point is done by masking a reflective marker to a single tiny point and moving the laser middle to the reflective marker (see Figure 4 and Figure 5).



Figure 4. Scanning of reflective marker and masking to single point of reference

The achieved calibration point exhibits a measurement deviation of 0.19mm. The ABC 2-Point calibration procedure is conducted relative to a precalibrated robot base frame, a planar 300mm x 300mm steel plate of thickness 4mm.



Figure 5. Determination of reflective marker center

The process involves aligning the laser consecutively with the planar surface edge for yaw axis (KUKA A), horizontal alignment in the scanned profile diagram over for pitch axis (KUKA C), and moving the robot along a trajectory, perpendicular to the laser projection, ensuring constant average height along the trajectory line for roll axis (KUKA B), as illustrated in Figure 6.



Figure 6. Calibration of orientation in Euler Angle

2.1.2 CRC Implementation of Robot

The objective of Cloud Remote Control (CRC) within the context of Industry 4.0 communication is to facilitate modular plug & produce work cells for robotic and other automated production systems. Therefore, the IoT part of CRC establishes a semantic networking interface for communication. This necessitates both seamless expandability and accessible adaptability of the automation equipment, such as KUKA industrial robot arm and various actuators as well as sensors. Central to this communication framework is the utilization of the JSON data format with standardized device-specific topics, enabling the seamless integration of novel devices into the system [8].

CRC establishes an Industry 4.0 communication or rather information layer based on the publish-subscribe model through MQTT (which can however be transferred to any publish subscribe communication protocols such as OPC-UA). This communication layer enables users to direct commands to specific devices by unique identifier while allowing devices to broadcast their statuses, facilitating coordination among various assets. The system integrates KUKA industrial robot arms, controlled via JSON strings sent to the "devicename/cmd" channel, managing diverse robot arm motions and commands. Moreover, device-specific states, such as selected tools, TCP velocity, and robot TCP consistently published frames. are to the "devicename/state" topic at approximately 50 Hz, ensuring effective synchronization and monitoring of device statuses within the defined environment.

2.1.3 CRC Integration of the Line Scanner

The software solution is designed to be integrated within the Cloud Remote Control (CRC) framework as an additional IoT device. As stated in the previous chapter, this framework enables seamless orchestration among individual assets within a workspace, facilitated by a control unit. Irrespective of the configuration or spatial arrangement of physical assets, the framework ensures wireless coordination among them.

For sensor systems the CRC differentiates between two basic types streaming and triggered sensors. As streaming sensors continuously stream their measurement data directly within their state message, these are only used for simple sensors that do not create too much load on the network. Triggered sensors can be identified by a command support, these sensors will only send out new data, if the appropriate command is sent. Figure 7 presents implemented commands that are supported by the processor for the 1D-Line sensor are as follows:

 "CapturePointCloud": Initiates the acquisition of a series of line scans to construct a 3D point cloud scan while the robot is in motion.

- "ExportPointCloud": Saves the recorded Point Cloud data locally in .txt and .pcd formats and generates a .stl mesh:
- "ClearPointCloud": Erases all recorded instances of point clouds.
- "CaptureSingleLine": Captures the current linear scan and stores the resultant point cloud data locally.
- "DetectEdge": Executes the edge detection script within the program, offering a preview of the results in an Open3D window.



Figure 7. Line-Scanner Processor software architecture through MQTT

2.1.4 Device Synchronization

The need to synchronize laser profiler linear scan data with the robot TCP to acquire accurate 3D point cloud scans of specified objects is accomplished by transforming the sensorial data to the robot tool position in close to real time. To facilitate this process, a dedicated class called 'Transformer' has been developed.

The CRC robot controller integrates the robot as part of the network and streams the TCP values of the active tool in the KUKA frame data format {"X":0, "Y":0, "Z":0, "A":0, "B":0, "C":0} under the topic *"robotname/state"* approximately every 20 milliseconds. The Transformer class is initialized with the current TCP. This incoming data is then translated into a transformation matrix, thus allowing the transformation of the currently received instance of point cloud data from the Open3D library to seamlessly align with the calibrated robot TCP. To synchronize the data from the different devices, the Point-Cloud-Processor adds arrival time stamps to the received data. Other forms of time synchronization using the NTP or PTP protocols are possible to account for network latency, but the results did not show the need to implement distributed clock synchronization. The use of timestamps at the time of measurement with synchronized clocks will be part of future research. However, the arrival timestamps are used to compare and align the laser profile data with the TCP.

2.1.5 Scanning and Meshing

The scanner was positioned approximately at a distance of 380mm from the geometrical center of the scan object. A total of 6 scanning paths are defined by a spherical coordinate system around this center with a range of $\pm 40^{\circ}$ in φ (KUKA C) and $\pm 20^{\circ}$ in θ (KUKA B) orientation. The scans were performed with a TCP speed of 3mm/sec (see Figure 8) and a linear path length of 150mm for each individual scan trajectory. The scanning trajectory, point cloud registration, and meshing is application and geometry dependent, the goal with the proposed node connection is to scan the intersections for the generation of welding paths.

First, the recorded mesh is down-sampled by a small value for the local deviation of each scan point (e.g., 0.5 mm) to account for scanner accuracy. Next, the point normals are estimated and aligned using Open3D functions to ensure consistent face orientation.

The mesh construction process uses the Poisson surface reconstruction method proposed by Kazhdan in 2006 [9] as implemented in Open3D. After the construction phase, the resulting mesh undergoes an additional smoothing process without compromising the authenticity of its shape [10]. Together, these filters contribute to the refinement of the mesh, ensuring a smoother appearance while preserving the true structure derived from the original point cloud data.



Figure 8. Constructed mesh preview with 6 scans, total computation time 8 minutes at robot speed of 3mm/sec.

2.2 Point Cloud Registration

The TCP shows an observable calibration error of 0.19 mm as further described 2.1.1 Tool Calibration. Despite its seemingly small value, this variance has significant implications for the alignment of the acquired data relative to the robot's base coordinates.

Figure 9 and Figure 10 show four different scans from two perspectives, each taken from previously mentioned 6 linear scan paths (2.1.5 Scanning and Meshing), to illustrate the noticeable differences in the results.



Figure 9. Comparison of four scans, from varying scanning angles, represented in red, orange, green and blue colored point clouds.



Figure 10. Close up comparison of four scans, from varying scanning angles, represented in red, orange, green and blue colored point clouds, visible unmatched points marked in black squares.

These visual representations indicate that all four scans are approximately aligned with the robot base frame in the context of world coordinates. While acknowledging the potential for improving global alignment through initial reference geometry, this discrepancy was ignored for the purposes of this work.

Upon closer inspection, as presented in Figure 10, there are noticeable misalignment of points between the

point cloud collections. Such difference of orientation induces noise during meshing, consequently amplifying inconsistencies between the physical model and its digital representation.

To ensure increased quality results with minimal noise, an introduced approach involves employing a local point cloud registration method between captures, specifically leveraging the Open3D Multiway Registration[11]. This implementation integrates voxel down-sampling to optimize computation time while preserving data intensity. The optimization process is iterated twice: first to identify and prune uncertain alignments, then to refine the graph alignment.

Given the nature of linear point cloud scans, this implementation favors point-to-plane transformation estimation over the point-to-point Iterative Closest Point (ICP) registration method.

Figure 11 below shows a comparison between the meshed representations of six individual scans (shown in blue) and their counterparts after registration (shown in green). The contrast is present, where the surface quality and differs significantly, between the noisy unregistered point cloud and the registered point cloud collection with increased cleanliness.



Figure 11. Mesh reconstruction without registration (top left) and with registration (top right) and at the bottom, corresponding deviation to the digital geometries presented in a gradient from green (less surface deviation) to red (more surface deviation)

The resulting registered 3D point cloud captures, and mesh construction provide a complex digital representation of the physical object and enable a wide range of data interpretation algorithms to be applied. In recent years, various researches have focused on the interpretation of point cloud data for gap detection[12], line segment extraction[13] and object detection[14]. Subsequent chapters will go through a simplified approach of edge detection algorithm for potential welding path generation as mentioned before.

2.2.1 Development of Edge Detection Approach

Integration of an edge detection process as described in [15][16]enables the extraction of vital information from the recorded point cloud data. As for the proof of concept, an edge detection approach is implemented within this project. The applied methodology within this paper leverages the eigenvalues of individual points, considering a specified number of neighboring points, to identify directional changes present at the edges within the collection [17].

A number of neighbor correspondences assigned as property for each point within the cloud and eigenvalues were calculated in each neighborhood, to be filtered by the distinctive eigenvalue changes, and masked with color red as to be seen in Figure 12.



Figure 12. Potential candidates for detected edges depicted in red colored points.

Subsequently, potential edges are clustered using the Density-Based Scan Algorithm [18] implemented as part of Open3D, facilitates the grouping of the best-suited candidates in close proximity. The parameter of maximum correspondence distance to neighbors within a potential cluster is set to factor of 2.1 of the down-sampling value. Colored clustering is presented in Figure 13.



Figure 13. Density-based clustering of potential edges, represented in varying colors.

3 Outlook and Future Work

This proposed system presents an ideal solution for conducting detailed scans within smaller-scale assemblies. It can be seamlessly integrated with widerange, 3D point cloud cameras mounted on robots to facilitate initial determination -offering global awareness to the robot- of the precise locations where detailed scans are to take place. An exemplar 3D detection workflow can be determined as Figure 14.



Figure 14. Potential 3D scanning workflow based on the relation between scan precision and range.

Beyond its visual quality assurance capabilities, this system opens doors for further advancements through the introduction of various algorithms tailored for detailed scanning. Potential optimizations of the system as well as a more comparative analysis of the chosen algorithms for each step will be part of future research. Beyond this some proposed developments could include:

- 1. Layer Inspection and Height Comparison: Specifically designed for wire-arc additive manufacturing applications, this algorithm can focus on analyzing sub-millimeter differences present in each layer, ensuring precise monitoring and quality control.
- 2. Crack Detection and Repair: A potential algorithm can address critical applications such as detecting cracks on steel bridge trusses and subsequent repair through welding, enhancing structural integrity and safety.
- 3. **Deflection Analysis:** Targeting welded workpieces, this application can conduct comprehensive deflection analysis. It identifies and analyzes deformations in assembly parts caused by high heat output, ensuring the structural integrity of the affected components.
- 4. **Individualized Scanning path planning:** Due to the directional and scalar limitations of the linear scanner, a path planning approach based on geometrical properties of the scanned objects, can be implemented to overcome the occluded parts at each individual scan.

4 Conclusion

Research and implementation of visual systems in robotics area all over the industries have been predominantly focused on global localization of mobile robotics. Such advancements benefit the construction industry as well, however, there's a noticeable absence of specific inspection methods catering to detailing, joining, and assembly in smaller scale. Additionally, the dynamic nature of construction environments necessitates adaptable process setups.

This paper presents a promising approach to enhance the capacity of a high-precision, single-dimensional laser profiler sensor by integrating it with a 6-DoF motion system, a robotic arm. The incorporation of IoT enables both systems to function independently and facilitates potential reconfiguration of the motion system, allowing for the potential of different types of kinematics-capable robots integration.

The results of the paper proved that the consistency in scale and detail as well as the location, relative to the robot base, of the scanned object with the physical world conditions. Additionally, the IoT infrastructure showed that the integration of various systems can be standardized, through a simplified M2M framework. While this proposed application lacks the real-time capability of model reconstruction, due to the high-loads of point cloud computation and the software only being deployed on prototyping level, the potential of implementing the controller software on cloud level and communication layer through potential 5G infrastructure remains to be implemented in the future development of this ongoing research project.

The precise comparison of digital to physical world can lead to new paradigms of individualized production in steel prefabrication. Enhanced automated inspection and quality control in steel construction assembly can streamline the regulation and safety assurance that is widely done manually in construction industry. This paper contributes to the foundational aspects of automating the digitalization process of real-world conditions, presenting a pathway for potential advancements in steel prefabrication within the construction sector in near future.

Acknowledgements

This work is part of the research project Cloud56 funded by the Federal Ministry for Digital and Transport of Germany within the initiative InnoNT (funding number 190122018C).

References

- Chan, A. P. C., Scott, D., & Chan, A. P. L. (2004). Factors Affecting the Success of a Construction Project. Journal of Construction Engineering and Management, 130(1), 153-155. DOI: 10.1061/(ASCE)0733-9364(2004)130:1(153)
- [2] Q. Wang, M.-K. Kim, J. C. P. Cheng, & H. Sohn, "Automated quality assessment of precast concrete elements with geometry irregularities using terrestrial laser scanning," *Automation in Construction*, vol. 68, pp. 170-182, 2016. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2016.03.014.
- M.-K. Kim, J. C. P. Cheng, H. Sohn, & C.-C. Chang, "A framework for dimensional and surface quality assessment of precast concrete elements using BIM and 3D laser scanning," Automation in Construction, vol. 49, pt. B, pp. 225-238, 2015. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2014.07.010.
- [4] D. Rebolj, Z. Pučko, N. Čuš Babič, M. Bizjak, & D. Mongus, "Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring," *Automation in Construction*, vol. 84, pp. 323-334, 2017. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2017.09.021.
- [5] G. Flandin, F. Chaumette, & E. Marchand, "Eye-inhand/eye-to-hand cooperation for visual servoing," in Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), San Francisco, CA, USA, 2000, pp. 2741-2746 vol.3. DOI: 10.1109/ROBOT.2000.846442
- [6] Wendel A, Underwood J. Extrinsic Parameter Calibration for Line Scanning Cameras on Ground Vehicles with Navigation Systems Using a Calibration Pattern. Sensors (Basel). 2017 Oct 30;17(11):2491. DOI: 10.3390/s17112491.
- [7] Lim, M.-S., & Lim, J. (2008). Visual Measurement of Pile Movements for the Foundation Work Using a High-Speed Line-Scan Camera. Pattern Recognition, 41(6). ISSN: 0031-3203. Elsevier Science Inc. DOI: 10.1016/j.patcog.2007.10.025.
- [8] Stumm S., Kerber E., Krishnan A. (2024). Robotergestütztes Schweißen – Verteilte Produktionstechnik für dynamische Automatisierung – Chapter 15 in Springer, Deutschlang. ISBN: 978-3-658-42543-2 IoC - Internet of Construction, Hrsg. Brell-Cokcan S, Schmitt R.
- [9] Kazhdan, M., Bolitho, M., & Hoppe, H. (2006). Poisson Surface Reconstruction. In Symposium on Geometry Processing (Eds.: Alla Sheffer & Konrad Polthier, pp. 061-070). The Eurographics Association. ISSN: 1727-8384. ISBN: 3-905673-24-X. DOI: 10.2312/SGP/SGP06/061-070.

- [10] Taubin, G. (1995). Curve and surface smoothing without shrinkage. In Proceedings of IEEE International Conference on Computer Vision (pp. 852-857). DOI: 10.1109/ICCV.1995.466848.
- [11] Choi, S., Zhou, Q.-Y., & Koltun, V. (2015, June). Robust Reconstruction of Indoor Scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2015.7299195.
- [12] Alsadik, B., Gerke, M., & Vosselman, G. (2014, June). Visibility analysis of point cloud in close range photogrammetry. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II-5. DOI: 10.5194/isprsannals-II-5-9-2014.
- [13] Lin, Y., Wang, C., Cheng, J., Chen, B., Jia, F., Chen, Z., & Li, J. (2015). Line segment extraction for large-scale unorganized point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 102. DOI: 10.1016/j.isprsjprs.2014.12.027.
- [14] Zhou, Y., & Tuzel, O. (2018). VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4490-4499). DOI: 10.1109/CVPR.2018.00472.
- [15] Sampath, A., & Shan, J. (2010, March). Segmentation and Reconstruction of Polyhedral Building Roofs From Aerial LiDAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing, 48, 1554-1567. DOI: 10.1109/TGRS.2009.2030180.
- [16] Huang, X. (2013). Building reconstruction from airborne laser scanning data. Geo-spatial Information Science, 16(1), 35-44. Taylor & Francis. DOI: 10.1080/10095020.2013.774104.
- [17] Bazazian, D., Casas, J., & Ruiz-Hidalgo, J. (2015, November). Fast and Robust Edge Extraction in Unorganized Point Clouds. In Proceedings of the 2015 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). DOI: 10.1109/DICTA.2015.7371262.
- [18] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). Α Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (pp. 226-231). Portland, Oregon: AAAI Press. KDD'96. DOI: Series: 10.5555/3001460.3001507.

Instance Segmentation of Exterior Insulation Finishing System using Synthetic Datasets

Mingyun Kang¹ Sebeen Yoon¹ Juho Han¹ Sanghyeon Na¹ and Taehoon Kim¹

¹Architecture Engineering, Seoul National University of Science and Technology, Republic of Korea <u>kmg4312@seoultech.ac.kr</u>, <u>paul6452@seoultech.ac.kr</u>, <u>hanjooh00421@seoultech.ac.kr</u> <u>nsh143@seoultech.ac.kr</u>, <u>kimth@seoultech.ac.kr</u>

Abstract –

The quality inspection of adhesive of Exterior Insulation Finishing System (EIFS) is important because poor adhesive can lead to detachment of the insulation. Computer vision-based inspection stands out as a notable alternative. Recently, imaged-based deep learning model are widely used for the automated monitoring and inspection in construction field. To train the model, the relevant large datasets are essential. However, collecting datasets in the construction site is hazardous because of inherent risk of accidents. Also, synthetic datasets method which is one of alternatives to solve this problem are focused on fixed and regular shaped objects. To address these challenges, this study analyses the validity of synthetic datasets in terms of segmentation of adhesive in EIFS, which has irregular shape. For instance segmentation, the datasets were divided into two groups: (1) real datasets, composed of 100 actual photos, (2) mixed datasets, which combined 50 randomly sampled images from both synthetic datasets and real datasets. The mAP@50 of instance segmentation for real datasets and mixed datasets is 87% and 99%, respectively. This study prove that synthetic datasets can effectively train segmentation models, enabling the recognition of irregularly shaped objects and enhancing overall performance.

Keywords – Synthetic datasets, Image-based deep learning, Instance segmentation, Exterior insulation finishing system

1 Introduction

Exterior Insulation and Finishing System (EIFS) is a method of covering the entire exterior of building. There are primarily two methods for applying EIFS: the dry and wet processes. The wet process, the most common method, involves attaching insulating material to the structure, which is coated with an adhesive such as mortar [1,2]. Poor adhesive performance can lead to

detachment of the insulating material. However, inspecting the application quality with limited supervisory personnel at construction sites is laborintensive and virtually impractical in terms of time and cost. To address this challenge, automating the inspection of adhesive using deep learning-based computer vision can significantly enhance productivity by minimizing manpower and costs [2,3].

The relevant large datasets are essential for the deep learning. Existing open-source image datasets such as Common Objects in Context (COCO) [4] and the Stanford 2D-3D-Semantics Dataset [5] are available but, there are no datasets related to construction site, particularly, EIFS. Additionally, acquiring image datasets in construction site is difficult due to the inherent risk of accidents at construction sites. Moreover, the images obtained are often disordered and cluttered [6]. As an alternative, using web crawling techniques to acquire images allows the rapid collection of a large number of photos without visiting construction sites. However, a significant issue arises with the timeconsuming process of labeling these images to transform them into a dataset suitable for training in computer vision [6,7].

Recently, a novel approach has been proposed to overcome the challenges previously mentioned: synthetic image modeling, which involves creating datasets for specific fields. By synthesizing in artificial modeled virtual environment, desired scenes can be created. The synthesized scenes are already data-encoded, thus reducing efforts for labeling and annotation [3,7,8]. However, existing studies have focused on fixed and regular shaped objects, thus applying instance segmentation with synthetic datasets of irregular shaped object is not explored well.

The purpose of this study is to conduct foundational research on the effectiveness of training an image segmentation model for segmenting dab and insulation in EIFS utilizing synthetic data, comparing the performance between real datasets and mixed datasets combining synthetic datasets.

2 Related Work

2.1 Computer vision in construction.

With the advancement of datasets and the corresponding development of computer vision, applications in the construction industry have also evolved. AI, along with more sophisticated object detection and segmentation techniques, not only reduce tedious, human-error-prone tasks but also lead to improved research in terms of labor intensity and time consumption.

Computer vision facilitates rapid and accurate material quantity estimation, thereby aids progress monitoring. Li et al. (2021) [9] proposed a real-time, cost-effective rebar counting scheme using the YOLOv3 detector. Wang et al. (2021) [10] utilized surveillance video to track information of precast walls such as numbers and position, transmitting the data in the Building Information Model (BIM) to achieve automatic progress using timestamp methods.

In the early stages of computer vision application in construction safety management, it primarily focused on determining the use of personal protective equipment such as workers' hard hats. Fang et al. (2018) [11] conducted research on non-hardhat-use detection using 100,000 photos from far-field surveillance at construction sites. Additionally, research on classifying cracked or damaged buildings has been actively pursued. Silva et al. (2018) [12] proposed a model that uses deep learning image classification to differentiate between concrete photos with and without cracks. Gao et al. (2018) [13] conducted structural damage recognition through spalling condition checks and evaluation of damage levels.

In the construction field of quality inspection, Xia et al. (2024) [14] conducted research on First Floor Elevation estimation using the YOLOv5 model and mobile LiDAR point clouds. Chen et al. (2021) [15] investigated quality of building façade using photos obtained from unmanned aerial vehicles.

2.2 Synthetic datasets related to construction industry

While computer vision has evolved alongside the advancement of datasets, there has been a shortage of datasets related to the construction field [6,7]. Acquiring real image data poses risks due to the nature of construction sites, also labeling acquired images manually requires significant time and labor.

To address this challenge, lots of research has been conducted to train computer vision models using synthetic datasets created using various methods related to construction sites. Soltani et al. (2016) [3] proposed a method to separately synthesize excavators and backgrounds. They found that, compared to traditional annotation methods, auto-annotation reduced the required time. Neuhausen et al. (2020) [8] aimed to create synthetic datasets using Blender to track worker productivity and safety management. Hong et al. (2021) [16] utilized the BIM model to generate synthetic datasets for infrastructure scenes. While these studies provided direction and validated the utility of synthetic datasets for computer vision, they focused on large objects. Lee et al. (2023) [17] conducted research of safety monitoring through synthetic datasets focusing on small-sized personal protective equipment detection using the Unity game engine.

Particularly, Wang et al. (2023) [18] studied synthetic datasets for rebar instance segmentation. They compared the performance using a Mask R-CNN model with various combinations of real datasets and synthetic dataset, concluding that experimental groups composed with both real and synthetic datasets appropriately are more robust for instance segmentation than using only real datasets.

Although these studies sufficiently validated the utility of synthetic datasets in the construction industry for computer vision, they targeted fixed and regular shapes (e.g., excavators, rebar, etc.). Therefore, in this study, we aim to create synthetic datasets for EIFS instance segmentation, targeting irregularly shaped objects that may vary in appearance over time, and validate their utility accordingly.

3 Methodology

This study analyses the effectiveness of training an image segmentation model using mixed datasets combined with synthetic data, compared to real datasets. Instance segmentation model was applied to segment dab and insulation in an image, thereby the classes were set as insulation and dab. Figure 1 illustrates the shape of EIFS.



Figure 1. Ribbon and dab method


Figure 2. Framework of Study

As illustrated in Figure 2, the framework can be divided into two main phases. First, during the dataset manufacturing stage, photographs are collected from both real EIFS construction sites and virtual construction sites where EIFS is ongoing. Then, mixed datasets are made with synthetic datasets and randomly sampled real datasets. Both real datasets and mixed datasets are augmented before instance segmentation stage to prevent overfitting. Second, in the instance segmentation stage, the YOLOv8 model is used to train the datasets. The trained weights are then utilized to compare area segmentation in actual EIFS construction sites.

3.1 Data Manufacturing

To test the efficacy of synthetic datasets, the experimental datasets were divided into two control groups: (1) 100 numbers of realistic images from the construction site, and (2) a mixed dataset of 50 realistic and 50 synthetic images. For the real datasets, we visited three different EIFS construction sites, to acquire images for train model and testing. Labeling and annotation was carried out using Roboflow [19], and augmentation was performed to prevent overfitting.



Figure 3. Images and annotation using Roboflow

Synthetic dataset environment was created using Unreal Engine 4, and the dataset synthesis model was developed utilizing the algorithm of NVIDIA's Deep Learning Dataset Synthesizer (NDDS) [20,21].

As illustrated in left side of Figure 4 virtual construction site was created using Unreal Engine 4. For the variation of datasets such as shade adding, point of view, worker's position were conducted. The right side of Figure 4 shows the annotation of the image. Every single color of image represents the class which is divided respectively and automatically connected the components of Unreal Engine 4.



Figure 4. Virtual construction site of Unreal Engine 4(left) annotation of the image(right)

3.2 Instance Segmentation

The You Only Look Once (YOLO) model is one of the most renowned models for one-step object detection capabilities, offering rapid detection speed and high accuracy. The latest version of the YOLO series is YOLOv8 [22], which replaces the C3-module with C2fmodule for robust gradient flow, adopts a discrete head structure, these modifications greatly improve the detection accuracy.

YOLOv8-seg is an instance segmentation model

derived from YOLOv8. YOLOv8-seg consists of five models: 8n-seg, 8s-seg, 8m-seg, l-seg, and x-seg. Starting with the lightest model, 8n-seg, the amount of computation increases gradually towards 8x-seg. Consequently, mAP and processing speed also increase. The selected model was YOLOv8x-seg in this study, which is the highest accuracy model in the aspect of mAP.

4 Experimental Study

4.1 Evaluation Metrics

In this section, we aim to discuss the results of tests conducted using YOLOv8 to evaluate the validity of instance segmentation in both real datasets and mixed datasets. The data were split into training, validation, and test datasets at a ratio of 80:10:10. The parameters for the training model were set as follows: 200 epochs, a learning rate of 0.01, and a batch size of 16.

The performance metric for the segmentation algorithm was mean Average Precision at 50 (map@50). Average Precision (AP) is the area under the precisionrecall curve for a specific class, and the mAP is the average of these AP values across all classes. The map@50 specifically refers to the mAP calculated with an Intersection over Union (IoU) threshold set at 0.5, meaning that predictions must overlap at least 50% with the ground truth to be considered correct. The equation

of AP, mAP are demonstrated below.

$$AP = \lim_{n \to \infty} \sum_{k=1}^{n} (R_k - R_{k-1}) P_k = \int_0^1 p(r) \, dr \qquad (1)$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i$$
 (2)

The Intersection over Union (IoU) is calculated as the ratio of the area of overlap between the Ground Truth and the Detection to the sum of their areas (Equation(3))

$$IoU = \frac{Area of Overlap}{Area of Union}$$
(3)

$$precision = \frac{IP}{\frac{TP}{TP} + FP}$$
(4)

$$\operatorname{recall} = \frac{TT}{TP + FN} \tag{5}$$

True positive (TP) refers to the count of instances correctly predicted as true by the model, false positive (FP) denotes instances wrongly predicted as true while they are false, false negative (FN) represents instances incorrectly predicted as false that are actually true. Precision (Equation(4)) is the proportion of correct predictions among the results predicted by the model, while Recall (Equation(5)) is the proportion of actual true instances that the model correctly predicts as true. Both Precision and Recall are adjusted based on the confidence level setting used as the model's decision criterion. Raising the decision criterion increases Precision, whereas lowering it enhances Recall.

4.2 Test of Real Datasets

The result of Real datasets at the best confidence level, the values of precision, recall and map@50 are 0.89, 0.96, and 0.87 respectively. The highest performance at the 198th epoch are depicted in the Confusion Matrix shown in the following Figure 5. According to the confluence matrix, the model trained by real datasets predicted the dab well at 1.0, while the prediction of insulation was relatively low at 0.71.



Figure 5. Confusion matrix of real datasets

4.3 Test of Mixed Datasets

The results of the 100 numbers of mixed datasets which is composed with 50 numbers of synthetic datasets and 50 randomly sampled real datasets. The Precision, Recall, and map@50 at the best confidence level are 0.99, 0.94, and 0.95, respectively. The confusion matrix in the subsequent Figure 6 illustrates the performance outcomes for the mixed datasets dataset using the model that achieved its peak performance at the 175th epoch. According to the confusion matrix, it can be seen that model trained by synthetic datasets predicted well as 1.0 for both dab and insulation.



Figure 6. Confusion matrix of mixed datasets

4.4 Comparative Analysis

The difference in map@50 between real datasets and mixed datasets was more marginal than expected. However, the trends observed in YOLOv8's validation, as shown in Figure 7, indicated distinct learning patterns for real datasets and mixed datasets. While the number of datasets may not be sufficient, obtaining real datasets involved visiting construction sites to capture photographs, as mentioned in the introduction, which presents hazards and challenges in image labeling. In contrast, with synthetic datasets, just a few clicks can modify the environment, allowing for the manufacture of datasets with varied characteristics.



Figure 7. Prediction of real EIFS image with model trained by real datasets (left) and mixed datasets (right)

5 Conclusions

This study analyzed the effectiveness of training a deep learning model for segmenting irregular shaped dab and insulation in EIFS utilizing synthetic datasets, compared to real datasets. In this study, images of EIFS in construction sites were acquired from both real world and virtual environments to create datasets. To evaluate the segmentation performance of EIFS, the subjects were categorized into two classes, dab and insulation, and trained using the YOLOv8x-seg model. The map@50 results used as performance indicators for real datasets and mixed datasets were 0.87 and 0.95, respectively.

Utilizing synthetic datasets can reduce the risk of visiting construction sites with safety accidents. Also, generating synthetic datasets enable the creation of various compositions, environments, and scenarios with little effort. Furthermore, research suggested segmentation of irregular shaped object and potential model robustness as well. Thus, proposed method can be utilized in other applications such as concrete crack.

For further research, the effectiveness of applying synthetic datasets in the instance segmentation on various objects will be explored with a large size of datasets. Additionally, the comprehensive automatic supervision model will be studied as well as segmentation of ribbons with dab and insulation, to calculate the area of EIFS segments in pixels.

6 Acknowledgements

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 and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 1711191456).

7 References

- [1] J. H. Kim, S. M. Kim, and J. T. Kim, "Evaluation on Weather Resistance of Dry and Wet External Thermal Insulation Walls according to EOTA Test Standard," Journal of Korean Institute of Architectural Sustainable Environment and Building Systems, vol. 15, no. 4, pp. 405-418, 2021.
- [2] S. Yoon, M. Kang, H. Jang, and T. Kim, "Computer Vision-based Automated Adhesive Quality Inspection Model of Exterior Insulation and Finishing System,". Journal of the Korea Institute of Building Construction, vol. 23, no. 2, pp. 165– 173, Apr. 2023.
- [3] M. M. Soltani, Z. Zhu, and A. Hammad, Automated Annotation for Visual Recognition of Construction Resources Using Synthetic Images, Automation in Construction, vol. 62, pp. 14-23, 2016.
- [4] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft COCO Common Objects in Context, in Computer Vision–ECCV 2014 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer International Publishing, 2014, pp. 740-755.
- [5] I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, Joint 2D-3D-Semantic Data for Indoor Scene Understanding, arXiv preprint arXiv1702.01105, 2017.
- [6] H. Tajeen and Z. Zhu, Image dataset development for measuring construction equipment recognition performance, Automation in Construction, vol. 48, pp. 1-10, 2014.
- [7] J. G. Lee, J. Hwang, S. Chi, and J. Seo, Synthetic Image Dataset Development for Vision-Based Construction Equipment Detection, Journal of Computing in Civil Engineering, vol. 36, no. 5, Art. no. 04022020, 2022.
- [8] M. Neuhausen, P. Herbers, and M. König, Using Synthetic Data to Improve and Evaluate the Tracking Performance of Construction Workers on Site, Applied Sciences, vol. 10, no. 14, Art. no.

4948, 2020.

- [9] Y. Li, Y. Lu, and J. Chen, "A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector," Automation in Construction, vol. 124, p. 103602, 2021.
- [10] Z. Wang, Q. Zhang, B. Yang, T. Wu, K. Lei, B. Zhang, and T. Fang, "Vision-based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase," Journal of Computing in Civil Engineering, vol. 35, no. 1, pp. 04020056, 2021.
- [11] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," Automation in Construction, vol. 85, pp. 1-9, 2018.
- [12] W. R. L. da Silva and D. S. de Lucena, "Concrete Cracks Detection Based on Deep Learning Image Classification," The 18th International Conference on Experimental Mechanics, Jun. 2018, doi: 10.3390/icem18-05387.
- [13] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 9, pp. 748-768, 2018.
- [14] J. Xia and J. Gong, "Computer vision based first floor elevation estimation from mobile LiDAR data," Automation in Construction, vol. 159, p. 105258, 2024.
- [15] 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, vol. 122, p. 103503, 2021.
- [16] Y. Hong, S. Park, H. Kim, and H. Kim, "Synthetic Data Generation Using Building Information Models," Automation in Construction, vol. 130, Art. no. 103871, 2021.
- [17] H. Lee, J. Jeon, D. Lee, C. Park, J. Kim, and D. Lee, "Game engine-driven synthetic data generation for computer vision-based safety monitoring of construction workers," Automation in Construction, vol. 155, p. 105060, 2023.
- [18] H. Wang, Z. Ye, D. Wang, H. Jiang, and P. Liu, "Synthetic Datasets for Rebar Instance Segmentation Using Mask R-CNN," Buildings, vol. 13, no. 3, p. 585, 2023.
- [19] Roboflow, <u>https://roboflow.com</u>, Accessed: 13/03/2024
- [20] T. To, J. Tremblay, D. McKay, Y. Yamaguchi, K. Leung, A. Balanon, J. Cheng, W. Hodge, and S. Birchfield, "NDDS: NVIDIA Deep Learning Dataset Synthesizer," 2018. [Online]. Available: <u>https://github.com/NVIDIA/Dataset_Synthesizer</u>
- [21] J. Tremblay, T. To, and S. Birchfield, "Falling

Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2018, doi: 10.1109/cvprw.2018.00275.

[22] YOLOv8, <u>https://github.com/ultralytics/ultralytics</u>, Accessed: 13/03/2024

Virtual Construction Equipment Sensor for Determining Soil Stiffness during Compaction

Maximilian Schöberl¹, Michael Schneider¹, Claus Lechner² and Johannes Fottner¹

¹Chair of Materials Handling Material Flow Logistics, Technical University of Munich, Germany ²Wacker Neuson Produktion Deutschland GmbH & Co. KG, Reichersthofen, Germany <u>max.schoeberl@tum.de</u>

Abstract – Processes on construction sites are characterized by a high proportion of manual work, rapid variability and harsh working conditions. As a result, the level of digitalization and automation as well as productivity is low. In the near future, digital assistance and documentation systems will be used to counteract these shortcomings. To achieve this, machines must be able to learn more about their work processes and pass this information on to information systems. Innovative sensor systems for recording machine performance are in demand. Digital (BIM) models of the objects to be built offer the possibility of providing target quality data for machine operation and storing the achieved quality data after construction.

In soil compaction using vibratory plates, soil stiffness is still recorded manually today, although automated recording would be conceivable. Relevant input and output variables for the compaction process are to be collected for the soil compaction application. On this basis, a sensor system for determining soil stiffness during compaction is designed and built as a prototype. This will result in a multivariable system with influencing, measured and target variables (virtual sensor), which can be used for the structured development of statistically meaningful test series to determine the correlation between measured values and the quality of the compaction process (soil stiffness). Robust algorithms for recording and transferring the degree of soil compaction during compaction were derived from the tests on the chair's own test site and demonstrate that measuring the absolute soil stiffness during compaction with vibratory plates is possible. This conclusion provides the basis for increased quality, automation and digital continuity of future earthworks.

Keywords -

Virtual Sensor; Compaction; Soil Stiffness; Automation

1 Introduction

Compaction is a quality-defining subprocess of earthworks. As other construction processes today, it involves multiple manual workflows and is rarely digitized. An excavator prepares the earth surface specified in building plans and the specifications that has to be compacted according to norms. Subsequently, the compactor has to be placed in the area that should be compacted. The area is compacted and after completion the resulting ground stiffness is measured. Depending on the measurements the compaction is continued or the measurement receipts are documented in the construction diary and the compactor is removed from the area. The following construction processes rely on the documented data e. g. in building on the created surface. Figure 1 visualizes the compaction process of today.



Figure 1: Current compaction process in earthwork construction

On top of the high degree of manual work in this process and the incontinuity of the information flow along the process, there is potential to fusion the process steps "receive compaction order", "compact surface", "measure ground stiffness", "accept surface", and "document ground stiffness", by developing a compaction quality control system. This system should receive the target ground stiffnesses from a BIM-model, measure the ground stiffness during compaction, and document the as-built ground stiffnesses in the BIMmodel. Therefore, the research question of this paper is "How can a virtual construction equipment sensor reliably determine the soil stiffness during compaction in earthwork construction?".

In order to answer the research question, this paper will present the State of Science and Technology (2) on the compaction process, compaction progress determination and technical functionality of equipmentintegrated methods. The state of the art is subsequently extended through the description of the development of a virtual sensor (3). The developed virtual sensor is validated in compaction test series (4). Finally, the paper closes with a discussion, conclusion and outlook.

2 State of Science and Technology

2.1 Compaction Process

According to Richter and Heindel [1] soil compaction is a process, reducing the material's porosity while simultaneously enhancing its load capacity. Thereby, many different parameters influence the compaction process. The most significant being the soil type and grain size distribution, as well as the water content inside the material and the compaction equipment used [2–4].

The compaction effect produced by vibratory plates is based on the short repeal of friction between individual grain particles because of the vibrations. This leads to a rearrangement of the material bulk into a tighter packed substrate [1, 4]. To generate the vibrations on such equipment, one or more excitation shafts with an eccentric mass are coupled to an engine and transfer them into the ground via the base plate. If two or more shafts are used, changing their respective phase angle to one another results in a change of working direction, because of a different resulting force vector. Therefore the worker can easily manipulate the driving direction of the compactor [2, 5, 6].

2.2 Compaction Progress Determination

Measuring the achieved soil stiffness after the compaction with vibratory plates is mostly done by hand after the job is completed. Therefore, if an insufficient ground stiffness value is detected, the affected area has to be reworked. The used ground stiffness measuring methods often require special equipment and trained workers to ensure an accurate result. Among the most used control methods are the Proctor-test according to DIN 18127 [7], the static plate load test (DIN 18134) [8], the dynamic plate load test with the light weight deflectometer [9] as well as various different field tests based on replacement procedures and cut out soil samples in DIN 18125-2 [10]. Furthermore, radioactive radiation deflection methods [2, 3] and the measurement of seismic waves according to [2] and [4] can be considered a valid option to determining the compaction state. The absolute compaction value of the underlying surface can be

evaluated using such methods.

The exception to the rule is made by an effort done by the different manufacturers of vibratory plates to incorporate soil stiffness measurement systems on their equipment [11–15]. These systems are derived from well-established continuous compaction control systems used on roller compactors [1, 16, 17].

However, due to the nondeterministic vibration behaviour of compaction plates [11] and the cyclic loss of ground contact, the methods on vibratory plates lack the precision and reliability of their roller compactor counterparts. Furthermore, in contrast to the previously described conventional systems, with machine-based approaches it is only possible to detect the soil stiffness increase from one pass to another. The measured compaction states are therefore called relative compaction values.

Because these systems provide the base of research in this paper, the next part of the state of technology is dedicated to the basic functionality of soil stiffness measurement systems on vibratory plates.

2.3 Technical Functionality of Equipmentintegrated Methods

The patents described in [11–14] are all based on the collection and analyzation of vibrations on various different parts of the equipment. For example, the relative movement of the upper mass to the lower mass. The measured accelerations are then further processed by different filters and mathematical methods to ensure a stable output value. However, the detailed execution of the system complexity and sensor architecture varies vastly between the different manufacturers. The Bomag system [13] is able to predict the relative compaction increase quite simply by calculating the quotient of two characteristic parts of the observed vibration. On the other hand, looking at the approach from Wacker Neuson [14] with their Compatec system, a compaction value is calculated from solution of the three dimensional differential equations of motion. For those, the expected contact force vector between the ground plate and the soil has to be predicted based on the contact area and the rotation of the base plate. Furthermore, the contact area relies on an estimation itself.

These highly complicated calculation algorithm and the complex construction circumstances led to an impractical result due to too many uncertainties. According to further research, the compate system is not used on any currently available vibratory plates from Wacker Neuson, but only in a much simpler version. An other approach by Weber [12] matches the current machine vibrations with the taught in values in an internal database. To ensure an accurate relationship between the measured and reference values, the exact correlation for each construction project has to be determined by a calibration test ahead of the compaction work.

In conclusion, none of the above-described systems are able to determine an absolute compaction value which can be used to validate the executed work for every part of the compacted area.

3 Virtual Sensor Development

This chapter explains the development process of the virtual sensor hardware and the corresponding software. In addition, the setup of a first test environment, allowing for simultaneous testing and development of the sensor, including the experimental procedure are discussed. At the end, the gathered data is processed and analyzed, to derive a compaction value based on the ground stiffness from the measured sensor data.

3.1 Hardware

To determine a suitable hardware setup for measuring parameters correlating with the soil stiffness, a compaction influence parameter overview was established. Afterwards, fitting sensor solutions measuring the individual aspects are researched and structured in a morphological box (see Figure 2). Other variables such as mounting position, additional user input and possible correlation partners are added.

Features		Solution approaches						
	1st order features	2nd order features			3			
	Sensor position	Sensor position	Upper mass	Lower mass	Upper and lower mass	Base plate		
		Physical measure	Vibrations	Water content	Soil density/ pore fraction	Body/ airborne sound	Grain sizes	Grain size distribution
		Sensor principle	Acceleration sensor	Ultrasonic sensor	Resistance measurement	Acoustic sensors	Moisture meter	Capacitance measurement
	Data management	Data acquisition	Continuous	Discrete measuring points	Measurement runs	Calibration drives		
		Operator input	None	Soil type	Calibration data (table)	Desired degree of compaction Calibration	Calibration measurement	Data acquisition with soil probe
		Possible correlation	Degree of compaction	Static soil stiffness	Dynamic soil stiffness	FDVK value	Proctor density	Compaction performance

Figure 2: Morphological box

A multicriterial evaluation compares the different possibilities regarding their cost, technical feasibility, and suitability for implementation in series production machines. On top, sensors for measuring moisture and the grain sizes of the material were adapted from similar use cases and bench tested in individual experiments.

In the end, two different sensor setups were derived from the morphological box. One uses inertia measurement units (imus) to gather vibration information of the upper and lower mass simultaneously. The other one measures the acoustic emissions of the plate during compaction with a highly sensitive microphone. For sensor communication and power supply, a microcomputer in form of a raspberry pi 4 is utilized. This also acts as a communication interface with an external computer via a Wi-Fi-hotspot and VNC-server. The mounting position of all sensors is in the middle of the backside of the machine to ensure short wires and a minimal influence of the exhaust system on the recorded sound data (see Figure 3).



Figure 3: Sensor placement on the vibratory plate

3.2 Software

The software package contemplating the physical sensors is composed of different parts for every application. As the imu's are coupled to the raspberry pi via a separate microcontroller based on an esp32-pico chipset, the work from Tanaka [18] is adopted to fit the needs of this research. In addition, the audio software 'audacity' is installed on the raspberry pi for capturing the noise emissions.

For recording the acceleration data transmitted via USB from the sensor's microcontroller, the raspberry pi uses the terminal emulator 'Putty'. This enables the storage of the six individual acceleration data in an excelfile.

3.3 Test Site Setup

Besides the hard- and software setup, at suitable test area is required for doing the data collection test runs. The test site consists of a temporary compaction field made up of mixed-particle sized gravel, located on the concrete floored outdoor test area of the chair. The compaction area is constrained on one side by large concrete blocks and sloped on the other. This ensures an easy loosening up of the compacted material with a wheel loader. For faster material handling times, the test setups are made up alternating both sides of the middle separation layer. The material height for the different test fields varies from 30 to 50 cm depending on the desired experiment goal.

In addition, ten marks spacing one meter each are placed on the concrete blocks to help guide the exact measuring position of the conventional compaction determination method as a reference. Therefore, a dynamic plate load test with the light weight deflectometer model ZFG 2000 by ZORN with 300 mm base plate diameter and a 10 kg drop weight is used, because its widely spread application on real world construction sites and its ease of use.

Figure 4 shows a sketch of the utilized test site setup and the position of the reference marks (MP).



Figure 4: Test site setup, machine sketches from [19] and [20]

For all the test runs in the initial batch of experiments, a heavy remotely controlled compaction plate of type Wacker Neuson DPU110rLec970 [21] with a maximum centrifugal force of 110 kN, an operating weight of around 810 kg and a base plate width of 970 mm was used. After placing the equipment at the end of the test field, back and forth passes are carried out. In between each alternation, the achieved ground stiffness values are measured at the reference points inside the lane.

3.4 Data Collection

The experiments took place in cold but mostly dry weather. In total, eight different test fields were created with a cumulated amount of 30 passes with the vibratory plate. Table 1 gives an overview of the individual setups for each experiment.

Table 1: Test fields and their setups

Name	Number of	Material	Direction of
	passes	height	first pass
Field 0	4	50 cm	Forwards
Field 1	6	50 cm	Backwards
Field 2	5	50 cm	Forwards
Field 3	4	50 cm	Backwards
Field 4	4	30 cm	Forwards
Field 5	2	30 cm	Backwards
Field 6	2	30 cm	Forwards
Field 7	3	30 cm	Backwards

3.5 Evaluation Algorithms

The first step of analyzing the measured data is

creating a reference value curve for the dynamic ground stiffness from the light weight deflectometer. The individual data points can be connected to form a continuous line, as it is not expected, that the soil stiffness jumps rapidly. Furthermore, outliers due to measurement errors have to be considered when evaluating the sensor data in comparison to the references.

To get a feel for the gathered sensor data, the acceleration values and the audio recording are first plotted in their time representation. From those depictions it is possible to determine the different work modes of the equipment, such as idle, engine run-up and compaction operation. The plot also reveals the overflow of the imu on the lower mass due to an insufficient measuring interval of the imus.

For further analysis, Kuttner and Rohnen [22] differentiate between methods in the time domain and methods in the frequency domain. The Fourier analysis is a prominent representative of a frequency domain method for analyzing dominant vibration parts and their harmonics. For best results, it is recommended to prefilter the signal with a lowpass filter whose parameters are set with the Nyquist-Shannon-Theorem in mind [22]. According to Werner [23], the resolution of the resulting frequency spectrum when looking at a time discrete signal can be improved by using zero-padding.

Other closer looked at methods include different envelope techniques, such as described by Kuttner and Rohnen [22], but also statistical procedures. In this case the focus lies on the distribution of the amplitude density depicted as an histogram [24] and counting methods based on the rainflow-principle [25].

In addition, Takami et. al. [26] as well as Kanokogi and Takami's [27] approach of matching sensor data into a given group or to a reference value via neural network machine learning were also explored. It is remarked by the authors, that due to Wolpert and Macready's [28] nofree-lunch-theorem, every machine learning network has to be adopted to the specific problem, otherwise the full potential cannot be exploited.

Most of the before mention methods applied to the measured sensor data, does not lead to any visible differences when looking at data from low and high ground stiffness. However, by examining the Fourier spectrum of the acceleration values, there is a significance in the vertical vibration of the upper mass in the low frequency area. Upon closer inspection, the dominant frequency may well be the eigenfrequency of the upper mass, when compared to the experiments and simulations from Lohr [29]. Furthermore, the frequency and amplitude of this deep vibration changes with the soil stiffness the plate is driven over and therefore is also visible in the audio data. For this main finding in the first set of experiments, the further evaluation of the correspondence shall be explained in more detail. The sensor data is lowpass filtered at first and then divided into sections of equal time length. Each section is analysed with a fast Fourier transform (fft) algorithm [22, 23] and the peaks and frequency of the dominant low end are extracted. Due to no direct analytical correlation between these peaks and the soil stiffness, a genetic algorithm (ga) is set up to optimize the data processing and the parameters of an analytic correlation function.

The ga is an evolutionary algorithm inspired by natural selection and survival of the fittest. It was developed by Goldberg in the 1980s and is widely used for such optimization problems [30].

When used on the data of one test field, the achievable correlation values are quite decent for the best parameter sets. Transferring the same algorithm to a yet unknown data set does unfortunately not provide the same performance. There were different promising parameter sets tested on all the available data, but the resulting values did not converge.

In summary, the first tests failed to meet the requirements in terms of a compaction value correlated from the sensor data. However, a lot of lessons were learnt about the test environment, as well as the sensor data and its analyzation methods. For example, the methods from Takami et. al [26] as well as their predecessors Kanokogi and Takami [27] seem to be a promising pre-processing for the measurement data. Overall, machine learning algorithms are most promising to classify the pre-processed data.

This knowledge is put to use when enhancing the test setup and sensor system for a follow-up experiment.

4 Verification & Validation

In this chapter, the changes to the test setup based on the gathered experience from the previous research is presented. Also, the second round of data collection and its analysis is discussed.

4.1 Improved test setup

The first major change for the second round of data collection is the switch to a smaller vibratory plate for higher resolution in the individual compaction passes. From now on, a preproduction sample from Wacker Neusons new DPU6560 range with a maximum centrifugal force of 65 kN, an operating weight of 475 kg and a base plate width of 600 mm is utilized. In addition, the overflowing imu-sensor on the lower mass is upgraded to a much more capable unit.

As the operating system on the raspberry pi is switched to ROS2, it now supports an interface to an existing GNSS position measuring system via real-time kinematics positioning. Through this add-on it is possible to exactly track the machine's position and correlate it with the measured sensor data and the reference values from the plate load test.

On the test site side, the total length of the area is enlarged to ensure adequate areas for the start and stop of the compaction plate. Therefore, the impact of the startup and stop processes on the measured parameters can minimized.

4.2 Data collection

As the evaluation of the first data set showed, that it is fortunate to have a large number of passes on each test field, the test procedure is slightly adopted. Instead of doing only 2 or 3 passes on a field with 30 cm of material, the bulk height is kept constant at 50 cm throughout the whole test. In combination with the lower compaction power of the smaller plate, more passes per field can be executed. Table 2 shows the chosen parameter sets for each individual test field. It should be highlighted, that the passes on field 4 and field 5 are all carried out in the same direction to eliminate one additional variable when analysing the data.

Table 2: Test field setups for the second round of data collection

Name	Number of	Direction of	Compaction
	passes	first pass	strategy
Field 1	12	Forwards	Alternating
Field 2	11	Forwards	Alternating
Field 3	8	Backwards	Alternating
Field 4	11	Forwards	Forwards
Field 5	7	Backwards	Backwards

4.3 Data evaluation

For the data analysis from the second set of tests, the methods from Takami et. al [26] as well as their predecessors Kanokogi and Takami [27] mentioned in subsection 3.5 are applied in more detail.

The goal is to discretise the measured soil stiffness values into different classes and matching the corresponding sensor data signal intervals from the imus with the help of machine learning algorithms.

Due to a different compaction behaviour depending on the working direction of the plate, two separate algorithms have to be trained in order to achieve the best possible results. For training, the data sets from field 4 and field 5 are used, as these are obtained while keeping a constant driving direction. In order to prepare the data, the signals are first filtered and then divided into 10 equal length section corresponding to the 10 measuring points for the reference values. Afterwards, each section is transformed into the frequency spectrum using the fftalgorithm.

To classify the data, the dynamic soil stiffness values measured with the light weight deflectometer are

grouped into four equal sized intervals spanning the whole measuring range. These classifications in combination with the pre-processed frequency spectra are then input into the algorithms as training data.

In an iterative measure, various different algorithm parameters are obtained, from which the most promising are used for validation of this method on the remaining data sets.

4.4 Algorithm validation

In this chapter, the previously derived parameter set for the algorithms are tested on the so far unused data sets from fields 1-3.

The results of this generalization are all quite close together with the best parameter set for the forward working direction having a correct classification rate of a little over 80 %. When looking at the backward working direction, the performance is with close to 70 % a little bit lower. This may be caused by a smaller available data set for training in the beginning and therefore not being able to find an algorithm configuration with the same level of sophistication.

When combining the performance of both algorithms, 76 % of all data generated in the second batch of experiments are classified correctly. Figure 5 shows the confusion matrix of the combined algorithms for data analysis. The diagonal shows all correctly classified data samples and below/on the right the performance of each row/column is summarized in percent.

	Pr				
	23	27	1		45,1
Class		224	43	9	81,2
True		29	108	4	76,6
		1	1	10	83,3
	100	79,7	70,6	43,5	

Figure 5: Confusion matrix of the combined algorithms

Looking at the confusion matrix, the validation of the developed algorithms can be seen as a success. The falsely classified data is close to the diagonal, which shows, that these samples were placed in the neighbouring class. This may happen due to the real value being close to one of the class borders. From a total of 480 data samples only 11 (2,29 %) are not located close to the diagonal.

5 Discussion, Conclusion & Outlook

As the algorithm is able to determine the interval of more than three quarters of all measurement data sets to their actual soil stiffness values, the previously set goal was achieved. The algorithm matches the reliability asked for in the research question "How can a virtual construction equipment sensor reliably determine the soil stiffness during compaction in earthwork construction?". The algorithms in combination with the sensor hardware (subsection 3.1) forms a virtual sensor, that is able to accurately determine the absolute ground stiffness of compacted soil in our test study.

The findings of this publication are limited in the amount of tests and data that were used. Exemplarily, only one kind of soil was used in the test runs and the sensor was mounted on two different vibratory plates. Until there are no further experiments with other vibratory plates and material, the significance of these findings is limited to the scope of the presented test runs. However, the findings of our research project indicate that measuring ground stiffness during compaction is possible with our current virtual sensor, which will be optimized in forthcoming studies and iterations.

The results of our research have shown that reaching the main goal of reducing the high degree of manual work in the compaction process and creating a continuous information flow along this process is possible. The virtual sensor for compaction quality control forms the basis for combining the process steps "receive compaction order", "compact surface", "measure ground stiffness", "accept surface", and "document ground stiffness". The functionality of the envisioned compaction quality control system is depicted in Figure 6.



Figure 6: Functionality of the envisioned compaction quality control system

Future research should focus on increasing the classification accuracy of the algorithms and a cost and quality optimization of the hardware in use. Additionally, more tests have to be carried out to determine the performance of the virtual sensor in other equipment and soil combinations. In order to transform the current compaction process into the envisioned system, BIM-

model interfaces have to be crated [31] and the value of the solution has to be demonstrated to practitioners in order for the system to be accepted by them.

Acknowledgements

The authors thank the Bavarian State Ministry of Economic Affairs, Regional Development and Energy and the Bavarian Centre for Digitization for funding this research through the research project "MiProcess2Twin" (DIK0415/08).

References

- D. Richter and M. Heindel, *Straßen- und Tiefbau*. Vieweg+Teubner, 2011.
- [2] W. Striegler and D. Werner, *Erdstoffverdichtung*, 2nd ed. Berlin: VEB Verlag für Bauwesen, 1976.
- [3] R. Voss and R. Floss, *Die Bodenverdichtung im Straßenbau*, 5th ed. Düsseldorf: Werner-Verlag, 1968.
- [4] W. Dachroth, *Handbuch der Baugeologie und Geotechnik*, 4th ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017.
- [5] Ammann Group. "Produktkatalog APH Rüttelplatten: APH Serie Hydrostatische Rüttelplatten." https://www.avesco.ch/fileadmin/ avescoproducts/ c3c0b8e4f1fc59d5d62df7cf4f8efcad.pdf (accessed Nov. 23, 2022).
- [6] Wacker Neuson SE, 80809 München, DE. "Broschüre Vibrationsplatten." https:// www.wackerneuson.de/media/contentserv/ 71960.pdf (accessed Jan. 11, 2023).
- [7] Baugrund, Untersuchung von Bodenproben -Proctorversuch, DIN-18127, DIN Deutsches Institut für Normung e.V., Berlin, Sep. 2012.
- [8] Baugrund Versuche und Versuchsgeräte -Plattendruckversuch, DIN-18134, DIN Deutsches Institut für Normung e.V., Berlin, Apr. 2012.
- [9] Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV), Technische Prüfvorschriften für Boden und Fels im Straßenbau - Teil B 8.3: Dynamischer Plattendruckversuch mit dem Leichtem Fallgewichtsgerät: TP BF-StB Teil B 8.3, 2012nd ed. (FGSV R1 - Regelwerke). Köln: FGSV Verlag GmbH, 2012.
- [10] Baugrund, Untersuchung von Bodenproben Bestimmung der Dichte des Bodens – Teil 2: Feldversuche, DIN-18125-2, DIN Deutsches Institut für Normung e.V., Berlin, Nov. 2020.

- [11] R. Anderegg, U. Drees, A. Kuban, and D. Von Felten, "Verfahren zur Ermittlung einer Bodensteifigkeit und Bodenverdichtungsvorrichtung," EP 1 516 961 B1.
- [12] W. Latt and A. Kurzweg, "Verfahren und Vorrichtung zum Messen von Bodenparametern mittels Verdichtungsmaschinen," DE 10 2010 060 843 B4.
- [13] N. Laugwitz, "Verfahren und Vorrichtung zur Bestimmung des Verdichtungsgrades bei der Bodenverdichtung," DE 10028949 A1.
- Wacker Neuson SE, 80809 München, DE,
 "Verfahren und Vorrichtung zum Messen von Bodenparametern mittels Verdichtungsmaschinen," DE 10 2006 008 266 B4.
- [15] M. Steffen, "Bodenverdichtungsvorrichtung zum Verdichten eines Bodenbereichs," DE 10 2019 107 219 A1.
- [16] J. Pistrol, F. Kopf, and W. Völkel, "Verfahren zur Ermittlung des Verdichtungszustandes eines Untergrunds," EP 3 176 324 B1.
- [17] P. Corcoran and L. Chi, "Device and process to measure ground stiffness from compactors," WO 2015/112412 A1.
- [18] M. Tanaka. "I2C MPU6886 IMU Library." https:// github.com/tanakamasayuki/I2C_MPU6886 (accessed Apr. 14, 2023).
- [19] Kramer Werke GmbH, 88630 Pfullendorf, DE. "EINSTEIGEN UND LOSFAHREN Die Radund Teleskopradlader 5035/5040/5050/5055/5065 /5065T." https://www.kramer-online.com/media/ contentserv/37131.pdf (accessed Apr. 16, 2023).
- [20] Wacker Neuson SE, 80809 München, DE. "Broschüre Verdichtung." https:// www.wackerneuson.de/media/contentserv/ 108754.pdf (accessed Apr. 16, 2023).
- [21] Wacker Neuson SE, 80809 München, DE. "Produktinformationsblatt DPU110rLec970." https://www.wackerneuson.de/media/contentserv/ 64139.pdf (accessed Apr. 23, 2023).
- [22] T. Kuttner and A. Rohnen, *Praxis der Schwingungsmessung*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019.
- [23] M. Werner, Digitale Signalverarbeitung mit MATLAB[®]. Wiesbaden: Springer Fachmedien Wiesbaden, 2019.
- [24] J. S. Bendat and A. G. Piersol, Random data: Analysis and measurement procedures (Wiley series in probability and statistics). Hoboken, N.J: Wiley, 2010. [Online]. Available: https:// onlinelibrary.wiley.com/doi/book/10.1002/ 9781118032428

- [25] Y.-L. Lee and T. Tjhung, "Rainflow Cycle Counting Techniques," in *Metal Fatigue Analysis Handbook*, Y.-L. Lee, M. E. Barkey, and H.-T. Kang, Eds., Elsevier, 2012, pp. 89–114.
- [26] G. Takami, M. Tokuoka, H. Goto, and Y. Nozaka, "Machine Learning Applied to Sensor Data Analysis," in *Yokogawa Technical Report English Edition Vol. 59 No. 1*, Yokogawa Corporation, Ed., 2016, pp. 27–30.
- [27] H. Kanokogi and G. Takami, "Machine Learning Applied to Sensor Data Analysis: Part 2," in *Yokogawa Technical Report English Edition Vol.* 60 No. 1, Yokogawa Corporation, Ed., 2017, pp. 35–38.
- [28] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization,", *IEEE Trans. Evol. Computat.*, vol. 1, no. 1, pp. 67–82, 1997, doi: 10.1109/4235.585893.
- [29] W. Lohr, Untersuchungen zum Schwingungsverhalten von Vibrationsplatten mit Hilfe der Mehrkörperdynamik (Zugl.: Siegen, Univ., Diss., 2004) (Berichte aus dem Maschinenbau). Aachen: Shaker, 2005.
- [30] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning,* 30th ed. Boston: Addison-Wesley, 1989.
- [31] M. Schöberl, S. Offinger, T. Goldfuß, S. Kessler, J. Fottner, "BIM-integration of light construction equipment" in *Proceedings of the 39th international symposium on automation and robotics in construction (ISARC 2022)*, Bogota, Colombia, International Association for Automation and Robotics in Construction (IAARC), 2022, https://doi.org/10.22260/ISARC2022/0027.

1189

Motives and Barriers for Offsite and Onsite Construction 3D Printing

Wassim AlBalkhy ^{1*}, Elias Hernandez Valera ², Dorra Karmaoui ³, Zoubeir Lafhaj¹, Thomas Linner ⁴, Syrine Ayadi ¹, Reda Zerrari ¹, Assia Boutabba¹

¹Centrale Lille, CNRS, UMR 9013-LaMcube, Lille, France ²CyBe Construction, Oss, The Netherlands ³Ateliers Bataillard, Paris, France ⁴OTH Regensburg, Regensburg, Germany *Corresponding Author: Wassim.al balkhy@centralelille.fr

Abstract -

This study aims to compare between onsite and offsite construction 3D printing (C3DP). For this purpose, the Systematic Literature Review approach (SLR) was employed. The review which was based on studying 48 sources identified six categories to classify the motives and barriers for the two C3DP techniques; cost, transportation, design flexibility, workers and materials, production process, and environmental considerations. The literature identified the cost as the main factor that influences the choice between the two techniques. The presented list of motives and barriers is helpful to support decision-making in C3DP projects.

Keywords -

Additive Manufacturing, 3D Printing, Construction Management, Modular Construction, Offsite Construction, Literature Review.

1 Introduction

Three-dimensional printing (3D printing) and Additive Manufacturing (AM) are synonymous terms referring to a process that builds objects layer by layer from 3D model data, deviating from subtractive manufacturing methodologies [1]. The origin of this technology can be traced to Japanese researcher Hideo Kodama's writings, with Charles Hull developing the first commercial 3D printing machine. the Stereolithography fabrication system [2]. Over time, 3D printing has transitioned from producing prototypes to facilitating the serial production of parts across diverse fields, resulting in the global 3D printing sector's revenue exceeding \$10 billion in 2021, with expectations to surpass \$50 billion by 2030 [3,4].

The construction industry, characterised by

challenges such as low productivity, safety concerns, and environmental issues, can benefit significantly from 3D printing applications [5-8]. Its capacity to create customized structures not easily achievable with conventional methods, coupled with enhanced design flexibility and adaptability to changes, positions 3D printing as a transformative solution [6,9–11]. Furthermore, its ability to save time in design and construction processes makes it an ideal choice for emergency shelters [12]. Despite the substantial initial capital required for 3D printing equipment, cost savings in materials management, labour, and other associated expenses contribute to its economic feasibility [6]. Reports suggest that using 3D printing in housing construction can result in savings of up to 35% of the total house price in the UK [13].

Beyond economic advantages, 3D printing offers substantial sustainability benefits. Opportunities to reduce waste generation, employ eco-friendly materials, and decrease greenhouse gas emissions contribute to a more environmentally friendly construction approach [14,15]. Socially, the comfort, safety, and working conditions of construction workers can be improved through the application of 3D printing technology [12,16,17].

Construction 3D Printing (C3DP) encompasses two distinct approaches: onsite and offsite. Onsite C3DP involves the direct fabrication of building units at the designated construction site, where 3D printers are strategically positioned, autonomously producing the structured building with automated control—a process resembling typical industrial production. In contrast, offsite C3DP takes place away from the construction site, with units or components manufactured in a separate factory using automated 3D printers. These prefabricated components are then transported to the construction site for assembly, marking a clear distinction between industrial products and construction products. The onsite method emphasizes in situ construction, allowing for immediate and tailored building creation, while the offsite approach prioritizes prefabrication within a controlled environment before assembly at the construction site [18,19]. Figure 1 distinguishes between the processes of offsite and onsite C3DP.

The growing interest in Construction 3D Printing (C3DP) is evident in the existing literature, which extensively covers various aspects such as materials, the automated process, challenges, and usability across different project types and locations [12,20,21]. However, there has been a notable gap in the exploration of the distinctions between offsite and onsite C3DP. Delving into this topic and comprehending the motivations and barriers associated with each type would contribute significantly to the existing body of knowledge. This exploration is crucial not only for advancing academic research but also for providing valuable insights to practitioners, clients, decision-makers, and companies involved in C3DP projects. Understanding the potential outcomes and implications of choosing between offsite and onsite C3DP can inform strategic decision-making, leading to the enhancement of project performance. Consequently, this study aims to bridge this gap by offering a comprehensive comparison of the factors influencing both offsite and onsite C3DP.



Figure 1. C3DP approaches (a) Offsite, and (b) Onsite [19,22].

2 Research Methodology

In pursuit of the study's objectives, the methodology utilized was the Systematic Literature Review (SLR). The SLR was chosen to guarantee an impartial selection of sources, transparency, and the ability to repeat the process [23–25]. Following the guidelines of the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA), the SLR is structured around three primary phases: identification, screening, and eligibility [25]. The procedural steps involved the exploration of sources using the following terms:

("3d Printing" OR "3D Print" OR "additive manufacturing" OR "additive construction" OR "concrete print" OR "concrete printing") AND construction AND (modular OR modularity OR offsite OR "off-site" OR "off site" OR onsite OR "on site" OR "on-site").

As depicted in Figure 2, the search yielded a total of 191 sources. Among them, 32 studies were either unrelated to construction or did not directly address 3D Printing (for instance, papers focusing on 3D printing in other fields or listing it among various applications in construction). After searching the keywords, titles, and abstracts, a full-text search was conducted. During this phase, any results lacking information on onsite and offsite 3D printing were excluded. Upon completing the eligibility phase, the final count of included studies stood at 48.



Figure 2. The SLR approach

3 Results

The results of the study were organized based on five main categories as explained below:

3.1 Cost

The cost of the two types of C3DP has received the highest attention in the literature. Yang et al [26] presented a detailed comparison between the expected cost in conventional construction, offsite, and onsite C3DP projects. Based on their study, the time cost of C3DP is an important motive that needs further investigations to identify its impact. This is because it is one of the main factors that affect the company's expenses and motivates its transition toward 3D Printing. Their study also provided the different composition of the 3D printing cost in onsite and offsite projects. Offsite C3DP cost consists of printing stage cost and assembly stage cost. The printing stage cost includes management costs, manufacturing costs, printing costs, transport costs, and value-added tax. The assembly stage cost includes

construction and installation expenses, procurement expenses of equipment, and other expenses for engineering construction. Onsite C3DP cost consists of procurement expenses of equipment, construction and installation expenses, and other expenses for engineering construction. In turn, Besklubova [27] focused on the comparison between logistics costs in onsite and offsite C3DP. The study identified different factors that affect the cost in the two types of projects such as the proximity of the supplier, the size of the building, and the need for customization. It highlights material transportation costs as the dominant factor in logistics expenses, with equipment transportation costs being less significant for low-story buildings. More specifically, Onsite 3D printing proves economically superior for low-story buildings situated near suppliers, while offsite is advantageous for high-rises and distant suppliers. Alternative 3D printer technologies, notably the robotic arm, demonstrate a remarkable 96% cost reduction for onsite printing. Offsite 3D printing excels in customized or remote projects, offsetting high transportation expenses. Conversely, onsite 3D printing is cost-effective for nearby projects with minimal customization. Material transportation costs dominate logistics expenses, making onsite 3D printing more feasible for low-story buildings.

3.2 Transportation

Transportation plays a pivotal role in the dynamics of 3D printing construction, both onsite and offsite. The type of printing method, the size of the project, available transportation infrastructure, and the location of the construction site all play a role in determining the most appropriate transportation strategy. Onsite printing eliminates numerous logistical processes and preparation tasks associated with traditional construction methods [28]. As a result, onsite production allows for the immediate construction of larger structures without transportation concerns [29]. Additionally, the use of prefabricated elements, often results in damaged parts during transit, requiring on-site repairs and increased labour. This explains why some contractors prefer on-site printing to avoid damages, over-engineering, and added costs associated with transportation [29,30]. However, it may be less stable due to changing environmental conditions [29]. Onsite 3D printing is highly dependent on weather conditions due to the impact on the drying and shrinkage processes [29]. Moreover, accidents while transporting the printer may happen [12,29,31].

In turn, offsite 3D printing introduces complexities due to increased transportation needs, impacting costs. Besklubova [27] found that transportation expenses could nearly double in the offsite scenario when compared to onsite printing. Despite this, offsite construction offers advantages such as improved process control, part quality, safety, and simultaneous manufacturing and construction schedules [12,29]. However, transportation constraints, especially in remote environments, remain a challenge. The impact of weather on the transportation routes is also nonnegligible [12].

3.3 Workers and Materials

On-site and off-site C3DP present distinct labour and safety considerations. On-site printing helps reduce the demand for on-site labour by theoretically allowing the printing of entire building structures. While traditional construction roles may decrease, the need for specialized personnel to install, calibrate, and oversee 3D printing equipment remains. As a result, onsite 3D printing might be faced with the challenge of the experts' availability or the increased cost due to the need to provide training for workers [31–33].

Automation in on-site printing is expected to enhance safety by reducing hazardous tasks and providing a controlled environment. In contrast, off-site 3D printing involves prefabrication in a factory, potentially impacting on-site labour less than traditional methods. Safety benefits are attributed to lower risks and reduced congestion in a factory setting [12,29]. It is worth mentioning here that safety improvement in both types of systems is one of the main advantages of 3D Printing; contrastingly, a reduction in labour force and job opportunities may lead to political instability [31].

Concerning materials, using printable material remains the topic that received the highest attention in the literature for both types of 3D printing. The research in this field covered several aspects aiming at utilizing locally sourced materials when accessible, minimizing the need for material transportation to remote areas and reducing inventory levels [5,8,34,35]. Onsite 3D Printing might help maintain natural local materials. However, the availability of local printable materials remains very challenging [12,19,31]. Additionally, the choice between offsite and onsite 3D printing heavily depends on the type of material. For instance, concrete is widely used in offsite 3D printing due to versatility, strength, and availability. Prefabricated 3D printed components' performance can be optimized using reinforcement or specialized mixtures to improve extrusion and layering and ensure proper flowability and structural integrity. They can also integrate additives such as fibres to improve crack resistance during transportation and enhance workability [12,19,35].

3.4 Design Flexibility

Design flexibility allows C3DP to adapt to the needs of every project. The ability to modify, alter, and adjust the design details in a significant reduction in costs, materials, and labour during manufacturing [31,36,37].

C3DP offers more flexibility in modelling compared

to traditional construction methods [20]. By providing the opportunity to adjust to every conceptual design and location, C3DP brings solutions to the local environment, positively impacting the regions [12,31].

Design flexibility plays a pivotal role in both on-site and off-site C3DP. Some of the advantages include the capacity to create customized elements, different shapes, features, and textures. C3DP has proven to be a valuable competitor to prefabrication in terms of costeffectiveness and time efficiency. Simultaneously, C3DP can develop shapes that would be difficult or impossible to create using traditional methods. The use of specialized software and tools to optimize design is emphasized, allowing C3DP to optimize design and reduce risks during production [19,35,38].

Another advantage of C3DP, in comparison with conventional construction, is its ability to manufacture complex designs [20]. Design flexibility has positive effects in terms of efficiency and process time [39]. Further considerations of the design reflect on the use of local materials and labour. The C3DP manufacturing process addresses challenges related to skill development and technology adoption [40].

3.5 **Production Process**

C3DP increases efficiency and productivity compared to conventional construction. It can be used to automate many tasks involved in the production process. An advantage of C3DP is the capacity to print entire wall elements or other structural components in one piece, reducing the need for traditional construction methods such as formwork and scaffolding [41–44].

During the production process, several distinctions between on-site and off-site are addressed. On-site C3DP shows a significant improvement in terms of time efficiency compared to traditional construction methods, based on the elimination of assembly work and the fact that C3DP elements can dry while other activities are performed simultaneously. Several authors reflect on the potential of C3DP to be used in remote locations, offering a solution for housing shortages in those areas. Remote environments pose a challenge for C3DP due to local weather conditions and logistical complexities. On the other hand, off-site C3DP has the potential to become a solution as a manufacturing method for pre-fabricated modular buildings [12,19,43].

Another advantage of C3DP in off-site manufacturing is the flexibility during the design and production process, allowing builders to create different geometries and automate C3DP elements for mass production [29]. Offsite C3DP manufacturing allows builders to control external conditions for better results during printing [19,43].

3.6 Environmental Considerations

Concerning environmental impact, only a few studies have evaluated the environmental life cycle of C3DP [12]. The primary concern regarding the environmental impact of C3DP lies in material development, with concrete representing a significant percentage of CO2 emissions in C3DP [26,27,29,45]. The creation of new materials is crucial for the success of this technology [40].

New applications for C3DP depend on material properties, such as strength and durability. Some authors reflect on the potential benefits of C3DP in reducing waste and the environmental footprint of concrete [45]. The use of this technology allows builders to develop materials locally by using local aggregates. C3CP also provides the opportunity to cycle waste streams from other industries, improving circularity in the system [46].

A reduction in environmental and economic impact is evident. More studies need to be developed around the environmental life cycle assessment of C3DP materials [47]. C3DP optimizes the construction process and reduces material usage, time, and transport, directly impacting a reduction in CO2 emissions [26,27,48].

Further research on the environmental impact of C3DP needs to be conducted to reflect the differences between on-site and off-site printing. Both scenarios represent an improvement in environmental impact [19]. On one side, off-site printing allows developers to control environmental conditions and reduce risks during manufacturing [19,35]. On the other hand, several limitations around on-site C3DP are illustrated, including environmental conditions, equipment and material transportation, the use of a mixing system, and material preparation at the location, posing challenges to the implementation of C3DP on-site [38,41,48].

4 Discussion and Conclusions

3D printing in construction represents a groundbreaking innovation that aims to build more efficient, flexible, and sustainable production practices in the construction sector. As the technology continues to evolve and gain wider adoption, its influence on the way we conceptualize, design, and construct structures is expected to expand significantly. This paper aims to support the decision-making process in the construction of 3D printing projects. This is by conducting a review of the related literature to compare the onsite and offsite 3D printing techniques. The analysis of the results of this study revealed various motives and barriers for both techniques (as shown below in Table 1 for Onsite C3DP and Table 2 for Offsite C3DP). These motives were organized based on six categories: cost, transportation, workers and materials, production process, design flexibility, production process, and environmental considerations.

Barriers

Lack of

redundancy in

process and

Table 1. Onsite C3DP Motives and Barriers

Motives

Reduction of

components

transportation

Category

Cost

of sustainable	noise pollution,
materials	dust, and debris
contributing to	on-site
environmentally	disruption
conscious	
practices	
praetices	

	costs minimization of	hardware, specialized	Table 2. Offsite C3DP Motives and Barriers				
	preparation	equipment	Category	Motives	Barriers		
	tasks	costs, maintenance and repair costs	Cost	Reduced labour costs, reduction	Transportation expenses for		
Transportation	Elimination of the risk of	Potential damage during		expenses, ability	components from off-site		
	damage to	the printer's		areas without	facilities to		
	prefabricated	transit,		transporting the	construction		
	components	shortages of		printer, avoiding	locations.		
	during	skilled labour,		stoppages due to	assembly costs		
	transportation,	workers' safety		needed printer	(e.g., crane		
	minimization of	considerations		maintenance,	costs), risk of		
	waste,	in managing and		especially in	damage to		
	conservation of	maintaining 3D		remote areas	prefabricated		
	natural	printing			components		
	resources	equipment, need			during		
		build skills			transportation, transportation		
Workers and	Enabling larger	Component			limitations for		
Material	structures to be	design			large-scale		
	built without the	limitations in			items/componen		
	need for	terms of size			ts/machines		
	transportation,	and complexity,	Transportation	Reduction in the	Worker safety		
	improvement of	interruptions if	1	need for	concerns in the		
	quality,	3D printing		transporting raw	factory, extra		
	production	experiences		materials and	cost for training		
	speed, and site	technical issues,		printers to	in the factory		
	safety	quality control		multiple	-		
Design	Onsita	Waathar		construction			
Flevibility	modification	dependency		sites, ability to			
riexionity	mouncation	leading to		reach remote			
		interruptions		areas without			
		noise pollution		transporting the			
		dust and debris		printer, avoiding			
		on-site		stoppages due to			
		disruption		needed printer			
Production	Elimination of	Interruptions if		maintenance,			
Process	assembly work	3D printing		especially in			
	5	experiences	TTTTTTTTTTTTT	remote areas	D '		
		technical issues,	Workers and	Availability of	Design		
		quality control	Material	workers and	constraints for		
		challenges		stability of staff,	transportation,		
Environmental	Reduction in	Weather		reduction of	aesign		
Considerations	carbon	dependency		hu optimizing	annuations on		
	transportation	leading to		by optimizing	component size		
	emissions, use	interruptions,		the production	component size		

	process, ability	
	to standardize	
	items	
Design	Ability to	Dependency on
Flexibility	standardize	off-site facilities
-	items	
Production	Repeatability	Risks related to
Process	and ability to	production
	make	delays, supply
	continuous	chain
	improvement,	disruptions, or
	improved site	changes in
	safety,	project
	scalability to	requirements
	accommodate	
	changes in	
	project scope	
Environmental	Reduction in	Weather
Considerations	waste	dependency
	generation and	during the
	energy	transportation or
	consumption	assembly,
		carbon
		transportation
		emissions

Analysing Table 1 and Table 2 shows that some of the presented factors are not exclusively related to 3D printing. They can be considered generally when comparing onsite and offsite practices in construction. Examples of these factors cover safety concerns and logistics arrangements in factory environment or onsite.

The review of the literature showed that cost is a significant factor in the decision of whether to use onsite or offsite 3D printing construction. However, it is not possible to say definitively that one technique is more cost-effective than the other, as the choice depends on a variety of factors, including the location of the project, the size of the project, the need for customization, the need for standardized components, weather conditions, quality control measures, safety considerations, and the availability of human resources.

For example, Onsite C3DP is generally more costeffective for low-story buildings located near suppliers, while off-site printing is typically more economical for high-rise buildings and projects in remote locations.

In addition to cost considerations, it is important to weigh other factors, such as the potential to empower local human resources and create new job opportunities for local citizens. This is particularly important in areas that require development strategies.

The current study has several limitations. Firstly, it is based on a review of the literature and does not include findings from other methods, such as interviews. Secondly, the study focuses on listing the main motives and barriers without detailing the specific scenarios behind these results. For example, the study does not provide a detailed comparison between the cost of different items in both techniques or the emissions associated with each technique. Thirdly, the study does not focus on a specific location or, printing method, or material. Neither does it focus on a specific type of material. The presented comparison between offsite and onsite 3D printing may cover many other items if focusing on concrete, for instance. Examples of these factors include the ability to improve the mechanical performance of concrete, availability of additives, possibility of reinforcement, recyclability, and usability of materials at the end of the project life. Future studies should address these limitations by focusing on specific materials or using a variety of methods, including case studies, to provide more detailed and nuanced insights into the decision of whether to use onsite or offsite C3DP.

References

- [1] The American Society for Material and Testing (ASTM). Standard Terminology for Additive Manufacturing Technologies. Online: <u>https://web.mit.edu/2.810/www/files/readings/A</u> <u>dditiveManufacturingTerminology.pdf</u>, Accessed: 12/04/2024.
- [2] Sini F. Chiabert P. Bruno G. and Ségonds F. Lean management in Additive manufacturing: a methodological proposal for quality control, Politecnico di Torino, 2020. Online: <u>https://webthesis.biblio.polito.it/14001/1/tesi.pd</u> <u>f</u>, Accessed: 12/04/2024.
- [3] Ning, X. Liu T. Wu C. and Wang, C. 3D Printing in Construction: Current Status, Implementation Hindrances, and Development Agenda. *Advances in Civil Engineering*, 2021.
- [4] Molitch-Hou M. Three Areas Holding Back The \$10.6B 3D Printing Industry. Online: <u>https://www.forbes.com/sites/michaelmolitch-hou/2022/04/25/three-areas-holding-back-the-106b-3d-printing-industry/</u>, Accessed: 12/04/2024
- [5] Lafhaj Z. Albalkhy W. and Karmaoui, D. Identification of Lean Waste in Construction 3D Printing Processes. In Proceedings of The International Inorganic-Bonded Fiber Composite Conference (IIBCC). Hamburg, Germany, 2022.
- [6] Hou L. Tan Y. Luo W. Xu S. Mao C. and Moon S. Towards a more extensive application of offsite construction: a technological review. *International Journal of Construction Management*, 22(11):2154–2165. 2022.

- [7] Besklubova S. Tan B. Q. Zhong R. Y. and Spicek N. Logistic cost analysis for 3D printing construction projects using a multi-stage network-based approach. *Automation in Construction*, 151. 2023.
- [8] Albalkhy W. Bing S. El-Babidi S. Lafhaj Z. and Ducoulombier L. The Analysis of Lean Wastes in Construction 3D Printing: A Case Study. *The* 40th International Symposium on Automation and Robotics in Construction (ISARC 2023), pages 621–628, Chennai, India, 2023.
- [9] Abdel–Rasheed I. El-Mikawi M. A. and Saleh M. Empirical Model for Prediction the Impact of Change Orders on Construction Projects- Sports Facilities Case Study. *The 10 International Conference on Civil and Architecture Engineering*, 10, pages: 1–11, Cairo, Egypt, 2014.
- [10] Assbeihat J. M. and Sweis G. Factors affecting change orders in public construction projects. *International Journal of Applied*, 5(6):56–63, 2015.
- [11] Sweis G. Sweis R. Abu Hammad A. and Shboul A. Delays in construction projects: The case of Jordan. *International Journal of Project Management*, 26(6):665–674, 2008.
- [12] Schuldt S. J. Jagoda J. A. Hoisington A. J. and Delorit J. D. A systematic review and analysis of the viability of 3D-printed construction in remote environments. *Automation in Construction*, *125*(103642): 1–16, 2021.
- [13] Sini F. Chiabert P. Bruno G. and Ségonds F. A Lean Quality Control Approach for Additive Manufacturing. *IFIP Advances in Information and Communication Technology*, *594*: 59–69, 2020.
- [14] Romdhane L. 3D Printing in Construction: Benefits and Challenges. *International Journal* of Structural and Civil Engineering Research, 9(4):314–317, 2020.
- [15] Dixit M. K. 3-D Printing in Building Construction: A Literature Review of Opportunities and Challenges of Reducing Life Cycle Energy and Carbon of Buildings. *IOP Conference Series: Earth and Environmental Science*, 290(1), 2019.
- [16] Krimi I. Lafhaj Z. and Ducoulombier L. Prospective study on the integration of additive manufacturing to building industry—Case of a French construction company. *Additive Manufacturing*, 16, 107–114, 2017.

- [17] Lafhaj Z. Rabenantoandro A. Z. El Moussaoui S. Dakhli Z. and Youssef N. Experimental Approach for Printability Assessment: Toward a Practical Decision-Making Framework of Printability for Cementitious Materials. *Buildings*, 9(12):245, 2019.
- [18] Xiao J. Ji G. Zhang Y. Ma G. Mechtcherine V. Pan J. Wang L. Ding T. Duan Z. and Du S. Largescale 3D printing concrete technology: Current status and future opportunities. In *Cement and Concrete Composites*, 122, 2021.
- [19] Ter Haar B. Kruger J. and van Zijl G. Off-site construction with 3D concrete printing. *Automation in Construction*, 152, 104906. 2023.
- [20] Ali M. H. Issayev G. Shehab E. and Sarfraz S. A critical review of 3D printing and digital manufacturing in construction engineering. *Rapid Prototyping Journal*, 28(7):1312–1324. 2022.
- [21] Shahrubudin N. Lee T. C. and Ramlan R. An Overview on 3D Printing Technology: Technological, Materials, and Applications. *Procedia Manufacturing*, 35:1286–1296. 2019.
- [22] Xu W. G. Huang S. Y. Han D. Zhang Z. L. Gao Y. Feng P. and Zhang D. B. Toward automated construction: The design-to-printing workflow for a robotic in-situ 3D printed house. *Case Studies in Construction Materials*, 17. 2022.
- [23] Albalkhy W. and Sweis R. Barriers to adopting lean construction in the construction industry: a literature review. *International Journal of Lean Six Sigma*, *12*(2):210–236, 2021.
- [25] Moher D. Liberati A. Tetzlaff J. Altman D. G. Altman D. Antes G. Atkins D. Barbour V. Barrowman N. Berlin J. A. Clark J. Clarke M. Cook D. D'Amico R. Deeks J. J. Devereaux P. J. Dickersin K. Egger M. Ernst E...and Tugwell P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine, 6(7):1–6. 2009.
- [26] Yang H. Chung J. K. H. Chen Y. and Li Y. The cost calculation method of construction 3D printing aligned with Internet of Things. *Eurasip Journal on Wireless Communications and Networking*, 2018(1):1–9, 2018.
- [27] Besklubova S. Tan B. Q. Zhong R. Y. and Spicek N. Logistic cost analysis for 3D printing construction projects using a multi-stage network-based approach. *Automation in Construction*, 151, 2023.

- [28] Kothman I. and Faber N. How 3D printing technology changes the rules of the game Insights from the construction sector. *Journal of Manufacturing Technology Management*, 27(7): 932–943, 2016.
- [29] Adaloudis M. and Bonnin Roca J. Sustainability tradeoffs in the adoption of 3D Concrete Printing in the construction industry. *Journal of Cleaner Production*, 307, 127201, 2021.
- [30] Vantyghem G. De Corte W. Shakour E. and Amir O. 3D printing of a post-tensioned concrete girder designed by topology optimization. *Automation in Construction*, 112, 103084, 2020.
- [31] Bazli M. Ashrafi H. Rajabipour A. and Kutay C. 3D printing for remote housing: Benefits and challenges. *Automation in Construction*, 148, 2023.
- [32] Khajavi S. H. Tetik M. Mohite A. Peltokorpi A. Li M. Y. Weng Y. W. and Holmström J. Additive Manufacturing in the Construction Industry: The Comparative Competitiveness of 3D Concrete Printing. *Applied Sciences*, 11(9), 2021.
- [33] Jipa, A. and Dillenburger B. 3D Printed Formwork for Concrete: State-of-the-Art, Opportunities, Challenges, and Applications. 3D *Printing and Additive Manufacturing*, 9(2): 84– 107, 2022.
- [34] Bar-Sinai, K. L., Shaked, T., & Sprecher, A. Robotic tools, native matter: workflow and methods for geomaterial reconstitution using additive manufacturing. *Architectural Science Review*, 64(6): 490–503, 2021.
- [35] Pasco J. Lei Z. and Aranas C. (2022). Additive Manufacturing in Off-Site Construction: Review and Future Directions. *Buildings*, *12*(1), 53, 2022.
- [36] Volpe S. Sangiorgio V. Petrella A. Coppola A. Notarnicola M. and Fiorito F. Building Envelope Prefabricated with 3D Printing Technology. Sustainability, 13(16), 2021.
- [37] Zhang X. Li M. Y. Lim J. H. Weng Y. W. Tay Y.
 W. D. Pham H. and Pham Q. C. Large-scale 3D printing by a team of mobile robots. *Automation in Construction*, 95: 98–106, 2018.
- [38] Poullain P. Paquet E. Garnier S. and Furet B. On site deployment of 3D printing for the building construction - The case of YhnovaTM. 8th Sci. Conf. Mater. Probl. Civ. Eng. 163, Cracow, Poland, 2018.
- [39] Sakin M. and Kiroglu Y. C. 3D Printing of Buildings: Construction of the Sustainable

Houses of the Future by BIM. *Energy Procedia*, *134*:702–711, 2017.

- [40] Rollakanti C. R. and Prasad, C. Applications, performance, challenges and current progress of 3D concrete printing technologies as the future of sustainable construction - A state of the art review. In Materials Today-Proceedings (Vol. 65, Issue International Conference on Advances in Construction Materials and Structures (ICCMS), Trivandrum, India, 995–1000, 2022.
- [41] Lim J. H. Panda B. and Pham, Q. C. Improving flexural characteristics of 3D printed geopolymer composites with in-process steel cable reinforcement. *Construction and Building Materials*, 178, 32–41, 2018.
- [42] Žujović M. Obradović R. Rakonjac I. and Milošević J. 3D Printing Technologies in Architectural Design and Construction: A Systematic Literature Review. *Buildings*, 12(9), 2022.
- [43] Placzek G. and Schwerdtner P. Concrete Additive Manufacturing in Construction: Integration Based on Component-Related Fabrication Strategies, *Buildings*. 13 (2023).
- [44] Bos F. Wolfs R. Ahmed Z. and Salet T. Additive manufacturing of concrete in construction: potentials and challenges of 3D concrete printing. *Virtual and Physical Prototyping*, 11(3): 209– 225, 2016.
- [45] Baduge S. K. Navaratnam S. Abu-Zidan Y. McCormack T. Nguyen K. Mendis P. Zhang G. M. and Aye L. Improving performance of additive manufactured (3D printed) concrete: A review on material mix design, processing, interlayer bonding, and reinforcing methods. *Structures*, 29: 1597–1609. 2021.
- [46] Valera E. H. Cremades R. van Leeuwen E. and van Timmeren A. Additive manufacturing in cities: Closing circular resource loops. In Circular Economy 2(3), 2023.
- [47] Gislason S. Bruhn S. Breseghello L. Sen B. Liu G. and Naboni R. Porous 3D printed concrete beams show an environmental promise: a cradleto-grave comparative life cycle assessment. *Clean Technologies and Environmental Policy*, 24(8): 2639–2654, 2022.
- [48] Batikha M. Jotangia R. Baaj M. Y. and Mousleh I. 3D concrete printing for sustainable and economical construction: A comparative study. *Automation in Construction*, 134, 2022.

BIM Model View Definition (MVD) for disassembly planning of buildings

Benjamin Sanchez¹, Srijeet Halder¹, Ranjith K. Soman² and OK-Youn Yu¹

¹ Department of Sustainable Technology and the Built Environment, Appalachian State University, USA ²Department of Management in the Built Environment, Delft University of Technology, The Netherlands <u>sanchezandradeb@appstate.edu</u>, <u>halderss1@appstate.edu</u>, <u>r.soman@tudelft.nl</u>, <u>yuo@appstate.edu</u>

Abstract

Buildings are responsible for a huge amount of negative environmental impacts at the End-of-Life (EoL) stage because of the massive global production of waste. Therefore, there is an urgent need for developing design systems to decrease waste generation by facilitating the reuse and recycling of building materials. Building Information Modeling (BIM) has been demonstrated to be an effective tool for enhancing building performance at the different life cycle stages. However, the implementation of BIM for assessing the EoL stage of building projects is still underdeveloped. One important activity on the EoL stage is disassembly planning for reuse of building components and building renovation. The aim of this study is to develop the framework for a BIM Model View Definition (MVD) for disassembly planning. First, we developed a framework for an MVD for the disassembly planning of buildings based on the literature review in the field. Then, we proposed a workflow for implementing the MVD, as well as the validation procedures. In the end, the MVD was validated by compiling the information of a BIMbased disassembly model. Results of the case study show that the presented approach is efficient for corroborating the syntax, semantic structure, and information content of BIM models.

Keywords -

Model View Definition, disassembly information modeling, Building Information Modeling, circular economy

1 Introduction

The significance of reusing building components and systems has increased because the construction industry is annually accountable for around 40% of the global consumption of natural resources and the redirection of 40% of waste to landfills [1]. To address this challenge, technological advancements such as Construction Waste Management (CWM), Materials Passports (MP), Product Recovery Management (PRM), and Life Cycle Assessment (LCA) have been introduced in the last decades [2]. These innovations aim to enhance the rates of reusing and recycling building components. Nevertheless, the adoption of building components and system reuse remains limited, primarily due to a lack of research on reclamation protocols and the absence of standardized procedures [3]. In this matter, disassembly planning is a strategic approach for the recovery of building components and systems for their future reuse or recycling [4-8]. Disassembly planning involves identifying the necessary sequential steps for taking apart a building, delineating deconstruction activities, and arranging them in a logical order.

Unfortunately, disassembly planning for buildings is not as advanced as in other industries such as manufacturing, automotive, and electronics [4,9]. The definition of the information and data structure that a disassembly model must contain are critical for the implementation of disassembly planning methods and theories. In the following sections we present a framework for defining the information definition for disassembly planning of buildings in the context of Building Information Modeling (BIM).

2 Background

2.1 Disassembly planning for reuse of building components and building renovation

The potential benefits of building renovation rely on the fact that it is possible to take away components from obsolete building and then an repair, reuse, remanufacture, or recycle Planning for them. disassembly plays a key role in the building renovation process where the disassembly planning sequence, as well as the disassembly methods to recover target components, have to be performed in an efficient way. The objectives are to reduce building costs and to increase the reuse times for building components. If the design for disassembly is too complex or timeconsuming, the associated economic and environmental

costs could be higher than installing new components. This field has been studied in the manufacturing industry since the last decade, concluding that disassembly planning can reduce the time and cost associated with disassembling products [10]. Figure 1 shows the comparison of the material flows through the building lifecycle stages for the traditional approach of building demolition versus the disassembly planning approach. Building demolition produces large amount of waste with the possibility to recycle some material. In contrast, selective disassembly enables the direct reuse of buildings in different levels of decomposition (e.g. reus of a building, a subsystem, a component). Table 1 shows the generic classification of product disassembly methods, as well as the literature review of the applied theories to find the optimum, or near optimum, disassembly path for non-destructive disassembly methods.



Figure 1. Building lifecycle stages: Demolition versus disassembly planning

Table 1. Disassembly planning methods and models for non-destructive disassembly optimization

	assembly	Complex wave propagation algorithms [11-13]
selective		Less complex transition matrices, mixed integer linear programming, and genetic algorithms [14,15]
e and	ial dis	Disassembly sequence structure graphs (DSSGs) and expert rules [16]
mplet	quenti	Disassembly precedence graphs and linear programming models [17]
3)	Se	Genetic algorithms optimisation [18,19]
oly		Ant colony optimisation [20-22]
eml		Directed AND/OR graphs and linear
ass	ly.	programming [23]
dis	isassemb	AND/OR graphs and constraint programming
ve		[24]
cti		Petri nets and network searching algorithms [25]
tru	il d	Liaison graphs and grouping genetic algorithms
des	alle	[26]
- H	ara	Fuzzy-rough set mapping model, Disassembly
ž	щ	factors and Disassembly membership functions
		[27]

In accordance with disassembly planning, the process of planning the dismantling of an existing asset is known as deconstruction project planning. Hübner et al. [28] provided a comprehensive literature overview of project planning methods for deconstruction of buildings as well as some research gaps in this field. The authors classified the existing deconstruction planning methods according to specific construction project objectives. The proposed project objectives are time, cost, resources, risk, and quality, which, in turn, have been the main performance indicators on the field of construction project management. Their study presents a full-range variety of methods for the decision-making process on strategic and operational deconstruction planning, based on the optimization of single or multiple project objectives. In comparison to the manufacturing industry, just few disassembly planning methods have been developed for buildings and building subsystems. This have limited the implementation of disassembly planning for building projects.

2.2 BIM Model View Definition use cases in construction

Building Information Modeling (BIM) is a data-rich, object-based, intelligent, and parametric digital representation of a building, from which views appropriate to various users' needs can be extracted and analysed to generate feedback and improvement of the facility design [29]. In the last decades BIM has been implemented in diverse research fields in construction such as design visualization, construction reviews, structural analysis, design coordination, planning of trades and systems, construction scheduling and sequencing, among others [2,30]. The standard BIM data structure is known with the name of Industry Foundation Classes (IFC) schema. IFC is an open, international standard (ISO 16739-1:2018), meant to be usable across a wide range of hardware devices, software platforms, and interfaces for many different use cases [31].

A Model View Definition (MVD) is a subset of the overall IFC schema to describe data exchange for a specific use or workflow, narrowing the scope depending on the need of the receiver [32]. An IFC View Definition, or MVD, defines a subset of the IFC schema, that is needed to satisfy one or many exchange requirements of the AEC industry. Some examples of MVDs for specific exchange requirements are quantity take-off view, structural design to structural analysis, indoor climate simulation to HVAC design, architectural design to thermal simulation, architectural design to circulation/security analysis, energy analysis view, among others [32].

For a long time, everyone could create their own MVD and approach software vendors to implement it. This created a situation with several MVDs that have

been created that are not interoperable with each other and need additional efforts for implementation in software tools. Therefore, the MVD structure is now regulated by the buildingSMART International (bSI) Standard [33] and has been implemented in numerous BIM interoperability studies [34]. The aim of MVD is to specify exactly which information is to be exchanged in each exchange scenario and how to relate it to the IFC model [33]. The development of MVDs for new purposes in the construction industry is highly recommended by bSI to help satisfy arising industry needs, explore limitations in current approaches, and propose new necessary extensions [34]. According to Son et al. [35], the number of MVDs is still too small considering the number of information exchange scenarios required by the industry. Even though, disassembly planning has been recently implemented using BIM in multiple tools and methods [36,37], none of them have described the structure for IFC data exchange. In this paper, we argue the need for the development of an MVD for disassembly planning for enabling the use of BIM for the assessment of dismantling operations. The conformation of an MVD can use existing extensions of the IFC schema, add new extensions, add additional restrictions, and overrule others.

3 A framework for a MVD for disassembly planning of buildings

The proposed research approach is developed in three main stages. In the first stage, a literature review was carried out to determine the necessary information for disassembly planning of buildings. The objective of this first stage is to define the subset of information needed for developing a disassembly planning assessment for buildings. In the second stage, we develop the MVD. Then we propose the workflow for the implementation of a disassembly planning MVD for BIM models. This workflow describes the main concepts, actors, and information flow for the use case of disassembly planning. In a final stage, we present a validation approach of the disassembly planning MVD. We present the results of a case study as a functional demonstration of the proposed methodology and the proposed technology. Figure 2 displays the proposed research approach as well as how the sections of this study are organized.

3.1 Parameters for disassembly planning for buildings

In previous studies, Sanchez et. al [36,37] developed a literature review for identifying the key parameters for BIM-based disassembly planning of buildings. In these studies, the authors identified three stages in the BIM-



Figure 2. Research approach for a MVD for disassembly planning of buildings (*MVD-Model View Definition, BIM-Building Information Engineering, IFC-Industry Foundation Classes).

based disassembly planning process. The three stages are the preprocessing stage, the analytical model, and the performance analysis. In each one of the stages, the authors identified the key parameters (information packages) for disassembly planning models. Figure 3 shows the workflow for implementing disassembly planning MVD for BIM models, and Table 2 shows the stages and parameters for BIM-based disassembly models. For the purposes of this study, we use the proposed key parameters in Table 2 as the basis for the information package for the MVD for disassembly building models. It is worthy to mention that this list of parameters can be expanded according to the requirements for BIM assessment.



Figure 3. Workflow for implementing disassembly planning MVD for BIM models (*MVD-Model View Definition, BIM-Building Information Modeling, IFC-Industry Foundation Classes)

3.2 Disassembly planning MVD development

The parameters of the MVD information package are described and mapped according to the existing IFC entities (IFC4) (see Table 2). The parameters can be categorized according to the nature of the source of the information such as existing (E), new (N), or retrievable (R). The existing parameters are embedded in the BIM

	Parameter	Description	Source	IFC Schema	Type*
	Global disassembly model	Unique descriptor for	[5,9,10,40]	<i>IfcGlobalUniqueID</i>	E
	-	identifying all the parts.			
	Disassembly part type	Component (c) or connection	[5,9,10]	IfcBuiltElement;	Ν
		(f).		IfcElementComponent	
	Disassembly part id.	Numerical identifier for part	[5,9,10,40]	IfcGlobalUniqueID,	Ν
		types (e.g., c_1, f_1).		IfcBuiltElement;	
	Hosted components	Components attached to a host	[5,9,40]	IfcRelConnectsElements	N
	Hosted components	component		ijener@onneetsEtements	1
ing.	Hosted connections	Connections attached to a host	[9,40]	IfcRelConnectsElements	N
ess	Hosted connections	component		I forter connects Lientenits	19
õ	Connection disassembly	Type of connection	[5,10,40]	IfcRelConnectsWithRea	N
epi	Connection disassembly	Type of connection.		lizingElements	1
\mathbf{Pr}	Fastener constraint type	Extraction constraints.	[9,10,40]	<i>IfcElementComponent</i>	Ν
	Physical interface	Type of connection.	[10,40]	IfcRelConnectsWithRea	Е
			[0 10 40]	lizingElements	-
	Global coordinate system	Global reference system of the	[9,10,40]	IJCCoordinateRejerence System	Е
	T 1 1 1 1 1	model.	[10]		Б
	Local coordinate system	Local reference system of each	[10]	IJCObjectPlacement	E
		part.	FQ 101	K - IIDI	
	Assembly elements location	Location of each part in the	[7,10]	IJCLOCAIPIAcement	Ν
		coordinate system.	[5 9 40]	K-Flow and Assessible	
	Structural composition	Structural interdependence of	[3,9,40]	IJCElementAssembly	Ν
		components.	[5 9 10 40]	<i>K</i> -D-1D	
	Graph data structure	Abstract representation of a	[3,7,10,40]	IJCReiDecomposes	Ν
5		disassembly model.	[9 10]	HoElow out Ang ombly	Б
ode	Extraction directions	Directions for removing any	[7,10]	IJCElementAssembly	E
ũ		part.	[9 10 40]	KoDuo du ot Donuog out ati	Б
cal	Object geometry (2D, 3D)	virtual representation of BIM	[3,10,10]	on	E
lyti		elements (parts)	[9,10]	If a Shan a Donne contation	N
na	Physical constraints	Constraints of a part in any	[,,10]	ijcsnapekepresentation	N
A	Madalan ankasanahlisa	Crease of contraction.	[10.40]	IfeProductPanrasantati	Б
	Wodular subassemblies	diagagembly module	[,]	on	E
		Change for a horizon module.	[5.9.10]	IfeFlowant Assambly	N
	working space	Space for a numan worker to	[*,,,:*]	ijeLiemeniAssembly	IN
	Disassembly to al	Equipment for disassembly	[10]	IfePalSpacePoundary	N
	Disassembly tool	works		ускезрисевоинии у	IN
	Disassembly method	works. Mode of disassembly works	[10,40]	IfcPronertvSet	N
	Environmental impacts (I CA)	Life Cycle Assessment (LCA)	[9,10,40]	lfcPropertySet	P
	Environmental impacts (ECA)	per building component		ijer ropertjiser	K
ce	Disassembly time	Time for disassembling	[5,10,40]	<i>IfcPropertvSet</i>	R
nan	Disassembly cost	Cost for disassembling	[5,9,10,40]	IfcPropertvSet	R
orn	Disassembly revenue Profit for disassembling		[5,10,40]	IfcPropertySet	R
erf	Disassembly distance	Distance moved in	[10]	IfcPropertySet	N
Р		disassembling a component			11
	Operation number	Number of activities developed	[10]	<i>IfcPropertySet</i>	Ν
	- r	for the disassembly process.		•	- •
	Disassembly energy	Energy consumption of	[10]	<i>IfcPropertySet</i>	R
	consumption	machinery for disassembly.		*	
	<u>+</u>				

Table 2. Key para	ameters for BIM-	based disassem	bly i	olanning	of building	ngs
21			~ 1			0

*Information type: existing (E), new (N), or retrievable (R)

1201

elements. The new parameters must be defined by the user. Also, these parameters can be generated automatically with customized subroutines. The retrievable parameters are properties of the BIM elements that can be retrieved from external databases. The full description of the parameters can be found in [36].

3.3 Workflow to implement disassembly planning MVD

Figure 3 shows the proposed workflow for the implementation of the disassembly planning MVD for BIM models. In the first step, a conceptual disassembly model design must be developed according to the building assembly under study. In the second step, the BIM disassembly model must be developed. In this step, the minimum IFC entities for disassembly models are defined as a template according to the MVD for disassembly planning. Also, in this second step, the BIM model must be semantically enriched according to the MVD. The semantic enrichment can be done manually (by the user) or through algorithms that retrieve geometrical and non-geometrical information from the BIM model. With a functional BIM disassembly model. in the next steps, it is possible to develop an appropriate analytical model, as well as a disassembly planning solution.

3.4 Validation approach for disassembly planning MVD

An MVD is the computational application of an information package that maps the data exchanges to a subset of the IFC schema and describes their needs in a computer-readable data model [34]. Figure 4 shows the workflow for the validation of the MVD for disassembly planning. The BIM-Q tool is proposed for MVD configuration and to generate the mvdXML file. The BIM-Q tool is a commercial software for supporting BIM use case requirements. The BIM-Q tool has useful MVD templates for the different IFC use cases (IFC models) such as structural analysis, building energy analysis, and cost estimation. Due to the specific characteristics of analysis and data structure for processing, a new IFC use case for disassembly planning implementation and standardization was necessary. The mvdXML format is a data schema used to specify the minimum exchange requirements and it is the currently recommended data schema for model validation [34]. The result of the MVD configuration is a description of the appropriate IFC information model and their exchange data requirements. In the final step, simplebim® software is proposed for validating the IFC file. This MVD validation approach has been used for other studies in the field of BIM data

exchange [34,38,39].

The disassembly planning MVD was validated by compiling the information of the BIM-based disassembly model of a case study described in a previous research [36] and by verifying the resulting IFC files using the simplebim® tool. To test the MVD construct, the information depicting disassembly planning characteristics was included in the IFC files for the



Figure 4. Disassembly planning MVD development and validation (*MVD-Model View Definition, BIM-Building Information Modeling, IFC-Industry Foundation Classes, mvdXML-MVD data format, BIM-Q-commercial software for supporting Building Information Modeling use case requirements, simplebim- Open BIM IFC application for verifying data accuracy)

disassembly model prototype. The simplebim® tool was implemented to corroborate syntax, semantic structure, and content of the output IFC file. This validation process was necessary to refine the enrichment process by detecting errors of semantic structure (e.g., erroneous property set labels, mistaken entity relationships) and missing information (e.g., missing exchange requirements, information lost in the enrichment process). Figure 5 shows part of the results of the MVD for disassembly planning for buildings.

4 Conclusions and future work

This research explores the necessity for standardized information in disassembly models of buildings within the BIM framework. Efficient management of asset data is essential for informed decision-making throughout all stages of a building's life cycle, including its End-of-Life phase. This paper introduces a formal specification of a MVD for disassembly planning. This approach aims to



Figure 5. MVD for disassembly planning: a) BIM disassembly planning parameters, and b) MVD information package using the BIM-Q interface

enhance the adoption of building disassembly by minimizing the effort associated with planning tasks. The framework outlined in this research has the potential to address the current technical obstacles that hinder the systematic and standardized implementation of disassembly planning in the construction industry.

A future phase for this study involves creating an automated semantic enrichment engine for disassembly models based on BIM. This strategy aims to enhance the generation of accurate and thoroughly semantically enriched BIM disassembly models. One drawback of the suggested approach is the substantial volume of data that needs to be incorporated into a BIM model at a component level. The high Level of Detail (LoD) may pose computational challenges, especially for large building assemblies. Nevertheless, it might be feasible to apply this approach for subassemblies of buildings. Future advancements in computational technology could make high LoD assessments technically feasible for complete building assemblages.

This study offers a contribution to the field due to key factors that distinguish it from alternative solutions. First, unlike some alternative solutions which may focus on specific aspects of disassembly planning, this paper presents a holistic approach leveraging BIM MVD. It considers various crucial factors such as material properties, structural integrity, spatial relationships, and environmental impact, offering a comprehensive framework for disassembly planning. Second, the BIM proposed in the paper promotes MVD data interoperability and standardization, which are essential for seamless communication and collaboration among stakeholders involved in disassembly projects. This ensures that relevant information is accurately exchanged and utilized throughout the lifecycle of the building,

leading to more efficient planning and execution. Finally, the BIM MVD framework presented in the paper is designed to be scalable and adaptable to different types and scales of buildings, making it suitable for a wide range of applications. Whether dealing with small-scale structures or large-scale complexes, the approach can be tailored to meet specific project requirements and constraints.

Reference

- S. Badi, N. Murtagh, Green supply chain management in construction: A systematic literature review and future research agenda, Journal of Cleaner Production. 223 (2019) 312-322, 10.1016/j.jclepro.2019.03.132 ISSN 0959-6526.
- [2] F. Jalaei, Integrate Building Information Modeling (BIM) and Sustainable Design at the Conceptual Stage of Building Projects, Doctoral Dissertation, 2015, University of Ottawa, .
- [3] A. Mahpour, Prioritizing barriers to adopt circular economy in construction and demolition waste management, Resources, Conservation and Recycling. 134 (2018) 216-227.
- [4] B. Sanchez, C. Rausch, C. Haas, Deconstruction programming for adaptive reuse of buildings, Automation in Construction. 107 (2019) 102921, 10.1016/j.autcon.2019.102921 ISSN 0926-5805.
- [5] F. Denis, C. Vandervaeren, N.D. Temmerman, Using network analysis and BIM to quantify the impact of Design for Disassembly, Buildings. 8 (2018) 113, 10.3390/buildings8080113 ISSN 2075-5309.
- [6] 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), 10.1016/j.resconrec.2021.106001.

- [7] 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 (2021) 307, 10.1108/ECAM-12-2020-1023 ISSN 0969-9988.
- [8] E. Durmisevic, Transformable building structures: design for dissassembly as a way to introduce sustainable engineering to building design & construction, (2006).
- [9] B. Sanchez, C. Haas, A novel selective disassembly sequence planning method for adaptive reuse of buildings, Journal of Cleaner Production. 183 (2018) 998-1010, 10.1016/j.jclepro.2018.02.201 ISSN 0959-6526.
- [10] Z. Zhou, J. Liu, D.T. Pham, W. Xu, F.J. Ramirez, C. Ji, Q. Liu, Disassembly sequence planning: Recent developments and future trends, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture. (2019), 10.1177/0954405418789975.
- [11] H. Srinivasan, R. Gadh, A geometric algorithm for single selective disassembly using the wave propagation abstraction, Computer-Aided Design. 30 (1998) 603-613.
- [12] H. Srinivasan, R. Figueroa, R. Gadh, Selective disassembly for virtual prototyping as applied to demanufacturing, Robotics and Computer-Integrated Manufacturing. 15 (1999) 231-245.
- [13] C. Mascle, B. Balasoiu, Algorithmic selection of a disassembly sequence of a component by a wave propagation method, Robotics and Computer-Integrated Manufacturing. 19 (2003) 439-448.
- [14] A. Elsayed, E. Kongar, S.M. Gupta, An evolutionary algorithm for selective disassembly of end-of-life products, Int J Swarm Intell Evol Comput. 1 (2012) 1-7.
- [15] S. Behdad, M. Kwak, H. Kim, D. Thurston, Selective disassembly and simultaneous end-of-life decision making for multiple products, , International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 49057, 2009, pp. 313-321, .
- [16] S. Smith, G. Smith, W.-. Chen, Disassembly sequence structure graphs: An optimal approach for multiple-target selective disassembly sequence planning, Advanced Engineering Informatics. 26 (2012) 306-316, 10.1016/j.aei.2011.11.003 ISSN 1474-0346.

- [17] H.J. Han, J.M. Yu, D.H. Lee, Mathematical model and solution algorithms for selective disassembly sequencing with multiple target components and sequence-dependent setups, International Journal of Production Research. 51 (2013) 4997-5010, 10.1080/00207543.2013.788794.
- [18] S.S. Smith, W. Chen, Rule-based recursive selective disassembly sequence planning for green design, Advanced Engineering Informatics. 25 (2011) 77-87, 10.1016/j.aei.2010.03.002 ISSN 1474-0346.
- [19] G.C. Smith, S.S. Smith, An enhanced genetic algorithm for automated assembly planning, Robotics and Computer-Integrated Manufacturing. 18 (2002) 355-364.
- [20] C. Lu, H.Z. Huang, B. Zheng, J. Fuh, Y.S. Wong, An ant colony optimization approach to disassembly planning, , 2008 International Conference on Apperceiving Computing and Intelligence Analysis, IEEE, 2008, pp. 81-84, .
- [21] J.F. Wang, J.H. Liu, S.Q. Li, Y.F. Zhong, Intelligent selective disassembly using the ant colony algorithm, Ai Edam. 17 (2003) 325-333.
- [22] X. Junfang, Planning of selective disassembly sequence based on ant colony optimization algorithm, JOURNAL OF COMPUTER AIDED DESIGN AND COMPUTER GRAPHICS. 19 (2007) 742.
- [23] J. Kang, D. Lee, P. Xirouchakis, J. Persson, Parallel disassembly sequencing with sequence-dependent operation times, CIRP Annals. 50 (2001) 343-346.
- [24] X. Zwingmann, D. Ait-Kadi, A. Coulibaly, B. Mutel, Optimal disassembly sequencing strategy using constraint programming approach, Journal of Quality in Maintenance Engineering. 14 (2008) 46-58.
- [25] X. Guo, S. Liu, M. Zhou, G. Tian, Disassembly sequence optimization for large-scale products with multiresource constraints using scatter search and Petri nets, IEEE transactions on cybernetics. 46 (2015) 2435-2446.
- [26] H. Tseng, C. Chang, C. Cheng, Disassemblyoriented assessment methodology for product modularity, International Journal of Production Research. 48 (2010) 4297-4320, 10.1080/00207540902893433 ISSN 0020-7543.
- [27] X.F. Zhang, G. Yu, Z.Y. Hu, C.H. Pei, G.Q. Ma, Parallel disassembly sequence planning for complex products based on fuzzy-rough sets, The International Journal of Advanced Manufacturing Technology. 72 (2014) 231-239.
- [28] F. Hübner, R. Volk, A. Kühlen, F. Schultmann, Review of project planning methods for deconstruction projects of buildings, Built Environment Project and Asset Management. 7

(2017) 212-226, 10.1108/BEPAM-11-2016-0075 ISSN 2044-124X.

- [29] US General Services Administration (GSA), BIM Guide Overview, (2007).
- [30] R. Volk, J. Stengel, F. Schultmann, Building Information Modeling (BIM) for existing buildings
 Literature review and future needs, Automation in Construction. 43 (2014) 204, 10.1016/j.autcon.2013.10.023 ISSN 0926-5805.
- [31] BuildingSMART International, Industry Foundation Classes (IFC) - An introduction, (2020) Available at <u>https://technical.buildingsmart.org/standards/ifc/,</u> last access: June/30.
- [32] BuildingSMART International, MVD Database, (2023) Available at <u>https://technical.buildingsmart.org/standards/ifc/m</u>vd/mvd-database/, .
- [33] National Institute of Building Sciences buildingSMART alliance, National BIM Standard-United States Version 2, USA, 2012, .
- [34] R. Sacks, A. Kedar, A. Borrmann, L. Ma, I. Brilakis, P. Hüthwohl, S. Daum, U. Kattel, R. Yosef, T. Liebich, B.E. Barutcu, S. Muhic, SeeBridge as next generation bridge inspection: Overview, Information Delivery Manual and Model View Definition, Automation in Construction. 90 (2018) 134-145, 10.1016/j.autcon.2018.02.033 ISSN 0926-5805.
- [35] S. Son, G. Lee, J. Jung, J. Kim, K. Jeon, Automated generation of a model view definition from an information delivery manual using idmXSD and buildingSMART data dictionary, Advanced Engineering Informatics. 54 (2022) 101731, 10.1016/j.aei.2022.101731 ISSN 1474-0346.
- [36] B. Sanchez, P. Herthogs, R. Stouffs, Identifying key parameters for BIM-based disassembly planning, Identifying key parameters for BIM-based disassembly planning, ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 40, IAARC Publications, 2023, pp. 32-39, .
- [37] B. Sanchez, C. Rausch, C. Haas, T. Hartmann, A framework for BIM-based disassembly models to support reuse of building components, Resources, Conservation and Recycling. 175 (2021) 105825, 10.1016/j.resconrec.2021.105825 ISSN 0921-3449.
- [38] M. Weise, N. Nisbet, T. Liebich, C. Benghi, IFC model checking based on mvdXML 1.1, in: Anonymous eWork and eBusiness in Architecture, Engineering and Construction, CRC Press, 2017, pp. 19-26.
- [39] AEC3 Ldt, BIM-Q compact knowledge: the little BMQ 1x1, (2021) Available at https://www.bimq.de/en/compact-knowledge/,.

[40] E. Durmisevic, A. Guerriero, C. Boje, B. Domange, G. Bosch, Development of a conceptual digital deconstruction platform with integrated Reversible BIM to aid decision making and facilitate a circular economy, Proc. of the Joint Conference CIB W78-LDAC, vol. 2021, 2021, pp. 11-15, .

Customer Satisfaction in Construction Robots: A Multi-Stakeholder Perspective

Siwei Chang¹ Kwok Tai Wai² Ching-Wei Chang¹ Yixing Yuan¹ Juhao Su³

¹Hong Kong Center for Construction Robotics, Hong Kong ²Wah Tung Façade Company Limited, Cheung Sha Wan, Hong Kong ³Academy of Interdisciplinary Studies, The Hong Kong University of Science and Technology, Hong Kong S.A.R.

siweichang@ust.hk, davidkwok@wahtungfacade.com, ccw@ust.hk, yixing@ust.hk, jsuan@connect.ust.hk

Abstract

Construction robotics, regarded as a key trigger to reshape the traditional construction industry, has continuously emerged. Although various construction robot prototypes have been developed, only some of them are widely adopted for on-site practices. To improve users' acceptance of this novel technology, this research aims to investigate customers' satisfaction levels utilizing construction robots to address construction constraint factors from a multistakeholder perspective. Specifically, ten constraint factors were identified through a systematic literature review. The satisfaction level for each constraint factor was then scored using a 1-5 Likert scale through a questionnaire survey. Based on the satisfaction scores, six key stakeholder groups were first segmented using the k-means algorithm to target acceptance patterns. Mean score ranking of the scores marked by stakeholders unwilling to adopt construction robots was conducted to identify current concerns. Results indicate that acceptance is more probable among experienced stakeholders and those familiar with construction robots, signifying the market's readiness for broader adoption. High costs and a lack of partnership cooperation emerged as primary constraint factors. Recommendations to solve these concerns are detailed. The results contribute to speeding up the development of construction robots by investigating and improving customer acceptance of this novel technology.

Keywords -

Construction robotics; Customer satisfaction; technology acceptance

1 Introduction

Construction robots, defined as intelligent machines equipped with sensors, actuators, and advanced control systems tailored for construction activities, have evolved significantly since the late 20th century [1]. This technological evolution has led to diverse applications, ranging from inspection and excavation to on-site construction and demolition.

For example, [2] developed a robotic vehicle system for unmanned excavations (depicted in Figure 1 (a)). This unmanned excavation paradigm integrates a vehicle chassis and 6 degrees of freedom (DOF) manipulators, augmented by sensor suites comprising cameras, realtime kinematic positioning (RTK), global positioning system (GPS), and inertial measurement unit (IMU) that are used for comprehensive position and attitude measurement. A 5G communication module was developed to facilitate the transmission of control signals. [3] advanced the field with a polyarticulated robot featuring an articulated arm affixed to an automatic guided vehicle (AGV) for precision-controlled automated concrete pouring tasks (Figure 1 (b)). Executing its operations based on trajectories derived from Building Information Modeling (BIM) geometry information, the system manipulates the articulated arm and AGV, facilitating precise concrete pouring at predetermined positions. [4] introduced a retractable robot for automated wall spraying (Figure 1(c)), demonstrating meticulous engineering to balance the imperatives of a compact robot and a large working area. Controlling algorithms, including a surface-to-surface parallel adjustment mechanism relying on laser ranging and a polar coordinate transformation method leveraging LiDAR data, confer the ability to maintain parallelism to the wall during spraying and autonomously identify working areas. Concurrently, [5] developed a robotic system tailored for the sorting of construction demolition waste (Figure 1(d)). This system, featuring a wheeled mobile chassis for precise navigation and a robot manipulator for waste handling, incorporates computervision technology to recognize and locate waste.

Rigorous validation through both laboratory testing and on-site validation substantiates the system's capacity for the autonomous execution of envisioned highefficient construction processes. However, despite the development of these pioneering prototypes, the on-site integration of construction robots remains gradual [6-7]. The examples mentioned are still in the initial design or field-testing stages and are yet to be accepted for practical uses.

This study aims to enhance users' acceptance of this novel technology by examining its customer satisfaction from a multi-stakeholder perspective. Here, customers refer to the one who may pay for the construction robots and then use or interact with them. The findings reveal an in-depth recognition of various customer requirements and expectations, which is an important foundation for the design and evolution of construction robots. These findings also provide insights into facilitating the strategic prioritization of features and functionalities that hold the utmost significance for users.



Figure 1. Construction robot prototypes. (a) Unmanned excavator. (b) Polyarticulated concrete pouring robot. (c) Wall painting robot. (d) Demolish waste collection robot

2 Literature review

In accordance with the Technology Acceptance Model (TAM), customer satisfaction, regarded as an indicator of perceived usefulness and ease of use, is increasingly employed to measure the acceptance level of advanced technologies [8]. For instance, [9] conducted a study to identify influential factors that affect customer satisfaction with building information modeling (BIM). to evaluate the success of BIM implementation. To stimulate user's acceptance of BIM, [10] employed the entropy method to establish a quantitative model for the measurement of BIM user satisfaction. The significance of examining customer satisfaction to enhance technology acceptance has been underscored by existed research studies [11]. Therefore, this study intends to enhance users' acceptance of construction robots by comprehensively understanding customer satisfaction, focusing on mitigating critical constraint factors within construction works.

In particular, this study explores customer satisfaction from a multi-stakeholder perspective, aiming to consider diverse needs and expectations across roles in the construction industry. A proactive approach to addressing concerns and barriers from a broad perspective is also expected to explore potential collaboration patterns that promote the integration of robotic technology into the traditional construction industry.

3 Methodology

An in-depth literature review was initially conducted to identify these constraint factors, such as time, cost, and quality [12]. Subsequently, a survey was conducted among six distinct stakeholder groups. The invited experts were asked to score their satisfaction with implementing construction robots for mitigating constraint factors using a 1-5 Likert scale. The experts were selected using the Delphi method [13]. The reliability of questionnaire responses was verified using Cronbach's α . During the data analysis phase, stakeholder groups were first clustered into segments based on similar scoring patterns using the k-means algorithm. This segmentation facilitated a focused and targeted finding by refining observations from individual stakeholder groups. Further, factor scores within stakeholder segments were ranked using the mean score ranking to elucidate primary concerns and interests. The detailed research methodology and findings are explained below.

3.1 Identify constrain factors

A systematic literature review was conducted in this section to search for constraint factors. The constraints are the factors that prevent construction process from progressing smoothly. Scopus was chosen for the following reasons: 1) Scopus has frequently been employed in construction-related review studies [14]. 2) Scopus covers a broader range of disciplines, including engineering, construction, and management, and has the potential to index construction constraint management papers [15]. 3) Scopus contains more recent citations and provides the most recent research findings, which is critical for staying current in rapidly evolving fields [16]. To narrow the search scope of construction constraints, the following search keywords were used: "construction," "industry," and "constraint." The search string used was "article title, abstract, keywords." A total of 1925 papers with constraint analysis were chosen to identify constraint factors.

VOSviewer, a popular bibliometric analysis tool, was

employed to extract and classify constraint factors efficiently. The searched literature was specifically exported from Scopus into a ".csv" file, then imported into VOSviewer for keyword analysis. The VOSviewer's keywords co-occurrence analysis section was used to extract and summarize keywords from the 1925 papers automatically. Figure 2 depicts the visualization of keyword occurrences. As can be seen, a comprehensive visualization of constraint words, such as "cost," "time," and "quality," can be obtained.



Figure 2. Keywords co-occurrence map

From the keyword co-occurrence map, a total of ten constraint factors were identified. Table 1 lists and describes the identified factors. The factor "time" refers to time delays that affect the smooth construction process. The factor "cost" refers to budgetary constraints and cost overruns that affect project profitability. The factor "quality" refers to the issue of quality compromise, which causes safety concerns. The factor "environment" includes concerns for the environmental impact of construction activities. The factor "safety" is concerned with ensuring the well-being of construction workers in order to prevent accidents and injuries. The term "labour shortage" refers to the shortages of skilled workers. The factor "communication" involves difficulties in the insufficient exchange of information among project stakeholders, which is critical for coordinating activities. The factor "partnership" refers to concerns about collaboration and coordination among various entities caused by the fragmented and diverse nature of construction activities and a lack of trust in cooperation. The factor "production" refers to the problem of inefficient material and machinery production to meet project demands. The factor "logistics" refers to the problem of inefficient material, equipment, and personnel transportation, which is critical for the timely delivery of resources to the construction site and minimizing disruptions to the workflow.

ID Factor Description Schedule delay problem F1 Time Budgetary limitation and F2 Cost cost overruns Quality compromise F3 Quality problem Concerns related to the F4 Environment ecological impacts Hazardous caused by F5 Safety accidents and injuries Lacking availability of Labour F6 skilled workers shortage Communicat Inadequate exchange of F7 information ion Hesitations on the F8 Partnership collaboration and coordination Inefficient production of F9 Production materials and machinery Lacking efficient F10 Logistics movement and transportation

3.2 Score satisfaction

To investigate customer satisfaction with employing construction robots to alleviate the identified constraints, a questionnaire survey was conducted. In Hong Kong in 2023, an electronic questionnaire was generated using Google Form and distributed to invited experts via e-mail and mobile messaging with a web link. Using a five-point Likert scale, the experts were asked to express their professional views on whether they satisfied with using construction robots to alleviate the ten constraint factors (1 = strongly agree; 2 = agree; 3 = neutral; 4 = disagree; 5 = strongly disagree). The Likert scale method was chosen due to its ease of quantifying and rating the level of satisfaction with individual constrain factors based on the opinions of multiple stakeholders.

56 effective respondents were collected. The 56 responses can be considered sufficient because it has been proven that a median of 32.5 to 40 participants is acceptable in phenomenological studies [17]. The response rates of government departments, developers, consulting firms, main contractors, sub-contractors, and universities are 40%, 35%, 40%, 40%, 65%, and 60%, respectively.

The invited experts involve six stakeholder groups: consulting firms, government departments, main contractors, real estate developers, subcontractors, and universities. The distributions of the six stakeholder groups are shown in Figure 3. All the stakeholder groups are regarded as the prominent participants in the construction industry [18]. Only those with experience in

Table 1 Constraint factors and their description

construction technology and advanced construction technology, including construction robotics, were invited. The reliability and credibility of the research can be ensured because: 1) Most of them, especially the university professors, have both industry and academic experience. Survey participants with multiple backgrounds across organizations provide more valid responses [19]. 2) 82% of the experts held top positions in their organizations, such as senior or executive level (30%). 3) 82% of the experts have more than five years of working experience in the construction industry and grasp knowledge of construction robotics.



Figure 3. Distribution of investigated experts

3.3 Data analysis

3.3.1 Reliability testing

Cronbach's alpha coefficient was first measured to test the reliability of the collected response. Cronbach's alpha presents response reliability by measuring the internal consistency of the satisfaction scores on each constraint factor. Cronbach's alpha measures the degree of consistency on a standardized scale ranging from 0 to 1. A higher Cronbach's alpha indicates higher reliability because a participant is likelier to provide similar scores for the assessment items [20].

The responses were first imported into the SPSS software to compute the Cronbach's alpha index. A Cronbach's alpha value of 0.89 was obtained, indicating the acceptable reliability of the responses.

3.3.2 Stakeholder segmentation

Stakeholder groups with similar characteristics were segmented to investigate overall concerns and interests among different stakeholder groups, explore potential collaborative possibilities, and provide targeted solutions. K-means, a widely used unsupervised machine learning technique, was employed to do so. The k-means clustering algorithm divides a dataset into distinct and non-overlapping groups by assigning data points to clusters iteratively based on their proximity to each cluster's mean (centroid). The k-means algorithm has been widely used in customer segmentation due to its simplicity, versatility for clustering numerical data, and intuitive approach. For instance, [21] utilized a k-means algorithm to cluster customers to study how well a specific product performs in terms of marketing. The findings were proven more accurate by introducing the kmeans algorithm, an unsupervised learning-based processing method. Because of the scattered nature of the scoring data, this study employed the k-means algorithm to cluster the six stakeholder groups to find hidden satisfaction patterns.

Specifically, the initial cluster centroids were determined by randomly selecting several K points from the dataset. Second, each data point was assigned to the cluster centroid that was closest to it, and the distance between the data point and the assigned cluster centroid was calculated using Euclidean distance (see Equation (1)). Finally, the centroid points were updated, the data points were reassigned, and the distance was calculated again. Steps 2 and 3 were repeated until the sum of the distances (computed using Equation (2)) reached the minimum to determine the cluster centroid and the specific cluster of each point.

$$d(d_i, c_i) = \sqrt{(x_{di} - x_{ci})^2 + (y_{di} - y_{ci})^2}$$
(1)

$$dj = \sum_{i=1}^{n} d(d_i, c_i) \tag{2}$$

Here: $d(d_i, c_i)$ refers to the distance between the data and centroid points. d_i and c_i refer to the i - th data point and centroid point, respectively. x_{di} and x_{ci} refers to the i - th x coordinates of the data point and the centroid point. y_{di} and y_{di} refers to the i - th ycoordinates of the data point and the centroid point. djrefers to the sum of the distance. k means there are k - th centroid points in total.

To determine the optimal number of clusters, the elbow method was employed. First, the within-cluster sum of squared errors (SSE) for various centroid point values was calculated. When the SSE value first begins to decrease, the optimal number of centroid points appears.

The algorithm was coded using the Python language. The KMeans package of the sklearn.cluster library was employed to execute the k-means algorithm.

3.3.3 Mean score ranking

The significance of factors was determined using a widely used data statistic approach, the mean score ranking method [22]. Using Equation (3), the 1-5 Likert scale scores, representing customers' satisfaction with using construction robots to mitigate constraints, are averaged across all score sets. The lower the mean score, the higher the satisfaction because the scoring was presented on a scale of 1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, and 5 = strongly disagree.

$$\overline{X_F} = \frac{1}{n} \sum_{i=1}^{n} X_{iF} \tag{3}$$

Here $\overline{X_F}$ refers to the mean score for factor F. *n* refers to the number of customers in a stakeholder segment. X_{iF} refers to the score for factor F of each stakeholder.

4 Results

Figure 4 depicts the computed SSE values, which indicate the optimal number of clustering centers. As can be seen, when the number of cluster centers reaches 3, the elbow point appears. Before that, the SSE value drops dramatically from 200 to 13.5. After that, the SSE value steadily decreases from 13.5 to 0.05 as the cluster center number increases from 3 to 6, after which the SSE value remains constant. As a result, the optimal cluster center number was determined to be 3, the SSE curve's elbow point.



Figure 4. Visualization of SSE result

Figure 5 depicts the results of segmentation for six different stakeholder groups. The initial number of cluster centers was set to three based on the SSE computation result. The six stakeholder groups are represented by the x-coordinate in the following order: 0consulting, 1-main contractor, 2-developer, 3subcontractor, 4-government, and 5-university. The mean scores of the ten constraint factors are represented by the y-coordinate. The mean values were used as a representation of all observations [23] and were fed into the k-means algorithm instead of the initial scores. The blue dots represent each stakeholder's mean score, while the red triangle represents the cluster centers.



Figure 5. Stakeholder segmentation.

Based on their preferences, the six stakeholder groups were categorized into three groups. Segments 1, 2, and 3 consist of government and university, developer and subcontractor, consulting and main contractors, with (0.53, 0), (2.68, 2) cluster centers, respectively. The majority of the stakeholders are confident of utilizing construction robots to alleviate constraint factors, as indicated by the distribution of their mean scores, which range from 1 to 3. To facilitate an in-depth discussion of customer satisfaction, the distribution of each stakeholder's satisfaction score is detailed below.

5 Discussion

5.1 Satisfaction analysis among stakeholders

Figure 6 demonstrates stakeholders' satisfaction scores for each constraint factor, reflecting how satisfied stakeholders are when employing construction robots to mitigate constraints. Overall, the majority of the stakeholders show a neutral or positive attitude. The areas with satisfaction scores of 2 overlap. the most, with nearly 48% of the stakeholders agreeing that using construction robots can help improve constraints to some degree. It is also worthwhile to be aware of any potential acceptance hesitations. The overlap area of scores of 3 ranks second, with nearly 26% of stakeholders holding a neutral attitude toward the effect of construction robots. A general industry concern about embracing rapid technological changes when there is no apparent productivity or financial benefit may result in neutral attitudes [24]. A lack of familiarity and hands-on experience may also contribute to a cautious attitude [25]. While nearly 7% of stakeholders scored a 5, they strongly opposed the adoption of construction robots.



Figure 6. Satisfaction scores of stakeholder groups on constraint factors

It can be seen that stakeholder segment 2 (developer and subcontractor) is currently the most satisfied with construction robots. Almost all of them showed a satisfaction score of 2. It's interesting that experienced experts are more likely to accept construction robots. All of the stakeholders in the second segment are experienced experts who hold senior or executive-level positions in their organizations. Most of them have more than 20 years of experience in the construction industry. Senior-level stakeholders with extensive construction industry experience have most likely witnessed and adapted to various technological advancements. They may be more open to accepting and embracing construction robots because of a history of adapting to new technologies [26]. Senior and executive-level stakeholders are often critical in shaping the strategic direction. Their long-term perspective allows them to see construction robots' potential transformative impact.

Another pattern is that stakeholders more familiar with construction robotics are more likely to accept it. Construction robots are well-known to nearly 84.2% of stakeholders in the second segment. Customers familiar with construction robots are likely to have firsthand knowledge of or exposure to the technology [27]. These customers may have received education or training on using and benefitting from construction robots. This hands-on experience allows them to comprehend the practical benefits, operational capabilities, and potential benefits that construction robots bring to projects.

The satisfaction of experienced and familiar customers may indicate that the market is ready for more widespread adoption of construction robots. Positive feedback from these customers validates the practical benefits and dependability of construction robots, indicating that it is an appropriate moment to invest in the product further. The finding is consistent with [28], which claims that familiar and experienced customers indicate product credibility. If the product, including all the forms of construction robot products, such as mechanical arm, elements, and the services, currently meets the needs of experienced and familiar users, it may be worth investigating possibilities to diversify its applications or functionalities. It is critical to continue collecting feedback to improve continuously [29]. Furthermore, the findings highlight the importance of educating potential users on the benefits and capabilities of construction robots. Increased awareness can lead to increased user satisfaction [30].

5.2 Satisfaction analysis of constraint factors

Although stakeholders in the second segment have a positive attitude, some experts, particularly in the third stakeholder segment, have expressed their unwillingness to accept construction robots. The satisfaction scores of 4 and 5 appear 21 times, indicating that these stakeholders disagree or strongly disagree about the effectiveness of using construction robots to mitigate constraints. This section discusses the mean score ranking of satisfaction scores for each constraint factor provided by the stakeholders in segment 3 (government and university) to identify key concerns that impede the acceptance of construction robots and suggest improvement possibilities.



Figure 7. Mean score ranking of satisfaction scoring on each constraint factor

Figure 7 depicts the ranking of satisfaction scores on each constraint factor by mean score. It can be seen that the top two concerns are constraint factors F2 (cost) and F8 (partnership). The findings show that some customers believe there is no obvious benefit to using construction robots to reduce construction costs and increase partnership collaborations. While it may appear counterintuitive for government and universities to be concerned about the high overall cost of using construction robots because they do not directly lead construction projects. There are several reasons behind: 1) Government is subject to budget constraints. Concerns about high costs reflect a sense of responsibility to taxpayers as well as the need to demonstrate responsible use to prevent public resources from overspending [31]. 2) Universities frequently collaborate on research projects with industry partners. If the cost of implementing construction robots is perceived to be high, it may influence industry partners' willingness to engage in collaborative ventures, affecting research opportunities and industry-academic partnerships [32].

Customers in government and universities are dissatisfied with the performance of construction robots in improving partnership collaborations, in addition to the high cost. Building strong social networks and enhancing collaboration is critical for universities to apply for research funding and conduct research [33]. As a result, they expect the construction robots to be able to improve collaboration among various stakeholders to ensure a smooth construction process. Governments are in charge of the regulatory frameworks that govern industries. If construction robots cannot improve partnership collaboration, it may impede the creation of an environment conducive to the successful integration of all participants [34].

Cost-cutting strategies such as flexible leasing and financing options are suggested to alleviate customer concerns about high costs. Offering alternatives to outright purchase can alleviate financial burdens and make construction robot adoption more accessible [35]. Designing construction robots in a modular manner to facilitate adaptability and scalability has the potential to increase customer acceptance. Customers may begin with a smaller investment and scale up as needed, matching the cost to the specific needs of their projects [36].

Focusing on the concept of human-robot teaming, in which construction robots supplement rather than replace human skills, may help address the partnership collaboration issue [37]. It is also suggested that features such as natural language processing, gesture recognition, and real-time communication interfaces be included to improve collaboration and provide advanced communication capabilities to construction robots. Developing construction robots that can adapt to the dynamic and fragmented nature of construction workflows is also recommended. Using machine learning algorithms and artificial intelligence (AI) to enable robots to understand and respond to changes in tasks, schedules, and team dynamics promotes better coordination.

6 Conclusion

Due to the limited adoption of various prototypes, a comprehensive investigation into customer satisfaction and acceptance of construction robots has been conducted. To do so, stakeholders' satisfaction level on using construction robots to mitigate constraints, such as cost and time, were analyzed. Ten constraint factors were identified through a systematic literature review. Their satisfaction scores were given using a Likert scale via a questionnaire survey. To target satisfaction patterns among multi-stakeholders, six stakeholder groups with similar preference were divided into three segments using the k-means algorithm. The average score ranking of constraint factors revealed existing concerns and improvement possibilities. These findings contribute to our understanding of customer satisfaction and acceptance for the developed construction robot prototypes while also providing actionable detailed insights for further improving their functionality. However, the results are more specific to the development of construction robotics within Hong Kong. Although the construction industry in Hong Kong is representative on a global scale, further study intends to extend scopes by conducting investigations across various regions to achieve comprehensive insights.

References

- [1] Walzer, Alexander N., et al. Beyond googly eyes: stakeholder perceptions of robots in construction. *Construction Robotics*, 6.3 (2022): 221-237.
- [2] Heikkilä, R., et al. Development of an earthmoving machinery autonomous excavator development platform. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 1005-1010, Alberta, Canada, 2019.
- [3] Furet, B., et al. 3D printing for construction based on a complex wall of polymer-foam and concrete. *Additive Manufacturing*, 28: 58-64, 2019.
- [4] Wang, Y., et al. Intelligent spraying robot for building walls with mobility and perception. *Automation in Construction*, 139:104270, 2022.
- [5] Chen, Xinxing, et al. Robot for automatic waste sorting on construction sites. *Automation in Construction*, 141: 104387, 2022.
- [6] Chang, S. et al. Evolution pathways of robotic technologies and applications in construction. *Advanced Engineering Informatics*, 51, 101529, 2022.
- [7] Pan, W., Hu, R., Linner, T., & Bock, T. A methodological approach to implement on-site construction robotics and automation: a case of Hong Kong. Proceedings of 35th International Symposium on Automation and Robotics in Construction, 362–369, 2018.
- [8] Ramadhani, F., et al. Determinants of web-user satisfaction: using technology acceptance model. In *MATEC Web of Conferences*, pages 05009, West Sumatra, Indonesia, 2018.
- [9] Song, J., et al. Exploring the influence of system quality, information quality, and external service on BIM user satisfaction. *Journal of Management in Engineering*, 33(6), 04017036, 2017.
- [10] Zhang, H. M., Chong, H. Y., Zeng, Y., & Zhang, W. The effective mediating role of stakeholder management in the relationship between BIM

implementation and project performance. Engineering, Construction and Architectural Management, 30(6), 2503-2522, 2023.

- [11] Sun, Y., et al. What motivates people to pay for online sports streaming? An empirical evaluation of the revised technology acceptance model. *Frontiers in Psychology*, 12, 619314, 2021.
- [12] Tanko, B. L. et al. Stakeholders assessment of constraints to project delivery in the Nigerian construction industry. *International Journal of Built Environment and Sustainability*, 4(1). 2017.
- [13] Lund, S., et al. Applying stakeholder Delphi techniques for planning sustainable use of aquatic resources: experiences from upland China, India and Vietnam. *Sustainability of Water Quality and Ecology*, 3, 14-24, 2014.
- [14] Utama, W., et al. Review of research trend in international construction projects: a bibliometric analysis. *Construction Economics and Building*, 16(2), 71-82, 2016.
- [15] El-adaway, I. H., et al. Analytic overview of citation metrics in the civil engineering domain with focus on construction engineering and management specialty area and its subdisciplines. *Journal of construction engineering and management*, 145(10), 04019060, 2019.
- [16] Pranckutė, R. Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1), 12, 2021.
- [17] Saunders, M. N., et al. (2016). Reporting and justifying the number of interview participants in organization and workplace research. British Journal of Management, 27(4), 836-852.
- [18] Law, K. et al. Factors influencing adoption of construction robotics in Hong Kong's industry: A multistakeholder perspective. *Journal of Management in Engineering*, 38(2), 04021096, 2022.
- [19] Mena, S. et al. Input and output legitimacy of multistakeholder initiatives. *Business Ethics Quarterly*, 22(3), 527-556, 2012.
- [20] Wang, X., et al. When public participation in administration leads to trust: An empirical assessment of managers' perceptions. *Public* administration review, 67(2), 265-278, 2007.
- [21] Narayana, V. L., et al. Mall customer segmentation using machine learning. In 2022 International Conference on Electronics and Renewable Systems (ICEARS), pages 1280-1288, Tuticorin, India, 2022.
- [22] Whitehurst, D. G., et al. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Medical Decision Making*, 31(6), E34-E44, 2011.
- [23] He, Q., et al. Machine condition monitoring using principal component representations. *Mechanical*

Systems and Signal Processing, 23(2), 446-466, 2009.

- [24] Uher, T. et al. Risk management in the conceptual phase of a project. *International journal of project management*, 17(3), 161-169, 1999.
- [25] Hopcan, S. et al. Exploring the artificial intelligence anxiety and machine learning attitudes of teacher candidates. *Education and Information Technologies*, 1-21, 2023.
- [26] Turja, T., et al. Robot acceptance at work: a multilevel analysis based on 27 EU countries. *International Journal of Social Robotics*, 11(4), 679-689, 2019.
- [27] Wu, Y. H., et al. Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the Living Lab setting. *Clinical interventions in aging*, 801-811, 2014.
- [28] Kharouf, H., et al. A signaling theory approach to relationship recovery. *European Journal of Marketing*, 54(9), 2139-2170, 2020.
- [29] Luchs, M. G., et al. Perspective: A review of marketing research on product design with directions for future research. *Journal of Product Innovation Management*, 33(3), 320-341, 2016.
- [30] Eom, S. B., et al. The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sciences Journal of Innovative Education*, 4(2), 215-235, 2006.
- [31] Listokin, Y. et al. I like to pay taxes: Taxpayer support for government spending and the efficiency of the tax system. *Tax L. Rev.*, 66, 179, 2012.
- [32] Freitas, I. et al. University–industry collaboration and innovation in emergent and mature industries in new industrialized countries. *Research Policy*, 42(2), 443-453, 2013.
- [33] Huang, J. et al. Building Research Collaboration Networks--An Interpersonal Perspective for Research Capacity Building. *Journal of Research Administration*, 45(2), 89-112, 2014.
- [34] Haley, U. et al. Government policy and firm strategy in the solar photovoltaic industry. *California Management Review*, 54(1), 17-38, 2011.
- [35] Bertini, M., et al. The ends game: How smart companies stop selling products and start delivering value. *MIT Press*, 2020.
- [36] Shao, Y., et al. Mitigating investment risk using modular technologies. *Computers & Chemical Engineering*, 153, 107424, 2021.
- [37] Kim, Y., et al. Delegation or collaboration: Understanding different construction stakeholders' perceptions of robotization. *Journal of Management in Engineering*, 38(1), 04021084, 2022.
PESTEL Analysis of Factors Influencing the Demand and Supply of Modular Construction: Perspectives from Hong Kong

Yidan Zhang¹, Wei Pan¹ and Mi Pan²

¹Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
²Department of Civil and Environmental Engineering, University of Macau, Macau, China ydzhang@connect.hku.hk, wpan@hku.hk, mipan@um.edu.mo

Abstract -

Modular construction (MC) shifts labourintensive site-based activities to an off-site controlled environment through prefabrication, and is deemed an innovative approach for industry transformation. To maximise the benefits of MC, there is a need to predict the demand volume and supply capacity to inform proactive and strategic supply chain planning and development. This study aims to provide a holistic exploration of the factors influencing the MC market, based on a literature review and a study of the case of modular integrated construction (MiC) in Hong Kong. A systematic framework was proposed with 15 demand-related factors and 10 supply-related factors to explain the complex influential mechanism from the political, economic, social, technological, environmental, and legal (PESTEL) perspectives, which contribute to the understanding of the dynamic evolution of the MiC market. The findings revealed that the influencing factors are manifold, with the political factors influencing the MiC demand the most the economic factors determining and the development pattern of the MiC supply. The presented study lays a foundation for the establishment of a practical MC demand-and-supply model for better market planning. Attention should be devoted to rationalising the quantification of qualitative factors in future research.

Keywords -

Modular Construction; Demand and Supply; Market Dynamics; Influencing Factors; PESTEL Analysis

1 Introduction

Housing shortage has plagued many metropolitan areas. To meet the growing demand for housing and satisfy citizens' wishes towards better living standards, modular construction (MC) has gained increasing attention for its various benefits. Pan and Zhang [1] revealed that both concrete and steel modular systems outperform their conventional counterparts, reflected by heightened environmental sustainability (e.g. 46-87% reduction in waste disposal), improved social harmony (e.g. no records of accidents in modular works), and significantly improved economic efficiency (e.g. 32-50% increase in speed of superstructure construction). Owing to the enhanced sustainability of the modular approach, the governments of many countries and regions have actively advocated its adoption through various terms, such as modular integrated construction (MiC) in Hong Kong [2, 3], prefabricated prefinished volumetric construction (PPVC) in Singapore [4], and permanent modular construction (PMC) in the United States [5].

Nevertheless, due to the infancy nature of MC and its incomplete supply chain in many areas, the prospects of the MC market are full of uncertainties in the form of demand variability and supply capability [6]. The lack of anticipation of the MC market can be fatal to both the demand and supply sides. The imbalance of MC demand and supply can lead to the bankruptcy of factories due to insufficient orders, or unfulfilled MC demand and high prices due to the lack of supplier alternatives [7]. In reality, modular firms have witnessed some publicised failures, such as Urban Splash and Katerra [7, 8]. Despite the importance of MC market analysis, previous research solely focused on the demand side, such as exploring the drivers, constraints, and strategies for MC adoption [4, 9], while neglecting the upstream supply side. Few papers displayed a sense of supply chain management and provided a broad analysis of the evolving MC demand and supply. To optimise mass production, it is imperative to have a clear vision of the MC market to support the formulation of proactive government strategies and corporate plans for MC development. Therefore, there is an urgent need to understand the factors influencing MC market dynamics, which have not yet been fully explored.

To address the research gap, this study aims to explore the factors influencing MC demand and supply

systematically through a literature review and a study of the case of MiC in Hong Kong. MiC encompasses all aspects of MC with the highest degree of integration of diverse construction activities [9]. Subsequently, Section 2 introduces the research methodology. Then, the influencing factors of the MiC demand and supply are structured into political, economic, social, technological, environmental and legal (PESTEL) aspects and delineated in Section 3. Section 4 discusses the findings of the study, followed by the conclusions in Section 5.

2 Methodology

The research was carried out in four main stages to analyse the factors influencing the MC demand and supply (see Figure 1).

The first stage was factor identification through a literature review. The reviewed literature covered construction demand prediction, housing supply modelling, diffusion of innovation, and MC adoptionrelated research. Factors were extracted through content analysis and those with similar meanings were merged.

The second stage was factor verification using a case study. As the relevant literature was limited and the influencing factors were context-sensitive, the case of MiC in Hong Kong and its supply chain in the Greater Bay Area (GBA) of China was used to supplement and adjust the factors [11]. The GBA is a mega-region that consists of nine major cities in Guangdong Province and two special administrative regions - Hong Kong and Macao. The case study was conducted using two interrelated methods: factory and site visits and interviews. Six group interviews were conducted during or after the visits with the managers and engineers in a semi-structured manner. All the respondents had over five years of working experience with MiC. Each interview lasted around 30 minutes to collect their basic and intuitive views on the topic. Table 1 presents the research activities conducted for the study. Data collected from the interviews were documented. Then, the factors were extracted from the records and synthesised.



Figure 1 Research stages and methods

Factory visit	Supplier 1	Supplier 2	Supplier 3
MiC supplier			
Product type	Concrete MiC	Steel MiC	Concrete and steel MiC
Location	Guangzhou, Guangdong	Foshan, Guangdong	Zhuhai, Guangdong
Follow-up interview	3 suppliers	2 suppliers	3 suppliers
Site visit	Project 1 (mockup)	Project 2 (mockup)	Project 3
Construction sites			
Project type	Public housing	Private residential building	Affordable housing
Location	Hong Kong	Hong Kong	Shenzhen, Guangdong
Follow-up interview	1 client and 2 contractors	1 client and 1 contractor	1 contractor and 2 from the industry institution

Table 1 Details of the case study

The third stage employed the PESTEL analysis. PESTEL stands for political, economic, social, technological, environmental, and legal, offering a holistic guiding tool for scrutinising a wide spectrum of 'climate' elements for marketing planning [12]. Political factors refer to government intervention in the MiC market. Economic factors can significantly affect MiC market sentiment due to the cost-conscious culture in construction. Social factors mainly concern labour conditions, health consciousness and cultural aspects. Technological factors focus on MiC itself and the changes in the technological landscape. Environmental factors reflect the sustainable concerns and geographical constraints in practice. Legal factors focus on the regulatory environment for MiC, such as the approval process and design standards. The identified factors were divided into the demand and supply sides, and further grouped based on the PESTEL perspectives.

The fourth stage proposed a PESTEL framework of the factors influencing the MiC demand and supply, generating valuable theoretical and practical implications for market forecasting and supply chain configuration.

3 PESTEL analysis of factors influencing the demand and supply of MC

The influencing factors identified from the construction engineering and management literature and the case study are summarised from the six PESTEL aspects in Table 2 and elaborated below.

3.1 Political perspective

Political factors refer to government intervention in the MiC market, taking the forms of tax policy, financial incentives, and mandatory requirements [13]. The current MiC demand in Hong Kong is primarily driven by government policies. Since 2017 when MiC was first included in the Chief Executive's Policy Address 2017 [2], a series of policies have been put in place to facilitate MiC development, reflecting a strong and pivotal push from the change agent. The most important political factor emphasised by the interviewees is the policy incentives for MiC implementation, such as the 10% gross floor area concession incentives and the Construction Innovation and Technology Fund (CITF) for MiC promotion. In addition to the financial incentives, the government has also formulated various policy initiatives to stimulate MiC adoption in the public sector. During the COVID-19 pandemic, MiC was promoted for the construction of modular quarantine camps due to its well-recognised advantages of fast delivery [14]. In 2020, the Development Bureau announced that government building projects with total construction floor area larger than 300m² under the Capital Works Programme are

mandated to adopt MiC [15]. The Policy Address 2022 advocated for MiC adoption in public housing and set an ambitious adoption rate of no less than 50% from 2028/29 to 2032/33 [16]. These policies have sparked investments in MiC in the public sector, evidenced by the increasing number of modular buildings commenced, while MiC adoption in the private sector is still trailing.

As for the MiC supply in the GBA, the national and local governments have issued relevant policy measures to augment the supply of prefabricated buildings, of which MiC is a kind of product with the highest degree of prefabrication and integration but not the mainstream one. In 2017, Shenzhen was included in the first batch of national prefabricated building demonstration cities, and in 2020, Foshan earned selection in the second batch [17, 18]. As of 2022, 22 companies have been selected as the national prefabricated industrial bases and 83 companies (78 in the GBA and 5 in the other cities of Guangdong) have been selected as the provincial ones [17-19]. In addition, the government announced that prefabricated buildings would account for more than 50% of new buildings in Guangdong by 2030 [19]. Noticeably, the interviewee from the industry institution indicated that the boom in prefabricated buildings as well as the emergence of MiC in the GBA would increase factories' supply capacity. This may be accompanied by a decrease in the proportion allocated to Hong Kong, but a high possibility for an increase in the absolute quantity of MiC supplied to Hong Kong. Overall, the government focused on leveraging the leading role of demonstration cities and industrial bases and the pull effects of the market demand to enhance the supply capability of prefabricated products including MiC, highlighting the whole supply chain synergy.

3.2 Economic perspective

Economic factors are the determining factors that affect the MiC market sentiment in the context of a costconscious culture in construction [9]. From the demand side, the perceived relative advantages of MiC in the economic aspect (e.g. reduced construction time and cost, improved quality, and enhanced productivity) are one of the most important driving forces for adopting this innovative approach [6, 9]. Besides, according to the law of supply and demand, the MiC demand will increase along with the improvement of MiC manufacturers' supply capability and its resulting cost-effectiveness. Another important factor is the construction demand in the various building sectors and the favourable economic environment for construction. As construction demand increases, there is a likelihood of a concurrent increase in the MiC demand. Factors may influence the construction demand include population and household, gross domestic products, interest rate, labour employment, and housing stock and price [20-23, 31].

Category	Factors from the demand side	Factors from the supply side
Political	• Economic incentives for MiC promotion (e.g. provision	Policy support for MiC
perspective	of innovation fund) [9, 13, 23, 29, 30, 33, 36, E]	manufacturers [E]
	• Policy initiatives for MiC promotion (e.g. setting targets	
	for MiC adoption rate) [9, 13, 26, 33, E]	
Economic	• Relative advantages in the economic aspect (e.g. time-	• MiC market demand [9,
perspective	and cost-efficiency, better quality performance) [6, 9,	22, E]
	25-28, 33, 34]	• Profit earning [E]
	• Supply capability of MiC manufacturers [6, 9, 23, 24,	Company strategic
	28-30, 33, 36, E]	planning [E]
	• Construction demand [29, 32, E]	• Market limitations (e.g.
	• Economic environment (e.g. interest rate) [20-22, 31]	finance) [32]
Social	• Relative advantages in the social aspect (e.g. improved	• Labour conditions in the
perspective	health, safety, and welfare) [9, 25]	other GBA cities (e.g.
	• Labour conditions in Hong Kong (e.g. labour size, skill	labour size, skill level)
	level) [9, 20-23, 29, 30, 31, 33, E]	[9, 20-23, 29, 31, 33]
	• End-user attitudes towards MiC [6, 9, 23, 26, 28, 30, E]	
Technological	• Complexities of MiC implementation [4, 6, 9, 23-25, 28,	Adoption of innovative
perspective	30, 36]	technologies in MiC [4,
	• Compatibility of MiC for different building sectors [25,	9, 23, EJ
	28, E]	• Sufficient eligible MiC
	• Successful MiC demonstration projects [9, E]	suppliers [6, 9, 23, 24,
Б. (28-30, 36, E]
Environment	• Relative advantages in the environmental aspect (e.g.	• Achievement of
perspective	nigher material utilisation) [9, 23, 27]	sustainable goals [20]
	• Geographical constraints on MiC implementation [4, 9, 23, 24]	
Legal	• Regulatory readiness for MiC implementation (e.g.	• Uniform standards and
perspective	improved approval procedures) [6, 9, 13, 23-26, 28-30.	norms applicable to both
r	36. El	Hong Kong and the
	/ 1	other GBA cities [E]

Table 2 PESTEL analysis of the key factors influencing MiC demand and supply

Note: E represents empirical evidence obtained from the case study.

Hwang et al. [4], Feldmann et al. [6], Pan et al. [9], Jin et al. [13], Goh [20], Fan et al. [21], Jiang and Liu [22], Rahman [23], Choi et al. [24], Rogers [25], Jiang et al. [26], Choi et al. [27], Pan et al. [28], Mao et al. [29], Wuni and Shen [30], Kim et al. [31], MacAskill et al. [32], Pan et al. [33], Shin et al. [34], Katebi et al. [35], Bello et al. [36], Pan et al. [37]

The development pattern of the MiC supply is determined by economic considerations. The stable market demand for MiC and manufacturers' capability of profit earning positively influence the strategic planning of the MiC suppliers [6]. Since MiC places special emphasis on 'integration', more high-value construction activities were shifted from sites to upstream factories [3], attracting great interest in increasing MiC supply capacity to grab more profits. Besides, the MiC supply is expected to expand in response to the growing demand for MiC after seeing the burgeoning MiC market in Hong Kong, exemplified by several major MiC suppliers (e.g. Suppliers 1 and 3) who have new factories in planning. In contrast, market limitations (e.g. availability of finance, regional space, planning and construction capacities) may constrain the expansion of MiC factories [32].

3.3 Social perspective

Social factors mainly concern labour conditions, health consciousness and cultural aspects that affect the MiC market. With regard to the MiC demand, the severe labour ageing and shortage in Hong Kong have spurred the industry to find a way out, and MiC's advantages in saving site labour have made the industry explore this innovative approach [9]. However, as MiC involves many challenging works (e.g. heavy crane operations), the lack of skilled workers with MiC experience may hinder the widespread of MiC [35]. In addition to the better economic performance of MiC, the perceived relative advantages in terms of the social aspect (e.g. improved health, safety and welfare for workers) are also one of the driving forces for MiC promotion. Besides, the interviewees emphasised the importance of end-user attitudes towards MiC, because positive feedback from end-users can provide more confidence to developers and drive them to implement MiC in more projects. However, due to the current market preference for customized design, concrete structural systems, and conventional construction approaches, raising public awareness of MiC, such as through organising propagation campaigns, becomes a fundamental strategy for changing people's perspectives and thus increasing the take-ups of MiC.

Regarding the MiC supply, the major social barrier remains the labour issue. While the labour shortage in the other GBA cities may not be as acute as that in Hong Kong, the problem of ageing workers has become increasingly prominent [17]. Besides, the MiC suppliers argued that there is a shortage of mechanical, electrical and plumbing technicians with Hong Kong licenses to work in the factories in the Mainland, negatively influencing the supply capacity and production cycles. As such, the provision of regular training for the workers either in Hong Kong or in the other GBA cities is necessary for the rapid development of MiC.

3.4 Technological perspective

Technological factors focus on MiC itself and the impacts of changes in the technological landscape on the MiC market. As for the MiC demand, the complexity of MiC implementation (e.g. early design freeze, just-intime delivery, heavy-duty machine operations) is the prominent technological consideration, since it may amplify the stakeholders' concerns about adopting MiC, thereby affecting the stakeholders' top management and investment in MiC [4, 30, 34]. Compatibility refers to the degree to which MiC is perceived to be consistent with the values, experiences, and needs of the potential adopters [17]. Normally, experiences in prefabricated construction should help MiC diffusion as these knowledge and skills are transferable to the MiC field. Whereas, since prefabricated construction has been extensively used in public housing construction and has demonstrated acceptable time and cost performance, part of the stakeholders display reluctance to embrace change and to invest in new equipment and worker training [29]. Besides, the compatibility of MiC in different building sectors varies. For example, as MiC benefits can be maximised through standardisation, public housing, elderly homes and staff quarters with repeated design were regarded as most suitable for MiC [28]. In contrast, as indicated by one of the manufacturers in the interview, the ongoing private residential building project in the factory is highly customised with no repeated module designs on a typical floor. This feature poses a challenge to the compression of production cost due to the need for more module moulds, leading to market hesitance to use MiC in the private sector. In addition, successful MiC demonstration projects allow potential adopters to

observe the MiC benefits and see the innovative approach being used by their peers [17].

Technological factors that influence the MiC supply were related to the integration of MiC and other innovative technologies (e.g. building information model, artificial intelligence, robotics), which has the potential to enhance operational efficiency and significantly elevate supply capacity. As observed in the factories, Suppliers 2 and 3 have introduced robotic arms to improve production. The other factor that influences the MiC supply is the availability of eligible MiC suppliers [6, 9, 23]. The respondent from the industry institution indicated that many manufacturers have extensive experience in prefabricated construction and are capable of, or intend to, shifting to producing MiC. However, the contractors are concerned about the availability of concrete MiC suppliers, because some markets such as housing prefer concrete MiC systems that are more technically demanding and can be manufactured by fewer suppliers [29].

3.5 Environmental perspective

Environmental factors reflect the sustainable considerations and geographical constraints in MiC implementation. Regarding MiC demand, the perceived relative advantages in the environmental aspect (e.g. higher material utilisation, reduced impacts of noise, traffic and dust on the local community) can improve stakeholder and community satisfaction, thus increasing public acceptance of MiC adoption [1]. Whereas, the narrow streets and undulating terrain in Hong Kong impose restraints on module size, which may limit their utility and applicability across various scenarios and thus affect the widespread adoption of MiC.

The environmental motivation for increasing MiC supply is from the need for sustainable development. As the construction industry is often characterised by high investments, elevated levels of waste and pollution, and diminished productivity, moving labour-intensive site works to a controlled manufacturing environment is regarded as a lean practice for achieving carbon neutrality. Under sustainable-driven urbanisation, the advantages of MiC may gain more policy support [26].

3.6 Legal perspective

Legal factors refer to the laws and regulations, but this study focuses on a suitable regulatory framework for MiC practice. From the MiC demand side, both the extant literature and the interviews revealed that the overstringent regulations and red tape in approval procedures pose a hindrance to the full realisation of MiC advantages, thereby impeding the formulation of market-driven MiC demand [9, 13]. According to an interviewee from a MiC manufacturer, the modules and connections are always over-designed given the current strict regulations and the conservative review approach adopted by the governments, leading to cost escalation and waste of materials. In addition, the approval procedures of various government departments are time-consuming, bringing great uncertainties to process control [24].

Regarding MiC supply, the absence of standards MiC comprehensive and codes for implementation leads to inefficient process operations and a lack of accountability structures [9]. Besides, due to the different regulations applied in Hong Kong and the other GBA cities (i.e. Hong Kong mainly follows the British standard system, but factories in Mainland China adopt the Chinese national standard), extra time and cost are spent on design, material sourcing and accreditation problems, hindering the efficient MiC supply.

4 Discussion

This study unveiled a variety of factors influencing MiC demand and supply from PESTEL perspectives, based on a comprehensive literature review and empirical evidence from Hong Kong. The identified influencing factors are manifold, which demonstrates the need for systems thinking in addressing the complicated MiC market issues. Therefore, a systematic framework is proposed for analysing the macro environment in terms of PESTEL aspects that influence MiC demand and supply in Hong Kong and the GBA of China (Figure 2).

The framework deconstructs the MiC market into the interacting demand and supply sides and elaborates on the PESTEL factors influencing these two sides. The demand-related factors mainly focus on the drivers that promote, and the constraints that inhibit, MiC adoption and the supply-related factors revolve around the supply capacity and the potential growth of MiC suppliers. The findings revealed the emphasis on political factors, especially in the early stage of MiC adoption. This is aligned with the findings by Pan et al. [9] and Mao et al. [29], which showed the current demand for MC as government-driven and the need for policy incentives to drive MC adoption. The rationale behind the emphasis on political factors is partly attributable to cost, as stakeholders expect more economic incentives to offset cost premiums [30]. The MiC supply evolves in tandem with the changes in MiC demand and is governed by technological and regulatory considerations. While Feldmann et al. [6] highlighted the significance of the availability of suitable suppliers, they did not regard regulation as a determinant of module supply in Germany. However, the pivotal role of regulatory factors has been recognized by many studies conducted within the Chinese context, e.g. [29]. In addition, the characteristics of different building sectors should be embedded in the analysis of the influencing factors, given the suitability of MiC for different building sectors. However, this study refrains from conducting an in-depth analysis of different building sectors due to page limits of the paper.

The proposed framework enhances the knowledge base of MiC market planning and its supporting supply chain design. Although some of the influencing factors were based on the context of MiC demand in Hong Kong and the supply in the GBA of China, they should provide valuable insights for other regions that rely on a crossboundary supply chain. Furthermore, other regions could build upon the factors listed in Table 2 and incorporate region-specific considerations into the framework to enhance its applicability and effectiveness.



Figure 2 A systematic framework of factors influencing the demand and supply of MiC in Hong Kong and the GBA of China

Nevertheless, the presented work is a fundamental step to establishing a simulation model to understand the evolution of the MiC market. Several improvements for future research have been identified. First, it is meaningful to analyse the interrelationships between the critical factors and the pathways through which the factors influence MiC demand and supply. Second, the analysis of the factors should further incorporate the characteristics of different building sectors for a more accurate vision of the sector-level MiC market. Third, a sophisticated model employing advanced simulation techniques should be developed to accurately forecast future market trends and dynamics in the short and long terms. As some influencing factors were challenging to quantify, this might make future efforts to develop a practical simulation model difficult. Coupled with detailed scenarios that fit the societal movement, tailored policy recommendations should be generated to guide the sustainable development of MiC.

5 Conclusions

MC has been experiencing a global revival recently due to its various benefits, but how this momentum will develop further remains unclear. Building on the results of the literature review and a study of the case of Hong Kong, a holistic framework was proposed with a total of 15 demand-related factors and 10 supply-related factors to explain the complex influential mechanism from the PESTEL perspectives.

The findings revealed that the influencing factors of the MiC market are multifaceted, with the political factors most influencing the MiC demand and the economic factors determining the development pattern of the MiC supply. Also, there remains ample room for mitigating the constraints imposed by technological and legal factors on the MiC supply, e.g. increasing the availability of concrete MiC suppliers, and improving the standards and codes for MiC design and implementation. Noticeably, other influencing factors are not secondary. For example, social factors play an increasingly important role in the move towards a people-centred Industry 5.0 era. In addition, attention must be paid to rationalising the quantification of the qualitative factors.

The developed framework contributes to a better understanding of the dynamic evolution of the MiC market. The revealed factors help policymakers develop strategies to underpin the sustainable development of MiC and support corporates to proactively respond to market dynamics.

Being aware that the presented study is yet far from delivering a practical model to simulate future MiC demand and supply, future research is needed to explore the critical influencing factors, demonstrate the interrelationships and the evolution of these influencing factors, simulate the dynamic evolution of the MiC market, and conduct detailed scenario analysis to help formulate appropriate strategies for improvement.

Acknowledgements

This work was supported by the Strategic Public Policy Research Funding Scheme (Project No.: S2019.A8.013.19S) of The Government of the Hong Kong Special Administrative Region.

References

- [1] Pan W. and Zhang Z. Benchmarking the sustainability of concrete and steel modular construction for buildings in urban development. *Sustainable Cities and Society*, 90, 2023.
- [2] HKSAR. The Chief Executive's 2017 Policy Address
 We Connect for Hope and Happiness. The Hong Kong Special Administrative Region. 2017.
- [3] Pan W. and Hon C.K. Briefing: Modular integrated construction for high-rise buildings. *Proceedings of* the Institution of Civil Engineers - Municipal Engineer, 173(2):64-68, 2020.
- [4] Hwang B.-G., Shan M., and Looi K.-Y. Key constraints and mitigation strategies for prefabricated prefinished volumetric construction. *Journal of Cleaner Production*, 183:183-193, 2018.
- [5] Smith R.E. and Rice T. Permanent Modular Construction: Process, Practice, Performance. University of Utah, 2015.
- [6] Feldmann F.G., Birkel H., and Hartmann E., Exploring barriers towards modular construction – A developer perspective using fuzzy DEMATEL. *Journal of Cleaner Production*, 367, 2022.
- [7] Green S.D. Modern methods of construction: reflections on the current research agenda. *Buildings and Cities*, 3(1):653-662, 2022.
- [8] Pullen T., Hall D., and Lessing J. A Preliminary Overview of Emerging Trends for Industrialized Construction in the United States. ETH Zürich, Zürich, Switzerland, 2019.
- [9] Pan W., Yang Y., and Pan M. Implementing modular integrated construction in high-rise high-density cities: perspectives in Hong Kong. *Building Research & Information*, 1-15, 2022.
- [10] Yin R.K. *Case study research: Design and methods* (5th ed.). Sage, Thousand Oaks, CA, 2014.
- [11] Aguilar F.J. Scanning the business environment. MacMilan Co., New York, 1967.
- [12] Jin X., Ekanayake E.M.A.C., and Shen G.Q.P. Critical policy drivers for modular integrated Construction projects in Hong Kong. *Building Research & Information*, 1-18, 2021.
- [13] Zhang Z. and Pan W. Fighting Covid-19 through fast

delivery of a modular quarantine camp with smart construction. *Proceedings of the Institution of Civil Engineers-Civil Engineering*, 174(2):89-96, 2021.

- [14] DEVB. Technical Circular (Works) No. 2/2020 Modular Integrated Construction (MiC). Development Bureau of HKSAR, Hong Kong, 2020.
- [15] HKSAR. The Chief Executive's 2022 Policy Address
 Charting a Brighter Tomorrow for Hong Kong. Hong Kong Special Administrative Region Government, Hong Kong, 2022.
- [16]NBS. *China Statistical Yearbook 2022*. National Bureau of Statistics of China, China Statistics Press, China, 2022.
- [17] MOHURD, Letter on identifying the first batch of prefabricated construction demonstration cities and industrial bases, The Ministry of Housing and Urban Rural Development, 2017.
- [18] MOHURD, Letter on identifying the second batch of prefabricated construction demonstration cities and industrial bases, The Ministry of Housing and Urban Rural Development, 2022.
- [19] GPDOHURD et al. Implementation Opinions on Accelerating the Development of New Building Industrialization, Guangdong Provincial Department of Housing and Urban-Rural Development and other Departments, 2022.
- [20] Goh B.H. Forecasting residential construction demand in Singapore: a comparative study of the accuracy of time series, regression and artificial neural network techniques. *Engineering, Construction and Architectural Management*, 5(3):261-275, 1998.
- [21] Fan R.Y.C., Ng S.T., and Wong J.M.W. Predicting construction market growth for urban metropolis: An econometric analysis. *Habitat International*, 35(2):167-174, 2011.
- [22] Jiang H. and Liu C.L. A panel vector error correction approach to forecasting demand in regional construction markets. *Construction Management and Economics*, 32(12):1205-1221, 2014.
- [23] Rahman M.M. Barriers of implementing modern methods of construction. *Journal of Management in Engineering*, 30(1):69-77, 2014.
- [24] Choi J.O., Chen X.B., and Kim T.W. Opportunities and challenges of modular methods in dense urban environment. *International Journal of Construction Management*, 19(2):93-105, 2017.
- [25] Rogers E.M. Diffusion of Innovations. The Free Press, New York, 1995.
- [26] Jiang, R., Mao C., Hou L., Wu C. and Tan J. A SWOT analysis for promoting off-site construction under the backdrop of China's new urbanisation. *Journal of Cleaner Production*, 173:225-234, 2018.
- [27] Choi J.O., O'Connor J.T., Kwak Y.H. and Shrestha B.K. Modularization business case analysis model

for industrial projects. *Journal of Management in Engineering*, 35(3), 2019.

- [28] Pan W., Yang Y., Zhang Z., and Chan S. Modularisation for Modernisation: A Strategy Paper Rethinking Hong Kong Construction. Development Bureau of HKSAR and The University of Hong Kong, HKSAR, China, 2019.
- [29] Mao C., Shen Q. Pan W., and Ye K., Major barriers to off-site construction: The developer's perspective in China. Journal of Management in Engineering, 31(3), 2015.
- [30] 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, 2020.
- [31] Kim K.-B., Cho J.-H., and Kim S.-B. Model-based dynamic forecasting for residential construction market demand: A systemic approach. *Applied Sciences*, 11(8), 2021.
- [32] MacAskill S., Mostafa S., Stewart R. A., Sahin O. and Suprun E. Offsite construction supply chain strategies for matching affordable rental housing demand: A system dynamics approach. *Sustainable Cities and Society*, 73, 2021.
- [33] Pan W., Wu J., Xie M., Zhang Y., Pan M., and Yang Y. Modular Integrated Construction Supply Chain in the Greater Bay Area for Hong Kong Development: Drivers, Opportunities, Constraints, Concerns, Measures and Strategies. The University of Hong Kong, HKSAR, China, 2021.
- [34] Shin J., Moon S., Cho B.-h., Hwang S. and Choi B. Extended technology acceptance model to explain the mechanism of modular construction adoption. *Journal of Cleaner Production*, 342:130963, 2022.
- [35] Katebi A., Homami P., and Najmeddin M. Acceptance model of precast concrete components in building construction based on Technology Acceptance Model (TAM) and Technology, Organization, and Environment (TOE) framework. *Journal of Building Engineering*, 45, 2022.
- [36] Bello A.O., Khan A.A., Idris A. and Awwal H.M. Barriers to modular construction systems implementation in developing countries' architecture, engineering and construction industry. *Engineering, Construction and Architectural Management*, ahead-of-print, 2023.
- [37] Pan W., Pan M., and Yang Y. A dialectical system framework for enhancing modular construction supply chain. *Proceedings of the Institution of Civil Engineers: Engineering Sustainability*, Ahead of Print, 2023.

Digital Transformation, Sustainability and Construction 5.0

Ming Shan Ng¹, Thomas Bock², Fabian Kastner³ and Silke Langenberg³

¹Center for the Possible Futures, Kyoto Institute of Technology, Japan ²Department of Architecture, Technical University of Munich, Germany ³3Institute for Preservation and Construction History, ETH Zurich, Switzerland

ng@kit.ac.jp, bockrobotics@web.de, kastner@arch.ethz.ch langenberg@arch.ethz.ch

Abstract -

The United Nations has set the 2030 Agenda for Sustainable Development for the society, the environment and the economy with the 17 Sustainable Development Goals (SDGs). Countries act in collaborative partnership to take transformative steps to shift industries and society onto a sustainable and resilient path. Among all, digital transformation is one key domain of transformation to achieve sustainability. In the architecture, engineering, construction and operations (AECO) sector, the industry is slowly adopting digital transformation to different extents such as the use of Build-ing Information Modelling (BIM), Internet of Things (IoT), computational tools, as well as automation in design and construction. Despite the commonly known practices and benefits of digital transformation such as productivity increase, the impacts of such transformation on sustainability have not yet been fully examined in research. The industry entails the corresponding digital transformation practices and their relationships with sustainability, so as to consider appropriate strategies. This work first investigates how the current practice in AECO adopts digital transformation, with case studies of the Japanese industry as examples; this is followed by the study of the relationships between digital transformation practices and the SDGs based on existing literature. Hence, the strategy propositions to assist the industry in current practice are elaborated. The research contributes to science by taking an initial step to examine the relationships between digital transformation and sustainability and present readyto-adopt strategy propositions. Future research includes indepth validations of the strategies and comparisons of the approaches in Japan with other countries.

Keywords -

Digital Transformation; Sustainability; SDGs; Construction 5.0

1 Introduction

1.1 Sustainability and Digital Transformation

The 2030 Agenda for Sustainable Development set by the United Nations for the three dimensions of sustainable development - society (people), the environment (planet) and the economy (prosperity) - for all countries and stakeholders to implement this plan and take transformative steps in collaborative partnership. The goals are to transform the industries and society and balance the three dimensions. In regards to the society, the concept of sustainability comprises human well-being and a healthy environment; regarding the environment, sustainability is related to sustainable consumption, production and management in response to climate change and the needs of the present and future generations; in regards to the economy, sustainability supports the economic, social and technological progress of the nature and ecosystems [1]. Furthermore, in the context of sustainability in organizations, industries, and society, Elkington [2] presents the concept of the Triple Bottom Line, which comprises the environmental bottom line, the economic bottom line and the social bottom line.

The United Nations proposes the 17 Sustainable Development Goals (SDGs) for countries to achieve sustainability in societies and industries, including the architecture, engineering, construction and operations (AECO) sector, and ensure the implementation of the 2030 Agenda. United Nations [1] defines the 17 SDGs and actions by 2030 as follows: SDG 1 ensures the mobilization of resources to end poverty in multiple ways; SGD 2 safeguards food security and sustainable food production to increase productivity and maintain ecosystems; SDG 3 strengthens healthiness and well-being with affordable access to healthcare; SDG 4 ensures inclusive quality education and access to technology for education; SDG 5 enhances gender equality through policy adoption and technologies; SDG 6 reduces pollution and maintains water-related ecosystems; SDG 7 ensures infrastructure for access to clean energy research and technology; SDG 8 promotes productive employment and a safe working environment; SDG 9 enhances industrialization and fosters innovation through technological capabilities; SDG 10 reduces inequality via regulation; SDG 11 ensures inclusive cities, housing, transport systems and urbanization, as well as the protection of cultural and natural heritage; SDG 12 ensures sustainable consumption, production and management of natural resources with innovation; SDG 13 strengthens resilience and adaptive capacity to climate-related hazards; SDG 14 conserves oceans, seas and marine resources; SDG 15 restores sustainable use of terrestrial and inland ecosystems, manages forests and land degradation; SDG 16 promotes inclusive societies and participatory decisionmaking; SDG 17 revitalizes the partnership in finance, technology and capability-building.

The World in 2050 [3] consolidates six key domains

of transformation for countries to provide pathways to achieve the 17 SDGs in ways that can be managed, in correspondence to appropriate infrastructure such as governance, values and policy tools. They are human capacity and demography, consumption and production, decarbonization and energy, food, biosphere and water, smart cities and digital transformation. Digital transformation can be defined as an "inter-connected set of discoveries and inventions" [3], as well as a form of business-model innovation or operational advancement through digitallyenabled provisions [4]. It involves emerging technologies such as robotics, artificial intelligence (AI), computer vision and remote sensing smart systems. It reshapes work, education and governance, as well as the production processes in every sector of the economy including the AECO sector. It can benefit productivity, accessibility, cost-effectiveness and material consumption to different extents. It could reduce lower-skilled jobs, incur inequality and bring challenges such as cybersecurity concerns, threats to the privacy of identities, digitally stored information, as well as social media manipulation etc. Moreover, digital transformation practices comprise people, technology and process to different extents [4, 5]. In general, it entails comprehensive governance to ensure that its impacts on our society can lead to sustainability for the next generation. Governance of technologies, AI and autonomous technical systems should align with the directions of the human-centered society so as to comprehend the potential benefits, liabilities and risks of the practices while avoiding the many potential downsides [3, 6]. Scholars such as Sachs et al. [7] and Vinuesa et al. [8] conducted literature reviews to study the relationships between digital transformation and each SDG and how the adoption of AI enables and inhibits each SDG respectively.

1.2 Digital Transformation in AECO



Figure 1. Robot-Oriented Design (ROD) first applications in Japan: ROD SMART Juroku Ginko in Nagoya in (a) 1992 and (b) 1993; (c) ROD SMART proto Shimizu Research Institute (SRI) in 1988; (d) ROD SMAS Building Research Institute (BRI) MOC in 1985-1988.

The AECO sector in the construction industry is one of the world's least digitized among others with fragmented design and construction processes and supply chains. The fragmentation provides little incentive to embrace systemic innovation for developing new methods and digital tools and implementing new inventions. In AECO, the success of a transformation can be influenced by how technologies enable new ways of working, new products and services, as well as operational models to different extents [4]. State-of-the-art digitally-enabled provisions on projects include the implementation of Building Information Modelling (BIM), computational design, digital fabrication technologies including robotics and additive manufacturing such as 3D printing, as well as BIM-based and computer vision construction site monitoring, construction process digital twins and management etc [9]. Also, reality gaming technology realizes modular construction and digital twins of robotic construction respectively [10, 11]. BIM in design and construction enables the exchange of project information at the right time to make the right decision to facilitate environmental sustainability considerations during a building design process [12]. On the process level, digital transformation adopts digitally-enabled industrialized construction methods such as Robot-Oriented Design (ROD) [13]. The notion of ROD enables immediate availability by rapid assembly, high accuracy by robotic precision motion planning and positioning to adjust and fix the workpieces, adaptability by robotic deconstruction, repair and upgrading, reconstruction, reassembly and reconfiguration. This thus achieves closed-loop construction for the highest sustainability by robotic reuse of any building, construction component or processes. Figure 1 shows the first applications of ROD in Japan. The Solid Material Assembly System (SMAS) verified this approach between 1984 and 1988 at the Building Research Institute (BRI) in Tsukuba. In 1988 the ROD concept was tested at the prototype set up of the SMART system at the Shimizu Research Institute (SRI) in Etchujima, Koto Ku and Tokyo. The research was first published in Japanese in 1988 by the first author of [13] with the analysis of the first 50 construction robots built by the Japanese general contractors from 1978 to 1988. ROD has been implemented at the first SMART site of the Juroku Ginko in Nagoya for a 20-storied office building in 1992/ 3 and various high-rise projects until 2019. Moreover, a BIM-based process enables integrated processes and the adoption of emerging technologies in organizations and fosters innovation in technology [14]. In addition, digital transformation is emerging on the urban scale in the built environment. Scholars such as Biljecki and Ito [15] and Ruhlandt et al. [16] study computer vision applications for Geographic Information Systems (GIS) research blockchain-enabled smart city governance respectively.

1.3 Society 5.0 and Construction 5.0

In Japan, the government promotes the national vision -Society 5.0 - to link science, technology and innovation efforts to SDGs and revolutionize social problems to achieve SDGs using AI and automation technologies. Society 5.0 comprises a "human-centered system that integrates cyberspace and physical spaces. [...] It is envisioned as a successor of transformations" [17]. It includes the transdisciplinary concepts of digitalization, integration and sustainability [5]. The Society 5.0 initiative in Japan requires interdisciplinary studies on multiple scales to comprehend and analyse the policy and legal opportunities of adopting AI and automation [6]. The overarching research can be included in the context of Construction 5.0, which comprises digitalization, integration and sustainability based on recent scholarship that studies new approaches of digital transformation in construction [18, 19].

1.4 Research Objectives and Research Methods

Despite the potential contributions of digital transformation to sustainability in the industry in AECO, it has not yet been fully researched how it can be implemented to foster sustainability. In Japan, the industry has been adopting digital technologies to different extents in AECO. Through implementing the appropriate digital transformation practices, stakeholders can achieve sustainability for society, environment and economy through processes, technologies and people to different extents. From an overarching perspective, digital transformation practices can help the industry to achieve the 17 SDGs and realize Construction 5.0. More in-depth research is needed to examine the corresponding digital transformation practices and their relationships with sustainability, as well as the potential strategy proposition. This work aims to answer the research question as follows: What digital transformation practices are related to sustainability in AECO? To answer the question, the following research objectives are addressed: (RO1) to investigate how the current practice in AECO adopts digital transformation; (RO2) to investigate relationships between digital transformation practices and the SDGs; (RO3) to elaborate on potential strategy propositions to assist the industry in current practice, so as to realize Construction 5.0.

To address RO1, this work identifies the appropriate digital transformation practices with examples from case studies in the Japanese AECO. The first author collected the data of the case studies through reviews of authority documents [20], firms' documents, presentations and websites, as well as semi-structured interviews with firms' digital transformation teams. The data have been cross-referenced, summarised and consolidated in this work. To address RO2, this work maps the identified digital transfor-

mation practices with the corresponding 17 SDGs respectively through a summary and review of existing literature presented in Sachs et al. [7] and Vinuesa et al. [8]. The results have been firstly presented in Ng [19]. Based on the findings from case studies and existing literature, the strategy propositions are presented in the discussion section to address RO3. The research contributes to the body of knowledge in the fields of automation in construction, sustainability, design, construction and operation management, as well as technology governance by taking an initial step to examine the relationships between digital transformation and sustainability and elaborate on ready-to-adopt strategy propositions.

This work is structured as follows. Section 1 presents the backgrounds of the research as the points of departure, the key research question, three research objectives and the research methods. Section 2 includes the research findings of the identified digital transformation practices in AECO with examples from the Japanese industry, as well as the relationship between digital transformation and sustainability, namely SDGs respectively. Section 3 discusses the strategy proposition as recommendations for stakeholders to adopt digital transformation in current practice. Section 4 concludes the work and proposes future research directions.

2 Research Findings

2.1 Digital Transformation Practices in AECO

Based on a thorough summary of existing research review from The World in 2050 [3], [20], [8] and [7], 27 digital transformation practices are identified as shown in Table 1. Under the categories of *people*, *technology* and *process*, how AECO can adopt these practices in current practice through case studies of the Japanese industry is further studied. The initial results can be found in the report Ng [19].

DXP1 - Invest in digital technologies and infrastructure can impact the economy through technology. It can influence microeconomics within a firm and/or macroeconomics for a country to invest in scientific institutions. Many companies have been investing in state-of-the-art technologies such as digital twin interactive platforms to improve process and building performances during design, construction and operations stages. For example, digital twin platforms investment (DXP1) facilitates DXP7 during the construction process, and DXP13 through information integration. Also, BIM-based digital twins can support DXP6 such as the use of blockchain and smart contracts. DXP2 - Promote collaboration and joint task development can improve society through people development. This practice can involve collaboration among institutions, companies and public-private partnerships. On Table 1. The identified 27 digital transformation practices in AECO under the major categories of *people (Pe), technology (Te)* and *process (Pr),* as well as *economy (Ec), society (So)* and *environment (En).* [19].

No. Digital Transformation Practice	Category 1	Category 2
DXP1 Invest in digital technologies and infrastructure	Te	Ec
DXP2 Promote collaboration and joint task development	Pe	So
DXP3 Conduct predictions to make better plans of action	Te	Ec, En
DXP4 Optimize resources in design and planning	Pr	Ec, En
DXP5 Foster policy development for more efficient actions	Pr	Ec, So
DXP6 Support affordable and trusted documentation	Te	Ec, So
DXP7 Thorough waste management	Te, Pr	En
DXP8 Design for human health and social well-being	Pe	En
DXP9 Adopt digital analytical tools to improve performances	Te	Ec
DXP10 Enable cloud-based reality environment for training purposes	Pe, Te	So
DXP11 Optimize environmental and heritage values of local communi	ities Pe	Ec, So, En
DXP12 Automate decision-making	Pr	Ec
DXP13 Optimize resources in consumption through data monitoring	Te	En
DXP14 Extend accessibility and improve revenue	Pr	Ec
DXP15 Facilitate R&D to foster innovation and adoption of technolog	y Te	Ec
DXP16 Scale creativity and innovation	Te	Ec
DXP17 Increase the demand for jobs related to automation	Pe	Ec, So
DXP18 Enforce effective financial risk assessment	Pr	Ec
DXP19 Foster circularization and enable circular economy	Pr	Ec, En
DXP20 Optimize procurement process and supply-chain management	Pe, Pr	Ec, En
DXP21 Responsible sourcing of resources	Pr	En
DXP22 Improve financial transparency and simplify transaction proces	ss Pr	Ec
DXP23 Enhance public and stakeholders engagements	Pe	So
DXP24 Develop specific digital tools to support sustainable practices	Te	En
DXP25 Optimize organizational structure	Pe	So
DXP26 Customize design tools and processes to optimize values	Te, Pr	En
DXP27 Provide cost-effective solutions	Pr	Ec

one hand, companies can collaborate with universities, institutions and organisations to jointly work on cutting-edge research and development; on the other hand, universities and educational institutes can collaborate with industry partners to promote industry-oriented educational activities or curricula to help educate the next-generation AECO professionals. DXP3 - Conduct predictions to make better plans of action can benefit the economy and the environment through technological development. This practice includes climate change forecasts using urban and geographic information such as urban building fabrics. In Japan, BIM-based simulation platforms are used to create virtual reality models to conduct predictions for assisting stakeholders to make better plans in design and construction. Also, a steel contractor uses customized computational tools can help to simulate robotic welding processes in 3D modelling and visualization for process prediction and real-time monitoring. DXP4 - Optimize resources in design and planning contributes to the economy and the environment through process development. This practice includes the optimization of design and resilience of infrastructure, buildings and construction, as well as the promotion of resource efficiency and the use of renewable materials based on informed decision-making already in the early design phase. In current practice, BIM-based digital platforms facilitate data integration and multi-disciplinary collaboration from the design phase to the building operation phase. When used correctly, BIM-based design processes and management can foster other practices such as DXP7 and DXP9, as well as DXP6, DXP12 and DXP13 to assist management during the construction and operations phases. DXP5 - Foster policy development for more efficient actions can benefit the economy and society through process improvement. For example, the Japanese government has been initiating the Society 5.0 framework

for efficient actions to achieve the 2030 Agenda and the SDGs in industries and society [17]. Collaboration between academics, industry practitioners and policy-makers (DXP2) can ensure appropriate policies and policy deployment that can deliver effective benefits to society. DXP6 -Support affordable and trusted documentation can impact the economy and society through technological development. This practice can assist stakeholders in construction process management and establish a common data environment with paperless recording to support trusted documentation. Also, BIM-based platforms also facilitate affordable and trusted real-time documentation among multi-parties on a project. DXP7 - Thorough waste management benefits the environment through technology and process development. In AECO; companies adopt this practice by assigning waste management specialists on projects to conduct evaluations of site wastes and waste disposal strategies. They monitor the proper collection of by-products generated on construction sites for recycling. Also, integrated BIM-based platforms can document material information and foster DX19. Moreover, the current practice has demonstrated that BIM-based design for manufacture and assembly (DfMA) processes reduce wastes at construction sites and save costs to different extents during the design and construction processes. DXP8 - Design for human health and social well-being supports the environment through the development of people. Some companies in Japan use digital systems, which are connected to a company's internal human resources systems, to detect and record the body temperature of workers with their masks and helmets on to enhance human health at construction sites during the pandemic period. DXP9 - Adopt digital analytical tools to improve performances improves the economy through technological development. The use of, for example, digital twin interactive systems help to improve performances in building design and construction in terms of efficiency and cost-effectiveness. This practice can foster other practices such as DX7 and DX13.

DXP10 - Enable cloud-based reality environment for training purposes can benefit society through people and technological developments. This practice can impact workforce development in society. In current practice, state-of-the-art digital twin platforms enable a cloud-based reality environment to help train complex fabrication processes. This not only enhances workers' capability, encourages the young generation to take part in the construction sector, but also assists pre-job site health and safety training. DXP11 - Optimize environmental and heritage values of local communities can impact the economy, society and environment through development about people. In Japan, contractors who specialize in cultural heritage retrofits using traditional fabrication methods such as prefabrication of carpentry work, state-of-the-art industrialized construction methods and computational design tools to optimize the heritage values of the historic built environment. Built heritage, which possesses historical values and cultural significance including but not limited to embodied historical craftsmanship and construction knowledge, enhances the cohesion of the local communities. DXP12 -Automate decision-making supports the economy through process development. In the current practice in Japan, contractors and sub-contractors invent their bespoke customized digital tools to facilitate simulations, building information analyses and visualization and other practices such as DXP9 and DXP26. This can automate decisionmaking to optimize design and construction processes. DXP13 - Optimize resources in consumption through data monitoring can benefit the environment through technological development. State-of-the-art digital twins platforms enable consumption-aware processes through realtime monitoring. This also facilitates DXP7 during construction and operations. Data monitoring can be further foster other practices such as DXP6 and DXP19. DXP14 -Extend accessibility and improve revenue can impact the economy through process development. It is commonly acknowledged in the industry that digital transformation can increase productivity and accessibility, and reduce production costs and material consumption [3]. Recent research such as Grüter et al. [21] shows how to scale up the competitiveness of circular construction through the use of digital building design optimization tools, reflecting a link between DXP13, DXP14 and DXP19. Also, recent scholarship studies the actor-network study of how digital design and digital fabrication can be implemented to potentially general profits in the current industry through a systemic literature review [22]. Both DXP15 - Facilitate research and development (R&D) to foster innovation and adoption of technology and DXP16 - Scale creativity and innovation can benefit the economy through the development of technology. This practice can be implemented with other practices such as DXP2 to promote collaboration and joint task development and spur innovation among organizations in the industry. DXP17 - Increase the demand for jobs related to automation facilitates the development of the economy and society through developing people-related aspects and workforce. In Japan, trade contractors have been adopting automated design tools and digitized life-cycle management systems to automate value chains from design and planning to fabrication phases. They increase the job opportunities for high-skilled professionals to facilitate automation processes. DXP18 -Enforce effective financial risk assessment can impact the economy through process development. In Japan, customized electronic commerce services have been currently used in practice for users to order, receive and make payments by scanning QR codes displayed on a terminal.

Digital services can help to improve financial transparency simplify transaction processes and enforce effective financial risk assessment for procurement.

DXP19 - Foster circularization and enable circular economy can impact on the economy and the environment through process development. In AECO, this practice promotes a circular economic model of reuse and recycling of existing resources. In current practice, for example, Japanese house-makers use BIM-based integrated systems for pre-design planning, design, construction, operations and maintenance and end-of-life management to manage resources throughout the building life cycle and facilitate material circularization. This practice not only saves materials and costs but also enables design to reduce the amount of building components for ease of disassembly and reassembly. DXP20 - Optimize procurement process and supply-chain management can benefit the economy and the environment through people and process developments. A Japanese house-maker has been transforming their "Scrap and Build" approach to a "Stock-type Society" approach by constructing houses with a longer lifespan. With appropriate design approaches for manufacture and assembly, as well as design approaches for disassembly, BIM-based platforms can help to facilitate DfMA to consider constructability during the early design phase and optimize the procurement process and supply chain management. DXP21 - Responsible sourcing of resources supports the environment through process development. In current practice, for example, a Japanese general contractor uses locally sourced cedar timber materials to fabricate cross-laminated timber (CLT) load-bearing walls and floors with digital fabrication technology through information integration in BIM. DXP22 - Improve financial transparency and simplify transaction process can benefit the economy through process development. A Japanese general contractor has been using a construction site cocreation platform, which provides electronic commerce services with financial transparency, This can simplify the transaction process and reduce financial risk for procurement to different extents. Also, a Japanese housemaker uses blockchain technology to ensure the security of trusted information exchange and financial transparency and simplify the transaction process with an effective endto-end solution for services ranging from the real estate lease agreement process to establishing insurance contracts. DXP23 - Enhance public and stakeholders engagements and DXP25 - Optimize organizational structure can impact on the society through development about people. State-of-the-art mixed reality technologies have been implemented to assist design-to-construction processes for off-site progress inspection remotely in an interactive and immersive reality environment. This can enhance stakeholders' engagement for design review and discussion during the early design phase. The Japanese house-makers optimize organizational structure with vertical integration and advance training to promote responsible procurement and co-creation for sustainable development through partnership and collaborative work. DXP24 - Develop specific digital tools to support sustainable practices supports the environment through technological development. On existing projects, building information analysis, simulation and visualization using state-of-the-art analytical computational design tools to optimize building performance such as energy consumption, daylight intake etc. during the early design phase. DXP26 - Customize design tools and processes to optimize values can impact the environment through process and technological developments. Customized tools can improve, for example, carbon emission during digital fabrication processes by precise calculation and design of structural and non-structural parts of the building components. DXP27 - Provide cost-effective solutions can benefit the economy through process development. For example, a Japanese supplier uses commercial BIM platforms for DfMA with QR-coded packaging to assist logistics, fabrication and on-site assembly process and reduce wastes on site (DXP7).

2.2 Digital Transformation Practices for 17 SDGs

Figure 2 shows the categorizations of the 27 digital transformation practices and the mapping of the unidirectional relationships from each of the 27 practices to the 17 SDGs to different extents through a summary of the literature review work mainly from [3], Sachs et al. [7], Vinuesa et al. [8] and other industry reports. The diagram and the work have been initially presented in Ng [19].

Among all, DXP3 and DXP4 can achieve ten SDGs. DXP3 can achieve, for example, SDG 13.1 target -Strengthen resilience and adaptive capacity to climaterelated hazards and natural disasters; DXP4 can address several targets such as SDG 9.1 target - Develop quality, reliable, sustainable and resilient infrastructure to support economic development and human well-being. Also, the practice can achieve SDG 8.4 target - Improve progressively. This facilitates global resource efficiency in consumption and production and endeavours to decouple economic growth from environmental degradation. Also, DXP13 can achieve eight SDGs; DXP1 and DXP7 are found to achieve seven and six SDGs respectively. Besides, it is identified that SDG 11 - Sustainable cities and communities and SDG 12 - Responsible consumption and production can be enabled by 11 practices. For example, DXP15 can achieve SDG 11; SDG 12 can also be enabled through DXP19; SDG 8 - Decent work and economic growth and SDG 9 - Industry, innovation and infrastructure can be enabled by eight practices. They can be enabled by, for example, DXP11. Also, SDG 17 - Partnership for the Goals can be enabled by seven practices including DXP23.

3 Discussion

This work presents an initial study of digital transformation and the corresponding impacts on sustainability in AECO. In current practice, existing adoptions of digitalization in planning, design, construction, operations and maintenance, as well as management, can foster sustainable development to different extents on project scale, building scale, urban scale, in the built environment and beyond the AECO sector. Construction 5.0 includes digitalization, integration and sustainability. It leads to a high-level concept of environmental, societal and economic advancements in industries and society, such as the Society 5.0 concept in Japan. Based on the research, five potential strategy propositions in Construction 5.0 are presented as follows. They can be implemented immediately by stakeholders on projects, organizational reforms and in the built environment. Also, they can lead to potential future research directions, which include a systemic literature review, in-depth investigations and case-based validation of the digital transformation practices and the relationships with SDGs.

- All-rounded solutions that integrate digital technologies development, organizational, information and process integration, as well as sustainability, are needed to shape Construction 5.0.
- Joint collaborations between research and practical implementation can be conducted on projects in the forms of, for example, integrated digital project de-livery business models, early contractor involvement in concurrent engineering, as well as rewards sharing mechanisms.
- State-of-the-art BIM-based approaches for sustainability, including resource circularization and circular economy business models, process and cost predictions for better estimations and optimizations in processes and building performances can be promoted.
- Process-based transition management in the AECO sector holds the potential for combined achievements in the key areas of a resource-efficient and decarbonized built environment.
- Appropriate business models are required in practice to invest and facilitate R&D, scale creativity and turn inventions into innovations, enforce effective financial risk assessments, extend accessibility of technologies, as well as provide cost-effective solutions.



Figure 2. A tripartite diagram presenting the digital transformation practices in the AECO sector for achieving the 17 SDGs under the holistic improvement framework of people, process and technology - and the three pillars of sustainability - economy, society and environment [19]. Data are based on The World in 2050 [3] and Vinuesa et al. [8]. SDG color courtesy of UN/SDG.

4 Conclusion

This work investigates digital transformation to achieve sustainability and the 2030 Agenda for Sustainable Development for society, the environment and the economy. To assist the industry in adopting digitalization in current practice, 27 digital transformation practices are identified. They include DXP4 - Optimize resources in design and planning and DXP19 - Foster circularization and enable circular economy that foster the economy and the environment, DXP6 - Support affordable and trusted documentation and DXP17 - Increase the demand for jobs related to automation that foster the economy and society, DX23 - Enhance public and stakeholders engagements that benefit the society, as well as DXP26 - Customize design tools and processes to optimize values to improve the environment. All practices are supported with the case studies of the Japanese industry to show how the current practice has been adopting digitalization to different extents and demonstrate the feasibility. For example, Japanese house-makers use BIM-based integrated systems to manage resources throughout the building life cycle to facilitate DXP19. Moreover, the relationships of how the 27 practices can foster the 17 SDGs respectively are investigated. The results show that DXP3 and DXP4 can

achieve ten SDGs including SDG 13.1 target - Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters. Also, SDG 11 - Sustainable cities and communities and SDG 12 - Responsible consumption and production can be enabled by 11 practices such as DX19 and DXP21 - Responsible sourcing of resources. Moreover, five strategy propositions in Construction 5.0 as future research directions that can be further investigated and validated are presented. They include processbased transition management for combined achievements to enable a resource-efficient and decarbonized built environment. This research presents an initial study of connecting tangible digital transformation practices that have been adopted in Japan as examples and explores novel research directions, which can be continued by scientists and practitioners to establish comprehensive concepts of Construction 5.0 for the coming ten years approaching 2030 and beyond. Also, the work aims to assist stakeholders in taking action in correspondence to the 2030 Agenda, achieving SDGs and realizing Society 5.0 through technological advancements to solve societal, environmental and economic challenges through appropriate strategies for process, technologiical and people.related developments. Future research includes in-depth quantitative and

qualitative validations of the strategies and thorough comparative studies of Japan's approaches with other countries for broader insights.

References

- United Nations. Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. Technical report, 2015.
- J. Elkington. Cannibals with forks : the triple bottom line of 21st century business. Capstone, 1997. ISBN 190096127X.
- [3] The World in 2050. Transformations to Achieve the Sustainable Development Goals. Report prepared by The World in 2050 initiative. Technical report, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 2018.
- [4] J. Koeleman, M.J. Ribeirinho, D. Rockhill, E. Sjödin, and G. Strube. Decoding digital transformation in construction. Technical report, 2019.
- [5] M.S. Ng. Design for Digital Fabrication (DfDFAB): management for a sustainable adoption of emerging technologies and innovations. PhD thesis, ETH Zurich, Zurich, 2022.
- [6] M.S. Ng, D.M. Hall, and S.-H. Hsieh. Liability factors and conceptual framework for contracts to manage design for digital fabrication in construction projects. J. Legal Affairs and Dispute Resolution, 2022.
- [7] J.D. Sachs, ..., and J. Rockström. Six Transformations to achieve the Sustainable Development Goals. *Nature Sustainability*, 2(9):805–814, 9 2019.
- [8] R. Vinuesa, ..., and F. Fuso Nerini. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1): 1–10, 2020.
- [9] T. Bock. The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in Construction*, 59:113– 121, 11 2015.
- [10] K.S.D. Ravi, M.S. Ng, J. Medina Ibáñez, and D.M. Hall. Real-time Digital Twin of Robotic construction processes in Mixed Reality. In *Proc. of the 38th ISARC*, pages 451–458. IAARC, 2021.
- [11] Ali Ezzeddine and Borja García de Soto. Connecting teams in modular construction projects using game

engine technology. *Automation in Construction*, 132: 103887, 2021.

- [12] M.A. Zanni, R. Soetanto, and K. Ruikar. Towards a BIM-enabled sustainable building design process: roles, responsibilities, and requirements. *Arch. Eng. and Design Mgmt.*, 13(2):101–129, 2017.
- [13] T. Bock and T. Linner. *Robot-Oriented Design*. Cambridge University Press, New York, 2015.
- [14] M.S Ng, Q. Chen, D.M. Hall, J. Hackl, and B.T. Adey. Designing for digital fabrication: an empirical study of industry needs, percived benefits and strategies for adoption. J. Mgmt. in Eng., 2022.
- [15] F. Biljecki and K. Ito. Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, 11 2021.
- [16] R.W.S. Ruhlandt, R. Levitt, R. Jain, and D. Hall. Drivers of Data and Analytics Utilization within (Smart) Cities: A Multimethod Approach. J. Mgmt. in Eng., 36(2):1–19, 2020.
- [17] JST (Japan Science and Technology). Mobilizing Science, Technology and Innovation for SDGs Japanese Actions in STI for SDGs. Technical report, 2021.
- [18] Farook Hamzeh, Vicente A. González, Luis F. Alarcon, and Salam Khalife. Lean Construction 4.0: Exploring the challenges of development in the AEC industry. In *IGLC 2021*, pages 207–216, 2021.
- [19] M.S. Ng. Designing for digital transformation to achieve the SDGs with examples from the Japanese construction industry. In *The Sustainable Development Goals in Context: SDG Blog*, pages 90–90. ETH Zurich, Zurich, 3 edition, 6 2022. URL https://wp-prd.let.ethz.ch/ sdgblog2022/chapter/dx-sdg-japan/.
- [20] JFCC (Japan Federal of Construction Contractors). Construction DX case studies. Technical report, 2022.
- [21] C. Grüter, M. Gordon, M. Muster, F. Kastner, P. Grönquist, A. Frangi, S. Langenberg, and C. De Wolf. Design for and from disassembly with timber elements: strategies based on two case studies from Switzerland. *Frontiers in Built Environment*, 2023.
- [22] M.S. Ng, D.M. Hall, M. Schmailzl, T. Linner, and T. Bock. Identifying enablers and relational ontology networks in design for digital fabrication: a systematic literature review. *Automation in Construction*, 2022.

A Safety Framework to Assess Autonomous Construction and Mining Equipment

Cynthia Brosque¹ and Bibhrajit Halder²

¹Ph.D., Product Manager SafeAI Inc., CA, United States ²Ph.D., Founder and CEO SafeAI Inc., CA, United States

cynthiab@safeai.ai, bibhrajit@safeai.ai

Abstract -

Artificial intelligence is set to transform the mining and construction industries by providing greater insights that will eventually create a safer, more productive, and more reliable environment. However, integrating autonomous technology and equipment in the field is still a complex task that necessitates a detailed safety study, analysis, identification, and mitigation of hazards. Before any autonomous operation can be realized, a safety plan needs to be executed by the technology provider and the site operator and/or subcontractors. This plan must be regularly assessed during the development and implementation phases of the technology on site. As the industry evolves to incorporate more autonomous systems, having a comprehensive and consistent safety framework to assess this technology becomes more relevant for innovators in the field. The industry can learn and adapt the analyses developed for other automation uses such as aviation, automobile, nuclear, and defense systems to consider not only the safety of technology, but also the interfaces with human operators and the impact of process changes. The purpose of this paper is to provide an introduction to a safety framework and workflow developed and followed by SafeAI for the application of its autonomous technology in construction and mining. This framework is applied across our global deployments. For example, we highlight its application for our California proving grounds.

Keywords -

Autonomous Ground Vehicle; Safety; Construction; Mining; Hazard Analysis; STPA.

1 Introduction

Increasing demand for safer, zero-entry worksites, higher productivity, and reduced costs are driving the need for innovative solutions across heavy industries, such as mining and construction.

The vision for an autonomous site of the future to achieve these goals involves a higher level of automation. However, the existing Safety frameworks in the industry do not apply the learnings on automation hazard analyses deployed in other fields such as aviation, automobile, nuclear, and defense systems. These industries consider automation as a system, including the interfaces with humans, processes, and change management. As equipment automation strategies evolve in construction and mining applications, so should the approaches for evaluating the hazards and safety of the technology.

SafeAI is a technology start-up based in Santa Clara, California. SafeAI's primary focus is to deliver safe, autonomous vehicle technology and solutions to heavy industry operators through robust computer and perceptionbased technologies, such as sensor fusion and deep learning, as well as cutting-edge modular and reconfigurable robotics software behavior frameworks.

SafeAI's autonomous solution retrofits existing construction and mining ground vehicles. The manufacturer and vehicle-agnostic solution utilizes a reusable hardware package that can be easily integrated with a broad range of vehicle types. The AI-powered autonomous software is developed to manage and operate vehicles autonomously in the toughest off-road environments (Figure 1).

2 Prior Work

Safety is integral to the development and deployment of autonomous technology at construction and mine sites. Hazard analysis techniques have been widely adopted by transportation industries where safety is critical as the first step to assess risk, i.e., investigating an incident before it occurs [1].

Common hazard analysis techniques involve Fault Tree Analysis (FTA), Event Tree Analysis (ETA), and Hazard and Operability Analysis (HAZOP), as well as their variants. FTA [1] focuses on understanding the logic leading to a top undesired event. It assists in designing a system or as a diagnostics tool and was originally developed by the aviation sector in the U.S [2]. Similarly, ETA condenses the FTA to make it more manageable to study complex designs such as nuclear power plants, chemical plants, and spacecraft analysis [3]. HAZOP is often used as a technique to identify operability hazards that can lead to product, environment, or other hazards broken down per module [4].



Figure 1. SafeAI Proving grounds

Failure Modes and Effects Analysis (FMEA) is also used as a bottom-up hazard analysis technique. This method is useful to analyze hardware failures of components like sensors, according to the information provided by their corresponding supplier. The failure modes from each component are associated with a severity scale based on its failure effect, probability of occurrence, and detection scale. However, this method is noted to have limited applicability for safety analysis at a Systems Level [5].

On the other hand, the System Theoretic Process Analysis (STPA) (Figure 2) is a relatively new hazard analysis technique developed by MIT based on an extended model of accident causation [6]. STPA [5] can be used at any stage of the system life cycle and when STPA is performed, it is assumed that the system design exists.

STPA [5] offers advantages over other conventional bottom-up safety analysis techniques that have been applied to the safety analysis of partially- or fully automated driving systems. STPA considers the unsafe interactions of system components by human interaction, software, etc. This allows an analysis of complex systems such as those found in aviation, spacecraft, automobile, nuclear, and defense systems.

STPA first identifies the potential for inadequate control

of the system that could lead to a hazardous state. A hazardous state is defined as one that violates the norms, rules, or constraints of the system. The method studies how each Unsafe Control Action (UCA) could occur and plans a safety requirement to mitigate or avoid the unsafe action. As the product design evolves, the safety requirements also get more detailed in an iterative process. Complete traceability is established between the requirements and the system deployment, which helps maintain the Quality Assurance (QA) process.

3 Method

Given the fact that automation in construction and mining entails a complex system review, SafeAI has integrated a framework to address a "safety first" approach which begins early in the design phase through to the deployment phase. During the design phase, the flexibility to make changes is at its highest while the cost is at its lowest (Figure 3).

Our Safety Framework is comprised of three main components, with the STPA method as the main foundation for hazard analysis. As far as the authors are aware, STPA has not been consistently applied to analyze Mining or





Figure 2. Safety Analysis timeline and type of study

 STPA

 Define
 Define
 Hizard
 Unsafe
 Causal

 Qaese
 ODD
 1/SA
 Actions
 Factor
 Safety Regs
 Verification

 Validation
 Validation
 Actions
 Actions
 Actions
 PAS

 Revise
 Revise
 Actions
 Actions
 Actions
 Pass

 Not Acceptable
 Management
 risks are
 acceptable
 Not Acceptable

 Component
 Maitunction
 Hazard
 A.B. C.
 Functional

 Definition
 Analysis
 Asit
 Safety

Figure 4. SafeAI Safety Framework

Figure 3. Flexibility to change design to incorporate safety assessment

Construction automation strategies. Unlike the traditional hazard analysis methods, this framework can be deployed in early technology development to define safety requirements which can mitigate hazards that would only arise during the operation phase. As the conceptual design is refined and more detailed design decisions are made, STPA maintains complete traceability of the system requirements including the human-in-the-loop analysis.

In addition to STPA, SafeAI implements the bottomup approach established by FMEA (Figure 4) to study component-level flaws and failures. Finally, it applies a Layer of Protection Analysis (LOPA) focusing on specific mitigation strategies for the risks identified during the Hazard Analysis phase.

This paper describes SafeAI Safety Framework for autonomous mining and construction equipment and shows its application in SafeAI's proving grounds quarry site in California.

4 SafeAI Safety Framework

This section addresses in more detail each step of SafeAI's Safety Framework:

1. STPA

2. FMEA

3. LOPA

4.1 STPA

The STPA risk assessment process follows the ISO 12100:2010 guidelines to:

(a) determine the limits of the machinery, which include the intended use and any reasonably foreseeable misuse thereof; (b) identify the hazards and associated hazardous situations; (c) estimate the risk for each identified hazard and hazardous situation; (d) estimate the risk for each identified hazard and hazardous situation; (e) evaluate the risk and take decisions about the need for risk reduction

The safety requirements generated from the STPA were directed to a risk reduction process during the development of the autonomous technology and software and are intended to continually occur during the development, operation, and improvement of the autonomous technology.

The risk reduction process according to ISO 12100:2010 aims to:

- determine the limits of the machinery, which include the intended use and any reasonably foreseeable misuse thereof;
- eliminate the hazard or reduce the risk associated with the hazard by means of protective measures.

Actions (a) to (d) are related to the risk assessment with STPA, while (e) is related to risk reduction or mitigation strategies addressed in the LOPA section.

The key steps of STPA include:

- 1. Define the Use Case (e.g., Load-Haul-Dump cycles with retrofitted vehicle)
- 2. Define the Operational Scenarios (i.e., the interactions between the vehicle and other equipment)
- 3. Define the Operational Design Domain (ODD) in which the AGV can operate
- 4. Assess potential System-Level Losses and Hazards
- 5. Identify Unsafe Control Actions that could lead the system to a hazardous state
- 6. Document 1st Level Safety Requirements
- Analyze the Causal factors for Unsafe Control Actions
- 8. Document 2nd Level of Safety Requirements
- Implement the Safety Requirements in the Development and Testing pipeline
- 10. Test and Validate Safety Requirements

As new functionality of the system is developed, this process becomes iterative with new layers of Safety requirements that feed into the development and the testing pipeline.

4.1.1 Use Case Definition

The framework begins by defining the Use Case intended for the Autonomous Ground Vehicle (AGV). The description of the Use Case provides inputs for the next steps which involve determining the Operational Scenarios and ODD.

In our case study, the Use Case Definition is focused on one Autonomous Haul Truck (AHT), also referred to as Autonomous Ground Vehicle (AGV) used in Load-Haul-Dump (LHD) cycles at SafeAI's California proving grounds. This site location includes a dedicated space for our Quality Assurance teams to test and progress the Autonomous development.

4.1.2 Operational Scenarios

Within the SafeAI framework, Operational Scenarios include the scenarios in which the autonomous vehicle interfaces with its environment. Adapted from Vehicle Interaction Systems [7], the interactions are defined as interactions between different kinds and types of equipment, obstacles, infrastructure, and/or humans, etc. in the given operational design domain of the AGV (Figure 5).

We identify the following AGV interaction scenarios applicable to the California LHD Use Case:

1. Control of AGV

The AGV movement (forward-backward direction, turns, maneuvers, and speeds) will be restricted to the ODD as described for various steps of the autonomy development.

2. AGV interacting with humans

The personnel involved in the site and task runs are authorized operators to change the mode from manual to autonomous, conduct any required scheduled or unscheduled maintenance and/or repair of the AGV, and act as remote (out-of-AGV) operators while the AGV is performing a task. Duties and responsibilities of all site personnel were detailed in the risk assessment according to the site regulations.

3. AGV interacting with other equipment (staffed or autonomous)

The AGV performs the loading tasks by interacting with a staffed loader via SafeAIFlux (Staffed Vehicle System) and ZENO (Fleet/Autonomous Management System) coordination. Other equipment in or around the task zone was detailed in the corresponding ODD. In this case, a light vehicle is also allowed in the Autonomous Operating Zone.

4. AGV interacting with the environment

The environment in which the AGV can interact within the approved Autonomous Operating Zone (AOZ), including ground conditions, weather, number and types of lanes, lane edges, grade, and obstacles as detailed in the corresponding ODD. Known obstacle types found in the environment are documented and tested in the ODD.



Figure 5. Illustration of Autonomous Ground Vehicle Operational Scenarios based on EMESRT[7]

4.1.3 Operational Design Domain

Operational Design Domain (ODD) specifies the boundary in which the AGV can safely operate. There-

fore, ODD provides the design constraints of the AGV. Per SAE J3016 [8] ODD is defined as the "operating conditions under which a given driving automation system or feature is specifically designed to function, including, but not limited to, environmental, geographical, and time-ofday restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics."

Defining an ODD early in the design process identifies the functional scope of the AGV and what conditions it should be able to handle safely. The ODD definition identifies where changes in system capabilities are required and can aid in generating AGV test cases with varying levels of complexity.

For our ground proving site in California, we designated an initial ODD where the vehicle is allowed to operate. This first ODD was defined as the tightest boundary or radius in which the AGV has been designed and tested to operate. By passing out a series of milestones, the AGV can operate in increasingly complex scenarios or areas of the site. ODD compliance is monitored during the operations to ensure that the AGV is working within the set boundaries.

In our example, the ODD includes:

- Private off-road, geo-fenced area with a Load Zone, a Dump Zone, and hauling roads defined for autonomous operation
- Operating hours between 8am to 5pm according to site shift (running only one shift at the moment with no night operations)
- Site maximum grades are less than 15 degrees
- Drivable area surface is loose gravel
- Drivable area features do not include icy, flooded or muddy surfaces on this site
- Fixed road structures include site office, storage container, and vegetation
- No low clearance areas in AOZ
- AGV can operate in rainfall as permitted by staffed site operations
- Minimum and maximum temperature allowed for operations are the same as for staffed operations
- Traffic rules according to USA and California driving code. Mining rules apply including radio communication and left-hand driving
- Site speed limit 15mph
- No humans allowed in vehicle path while the vehicle is in autonomous mode.
- Interacting road users are authorized mining and construction vehicles

4.1.4 Assess Potential System-Level Losses and Hazards

Defining System-Level Losses in the framework introduces the concept of unacceptable losses to internal and external stakeholders. These losses are documented for the ODD in which the AGV tasks take place.

Losses include cases where the AGV causes an unsafe, unplanned, or undesired scenario. For this use case, the scenarios identified are the damage to the autonomous vehicle, another entity, the environment, and humans or coming dangerously close to causing damage to the autonomous vehicle, another entity, the environment, and humans.

A System-level Hazard is defined as a set of conditions that together with a set of environmental conditions could lead to an accident or loss as defined above. Hazards are linked to each possible Loss to provide insights into the conditions or circumstances that could lead to unacceptable scenarios.

4.1.5 Identify Unsafe Control Actions

The next step of the framework identifies potential Control Actions in the autonomous system that may lead to the hazardous state(s) disclosed above, hence called Unsafe Control Actions (UCA). Hazardous states could result from such potential control actions because:

- A control action required for safety is not provided or not followed;
- An unsafe control action is provided;
- A potential safe control action is provided too early or too late, that is, at the wrong time or in the wrong sequence;
- A control action required for safety is stopped too soon or applied too long.

Table 1 illustrates the analysis of Unsafe Control Actions within the context of the autonomous hauling task in our proving grounds in California. In this example, the unacceptable loss is identified as AGV nearly colliding with one or more obstacles, terrain, or infrastructure. The hazardous circumstance that leads to the potential loss is that the AGV does not stop for obstacles in the road during the Load-Haul-Dump (LHD) cycle.

4.1.6 Safety Requirements (1st Level)

Completing the UCA analysis is useful to express a list of safety requirements for the AGV following established safety standards and prevention thresholds. Under SafeAI's framework, these requirements or Safety Conditions (SC) were implemented during product and software development workflow.

From the previous example, high-level safety requirements were elaborated as follows:

SC2.1: AGV shall maintain a minimum distance from other equipment/vehicle of at least 20m in zones. **SC2.2:**

 Table 1. Unsafe Control Action (UCA) Analysis

Control Loop/Action	CA not provided	UCA Provided	CA too late/early	CA delivered incorrectly
Stop for Obstacle in haul road	AGV does not stop for obstacle	AGV accelerates with obstacle	AGV stopped too late Min distance violation	AGV engages wrong brake to stop

AGV shall maintain a minimum distance from other obstacles of at least 20m. **SC2.3:** AGV shall maintain a safe distance from vulnerable road users according to the Use Case and ODD. **SC2.4:** AGV shall obey and follow the site rules.

4.1.7 Analyze Causal Factors

Each UCA is linked to one or more causal factors (CF). This process of determining causal factors is further broken down from the system level to the functional block level, until the algorithmic level.

For example, for our UCA "AGV does not stop for an obstacle in forward path", we identified the CF involving the software functional block of the Perception module as follows:

CF1.1: Perception did not identify the obstacle. **CF1.2:** Perception did not check for static objects in the path. **CF1.3:** Perception did not check for dynamic objects entering around/in the path. **CF1.4:** Perception did not identify dynamic objects entering around/in the path. **CF1.5:** Perception did not register the position of the objects in the path ahead.

4.1.8 Safety Requirements (2nd Level)

Based on the causes described in the previous subsection, a second level of safety requirements or constraints was defined specific to the Perception functions.

See the following requirement definition:

REQ-1: AGV shall perform a normal stopping procedure (e.g., retarder activation) when AGV reaches a minimum safe distance of 20m from an obstacle in a zone. **REQ-1.1:** Perception shall identify static equipment in the path within 80-40m limit. **REQ-1.2:** Perception shall check for static objects in the path within 80-40m limit. **REQ-1.3:** Perception shall check for dynamic entering around/in path within 80-40m limit. **REQ-1.4:** Perception shall identify for dynamic entering around/in path within 80-40m limit. **REQ-1.5:** Perception shall register position of the objects in path ahead within 80-40m limit.

As observed in this section, SafeAI has integrated STPA iterative approach in the Safety framework wherein we continue to identify a deeper causal factors for each identified causal factor that could lead to a hazardous state. Each new causal factor layer triggers the definition of new levels of safety requirements which are incorporated into the development pipeline.

4.2 FMEA

FMEA is deployed in the Safety Framework to assess hardware performance such as sensors and Drive-by-Wire. FMEA establishes Failure Mode Identifiers (FMI) for each component of the Autonomous kit.

This standard analysis established a *Severity Scale* for each component failure from very high/catastrophic to low or insignificant, the *Probability of Occurrence* of the failure according to the number of failures per day, month, or year, and finally the *Detection Scale* of the failure, i.e., the likelihood of the defect being detected by process controls or reported by the system.

FMEA guided the hazard analysis of individual component failures on the AGV system and provided controls or safety requirements to be implemented to prevent or detect the failure. For example, if the compute unit loses power, the potential effect of failure is loss of vehicle control, and hence the safety requirement is: AGV shall detect compute unit loss and stop immediately.

4.3 LOPA

As an outcome of the hazard analysis delineated in the STPA and FMEA subsections, we determine additional risk mitigation strategies to identify, plan, manage, reduce, or eliminate potential risks associated with each identified hazard.

The Layers of Protection Analysis (LOPA) [9] is a semiquantitative risk evaluation method that builds on a hierarchy of controls (as shown in Figure 6). Several safety systems or controls are arranged in a format from more effective and protective to less effective or reliant on human behavior.

One of AGV's key benefits for the Use Case involves moving the equipment operators out of the vehicle and the equipment operation zone. The risk to the driver is eliminated, which is the highest level of protection identified by the National Institute for Occupational Safety and Health (NIOSH) [10]. However, other operators, such as loading unit operators, ancillary unit operators, drilling operators, and technical services staff may need to access or operate within the autonomous operating zone. Therefore, it is recommended by the Department of Mines, Industry



Figure 6. NIOSH [10] Hierarchy of Controls

Regulation and Safety (DMIRS) that additional primary controls such as elimination and substitution be put in place to reduce risks to these employees [11]. SafeAI has integrated LOPA into the Safety framework to define protective layers for identified system-level hazard scenarios, and implement the protective layers as independent safeguarding barriers.

The first step in risk mitigation for the Use Case is to ensure that any form of risk or hazard is first prevented. The prevention process involves complying with applicable California and United States regulations, international safety standards as applicable, and site procedures defined by the site operation and management prior to the initiation of work with autonomous equipment at the site. The site requirements are periodically revised to ensure SafeAI management is up to date with the site requirements.

The standard ISO 17757:2019 [12] was referred to and studied for safety analysis and assessments for the deployment of the autonomous equipment at the site. This standard requires a risk assessment process for Autonomous and Semi-Autonomous Machine System Safety (ASAMS), which conforms to the principles of ISO 12100 [13] wherein, all identified risks shall be mitigated to acceptable risk levels. ISO 17757 also requires that safety-related parts of control systems shall comply with the appropriate functional safety performance level. Examples include ISO 13849 [14], ISO 19014 [15], or IEC 62061 [16].

The following risk mitigation steps included:

- **Identification of risks:** Risks associated with potential hazards are identified as part of the STPA and FMEA, and per risk, the process of prevention is followed.
- **Compliance checks:** The system must comply with international and local regulations (when applicable), and safety standards relevant to the system as a whole or to any system part.

- **Planning:** The system, its processes, and interactions are planned to ensure the prevention of each identified risk.
- Verification: The system is verified to ensure it meets the safety requirements, safe fallback actions and plans are in place, and it is prepared for ongoing maintenance implications. Fallback options include built-in redundancies to ensure that if one component or process fails, there is at least one method to bring the AGV to stop.

Each layer of protection integrated into the system should be independent of each other for effective risk reduction.

SafeAI defines layers of protection for each Use Case and Release from development to production (Figure 7). These layers are related to operational procedures or protection measures, as detailed by the corresponding Job Safety Analysis, handed to the Site Operator. Engineering controls arranged as layers of protection as mentioned in ISO 17757 and included in our Use Case in California are:

- 1) Site Procedures and Regulations
- 2) Remote Autonomous Stop (A-Stop)
- 3) Situational Awareness (AGV Perception)



Figure 7. Layers of Protection Analysis

4.4 Implement Safety Requirements

The last step of the Safety framework is to implement the requirements in the system design and validate that the requirements are passed. The requirements are documented internally along with the expected pass/fail criteria. The test results are linked to the requirements and any discrepancies are highlighted as part of this process. A decision on discrepancies might either be continued validation or issue a change request. The following section further details the testing and validation of the requirements as part of the SafeAI Safety Framework for the case study.

4.4.1 Verification and Validation

We have implemented rigorous verification and validation processes to ensure that our system meets the specifications outlined in our safety requirements. These activities are vital in developing autonomous products that are verifiable and traceable at all levels of analysis. Our SafeAI and testing site personnel are regularly trained in the latest iteration of these procedures and informed of their responsibilities concerning these activities.

Our verification and validation processes have incorporated the expectations outlined in relevant industry and safety standards. We also consult the regional safety regulations of our customers and recommendations from industry-specific organizations such as Earth Moving Equipment Safety Round Table (EMESRT) and Mine Safety and Health Administration (MSHA). As new guidelines are discussed and introduced, we reassess and realign our existing procedures accordingly.

The requirements we established at the beginning of the production lifecycle primarily inform our verification and validation activities at the site. Our quality assurance team then planned and coordinated the relevant tests for the system, module, and unit levels. These tests analyzed all levels of the component, from interfaces and boundary values to operational use cases and dependent failures.

Our verification strategy followed a process similar to that outlined by the "V-model" of development for functional safety (Figure 8). Software requirements fed into software specification, which detailed the system design, module design, unit design, and overall implementation.



Figure 8. V Development Model

Developers documented their work through issue tickets and merge requests. We utilized a number of methods to verify the code at all stages, such as: • An automated continuous integration and continuous development (CI/CD) pipeline that checks all pushed commits for compliance with Motor Industry Software Reliability Association (MISRA) and AUTomotive Open System ARchitecture (AUTOSAR) code guidelines and code styling rules; • A manual review and approval process with developers designated to have an appropriate degree of independence from the code in question; • Unit-level tests such as inspection, fault injection, and extended functional testing; • Branch coverage and multiple condition/decision coverage (MC/DC); and • Analyses of interfaces and resource usage.

When possible, we have also implemented tools that have been designated as safety certifiable to our desired standards and frameworks.

If errors or non-compliant code was detected at any stage, the request in question was blocked. All development occurs on separate branches and cannot be merged into production unless it has successfully passed the above checks and received all required approvals. Once all verification checks have been completed, safety requirements and goals were validated at the vehicle level to evaluate any residual risk that could potentially trigger hazardous behavior of the overall system. We performed validation in two main stages: initial simulation and on-site vehicle testing.

Simulation provided an opportunity to thoroughly test entire systems and/or sub-systems for adequate performance before deployment and testing on the real vehicle at the site. Simulation also provided unique features that would otherwise be extremely hard to achieve, such as edge cases and future prediction. SafeAI utilized the topology and map of the AGV actual working location in California to create the simulation environment. We then tested various operational scenarios over thousands of hours to ensure that the vehicle will encounter safety-critical scenarios multiple times under a variety of conditions.

5 Site Operation and Established Work Procedures

Safety in mining and construction operations is the top priority of all personnel and support teams. This includes all personnel directly or indirectly engaged in supporting autonomous solutions. It is critical that the autonomous solution deployment at construction and mining sites not only addresses existing safety rules and regulations, but also helps end-users increase safety performance with inbuilt features, processes, and technology.

This section outlines safe work practices and procedures for the California proving grounds. Any deviations from the safe work procedures, required a job safety or hazard analysis to capture the hazards of the task and ensure that adequate controls or change management actions are implemented and communicated.

Safe work practices included education and training, access to the AOZ, mode change procedures, and emergency response.

5.1 Education and Training

All supervisory and operating personnel were instructed on the system functionality and specific tasks to be undertaken, including the hazards and risks, the controls to be applied, and the job steps necessary to complete the tasks safely and correctly. Training (manuals, specifications, and instructions) covered the different job skills required, the operation's policies, applicable legislation and standards, site requirements for monitoring machine performance, and incident reporting.

All personnel successfully demonstrated evidencebased assessment of competency before working without supervision.

5.2 Access to the Autonomous Operating Zone (AOZ)

A clear visual indication of the AOZ was provided at each designated entry and exit point. The AOZ access control system was monitored, with appropriate actions in case of failure, based on safety plans and controls. Anyone entering the AOZ underwent a required AOZ induction and/or was escorted according to the risk assessment.

5.3 Mode Change Procedures

Manual to Autonomous mode change was represented and indicated to site personnel by clearly visible mode lights. Transitioning to Autonomous mode requires a series of gateways and steps, both in the vehicle or remotely (in close proximity) starting from the Manual mode to prevent a single human error from transitioning the vehicle to autonomous mode. The required steps were defined in the operating procedure for the site.

5.4 Emergency Response

All personnel must be familiar with the emergency response strategy, muster points, and emergency contacts before entering any site. Emergency response planning for autonomous operations was integrated into the comprehensive site emergency response planning.

In addition to the work procedures summarized above, reporting and communication execution, inspections, traffic management, and practices to monitor the environment were developed by SafeAI together with the site management to ensure a safe working environment for all.

6 Conclusions

Given the industry need to assess the impact of automation in construction and mining to increase safety and productivity, this document outlined a Safety Framework for comprehensive and systematic assessment of the safety hazards in Construction and Mining operations with Autonomous vehicles. The framework is mainly based on STPA, taking into account the latest approach developed for industries like automobile and aviation automation. Additionally, FMEA, and LOPA are integrated into the comprehensive safety approach. The requirements from this assessment and relevant local and international norms feed into SafeAI's product development and testing procedures. Application of the framework into our proving ground site is provided as an example of the practices SafeAI carries out at each new mine or construction site across the world. The risk mitigation strategies addressed in this document ensure a phased approach to a fully autonomous site while taking care of the development, operations, and QA teams on the field.

In sharing this framework, our goal is to advance the industry approach to evaluate the safety of new autonomous equipment.

References

- [1] Federal Aviation Administration. *Safety Risk Management Policy*, volume Order 8040.4A. U.S. Department of Transportation, https://www.faa.gov/documentLibrary/media/Order, 2012.
- [2] R. A. Evans. Engineering Design Handbook Design for Reliability, volume AMCP-706-196. US Army Materiel Command, https://apps.dtic.mil/dtic/tr/fulltext/u2/a027370.pdf, 1976.
- [3] Reactor safety study. an assessment of accident risks in u. s. commercial nuclear power plants. executive summary: main report. [pwr and bwr]. 1975. doi:10.2172/7134131.
- [4] IEC. Hazard and Operability Studies (HAZOP studies) – Application Guide. International Standard IEC 61882 (2.0 ed.), volume AMCP-706-196. International Electrotechnical Commission, ISBN 978-2-8322-3208-8, 2016.
- [5] J.P. Leveson, N.; Thomas. STPA Handbook. MIT Partnership for Systems Approaches to Safety and Security (PSASS), https://psas.scripts.mit.edu, 2018.
- [6] MIT. Mit partnership for systems approaches to safety and security (psass). On-line: http://psas. scripts.mit.edu/home/, Accessed: 12/15/2023.
- [7] EMESRT. Vehicle interaction. On-line: https:// emesrt.org/vehicle-interaction/, Accessed: 12/15/2023.
- [8] On-Road Automated Driving (Orad) Committee. Taxonomy definitions for operational design domain (odd) for driving automation systems j3259. SAE International Standards, 2021. doi:https://www.sae.org/standards/content/j3259/.
- [9] Ronald J. Willey. Layer of protection analysis. *Procedia Engineering*, 84:12–22, 2014. doi:https://doi.org/10.1016/j.proeng.2014.10.405.

- [10] The National Institute for Occupational Safety and Health (NIOSH). Hierarchy of controls. On-line: https://www.cdc.gov/niosh/ topics/hierarchy/default.html/, Accessed: 12/15/2023.
- [11] Mining Industry Advisory Committee, Work Health, and Safety Commission. *Mine safety management system: Code of practice, Department of Mines, Industry Regulation and Safety.* Government of Western Australia, ISBN 978 1 922873 01 9, 2022.
- [12] International Organization for Standardization. Earth-moving machinery and mining autonomous and semi-autonomous machine system safety. Requirements with guidance for use (ISO Standard No. 17757:2019) https://www.iso.org/standard/ 76126.html/, 2019.
- [13] International Organization for Standardization. Safety of machinery. general principles for design. risk assessment and risk reduction. Requirements with guidance for use (ISO Standard No. 12100:2010) /https://www.iso.org/ standard/51528.html, 2010.
- [14] International Organization for Standardization. Safety of machinery. safety-related parts of control systems part 1: General principles for design. Requirements with guidance for use (ISO Standard No. 13849-1:2023) /https://www.iso.org/ standard/73481.html, 2023.
- [15] International Organization for Standardization. Earth-moving machinery. functional safety. part 1: Methodology to determine safety-related parts of the control system and performance requirements. Requirements with guidance for use (ISO Standard No. 19014-1:2018) /https://www.iso.org/ standard/70715.html, 2018.
- [16] International Electrotechnical Commission. Safety of machinery - functional safety of safety-related control systems. Requirements with guidance for use (IEC No. 62061:2021) /https://webstore. iec.ch/publication/59927, 2021.

Human-Robot Partnership: An Overarching Consideration for Interaction and Collaboration

Jingshuo Yu¹, Qian Chen¹, Samuel A. Prieto², and Borja García de Soto²

¹School of Engineering, University of British Columbia, Okanagan Campus, Canada ²S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD),

United Arab Emirates

jensonyu@mail.ubc.ca, gian.chen@ubc.ca, samuel.prieto@nyu.edu, garcia.de.soto@nyu.edu

Abstract –

The construction industry is looking at automation and robotization to enhance productivity and reduce the safety risks of various tasks during the construction, operations and maintenance (O&M) phases. In that context, human-robot interaction (HRI) and human-robot collaboration (HRC) are highly relevant. Although those terms are different, there is some misconception, and, in some cases, they have been used interchangeably. To address that, this study clarifies the meanings of each. We collected the existing explanations, reviewed the keywords, and completed a literature review of HRI and HRC applications in construction and O&M, and included definitions for HRI and HRC in the context of construction and O&M needs. Based on the review, we summarized the key elements to differentiate HRI and HRC and promote the concept of human-robot partnership (HRP) as a potential solution to overcome the identified limitations seen from pure HRI and HRC literature, followed by the future directions of HRP applications proposed for construction and O&M activities and needs.

Keywords -

Human-robot Collaboration; Human-robot Interaction; Human-robot Partnership; On-site Construction; Operations & Maintenance; Robotics

1 Introduction

While accounting for 13% of the world's gross domestic production (GDP), the construction industry suffers from low productivity, labor shortage, and high risks [1]. The case is similar for the O&M of buildings and infrastructures, which often costs more than the expenditure of initial construction and contains dangerous tasks (e.g., the maintenance of nuclear power plants and bridges) [2], [3]. As a potential solution for addressing these problems, robots are adopted in construction and O&M fields. However, even though robots for construction and O&M have progressed a lot due to the development of hardware, software, and artificial intelligence, it is still extremely challenging to meet the requirements of full autonomy or no human intervention [4]. The concepts of human-robot interaction (HRI) and human-robot collaboration (HRC) are introduced to make robotics applications in construction and O&M more adaptable to the unstructured and complex working environment [4]. Progress has been made in the intuitive interface for effective communication [5], task allocation to overcome constraints [6], and accurate detection of worker's cognitive load [7]. However, if we go back to the terms HRI and HRC, few researchers defined them and explained why they want to use one or both of them, as the topic of their works. In many scenarios, they were even used interchangeably, which makes it crucial to differentiate the two terms to avoid misuse and confusion. Many applications contain both interaction and collaboration between humans and robots, such as the interaction with collaborative robots [8], [9], [10]. And it is quite difficult to classify many works into HRI or HRC according to existing definitions. For example, according to Liu et al. [11], humans can request robotic assistance by remote control through the proposed brain-computer interface, which not only satisfies the information exchange between human and robot stated in ISO's definition of HRI [12] but also meets the requirement that human and robot work together to achieve the shared goals in the definition of HRC suggested by [13]. This shows the necessity of proposing a term that is feasible for works that concern both HRI and HRC.

To address this gap, in this study, we selected papers related to HRI and HRC applications in construction and O&M by keyword search, filter setting, manual screening, and representative selection, as described in Section 2. Combining the existing explanations, keywords analysis of all relevant publications, and selected papers in Section 2, we offer definitions for HRI and HRC in the scope of construction and O&M, with which the key elements in HRI and HRC applications are analyzed in Section 3, and HRP is proposed as the combination of HRI and HRC and suggested to be used in future works that concern both HRI and HRP for standardization purposes. Section 4 offers potential future directions, followed by Section 5, which concludes the work with the highlights of our findings.

2 Materials Selection

2.1 Materials Selection

As displayed in Figure 1, we selected papers from the Scopus database with four keywords combinations: "human robot interaction construction", "human robot interaction operation maintenance", "human robot collaboration construction", and "human robot collaboration operation maintenance" with an intention to cover all the HRI and HRC applications in construction and O&M. Then we narrowed down the review scope to journal and conference papers in the last decade. The Scopus database was chosen for being the largest database of peer-reviewed articles. Other indexed databases and books were not selected because they did not provide useful additions [5]. In addition, the subject was limited to engineering, papers published in the last 10 years, and only publications in English were selected. After manual screening, 117 papers were included, of which 8 representative papers [1], [6], [7], [11], [14], [15], [16], [17] that involve the key elements in HRI and HRC applications in construction and O&M were used as the references of the focus review.



Figure 1. Selection process of relevant literature ([1], [6], [7], [11], [14], [15], [16], [17]) for this study

3 Focus Review

3.1 Key Definitions

To get a definition for HRI and HRC in the scope of construction and O&M, we referred to the existing explanations of the two terms. According to ISO 8373:2021 [12], HRI is the information and action exchange between humans and robots to perform a task by means of a user interface. HRC is described as the study of collaborative processes in human and robot agents working together to achieve shared goals in [13]. Based on that, we searched for HRI (without HRC) and HRC (without HRI) applications in the Scopus database with the keywords combinations "human robot interaction" AND NOT "collaboration" and "human robot collaboration" AND NOT "interaction". A keyword analysis was conducted for the results of the two searches with the open-source software VOSviewer [14]. For both HRI and HRC results, keywords were extracted from the paper title and abstract. The top 20 keywords of HRI and HRC were selected from the items that occurred more than 1700 times for HRI and 250 times for HRC. A list of keywords was obtained after removing overlapping and non-relevant items (e.g., technology, research, and problem) (Table 1). From that, it can be seen that a social robot that interacts and communicates with humans is an important application of HRI. In general, people participate in the process as the user, and

the robot functions as the service provider. HRC, however, is mostly used in the context of industrial robots, where the human plays the role of the worker to execute tasks with the robot peer together. Combining the existing explanations, keywords analysis, and literature on HRI and HRC in construction and O&M, we offer the following definitions for HRI and HRC in the scope of construction and O&M:

- HRI is the process in which humans exchange information with robots through a user interface for control or communication, by which they indirectly participate in the tasks.
- HRC is the state in which human workers and robots work together to achieve a certain task.

Based on the above work and definitions, we concluded the key elements of HRI and HRC and proposed human-robot partnership (HRP) as the combination of them in Figure 2, which will be discussed in the following sections in detail.

Table 1. Keywords of HRI and HRC

Туре	Keywords
	Social robot, User,
HRI	Environment,
	Behaviour, Framework
IDC	Cobot, Industry, Safety,
HKC	Worker



Figure 2. Key elements of HRI and HRC in construction and O&M

3.2 Human

According to the above definitions, HRI and HRC can be distinguished by the role of humans in the process. In HRI, humans are the users of service robots, they do not directly participate in the work, while they function as the workers together with robots in HRC. For both HRI and HRC, in the face of exposure to robots, the key element of humans is their well-being. Besides the potential physical risks like collisions, which can be addressed from the design perspective (e.g., using algorithms to predict workers' movements [17]), human workers may also suffer from cognitive load derived from the lack of confidence and predictability in the robot peers. Towards this end, Liu et al. [7] proposed a brainwave-driven worker-centred framework equipped with a wearable electroencephalograph (EEG) to measure workers' task-related cognitive load, with which robots' behavior can be adjusted accordingly. After a test with 14 subjects who worked with a terrestrial robot under different cognitive loads, the results showed that the robot could adjust its working space with 81.91% accuracy with the help of humans' brain signals.

3.3 Interaction

To enable the information exchange in our definition of HRI, various types of interfaces are used as the interaction bridge between humans and robots. However, the applicable scenarios of many of these interfaces are limited. For example, position and force-based interfaces do not allow human workers to teleoperate the robots in a hands-free manner, which is inconvenient when their hands are occupied. Also, the required headsets of visionbased interfaces (e.g., Virtual Reality, Augmented Reality) may reduce visibility at the actual sites and increase safety risks. Similarly, voice control interface is significantly affected by noise on construction sites. To expand the application scope of robots and better serve human workers, Liu et al. [11] proposed a braincomputer interface (BCI) with hands-free and nonmuscular interaction between humans and robots can be achieved. The interface enables users to control robots by transforming users' brain signals from a wearable EEG device into robotic commands with satisfactory accuracy.

In our definition of HRI, humans do not participate in the work directly; instead, they provide task-related information to robots by instructing or communicating with them. Robots are service providers who take the responsibility of completing the tasks physically, following the information from humans and their interpretation of it. To effectively interact with humans, robots must be capable of recognizing the information from humans. To enable robots to capture and interpret the hand gestures of users (construction workers), Wang et al. [14] developed a vision-based framework consisting of three working components: (1) workers detection and tracking, (2) recognition queues formulation and (3) hand gesture recognition. The first component detected the user giving hand gestures and creating their bounding box. With the second component, the region of the user cropped from the original frame was firstly expanded horizontally by 25% to capture more information and avoid errors; then, the extracted frames were compiled to form the hand gesture detection and classification queues for the last component. A hierarchical convolutional neural network (CNN) architecture was applied in the last component to detect and classify the user's hand gestures with 87.0% precision in the implementation.

3.4 Collaboration

To enable humans and robots to work together for a task as described in our definition of HRC, successful HRC applications in construction and O&M must be based on the reasonable task allocation between human workers and robots that takes the requirements of different subtasks and the constraints of both humans and robots into consideration. For example, humans have strong comprehensive analysis ability (lacked by robots), which makes them more suitable for decision-making work, while robots are more efficient in repetitive and labor-intensive work like material and component delivery. To enable the application of human-robot collaboration in building large erectable truss structures in a space station, Zhu et al. [6] decomposed the task process into a certain number of basic actions named therblig. Then, hierarchical task analysis was adopted to establish the task model, based on which task allocation was conducted according to the limitations and abilities of humans and robots. Finally, a virtual simulation was applied to verify the feasibility of the proposed method.

For a certain task of construction or O&M, the robots to be adopted should be capable of doing the allocated tasks, which imposes a challenge for the design and programming of robots. For instance, robots should be equipped with enough force capability required by the assigned tasks, should be able to reach all the task locations, should be capable of recognizing the obstacles along the defined path, should be prepared with necessary hand tools, and should be capable of being stopped in an emergency situation. In the work of Gautam et al. [15], a test of a construction robot was conducted for the installation of gypsum board panels, which is identified to be injury-prone from the perspectives of construction professionals. 3D scanning, reverse engineering, and 3D printing are applied to fabricate toll adaptors that enable robots to use human tools and target markings are used to indicate the robots to work at the right location and distance. The results of the experiment in a wooden house construction site showed that gypsum board installation can be achieved by the collaboration of the robot and human worker, and workers can benefit from ergonomics enhanced by collaborative robots.

3.5 Partnership

Based on our definitions of HRI and HRC in the scope of construction and O&M, pure HRI cannot exist alone because the aim of enabling interaction between humans and robots is to offer better collaboration. In other words, information exchange (interaction) alone is just a process as described by the definition in Section 3.1, and it is not meaningful until it realizes the aim of instructing robots to better deliver the task. In this

scenario, humans offer instructions, which are followed by robots to guide the physical work, resulting in them working together (collaboration). Without interaction, pure HRC is simply letting humans and robots take the assigned work from task allocation and thus lacks flexibility. In a dynamic and unstructured working environment (like construction sites), a minor change in the task may cause the robot to lose the capability to work, and humans have no access to manipulate or teach the robot to complete the task. Also, the human cannot modify the robot's behavior when an error occurs.

As shown in Figure 3, due to the above drawbacks, actual needs cannot always be fully satisfied (represented by the bland expression on the human's face) by pure HRI or HRC. This representation also shows how the robot is able to interact with the human by asking a question in the HRI, but not in the HRC. A feasible solution to address the issue is combining HRI and HRC to form a novel and flexible relationship between humans and robots that is defined as HRP (represented by a content expression in the human, the interaction part in the robot, and the collaboration in the hands). Actually, many works fit the concept of HRP. For example, Zhou et al. [16] proposed a visual-haptic interface to provide high-fidelity task scenarios and enhance control feedback. Typical physical interactions like weight, texture and inertia are captured with this method. In this work, humans provided instructions by operating an interface and received tactile feedback from the robots performing the physical tasks. However, this form of partnership is relatively low-level as humans still needed to fully control every working process, and robots did not learn how to do the work. Wang et al. [1] improved this using a learning from demonstration (LfD) method that enables humans to flexibly teach robots how to deliver tasks by applying intuitive demonstration. In this method, robots are equipped with basic skill primitives that are required by different construction tasks, such as reaching and nailing. A VR interface is provided to enable human workers to be aware of the on-site status and demonstrate task delivery through the selection of the robot's skill primitives.



Figure 3. Human-robot interaction, collaboration and partnership

4 Future Directions

4.1 Multi-sensor Fusion of Physiological Signals

With more and more focus imposed on human workers' mental stress during the HRP process, the method for stress measurement has transformed from collecting questionnaires from participants to a more reliable electroencephalogram (EEG) method that captures human brain signals directly. While much work has been done with wearable EEG devices, the concern is that the devices used in the construction field with a small number of electrodes (14-32) may not be capable of collecting reliable EEG signals compared with the 64-256 electrodes used in the clinical field. To address this issue, Liu et al. [15] applied a generative adversarial network to produce high-quality EEG signals. Another potential solution could be to leverage multi-sensor fusion. Besides EEG, there are already various kinds of methods for collecting human physiological signals, such photoplethysmogram (PPG), electrocardiogram as (ECG), and electrooculogram (EOG). By fusing signals from multiple sensors, the uncertainty of data will decrease, and more reliable information will be gained. Thus, we can have a deeper insight into the human worker's mental load.

4.2 Multi-source Data Fusion for HRP-based Predictive Maintenance

For current practices of O&M of buildings and infrastructure assets, robotic applications are mainly devised using vision-based robots for automatic crack detection. However, not all maintenance requirements can be detected with robotic vision, such as the detection of internal cracks in pipes and operating anomalies in building facilities. To achieve early-stage maintenance for these deteriorations and prevent them from further developing into failures, many works have been done in predictive maintenance (PdM) with various sensors. Human expertise [20] can also be fused with sensor data to accommodate the drawbacks of each other, and robots should be capable of multi-source data fusion and anomaly detection to predict the facility's health state in the future. In maintenance activities, given the amount and complexity of maintenance work in modern buildings, it is hard to enable robots to perform tasks by pre-programming them. A feasible solution could be the LfD [1] method mentioned before, which enables humans to teach robot task delivery by intuitive demonstration. In this way, instead of the low-level manual manipulation, humans can conduct high-level experience transferring in the process of HRP-based predictive maintenance, which better leverages the strengths of both humans and robots.

5 Conclusions

This study provided clear definitions of HRI and HRC in the context of construction and O&M applications by reviewing existing definitions, keyword analysis, and a systematic literature review. HRI was defined as the process where humans exchange information with robots through a user interface for control or communication by which they indirectly participate in the tasks. Slightly different is the HRC concept, in which human workers and robots work together to deliver a certain task. The key elements for interaction and collaboration between human users and robots are the interface and task allocation, respectively. For robots, HRI concentrates on their recognition of information from humans, and HRC focuses on their task capability. The crucial point for humans in both processes is the well-being problem. Since HRI cannot exist alone because it serves HRC, and HRC will be rigid and inoperable without HRI, we proposed the relationship defined as HRP to integrate features of HRI and HRC for improved task performance in construction and O&M and suggested to use of HRP in works that concern both HRI and HRC features. Finally, data fusion could be the future direction to innovate HRP research. With multi-sensor fusion of physiological signals (e.g., EEG, PPG, ECG, EOG), more accurate detection of a worker's mental load can be derived to better ensure the human's well-being, a key element for HRP. With multisource data fusion, human expertise can be included as the guidance of robotic maintenance work, which better leverages human's edge in intellectual work and the robot's ability for physical work.

Acknowledgments

This work benefited from the collaboration with the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001 and the support from the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES) funded by Tamkeen under the NYUAD Research Institute Award CG013.

References

- 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*, vol. 155, p. 105071, Nov. 2023, doi: 10.1016/j.autcon.2023.105071.
- [2] H. Oyediran, P. Ghimire, M. Peavy, K. Kim, and P. Barutha, "Robotics applicability for routine operator

tasks in power plant facilities," presented at the 38th International Symposium on Automation and Robotics in Construction, Dubai, UAE, Nov. 2021, doi: 10.22260/ISARC2021/0091.

- [3] H.-D. Bui, S. Nguyen, U.-H. Billah, C. Le, A. Tavakkoli, and H. M. La, "Control framework for a hybrid-steel bridge inspection robot," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 2585–2591. doi: 10.1109/IROS45743.2020.9340637.
- [4] Y. Leng, X. Shi, F. Hiroatsu, A. Kalachev, and D. Wan, "Automated construction for human–robot interaction in wooden buildings: Integrated robotic construction and digital design of iSMART wooden arches," *Journal of Field Robotics*, vol. 40, no. 4, pp. 810–827, Jun. 2023, doi: 10.1002/rob.22154.
- [5] M. Zhang, R. Xu, H. Wu, J. Pan, and X. Luo, "Human-robot collaboration for on-site construction," *Automation in Construction*, vol. 150, p. 104812, Jun. 2023, doi: 10.1016/j.autcon.2023.104812.
- [6] X. Zhu, C. Wang, M. Chen, S. Li, and J. Wang, "Concept plan and simulation of on-orbit assembly process based on human–robot collaboration for erectable truss structure," in *Man-Machine-Environment System Engineering*, vol. 645, Singapore: Springer Singapore, 2020, pp. 683–691, doi: 10.1007/978-981-15-6978-4_78.
- Y. Liu, M. Habibnezhad, and H. Jebelli, "Brainwave-driven human-robot collaboration in construction," *Automation in Construction*, vol. 124, p. 103556, Apr. 2021, doi: 10.1016/j.autcon.2021.103556.
- [8] R. R. Galin and R. V. Meshcheryakov, "Humanrobot interaction efficiency and human-robot collaboration," in *Robotics: Industry 4.0 Issues & New Intelligent Control Paradigms*, vol. 272, A. G. Kravets, Ed., in Studies in Systems, Decision and Control, vol. 272., Cham: Springer International Publishing, 2020, pp. 55–63. doi: 10.1007/978-3-030-37841-7 5.
- [9] L. G. Christiernin, "How to describe interaction with a collaborative robot," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna Austria: ACM, Mar. 2017, pp. 93–94. doi: 10.1145/3029798.3038325.
- [10]J. E. Michaelis, A. Siebert-Evenstone, D. W. Shaffer, and B. Mutlu, "Collaborative or simply uncaged? Understanding human-cobot interactions in automation," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12. doi: 10.1145/3313831.3376547.

- [11] Y. Liu, M. Habibnezhad, and H. Jebelli, "Braincomputer interface for hands-free teleoperation of construction robots," *Automation in Construction*, vol. 123, p. 103523, Mar. 2021, doi: 10.1016/j.autcon.2020.103523.
- [12] *Robotics Vocabulary*, ISO 8373:2021, 2021.
- [13] A. Bauer, D. Wollherr, and M. Buss, "Human-robot collaboration: A survey," *Int. J. Human. Robot.*, vol. 05, no. 01, pp. 47–66, Mar. 2008, doi: 10.1142/S0219843608001303.
- [14] X. Wang and Z. Zhu, "Vision–based framework for automatic interpretation of construction workers' hand gestures," *Automation in Construction*, vol. 130, p. 103872, Oct. 2021, doi: 10.1016/j.autcon.2021.103872.
- [15] M. Gautam, H. Fagerlund, B. Greicevci, F. Christophe, and J. Havula, "Collaborative robotics in construction: A test case on screwing gypsum boards on ceiling," in 2020 5th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, Vietnam: IEEE, Nov. 2020, pp. 88–93. doi: 10.1109/GTSD50082.2020.9303061.
- [16] T. Zhou, P. Xia, Y. Ye, and J. Du, "Embodied robot teleoperation based on high-fidelity visual-haptic simulator: pipe-fitting example," *J. Constr. Eng. Manage.*, vol. 149, no. 12, p. 04023129, Dec. 2023, doi: 10.1061/JCEMD4.COENG-13916.
- [17] J. Cai, A. Du, X. Liang, and S. Li, "Prediction-based path planning for safe and efficient human-robot collaboration in construction via deep reinforcement

learning," J. Comput. Civ. Eng., vol. 37, no. 1, p. 04022046, Jan. 2023, doi: 10.1061/(ASCE)CP.1943-5487.0001056.

- [18] N. J. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523– 538, Aug. 2010, doi: 10.1007/s11192-009-0146-3.
- [19] Y. Liu and H. Jebelli, "Enhanced robotic teleoperation in construction using a GAN-based physiological signal augmentation framework," in *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, vol. 239, S. Walbridge, M. Nik-Bakht, K. T. W. Ng, M. Shome, M. S. Alam, A. El Damatty, and G. Lovegrove, Eds., in Lecture Notes in Civil Engineering, vol. 239., Singapore: Springer Nature Singapore, 2023, pp. 295–307. doi: 10.1007/978-981-19-0503-2 24.
- [20] Z. Liu, N. Meyendorf, and N. Mrad, "The role of data fusion in predictive maintenance using digital twin," presented at the 44th Annual Review of Progress in Quantitative Nondestructive Evaluation, Volume 37, Provo, Utah, USA, 2018, p. 020023. doi: 10.1063/1.5031520.
- [21] C.-J. Liang, X. Wang, V. R. Kamat, and C. C. Menassa, "Human-robot collaboration in construction: classification and research trends," *J. Constr. Eng. Manage.*, vol. 147, no. 10, p. 03121006, Oct. 2021, doi: 10.1061/(ASCE)CO.1943-7862.0002154.

3D Printing vs. Traditional Construction: Cost Comparisons from Design to Waste Disposal Stages

Svetlana Besklubova¹, Muhammad Huzaifa Raza¹, Ray Y. Zhong^{1*}, and Miroslaw J. Skibniewski²

¹Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong

²Professor, Department of Civil and Environmental Engineering, University of Maryland, *College Park, MD 20742-3021*

lanabess@hku.hk, huzaifa@connect.hku.hk, zhongzry@hku.hk, mirek@umd.edu

Abstract -

Construction 3D printing (3DP) has the potential be cost-effective compared to traditional to construction, due to its shorter supply chain and higher level of automation. However, there is a lack of comprehensive studies comparing the costs of traditional construction and 3DP across various stages, from design to waste disposal. Therefore, this study investigates the overall cost of traditional and 3DP construction, considering both off-site and onsite 3DP techniques. Mathematical models are developed to analyse costs, including research and development (R&D), Architecture, Engineering, and Construction (AEC), transportation, warehousing, setup installation, printing and on-site assembling processes, waste disposal, and environmental costs. Real-time case studies have been investigated to validate the feasibility and practicality of the models. The comparative analysis revealed that the design stage cost of 3DP is 19 times higher than those of traditional construction due to significant investments in R&D. Off-site 3DP incurs higher logistics costs primarily due to holding costs. Nevertheless, it becomes a more viable option for small-scale projects (less than 35t for the considered cases), like printing architectural elements, as the expense of delivering printed components to the construction site is lower compared to transporting bulky 3DP equipment and raw materials. On-site 3DP presents a competitive alternative to traditional construction methods, both for individual projects and large-scale developments. The findings of this research provide valuable insights that can help the construction industry optimize cost-effectiveness and enhance efficiency in construction practices.

Keywords -

3D printing, logistics, economic assessment, additive manufacturing, case study

1 Introduction

The fourth industrial revolution, known as Industry 4.0, is characterized by the digitization of complex industrial tasks. One of the key technological advancements driving this digital transformation is 3DP, also referred to as additive manufacturing or rapid manufacturing [1]. 3DP involves the layer-by-layer joining of materials to create objects based on threedimensional models [2]. The global 3DP market experienced significant growth in 2020, with a remarkable 21% increase compared to 2019, reaching an estimated value of \$12.6 billion [3]. This growth highlights the potential of 3DP technology to drive various industries towards digitalization. In the construction industry, 3DP technology shows great promise in reducing labor costs, construction time, risky human operations, material usage, and waste [4, 5]. Additionally, it offers the advantage of architectural design flexibility and provides social and environmental benefits [6]. Consequently, significant efforts have been made by stakeholders to advance the construction 3DP industry. These efforts include the development of 3D printer robotic systems, suitable printed materials, and new applications. Various organizations worldwide have completed numerous single projects, ranging from smallscale structures to entire low-rise buildings. However, 3DP has yet to be thoroughly tested in mass-production scenarios [7, 8].

Despite growing interest and knowledge in 3DP technologies, their adoption in the construction industry lags behind the manufacturing sector [9]. Challenges such as technological feasibility, cost and time benefits, user training, safety considerations, and compliance with contractual and standard requirements impede widespread adoption [10]. These concerns create doubts among potential adopters about the value of implementing 3DP technology [11].

Several research studies have focused on the economic analysis of construction 3DP. Tobi, et al. [12] found that 3DP has the potential to reduce construction costs by 30% compared to conventional techniques. Yang, et al. [13] developed a cost calculation method considering various factors such as labor, material, machine, management, safety, and environmental costs for both off-site and on-site 3D printing. Aghimien, et al. [14] gathered insights from construction industry professionals and demonstrated that construction 3DP offers improved cost efficiency in housing projects and enhances productivity. Weng, et al. [15] conducted a comparative economic analysis of 3D concrete printing and precasting, observing a 34% reduction in overall costs for 3DP construction. Allouzi, et al. [16] compared 3DP with conventional construction techniques for a single-story building, finding a 65% reduction in material costs for 3DP. Markin, et al. [17] estimated the cost of foam concrete exterior walls produced through 3DP, with material costs contributing 70% to the total direct cost. Han, et al. [18] conducted a comparative analysis of 3DP and conventional construction methods for a hypothetical cylindrical silo, highlighting that material costs accounted for 83% of the overall cost in 3DP. Abdalla, et al. [19] compared the costs of formwork and raw materials between construction 3DP and conventional construction, reporting a 78% lower capital cost for 3DP due to the absence of formwork and concrete. The abovementioned studies discuss various cost components related to the construction phase, with some addressing environmental factors during the construction stage and resource utilization. Besklubova, et al. [20] conducted a detailed analysis of logistics costs for construction 3DP of low-story buildings, covering the entire process from suppliers to waste disposal.

While various studies have focused on cost analysis in construction 3DP, they often consider specific stages such as construction or logistics individually. As a result, obtaining a comprehensive understanding of the cost structure for the entire 3DP project, from design to waste disposal, continues to be a challenge. To address this gap, the current study provides a comprehensive cost analysis throughout the entire process, including environmental factors. This study considers two different 3DP construction techniques including off-site and on-site 3DP, to comprehensively assess their cost-effectiveness.

2 Model development

The research methodology consists of four key steps. Firstly, a comprehensive literature review was conducted. This literature review provided insights into the processes involved in each stage of construction projects and established a theoretical foundation for developing a cost estimation model to assess construction projects from the design to realization stage.

Secondly, the development of the cost estimation model began by selecting parameters based on the involved processes. The parameter selection followed the Activity-Based Costing (ABC) approach, which aims to accurately allocate overhead costs and resources, such as labor, materials, and equipment costs, to processes based on their actual consumption of resources. In comparison to traditional cost analysis approaches (uses machine hours or man-hours consumed as the basis for estimating costs), ABC method provides more accurate cost information, which enables effective monitoring of supply chain and production strategies [21]. Subsequently, the model assumptions were defined, and the parameters were quantified and formulated.

Thirdly, three comparable case studies were selected to ensure the feasibility and practicality of the proposed model. These case studies represented on-site 3DP, offsite 3DP, and traditional construction. A three-step data collection approach was utilized to gather a comprehensive dataset, employing triangulation of evidence from three interrelated methods [22].

Finally, a comprehensive cost breakdown analysis was carried out to evaluate and understand the individual cost components associated with both 3DP and traditional construction processes. Additionally, a sensitivity analysis was conducted for critical cost components, followed by data processing. Sensitivity analysis involves examining how uncertainty in model output, whether numerical or otherwise, can be attributed to various sources of uncertainty in the model input.

2.1 Parameters selection

The low-rise building project process involves different stages, including the design stage, logistics, and construction process (printing and assembly). The design stage includes R&D and AEC activities. In the traditional construction supply chain (CSC) flow, raw materials are obtained from suppliers and processed by manufacturers to create construction materials and elements. These materials and elements are then transported to the construction site, and any waste generated during the construction process is transported to disposal facilities (Figure1 (a)) [4, 23]. In the case of construction 3DP technology, there are two main supply chain configurations: on-site and off-site [4, 24]. On-site printing involves moving raw materials and 3DP equipment directly to the construction site (Figure 1 (b)). Although 3DP is often touted as a zero-waste technology [25], waste disposal facilities are still a part of the supply chain at its current stage of development. This is because the technology, with its imperfections such as failed components and inaccurate material calculation, still generates waste during the printing process [4]. Off-site printing involves moving materials from suppliers to a

3DP manufactory facility, and then to the construction site for assembly (Figure 1(c)). Similar to the above, construction waste is transported to disposal facilities. The logistics cost components have been discussed in detail by Besklubova, et al. [20]. Construction process costs encompass materials, machinery, equipment, and environmental factors.



Figure 1. Construction supply chain configurations for small and medium scale projects: (a) traditional, (b) onsite 3DP, and (c) off-site 3DP

2.2 Model assumptions

The model assumes a unidirectional flow, specifically from the supplier of raw materials to the off-site manufacturing facility, then from the manufacturing facility to the warehouse, and finally from the warehouse to the construction site and then to disposal facilities. In the case of off-site 3DP, there is no need to transport equipment because the 3DP equipment is owned by the off-site manufacturer. There is no limit to the total number of facilities at each node. To estimate the distances between the facilities, Google Maps is utilized [26]. The labour wage rates used in the analysis are based on the latest standard rates obtained from reliable online sources. The printing process considers the times required for installation, printer setup, material preparation, and actual walls printing. The 3DP equipment consists of a 3D printer, batching plant, concrete piston pump, and concrete silo. In the case of off-site 3DP, a crane is also utilized for the assembly of printed components. It should be noted that the development of the pricing policy did not consider weather conditions.

2.3 Quantification of parameters

This section explains how to quantify the total construction project cost and its related cost components. The activities involved in each construction stage were assessed using the ABC approach. The cost component parameters are summarized in Table 1 and expressed via equations 1-4 below.

	Table	. List of parameters
--	-------	----------------------

Daramatars	Description
rd	Description
n'u	Manpower for R&D
nus	Manpower at design stage
wra	Average salary for R&D
w ^{ds}	Average architect's salary
t rd /t ^{ds}	Project preparation period (R&D/ design)
HC_i^{mat}	Handling cost per unit of material at facility i
IHC ^{mat}	Inventory holding cost per unit of
IHC _i	materials/printed components at facility i
DCiwaste	Disposal cost per unit of waste
TC _{ij}	Transportation cost from i to j, $i \neq j$
Q_i^{mat}	Quantities of materials at facility i
$Q_{ij}^{mat} \\$	Quantities of materials transported from i to j, $i \neq j$
V_{ij}^{eq}	Volume of equipment transported from <i>i</i> to <i>j</i> , $i \neq j$
${\rm SUt}_{\rm i}^{\rm eq}$	Time required for each set up of the 3D printer at facility i
SUn _i ^{eq}	Number of the 3D printer set up times at facility i
E _{ij}	Carbon emissions factor for diesel consumption for material/ equipment transportation from i to j , $i \neq j$
CE _{CO₂}	Environmental cost of CO ₂ emissions
dis_{ij}^{mat}	Distance from <i>i</i> to <i>j</i> for material supply, $i \neq j$
dis _{ij} ^{eq}	Distance from <i>i</i> to <i>j</i> for equipment supply, $i \neq j$
$n_{i}^{SU} \\$	Number of workers for equipment set up at facility i
w_i^{SU}	Construction site worker wage rate for equipment set up at facility i
n ^{truck}	Number of 18 tonnes-trucks for materials transportation from i to i $i \neq i$
n_{ij}^{cont}	Number of 20-foot 3D printer transport containers from i to i $i \neq i$
that	Holding time of materials/printed
	components
t ^c	Time required for construction
w ^c	Average salary for construction workers
n ^c	Manpower at construction stage
MC	Cost of one tonne of material
EaC	Equipment cost (3D printing equipment
	and crane) i, $i = 1 \dots I$
EC	Unit energy cost
t ^{eq}	Time for the equipment operation i, i = 1 I during 3DP/assembly
(2)

PC_i	Power capacity (kW) of equipment i, $i =$
	11
EL_i	Useful life of equipment i, $i = 1 \dots I$
FC	Unit fuel (diesel) cost per liter
FCon _i	Fuel consumption factor for diesel machinery (e.g., crane) per hour i, $i = 1 \dots I$
$E_{i}^{el} \\$	Carbon emissions factor for electrical equipment i, $i = 1 \dots I$
E ^{dis}	Carbon emissions factor for diesel machinery $i, i = 1 \dots I$

$$= \begin{bmatrix} Research and Development + \\ Project Design \end{bmatrix} = \begin{cases} n^{rd} \times t^{rd} \times w^{rd} \\ + \\ n^{ds} \times t^{ds} \times w^{ds} \end{cases}$$

$$\begin{bmatrix} HandlingCost^{mat} + \\ HoldingCost^{mat} + \end{bmatrix}$$
(3)

$$\begin{aligned} \text{Logistics cost} &= \begin{bmatrix} \text{HoldingCost}^{Waste} + \\ \text{DisposalCost}^{waste} + \\ \text{SetUpCost}^{eq} + \\ \text{TransportationCost} + \\ \text{EnvironmentalCost} + \end{bmatrix} = \\ &= \begin{bmatrix} \sum_{i=1}^{I} \text{IHC}_{i}^{mat} \times Q_{i}^{mat} \times Q_{i}^{mat} + \\ \sum_{i=1}^{I} \text{IHC}_{i}^{comp} \times Q_{i}^{mat} \times T_{hold}^{mat} + \\ \sum_{i=1}^{I} \sum_{j=1}^{J} \text{DC}_{i}^{waste} \times Q_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \text{SUt}_{i}^{eq} \times \text{SUn}_{i}^{eq} \times n_{i}^{SU} \times w_{i}^{SU} + \\ \sum_{i=1}^{I} \sum_{j=1}^{J} \text{TC}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{TC}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{TC}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{E}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{E}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{E}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{mat} + \\ &= \begin{bmatrix} \sum_{i=1}^{I} \sum_{j=1}^{J} \text{E}_{ij} \times n_{ij}^{cont} \times \text{dis}_{ij}^{eq} \end{bmatrix} + \end{aligned}$$

$$\begin{aligned} \text{Construction cost} &= \\ & \begin{bmatrix} Labour \ cost \ + \\ Material \ cost \ + \\ Equipment \ cost \ + \\ EnergyConsumption \ + \\ Environmental \ cost \ (constr) \end{bmatrix} = \\ & \begin{bmatrix} t^c \times w^c \times n^c \ + \\ MC \times Q_i^{mat} \ + \\ \sum_{i=1}^{l} \frac{\text{EqC}_i}{EL_i} \times t_i^{eq} \ + \\ & \begin{bmatrix} \sum_{i=1}^{l} EC \times t_i^{eq} \times PC_i \ + \\ \sum_{i=1}^{l} FC \times t_i^{eq} \times FCon_i \end{bmatrix} \ + \\ & \begin{bmatrix} \left(\sum_{i=1}^{l} E_i^{el} \times t_i^{eq} \times PC_i \ + \\ \sum_{i=1}^{l} E_i^{dis} \times t_i^{eq} \times FCon_i \end{bmatrix} \right) \\ & \times CE_{CO_2} \end{aligned} \end{aligned}$$

(4)

3 Case study

This section illustrates the implementation of the proposed model to evaluate the feasibility of 3D printed projects compared to traditional construction methods. Case studies of residential buildings in Berlin and Beckum (Germany) are selected. The on-site 3D printed project, a collaborative effort involving PERI, COBOD, etc., was chosen due to its non-standard shape and on-site construction, aligning with industry interests. The project holds a building permit from the office in Beckum, providing valuable insights. The data from this project is sufficient for the case of 3DP off-site. A traditionally constructed house in Berlin, which utilized sand-lime blocks and had a similar floor area, was chosen for the comparison. These projects meet the selection criteria for the case study [27] because they are real examples from the same country, with similar currency, and quantum of work, and provide comprehensive data on various aspects. Even though the projects were implemented in differentsized cities (Berlin and Beckum), the cost calculations were based on average values for Germany. This includes factors such as material, equipment, and waste transportation, which often involve intercity logistics.

3.1 Data collection

To obtain a comprehensive set of data, a three-step data collection approach was employed [22]. The approach includes gathering data from open-source documents and articles, conducting surveys using a survey form, and conducting open-question interviews. The 3DP cases in this study utilized the BOD2 concrete printer, which was purchased by PERI. This printer model consists of a gantry system with multiple modules, enabling its use for both on-site concrete projects and offsite element production. The gathered data was then prepared for use in the model, which involved matching the obtained data with the parameters in the model.

3.2 Data processing

The prepared data is utilized in the model to calculate the different cost components. The cost of the project design and preparation stage is determined using Eq. (2), logistics cost is calculated using Eq. (3), and the construction cost is derived from Eq. (4). The calculated costs are summarized in Table 2. A pie chart is drawn to visualize the percentage of each cost component within the total cost for each case (Figure 2). Three cost elements (holding, set up, and construction environmental costs) are not included in the pie chart legend as their contribution to the total cost is negligible, close to or equal to zero.

The total cost for traditional construction amounts to €93,480, while on-site 3DP reaches €418,489 and off-site 3DP reaches €472,581. Upon conducting a comparative analysis, it becomes evident that approximately 80% and 71% of the total costs are attributed to the project preparation stage, which involves R&D. During this stage of construction 3DP development, each project is treated as unique and requires extensive preparation from the initial phase. Additionally, the active development of 3DP equipment necessitates significant investments in software and hardware updates for each project. However, as the technology matures and becomes more widely used, project designs can be applied to multiple housing projects using the same 3DP equipment, making the project preparation stage comparable to that of traditional construction. For example, establish a library of reusable designs from which the chosen project can be downloaded [28]. The logistic cost for off-site 3DP is higher due to the additional transportation of prefabricated components. On the other hand, on-site 3DP has a higher cost compared to traditional construction due to the transportation of massive 3DP equipment. However, more compact 3D printing robotic solutions are available in the market for use [20]. Moreover, despite the lower amount of materials involved, the material transportation costs of on-site 3DP are higher compared to traditional construction. This observation indicates that despite claims of traditional construction materials, such as cement, being suitable for 3DP technology [7], companies often procure materials from specialized 3DP organizations (e.g., equipment vendors often serve as the suppliers of raw materials) instead of purchasing them from the nearest convenient

company [29]. Previous studies have also highlighted the insufficiency of printing materials (e.g., Hossain, et al. [4], Zhang, et al. [30]).

In terms of construction cost, it accounts for approximately 7%, 10%, and 70% of the total cost in the cases of on-site 3DP, off-site 3DP, and traditional construction, respectively. Specifically, the construction cost of traditionally built projects is approximately 2.5 times higher than that of on-site 3DP construction due to its longer duration, which requires more manpower usage.



igure 2. Cost profile for three scenarios

Cost	3DP on-	3DP off-	Traditional
components	site, €	site, €	construction,€
Project	335 000	335 000	17 866,7
preparation	(80%)	(71%)	(19%)
Handling cost	852.6	2305.2	2091.6
Holding cost	91.8	248.4	91.8
Disposal cost	2121.2	284.1	3787.9
Set up cost	576.0	0.0	0.0
Transport. cost	47998.9	74028.5	3329.1
Environmental	5717.0	12596.2	568.1
cost(transport.)			
Logistics cost	57357.4	89462.5	9868.5
	(13%)	(19%)	(11%)
Labor cost	5280.0	7892.5	47025.0
Material cost	14976.0	15050.9	18720.0
Equipment	2400.0	18550.0	0.0
cost			
Energy	3226.9	5917.3	0.0
consumption			
Environmental	249.0	708.1	0.0
cost (constr.)			
Construction	26131.9	48118.8	65745.0
cost:	(7%)	(10%)	(70%)
Total:	418489.4	472581.2	93480.1

Table 2. Costs calculation

3.3 Sensitivity analysis using alternative scenarios



Figure 3. Influence of the variation of raw material quantity on its logistics cost



Figure 4. Influence of the variation of raw material quantity on its transportation cost



Figure 5. Influence of the distance variation on its transportation cost

The results presented in Table 2 indicate that, apart from the project preparation stage, logistics costs represent the largest proportion of the total cost in 3DP cases. The primary factor influencing logistics costs is the quantity of materials. Therefore, this variable was selected as the independent variable for sensitivity analysis. Recognizing the considerable range of distances between each pair of facilities in the three cases (spanning from 11 to 525km), a decision was made to standardize the distance between each pair of facilities to 100km. This was done to ensure that the calculations are not overly dependent on distance. Figure 3 illustrates the cost elements using clustered columns, while the total logistics costs are represented by lines. In terms of overall logistics cost, off-site 3DP exhibits significantly higher costs compared to the other two cases. It can be observed that the logistics cost for traditional and on-site 3DP is quite similar. Therefore, in this regard, 3DP technology can be considered very comparable despite the need for extensive equipment transportation. Regarding the cost components, transportation and holding costs contribute the most to logistics costs. In real case studies (Table 2), transportation costs tend to dominate due to the large distances between facilities. However, the sensitivity analysis reveals that the holding cost for off-site printed components is quite high. This is associated with the need to hold prefabricated components for 28 days to allow for sufficient strength gain before installation [21]. Implementing effective inventory management practices can help reduce holding costs. Additionally, the use of concrete additives can decrease the time required for strength gain, thus reducing holding costs. The charts indicate an intersection point when considering a small quantity of transported materials, where off-site 3DP demonstrates a more feasible result compared to on-site 3DP, with transportation costs slightly outweighing holding costs.

Therefore, further sensitivity analysis was conducted, focusing solely on transportation costs, including total transportation costs and costs elements associated with material and equipment transportation. Similarly, Figure 4 illustrates the cost elements of material and equipment transportation using clustered columns, while total transportation costs are represented by lines. The intersection between the lines representing on-site and off-site 3DP transportation costs indicates that off-site 3DP techniques can be considered feasible for small object printing, specifically for objects weighing less than 35 tonnes. In general, off-site 3DP becomes more attractive when the cost of delivering printed elements to the construction site is lower than transporting the massive 3D printing equipment along with raw materials.

Another significant factor that impacts transportation costs is the distance traveled. To evaluate the influence of distance variation on the total transportation costs, the distance between any two facilities is designated as the independent variable. Figure 5 presents the results, demonstrating the comparable feasibility of on-site 3DP and traditional construction methods.

4 Conclusions

This study aims to assess the costs associated with

3DP projects compared to traditional construction for low-story buildings. The cost assessment encompasses the project design stage, logistics, construction process with resources used, and waste disposal. Moreover, this model accounts for environmental factors. At the present stage of construction 3DP technology development, the cost of the project preparation stage is 19 times higher than that of traditional construction. However, the construction cost is 2.5 times lower compared to traditional methods. Sensitivity analysis, when compared to comparative analysis (Table 2), provides valuable insights revealing that logistics costs for traditional construction and on-site 3DP are relatively similar. The higher logistics costs of on-site 3DP observed in the comparative analysis can be attributed to the significant distances between suppliers and construction sites. This highlights the need to expand the 3DP materials market to encompass a broader range of traditional construction materials readily available within specific regions. By doing so, transportation costs can be reduced by sourcing materials from local suppliers near the construction site. Furthermore, the sensitivity analysis emphasizes that transportation and holding costs contribute the most to overall logistics costs, highlighting their significance as cost components. The off-site 3DP exhibits higher logistics cost due to holding costs. However, off-site 3DP is more feasible for small-scale projects, such as architecture elements printing, as the cost of delivering printed elements to the construction site is lower compared to transporting the massive 3DP equipment and raw materials. On-site 3D printing can be a competitive alternative to traditional construction methods for individual projects as well as large-scale developments. It is important to note that the model used in this study does not account for breakdowns caused by environmental conditions, equipment failures, or other technical faults.

This study benefits relevant stakeholders in the construction industry, as well as researchers, in the following ways: (1) it provides a comprehensive and structured approach for developing a cost estimation model; (2) the study develops a comprehensive model that plays a pivotal role in evaluating the feasibility of 3DP projects, covering all stages from design to realization; (3) by comparing 3DP scenarios with traditional construction methods, the study highlights the practicality and potential advantages of 3DP projects.

Future advancements in model development can incorporate specific project characteristics, such as geographical location, weather conditions, or remote accessibility, by including surcharge rates tailored to address these factors. To advance the integration of 3DP technology, the subsequent phase involves the technology sustainability assessment and creation of a strategic roadmap. This roadmap will incorporate scientific solutions and serve as a systematic guide, outlining the sequential utilization of the proposed mathematical models to evaluate their cost structure for technology evaluation.

Acknowledgement

This work was supported by the Guangdong Special Support Talent Program—Innovation and Entrepreneurship Leading Team under Grant [2019BT02S593], RGC Research Impact Fund [R7036-22] and RGC Theme-based Research Scheme [T32-707-22-N].

References

- S. El-Sayegh, L. Romdhane, and S. Manjikian, "A critical review of 3D printing in construction: Benefits, challenges, and risks," *Archives of Civil and Mechanical Engineering*, vol. 20, no. 2, pp. 1-25, 2020.
- [2] A. Paolini, S. Kollmannsberger, and E. Rank, "Additive manufacturing in construction: A review on processes, applications, and digital planning methods," *Additive manufacturing*, vol. 30, 2019, Art no. 100894.
- [3] Q. Munir and T. Kärki, "Cost analysis of various factors for geopolymer 3D printing of construction products in factories and on construction sites," *Recycling*, vol. 6, no. 3, 2021, Art no. 60.
- [4] M. A. Hossain, A. Zhumabekova, S. C. Paul, and J. R. Kim, "A Review of 3D Printing in Construction and its Impact on the Labor Market," *Sustainability*, vol. 12, no. 20, 2020, Art no. 8492.
- [5] P. Bedarf, A. Dutto, M. Zanini, and B. Dillenburger, "Foam 3D printing for construction: A review of applications, materials, and processes," *Automat Constr*, vol. 130, 2021, Art no. 103861.
- [6] C. Núñez, M. Regalado, and A. Gago, "3D Printing: An opportunity for the sustainable development of building construction," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2023, vol. 40: IAARC Publications, pp. 691-698.
- [7] A. Al Rashid, S. A. Khan, S. G. Al-Ghamdi, and M. Koc, "Additive manufacturing: Technology, applications, markets, and opportunities for the built environment," *Automat Constr*, vol. 118, 2020, Art no. 103268.
- [8] M. H. Ali, G. Issayev, E. Shehab, and S. Sarfraz, "A critical review of 3D printing and digital manufacturing in construction engineering," *Rapid Prototyping Journal*, vol. 28, pp. 1312-1324, 2022.
- [9] D. Won, B.-G. Hwang, S. Chi, and J.-L. Kor, "Adoption of Three-Dimensional Printing

Technology in Public Housing in Singapore: Drivers, Challenges, and Strategies," *Journal of Management in Engineering*, vol. 38, no. 4, p. 05022010, 2022.

- [10] M. Despeisse, M. Baumers, P. Brown, F. Charnley, S. J. Ford, A. Garmulewicz, S. Knowles, T. Minshall, L. Mortara, and F. Reed-Tsochas, "Unlocking value for a circular economy through 3D printing: A research agenda," *Technological Forecasting and Social Change*, vol. 115, pp. 75-84, 2017.
- [11] C.-C. Yeh and Y.-F. Chen, "Critical success factors for adoption of 3D printing," *Technological Forecasting and Social Change*, vol. 132, pp. 209-216, 2018.
- [12] A. L. M. Tobi, S. A. Omar, Z. Yehia, S. Al-Ojaili, A. Hashim, and O. Orhan, "Cost viability of 3D printed house in UK," *IOP Conference Series: Materials Science and Engineering*, vol. 319, no. 1, p. 012061, 2018.
- [13] H. Yang, J. K. H. Chung, Y. Chen, and Y. Li, "The cost calculation method of construction 3D printing aligned with internet of things," *EURASIP Journal* on Wireless Communications and Networking, vol. 2018, no. 1, p. 147, 2018.
- [14] D. Aghimien, C. Aigbavboa, L. Aghimien, W. D. Thwala, and L. Ndlovu, "Making a case for 3D printing for housing delivery in South Africa," *International Journal of Housing Markets and Analysis*, 2020.
- [15] Y. Weng, M. Li, S. Ruan, T. N. Wong, M. J. Tan, K. L. Ow Yeong, and S. Qian, "Comparative economic, environmental and productivity assessment of a concrete bathroom unit fabricated through 3D printing and a precast approach," *Journal of Cleaner Production*, vol. 261, p. 121245, 2020.
- [16] R. Allouzi, W. Al-Azhari, and R. Allouzi, "Conventional Construction and 3D Printing: A Comparison Study on Material Cost in Jordan," *Journal of Engineering*, vol. 2020, p. 1424682, 2020.
- [17] V. Markin, M. Krause, J. Otto, C. Schroefl, and V. Mechtcherine, "3D-printing with foam concrete: From material design and testing to application and sustainability," *J Build Eng*, vol. 43, 2021, Art no. 102870.
- [18] Y. Han, Z. Yang, T. Ding, and J. Xiao, "Environmental and economic assessment on 3D printed buildings with recycled concrete," *Journal of Cleaner Production*, vol. 278, p. 123884, 2021.
- [19] H. Abdalla, K. P. Fattah, M. Abdallah, and A. K. Tamimi, "Environmental Footprint and Economics of a Full-Scale 3D-Printed House," *Sustainability*, vol. 13, no. 21, 2021.
- [20] S. Besklubova, B. Q. Tan, R. Y. Zhong, and N. Spicek, "Logistic cost analysis for 3D printing construction projects using a multi-stage network-

based approach," Automat Constr, vol. 151, p. 104863, 2023.

- [21] Y. Fang and S. T. Ng, "Applying activity-based costing approach for construction logistics cost analysis," *Construction Innovation*, vol. 11, no. 3, pp. 259-281, 2011.
- [22] N. Spicek, S. Besklubova, and M. J. Skibniewski, "Benchmarking critical success factors in construction projects utilizing 3D printing technology," *International Journal of Applied Science and Engineering*, vol. 20, no. 4, pp. 1-18, 2023.
- [23] P. Krupík, "3D printers as part of Construction 4.0 with a focus on transport constructions," in *IOP Conference Series: Materials Science and Engineering*, High Tatras, Slovakia, 2020, vol. 867(1): IOP Publishing,
- [24] S. H. Khajavi, M. Tetik, A. Mohite, A. Peltokorpi, M. Li, Y. Weng, and J. Holmström, "Additive manufacturing in the construction industry: The comparative competitiveness of 3D concrete printing," *Applied Sciences*, vol. 11, no. 9, 2021, Art no. 3865.
- [25] P. Kerdlap, J. S. C. Low, and S. Ramakrishna, "Zero waste manufacturing: A framework and review of technology, research, and implementation barriers for enabling a circular economy transition in Singapore," *Resources, conservation and recycling,* vol. 151, 2019, Art no. 104438.
- [26] E. Kravchenko and S. Besklubova, "Value stream assessment of the sustainable concrete recycling process with sequestration of CO2 from flue gases," *Sustainable Production and Consumption*, 2024.
- [27] R. R. Ahmed and X. Zhang, "Multi-stage networkbased two-type cost minimization for the reverse logistics management of inert construction waste," *Waste Management*, vol. 120, pp. 805-819, 2021.
- [28] M. V. López, M. Carmen Sanz Contreras, and J. Ramon Osanz Díaz, "3D Printing in Presentation Architecture Projects," in *Putting Tradition into Practice: Heritage, Place and Design: Proceedings* of 5th INTBAU International Annual Event 5, 2018: Springer, pp. 636-644.
- [29] G. Sköld and H. Vidarsson. "Analyzing the potentials of 3D-printing in the construction industry." <u>https://publications.lib.chalmers.se/records/fulltext/</u>

218418/218418.pdf (accessed 30 September 2022.

[30] J. Zhang, J. Wang, S. Dong, X. Yu, and B. Han, "A review of the current progress and application of 3D printed concrete," *Composites Part A: Applied Science and Manufacturing*, vol. 125, 2019, Art no. 105533.

Predicting Indicators of Design Quality for Cast-in-Place Reinforced Concrete Structures Using Logistic Regression

Leonardo Garcia-Bottia¹ and Daniel Castro-Lacouture²

¹ Augment Digital Facilities Services, Atlanta GA, USA
²School of Construction Management Technology, Purdue University, USA leonar92@gmail.com, castro98@purdue.edu

Abstract -

This paper presents the application of the information contained in exchange standards to predict indicators of design quality for concrete-inplace reinforced concrete (CIP RC) structures early in the design process. A logistic regression model is applied to each node type of a frame structure: beamcolumn, slab-column, beam-slab, and beam-beam. All model results present the significance of the variable chosen, as well as the classification table with very high values of prediction accuracy. The results show how well the obtained models fit the data, and therefore may be used to estimate potential construction issues early in the process, based on the parameters of the design intent standard exchanges.

Keywords -

Design computing; Reinforced Concrete; Constructability; Logistic regression

1 Introduction

While there have been multiple studies around the evaluation of design indicators, [1] performed a comprehensive research and categorization effort of most of those available. Using three categories: functionality, build quality and impact, they identified multiple performance indicators of what potentially constitutes a good design, including layout, lighting and ventilation, energy, structural elements, building stability, comfort, and many others. The category that relates the most to the structure of the building from a design and construction standpoint is "Build Quality". However, these indicators were mostly developed to evaluate the performance of a design after it has been completed, and do not consider the valuable information available during design, particularly the one contained in intermediate model exchanges done during the design process. Regarding the category of build quality, research performed around the efforts of design professionals to purse enhanced effectiveness of their designs during construction, found design professionals that most consider that

constructability is a key indicator of the quality of the finished product or building [2].

The concept of "Constructability" or "Buildability" refers to the application of construction knowledge during the planning and design phases to make the construction process more efficient, practical, or sometimes even realistic [3]. This concept has been around for several years, and while the focus has changed through time, a review on the previous, current and future research done around it found that its application today is as important as ever for reasons including increased project complexities, great amount of ambiguous information, new relationships between stakeholders, and increased use of powerful methods and software tools [4]. Constructability can be approached from several angles, and pursue different benefits, including costs, time, labor, efficiency, and others. [5] grouped in seven themes the Construction Industry Institute constructability principles, and conducted a survey to estimate the potential and realized value of each of these groups. The group considering principles about designs that facilitate construction efficiency was ranked amongst the three with the highest potential value, which shows how much industry professionals value the positive impact that informed design decisions may have on the efficiency of the construction process.

Constructability is particularly important for Cast-in-Place Reinforced Concrete (CIP RC) buildings, because as a process that is very labor-intensive, it can benefit greatly from considerations taken during design that lead to a more efficient construction process [6]. [7] developed a constructability adviser system based on an object-oriented enriched CAD tool (a predecessor of BIM tools), to provide constructability feedback for CIP RC structures using criteria such as layout, dimensioning and construction methods. The paper identified two levels of reasoning when performing constructability analyses: reasoning about attributes of objects, and reasoning about relationships between attributes of objects. Although the research focused mostly on elements' dimensioning and forming methods due to their high impact on the costs, it identifies the most

important preliminary design variables that may be constrained or considered for constructability analyses: dimensions of elements, distance between elements, changes in dimensions and distances, concrete strength, quantity and type of reinforcement, and modularity. Out of these, the dimensions of elements and the quantity and type of reinforcement are applicable and relevant if the design intent and construction planning standards want to be used for analysis.

Section 2 of this paper presents a literature review of design indicators and constructability. Section 3 describes the research methodology used, including the data types, the development of the training database, and the development and results of the logistic regression model. Finally, section 4 presents the study conclusions.

2 Literature Review

Most of the research CIP-BIM oriented has focused on the reinforcement optimization of the elements using BIM models as noted in [8], on the assessment and recommendations of BIM capabilities to handle the concrete reinforcement supply chain [9], [10] and on defining the unique requirements CIP RC has regarding its modeling and processes on BIM [11]. In the assessments performed for the BIM capabilities of current tools, from the evaluated categories of design and modeling, editing, project management and interoperability, interoperability proved to be the weakest because of the lack of a standardized way to document and translate the information [9]. Nevertheless, these assessments were performed before the release of the latest ACI 131 documents which propose a standard way to exchange concrete reinforcement information [12]. Furthermore, although in reality CIP RC is monolithic, during the modeling it has to be broken down into members, which means that the delineation between such members is conceptual and not physical. There has also been research focused on reinforcement bars, particularly related to the impact of design on rebar productivity [13], and on optimizing cost by integrating rebar design and construction [14], [15].

Previous research [16], shows [17] that constructability of the design can be seen as a good quality indicator for CIP RC design and planning, particularly because it can use the information available during design intent and construction planning model exchanges, to contribute to efficiently achieve the intent during construction. Since the information about connectivity, dimensions and design intent reinforcement is something that is now available as part of the exchange models in a standard way, the congestion of the reinforcement, particularly in the areas between interacting elements, appears as an excellent alternative to measure the constructability of the design and planning, and to use as an indicator to develop predictions on potential future issues the design may encounter once it reaches more detailed stages. Current design tools allow the engineer to use the design intent to perform reasoning about attributes of objects as shown in blue in Figure 1 (such as a column, or a beam), but do not typically perform reasoning about relationships between attributes of objects such as the ones shown in yellow and red (beam-column or beam-slab interactions for instance). These are types of analyses that could be performed now that the design intent is available as part of a BIM model that holds the information about objects' connectivity and interacting volumes.



Figure 1. Types of design intent constructability analysis

Since the steel reinforcement ratio is typically a design decision based on code requirements and load demands, it is not a variable that can be modified for enhanced constructability. However, the way the steel ratio is achieved through the selection of diameter of bars, number of bars and bar separation is something more easily modifiable that has a direct impact on constructability; the use of fewer bars for a same ratio would derive in arrangements of larger diameters and spacings, thus reducing congestion and making the number of bars per volume of concrete a good estimator of congestion [18]. Therefore, the design indicator selected, "Constructability", will be estimated in terms of congestion as done by [18]. To create an estimating method applicable to several types of occurrences (element interaction types) with varying concrete element dimensions and steel distribution, the number of bars alone is not enough. Consequently, a similar concept to the steel volumetric ratio is proposed as the independent variable but using the number of bars per volume of concrete, thus accounting for most of the parameters aforementioned.

3 Research Methodology

A logistic regression model is developed, allowing the use of rich data contained within the exchange standards to estimate design indicators that inform the design and coordination processes of potential design issues during the exchanges. The focus of this application consists of identifying indicators that could be estimated from typical parameters available in the exchange files, create a database to train the model, and use it to inform the design process early on about issues that may arise during further phases of the project.

The study considers the intersections of pairs of elements for framed structures, including beams, columns, and slabs. The specific interactions considered are beam-column, beam-slab, beam-beam, and columnslab, and are shown in Table 1. transversal reinforcement, up to the heigh of the secondary beam, plus the secondary beam longitudinal reinforcement that enters and anchors in the main beam. The column-beam interaction considers all the column reinforcement, plus the beam longitudinal reinforcement that continues through the column. The column-slab interaction considers all the reinforcement of both elements. The same method considered for slabs could be easily extrapolated to footings and pile caps, since the reinforcement distribution is not that different between these elements. The properties required to estimate the congestion of the intersection are the number of bars each element contributes, and the volume of the intersection itself. As shown in the simplified data structure on Figure 2, the number of bars is derived from the design intent property set containing the design intent reinforcement information, and the volume intersection is derived from the geometric representation of the elements.



The beam-slab interaction considers all the beam longitudinal and transversal reinforcement, up to the thickness of the slab, plus the slab reinforcement that enters and anchors in the beam. The beam-beam interaction considers all the mean beam longitudinal and



Figure 2. Data structure of parameters required for indicator estimation

These properties could be easily extracted from an IFC file because of the way they have been standardized as proposed and aligned with ACI efforts. Different types of elements will have some of the properties listed that contribute to the number of bars (for example, slabs will have top and bottom bars and rebar mesh, while beams will have longitudinal bars, stirrups and ties). The following sections provide the detail of what reinforcement and parameters are considered for each type of intersection to estimate the indicator.

3.1 Training Database

The first step to generate the database was to define a representative number of reinforcement distributions for each element. Three beam sections were considered: small, medium, and large. For each of these sections, several options were generated varying the top and

1257

bottom reinforcement ratio (in one and two lines), the stirrup spacing, and the number of vertical legs. Combinations of these parameters were based on typical occurrences in practice, for example: stirrup spacings will typically be smaller where top reinforcement ratios are higher, which is near the supports. For each ratio, two alternatives were proposed: more smaller bars, or fewer bigger bars. This is a concept directly related to constructability: several times it will be more constructible to use fewer bigger bars that allow more spacing and lead to less congestion. Table 2 shows the three slab thicknesses considered: small, medium, and large. For each of these sections, several options were generated varying the top and bottom reinforcement ratios (assumed equal in both directions). Combinations of these parameters were based on typical occurrences in practice. For each ratio, two alternatives were proposed: more smaller bars, or fewer bigger bars.

A similar approach was followed for beams, where three typical section sizes were considered, each with three top and three bottom reinforcement ratios (minimum, average, and maximum), each with three to four typical stirrup spacings for the section, and each with one to two number of legs, for a total of 64 combinations (beams). For columns, three typical section sizes were considered, each with three reinforcement ratios (minimum, average, and maximum), each with three typical stirrup spacings for the section, and each with one to three number of tie legs, for a total of 54 combinations (columns). See Table 2. The combinations of the parameters were based on typical occurrences in practice, for example higher number of legs for higher reinforcement ratios on columns, or smaller stirrup spacings for higher top reinforcement ratios on beams.

 Table 2. Representative Slab Sections and Parameters for Database

	Slabs	5			Be	ams				Co	lumns		
Thick	ρ	# of	Section	Тор	Bott	Stirrup	Vert.	# of	Section	ρ	Tie	Tie	# of
	(%)	slabs		ρ	.ρ	space	legs	beams		(%)	space	legs	col.
				(%)	(%)								
4"	0.5	9	8"x12"	0.5	0.5	3"	2	16	12"x12"	1.5	2"	2	18
	1.0			1.0	1.0	6"				5.0	4"		
	1.5			2.0	2.0	12"				8.0	6"		
6"	0.5	9	16"x24"	0.5	0.5	3"	3	24	18"x18"	1.5	2"	2	18
	1.0			1.0	1.0	6"	3			5.0	4"	3	
	1.5			2.0	2.0	12"	4			8.0	6"	4	
						24"							
8"	0.5	9	24"x36"	0.5	0.5	3"		24	24"x24"	1.5	3"	3	18
	1.0			1.0	1.0	6"	4			5.0	6"	4	
	1.5			2.0	2.0	12"	5			8.0	12"	5	
						24"							

Afterwards, logical occurrences of intersections of these elements were created. If, for example, the 8"x12" beam section was combined with the 12"x12" column section, this generated $16 \times 18 = 288$ possible interactions. Some combinations were not considered because they would not normally occur in practice, such as a 24"x24" column with a 4" slab. Once the database was built, the

value of congestion as defined previously (number of bars in the intersection divided by the concrete volume of the intersection) was calculated for each of the interactions, using the parameters and relationships illustrated in Figure 2. Since these points will constitute the base to build the model, it is necessary to identify whether or not each of them is considered to have or not constructability issues.

A value of 1 is assigned to those occurrences with constructability issues, while a value of 0 is assigned to those without constructability issues. Table 3 presents the total number of interactions evaluated, the number of interactions considered to have constructability issues per criteria, and number of interactions without issues. The next sections explain in detail the three criteria used to determine whether each of these interactions was constructible or not.

Table 3.	Database	interactions	classification

T	Total	Iı Con:	nteractio structab	Interactions without		
Interaction	actions	Smin Criteria	min ρmin iteria Critoria (Total	Issues
Beam-Col	2,736	144	120	208	336	2,400
Col-Slab	1,134	150	0	333	414	720
Beam-Slab	1,440	164	54	75	245	895
Beam-Beam	2,752	128	0	224	322	2,430

3.1.1 Minimum Separation (Smin)

This criterion evaluated for each of the interactions that the reinforcement could physically and logically fit within the node, by ensuring minimum spacing was provided in critical cases. For the beam-column interaction, it was evaluated whether the longitudinal beam reinforcement could fit through the column reinforcement and ties, with a 1/8" tolerance. For the column-slab interaction, it was evaluated whether the slab reinforcement could fit through the column reinforcement and ties, with a 1/8" tolerance. For the beam-slab interaction it was evaluated whether the spacing between beam stirrups and anchoring slab reinforcement was at least 1", to allow the concrete to be placed and the largest size of aggregate to pass. For the beam-beam interaction, it was evaluated if the secondary beam anchoring reinforcement would fit through the main beam reinforcement and stirrups, with a 1/8" tolerance. Any interaction that did not satisfy these conditions, was assigned a value of 1, thus classifying it as an interaction with potential constructability issues.

3.1.2 Maximum Volumetric ratio (pmin)

This criterion was based on the ACI maximum ratio for column reinforcement. The ACI sets a maximum 8% steel ratio reinforcement in columns for longitudinal rebars mainly because above this number they consider the element to be hardly constructible [19]. If this limit is added to the maximum shear reinforcement caused by the minimum allowed separation, a value between 16% to 20% is obtained. Therefore, any intersection with a volumetric steel ratio greater than 16%, was assigned a value of 1, thus dimming it as an intersection with potential constructability issues.

3.1.3 Visual/Manual Inspection

Finally, the remaining intersections were visually inspected to determine whether the node or edge would present constructability issues based on the number of bars. It was found that intersections tend to present constructability issues at numbers greater than 60 bars per cubic feet. These intersections found to have potential constructability issues were assigned a value of 1.

3.2 Design Indicator Estimating Model

The model selected was logistic regression, because it fits the goal of the study: to estimate whether there will be an issue or not with an indicator based on parameters obtained from the standardized exchange models. More specifically, to estimate the probability that for a certain type of intersection, there will be a constructability issue based on the design intent. The procedure finds the best fitting curve by transforming the y-axis, odds of congestion, to a transformed logarithm log(odds of congestion / (1 - odds of congestion)). This new axis now goes from -infinity to +infinity, with all the data, previously lying at 1 or 0, now lying at +infinity and infinity. Then a line is fit to this data, and its coefficients are determined based on a linear model using the transformed y-axis. To transform the line from the transformed y-axis to the initial y-axis, the transformation $y = e^{\log(odds)}/1 + e^{\log(odds)}$ is used. After this transformation, the line becomes an s-shaped curve. To find the best fitting line, the method uses the concept of maximum likelihood. The procedure projects the original data points (located at - and + infinity) onto the candidate line, and is then transformed to the original axis. The likelihood of the line is the sum of the probabilities of the points after being projected onto the curve and transformed to the original axis. This line is rotated multiple times recording its likelihood, after which the best fitting line is obtained by selecting the model with the highest likelihood. Finally, since this is a classification problem (1 or 0), a threshold value, typically 0.5, is used to classify a new point as 1, congested, or 0, not congested. Based on this threshold value a weighted accuracy is calculated, which indicates the accuracy of the model to predict the points in the database as they were defined.

3.2.1 Beam-Slab Model

Figure 3 shows the logistic regression model for the beam-slab intersection.



intersection

The coefficients of the regression are shown in Table 4, along with the standard error, the Wald number (a measurement of the precision of the estimate), and the p-value.

Table 4. Regression Coefficients for Beam-Slab Intersection Model

	# Iter		20	20		0.05	
	coeff	s.e.	Wald	p-value	exp(b)	lower	upper
intercept	-9.001	0.486	342.367	0.000	0.000		
var 1	0.199	0.012	283.554	0.000	1.220	1.192	1.249

These values show that the variable chosen (number of bars per CF of concrete at the intersection) is statistically significant for this model. Equation 1 describes the model (best fitted curve), and can be used to calculate the probability of congestion, PC, based on the number of bars per cubic feet at the intersection, n. In other words, this model allows to estimate the probability that the intersection will present a constructability issue, which is the selected indicator.

$$PC(n) = \frac{e^{(-9.001+0.199n)}}{1+e^{(-9.001+0.199n)}}$$
(1)

Table 5 shows the classification table for the model, based on a cutoff value of 0.5. This value is the threshold value, above which points are classified a success, or with constructability issues, and below which points are classified a failure, or without constructability issues. The values shown correspond to a typical cutoff value of 0.5 or 50%. The weighted accuracy of the model at predicting success and failure is 88%, which is a good indicator of how well the model fits the behavior of the data.

Table 5. Classification Table for Beam-SlabIntersection Model

Intel section wroaci						
	Obs	Obs	Total			
-	Succ	Fail	Totai			
Pred Succ	180	72	252			
Pred Fail	108	1080	1188			
Total	288	1152	1440			
Accuracy	63%	94%	88%			
Cutoff	0.5					
AUC	0.944					

3.2.2 Beam-Column Model

Figure 4 shows the logistic regression model for the beam-column intersection.



Figure 4. Logistic regression model for beamcolumn intersection

The coefficients of the regression are shown in Table 6, along with the standard error, the Wald number, and the p-value.

Table 6. Regression Coefficients for Beam-ColumnIntersection Model

	# Iter		50	50		0.05	
	coeff	s.e.	Wald	p-value	exp(b)	lower	upper
intercept	-8.163	0.391	435.032	0.000	0.000		
var 1	0.164	0.008	381.940	0.000	1.178	1.159	1.197

These values show that the variable chosen (number of bars per CF of concrete at the intersection) is statistically significant for this model. Equation 2 describes the model (best fitted curve), and can be used to calculate the probability of congestion, PC, based on the number of bars per cubic feet at the intersection, n. In other words, this model allows to estimate the probability that the intersection will present a constructability issue, which is the selected indicator.

$$PC(n) = \frac{e^{-8.163 + 0.164n}}{1 + e^{-8.163 + 0.164n}}$$
(2)

Table 7 shows the classification table for the model, based on a cutoff value of 0.5. The weighted accuracy of the model at predicting success and failure is 94%, which is a good indicator of how well the model fits the behavior of the data.

Table 7. Classification Table for Beam-Column Intersection Model

	Oha	Oha	
	Obs	Obs	Total
	Succ	Fail	Iotai
Pred Succ	252	84	336
Pred Fail	84	2316	2400
Total	336	2400	2736
Accuracy	75%	97%	94%
Cutoff	0.5		
AUC	0.977		

3.2.3 Column-Slab Model

Figure 5 shows the logistic regression model for the column-slab intersection.



The coefficients of the regression are shown in Table 8, along with the standard error, the Wald number, and the p-value.

Table 8. Regression Coefficients for Column-Slab Intersection Model

		# lter		20		0.05	
	coeff	s.e.	Wald	p-value	exp(b)	lower	upper
intercept	-6.094	0.329	342.209	0.000	0.002		
var 1	0.115	0.006	318.458	0.000	1.122	1.107	1.136

These values show that the variable chosen is again statistically significant for this model. Equation 3 describes the model for this intersection type and can be used to calculate the probability that the intersection will present a constructability issue.

$$PC(n) = \frac{e^{-6.094 + 0.115n}}{1 + e^{-6.094 + 0.115n}}$$
(3)

Table 9 shows the classification table for the model, based on a cutoff value of 0.5. The weighted accuracy of the model at predicting success and failure is 94%, which is a good indicator of how well the model fits the behavior of the data.

 Table 9. Classification Table for Column-Slab

 Intersection Model

	Obs Succ	Obs Fail	Total
Pred Succ	340	83	423
Pred Fail	74	637	711
Total	414	720	1134
Accuracy	82%	88%	86%
Cutoff	0.5		
AUC	0.938		

3.2.4 Beam-Beam Model

Figure 6 shows the logistic regression model for the beam-beam intersection.



The coefficients of the regression are shown in Table 10, along with the standard error, the Wald number, and the p-value.

Table 10. Regression Coefficients for Beam-Beam Intersection Model

	# lter		20	20		0.05	
	coeff	s.e.	Wald	p-value	exp(b)	lower	upper
intercept	-7.587	0.339	501.575	0.000	0.001		
var 1	0.182	0.009	417.556	0.000	1.200	1.179	1.221

These values show that the variable chosen is again statistically significant for this model. Equation 4 describes the model for this intersection type and can be used to calculate the probability that the intersection will present a constructability issue.

$$PC(n) = \frac{e^{-6.094 + 0.115n}}{1 + e^{-6.094 + 0.115n}}$$
(4)

Table 11 shows the classification table for the model, based on a cutoff value of 0.5. The weighted accuracy of the model at predicting success and failure is 94%, which is a good indicator of how well the model fits the behavior of the data.

Table 11. Classification Table for Column-Slab Intersection Model

	Obs Succ	Obs Fail	Total
Pred Succ	230	42	272
Pred Fail	92	2388	2480
Total	322	2430	2752
Accuracy	71%	98%	95%
Cutoff	0.5		
AUC	0.960		

4 Conclusions

This paper presented the application of the information contained in the exchange standards to predict indicators of design quality for CIP RC structures early in the design process.

The paper started with a review of applicable design indicators for CIP RC related to the design intent and construction planning communication. Based on the review, constructability was found to be a good indicator of design quality, given that it relates the design result to how efficient is it to achieve it during construction and ensure the good performance of the structure as specified by the design.

To measure the constructability the parameter of congestion was proposed, given that more congested nodes tend to be harder to fabricate and place. Congestion is defined as the number of bars in the node per unit of volume of the node. This parameter can be calculated based on the parameters and properties shared during the design intent and construction planning exchanges. Afterwards, a database of representative beams, columns and slabs was generated to train the predictive algorithm. For each node in the database, geometric, volumetric, and engineering criteria were used to define whether the node was likely to have issues with construction, which constitutes a binary classification model.

Finally, a logistic regression model was applied to each node type of a frame structure: beam-column, slabcolumn, beam-slab, and beam-beam. All model results presented the significance of the variable chosen, as well as the classification table with very high values of prediction accuracy. The results obtained show how well the obtained models fit the data, and therefore may be used to estimate potential construction issues early in the process, based on the parameters of the design intent standard exchanges.

Future work involves the development and inclusion of further exchange models used in other parts of the CIP RC supply chain. The methodology can also be applied to other projects and CIP RC structures in order to extend the reach of the findings and develop more comprehensive implementation methods, and size-base estimations of the value of implementation. The methods are easily extensible to other tools and platforms, since they are developed with a generic approach and only the testing is done using specific tools. Furthermore, these methods may be adapted to other contexts, such as countries where BIM implementation has not been as advanced as it has in companies with heavy IT capabilities; or CIP RC bridges, where the development of standards poses other challenges and requirements. The model for prediction of constructability issues may be extended to include more CIP RC element interactions, and further refined as it is used in practice and more data becomes available.

5 References

- A. Suratkon, C.-M. Chan and S. Jusoh, "Indicators for Measuring Satisfaction Towards Design Quality of Buildings," *International Journal of GEOMATE*, 2016.
- [2] D. Arditi, A. Elhassan and Y. C. Toklu, "Constructability Analysis in the Design Firm," *Journal of Construction Engineering and Management*, pp. 117-126, 2002.
- [3] Construction Industry Institute, "Constructability: a primer," Publication 3-1, Austin, Texas, 1986.
- [4] D. Kifokeris and Y. Xenidis, "Constructability: Outline of Past, Present and Future Research," *Journal of Construction Engineering and Management*, p. 143(8): 04017035, 2017.
- [5] G. Jergeas and J. V. d. Put, "Benefits of Constructability on Construction Projects," *Journal of Construction Engineering and Management*, pp. 127(4): 281-290, 2001.
- [6] M. Fischer and C. B. Tatum, "Characteristics of Design-Relevant Constructability Knowledge," *Journals of Construction Engineering and Management*, pp. 132(3): 253-260, 1997.
- [7] M. Fischer, "Automating Constructability Reasoning with a Geometrical and Topological Project Model," *Computing Systems in Engineering*, pp. (4); 179-192, 1993.
- [8] M. Mangal and J. Cheng, "Automated Optimization of Steel Reinforcement in RC Building Frames Using Building Information Modeling and Hybrid Genetic Algorithm," *Automation in Construction*, pp. 39-57, 2018.
- [9] S. Aram, C. Eastman and R. Sacks, "Utilizing BIM to Improve the Concrete Reinforcement Supply Chain," *Computing in Civil Engineering ASCE*, pp. 333-340, 2012.

- [10] S. Aram, C. Eastman and R. Sacks, "Requirements for BIM Platforms in the Concrete Reinforcement Supply Chain," *Automation in Construction*, pp. 1-17, 2013.
- [11] R. Barak, Y. Jeong, R. Sacks and C. Eastman, "Unique Requirements of Building Information Modeling for Cast-In-Place Reinforced Concrete," *Journal of Computing in Civil Engineering ASCE*, pp. 64-74, 2009.
- [12] ACI Committee 131, "Guide to Use of Industry Foundation Classes in Exchange of Reinforcement Models ACI 131.2R-17," American Concrete Institute (ACI), 2017.
- [13] A. Mallya, K. Varghese and V. Reja, "Impact of Reinforcement Design on Rebar Productivity," 40th International Symposium on Automation and Robotics in Construction (ISARC 2023), 2023.
- [14] Z. Rahimi and M. Maghrebi, "Minimizing rebar cost using design and construction integration," *Automation in Construction*, p. 104701, 2023.
- [15] C. Zheng, C. Yi and M. Lu, "Integrated optimization of rebar detailing design and installation planning," *Automation in Construction*, pp. 32-47, 2019.
- [16] T. Hartmann and M. Fischer, "Supporting the constructability review with 3D/4D models," *Building Research & Information*, pp. 35(1): 70-80, 2007.
- [17] Y. Zhong and P. Wu, "Economic sustainability, environmental sustainability and constructability indicators related to concrete and steel projects," *Journal of Cleaner Production*, pp. (108) 748-756, 2015.
- [18] A. Kaveh, M. Kalateh-Ahani and M. Fahimi-Farzam, "Constructability optimal design of reinforced concrete retaining walls using a multiobjective genetic algorithm," *Stuctural Engineering and Mechanics*, pp. Vol 47 No 2: 227-245, 2013.
- [19] ACI, "Building Code Requirements for Structural Concrete," American Concrete Institute, Farmington Hills, MI, USA, 2014.

A Generic 3D Printing Life Cycle Assessment (LCA) Framework for AEC Applications

Bharadwaj R. K. Mantha¹, Ala Sati¹, Fatma Hosny¹, Mohamed Abdallah¹, and Saleh Abu Dabous¹

¹Department of Civil and Environmental Engineering, University of Sharjah, United Arab Emirates. rmantha@sharjah.ac.ae; asati@sharjah.ac.ae; fmohamed@sharjah.ac.ae; mabdallah@sharjah.ac.ae; sabudabous@sharjah.ac.ae;

Abstract -

Amidst the increasing adoption of threedimensional printing (3DP) in the architecture, engineering, and construction (AEC) industry, there exists a notable research gap concerning the quantification of its environmental impact. More importantly, there is a lack of a generic framework that can be employed for different material types and methods. Therefore, this research aims to develop a generic 3DP life cycle assessment (LCA) framework pertaining to the AEC industry. To substantiate the viability of the proposed framework, a meticulous case study was conducted. Focused on the LCA of a concrete 3DP process employed in constructing a twostory residential villa in the United Arab Emirates (UAE), the case study employed the contour crafting 3DP technique. A detailed cradle-to-construction LCA was executed using a functional unit of 1m³. Data compilation involved synthesizing information from literature sources and utilizing DesignBuilder™ software. The embodied carbon analysis of the case study vielded insightful results, indicating that the contour crafting technique emitted approximately 103,135 kg of carbon. Significantly, concrete emerged predominant construction material, as the contributing approximately 52% to the total equivalent over the various life cycle stages. Future work warrants further investigation into the mitigation measures to enhance the environmental performance of 3DP within the AEC industry. Additionally, the research encourages the exploration of alternative 3DP construction techniques and diverse project types, thereby broadening the applicability of the developed framework. This research provides a foundation for more sustainable practices and fosters further exploration of the 3DP implementation within the AEC industry. Keywords -

3D printing; LCA; Construction industry; Embodied carbon

1 Introduction

The inherent nature of the architecture engineering and construction (AEC) industry makes it one of the significant contributors to resource utilization. Therefore, understandably, the environmental impact is considerably high when compared to other industries. According to the statistics, it accounts for about 40% of global energy consumption, 28% of global greenhouse gas (GHG) emissions, and a significant amount of waste is generated [1]. The core of the AEC industry is the use of cement-based and concrete materials. With the growing need for new structures, the utilization of these materials is increasing. The production and utilization of concrete materials in the AEC have been shown to result in detrimental environmental impacts. According to Andrew [2], during concrete production, a large amount of carbon dioxide (CO₂) is released and represents 4% to 5% of worldwide emissions of CO₂.

In recent years, there has been a rise in interest in three-dimensional printing (3DP) technology for automating concrete construction. 3DP is the process of slicing a three-dimensional (3D) computer-aided design (CAD) model into two-dimensional (2D) layers and sequentially printing the materials to construct the full product, layer upon layer. The 3DP process has been widely used in various disciplines and enterprises. However, studies have revealed that 3DP is not commonly employed in buildings, and its applications remain limited [3]. 3DP allows for faster structure development while also reducing building time, labor costs, and waste generation [4]. According to Tinoco et al., [3], it can reduce construction time by 50 to 70 %, labor costs by 50 to 80 %, and waste production by up to 60%. All of these capabilities have raised interest in 3DP in the AEC industry. Despite the research and implementation of 3DP in the field and the documentation of its technological and economic advantages, there has been limited quantitative research on 3DP's environmental performance. Several existing review studies, such as [5], have attempted to examine

the applicability of life cycle assessment (LCA) for 3DP.

To further emphasize the need and highlight the gap in the existing literature, this research first systematically and critically analyzed the extant literature on 3DP's LCA, the summary of which was presented earlier. Quantifying this further, several recently published articles based on data from scientific databases such as Web of Science and Scopus were retrieved since 2011. It was observed that approximately 40 research articles were published on the topic of LCA in 3DP. The search strategy employed a comprehensive set of keywords to obtain relevant research articles. For example, the keyword search string criteria used are as follows: Keywords ("LCA" OR "life cycle" OR "life cycle analysis" OR "Environment*" AND "assess*") AND ((3d OR 3-d OR 3d-) AND print*) OR ("Additive manufacturing") AND ("Construct*" OR "Build*") AND ("Concrete" OR "cement-based" OR "cementitious" OR "geopolymer" OR "cement")]. The 'OR' and 'AND' are boolean operators used to combine or exclude specific terms to refine search results. Specifically, the usage of 'OR' between two terms will return results that include either one of the terms or both. For example, "LCA" OR "life cycle" will return results containing either "LCA", "life cycle" or both. On the other hand, the usage of "AND" between two terms will return results that include both terms. For example, "3d AND print" will return results that contain both "3d" and "print." To summarize, the usage of "OR" broadens the search whereas the usage "AND" narrows down the search. The overarching idea of the designed search string is to retrieve literature specifically focused on the LCA of the concrete 3DP process. It encompasses various terms related to life cycle analysis, environmental assessment, 3DP technologies (including alternative spellings), additive manufacturing, and specific materials like concrete, cement-based materials, and geopolymers. This comprehensive search strategy aims to identify the most relevant publications related to the environmental impact of 3DP in the context of concrete construction processes. However, the number of papers containing LCA results is less than 30, most of which were focused on specific material types. Therefore, this research concentrated on LCA for 3DP regardless of material type.

Additionally, [3] a recently published review research focused on the LCA of cementitious materials for concrete 3DP. They found that there are still very few published papers with LCA results of concrete 3DP. The total number of papers from 2016 to 2021 is only 15 papers, which are [1,6–19]. After reviewing these articles, it was observed that the majority of 3DP LCA investigations were conducted recently, beginning in 2016. Europe, where 3DP research is in its mature phases, has made the largest contributions to the field. China is the second most important contributor to this sector. Other countries appear to be less prominent, and only two studies were conducted in the Middle East region.

For the LCA details, most articles employed cradleto-gate system boundaries. There was, however, much diversity in the selection of the functional unit, making it difficult to compare based on this factor. Despite the minimal number of research articles, there was a wide variety of applications. However, the majority of the research used 3DP to print walls. The most researched material was concrete with cement, aggregate, additives, and water. The most popular database utilized in the studies for the life cycle inventory (LCI) is Eco Invent (https://ecoinvent.org/), which has different versions. Gabi database (https://sphera.com/productsustainability-gabi-data-search/) is the second most used one. In addition, some investigations utilized data from the literature as well. It is thus evident that the existing literature lacks a generic framework to conduct LCA for 3DP in the AEC industry that is not specific to a material type, functional unit, and application.

To address the gap in the literature, this research proposes a generic framework methodology to perform a Life Cycle Assessment (LCA) of the concrete 3DP process in the AEC industry to evaluate its environmental impacts. The evaluation will analyze one specific technique of 3DP used widely in the AEC industry. Therefore, the objectives of this research are to a) Develop a generic framework to conduct LCA of concrete 3DP pertaining to the AEC industry; b) Implement and validate the developed framework through a case study to investigate the CO_2 emissions and identify the materials that contribute most significantly to CO_2 emissions.

2 Proposed 3DP LCA Framework

Figure 1 shows the overview of the proposed research framework to achieve the above-mentioned objectives. Broadly categorized, the framework follows a three-stage procedure. The first stage is to define the purpose of the assessment, the system boundaries, and the functional unit for comparison. The second stage is to collect and quantify data regarding the inputs and outputs of the system. This includes raw material extraction, energy use, water use, emissions, and waste generation. The third and the final stage is to assess the potential environmental impacts. Each of the following tasks involved in these stages is discussed in detail in the sub sections below.

2.1 Define Goal and Scope

The goal and scope definition stage is the foundational stage of an LCA, where the parameters of the system are established. Literature reviews and indepth analyses of previous studies can be used as tools to identify the construction scenarios. In this stage, the goals of the assessment are clearly stated, which includes defining the reasons for conducting the research and the intended application of the results as suggested in [8]. In addition, define the system boundaries, which may range from the materials production to the end-of-life disposal. This delineation is crucial as it sets the limits for the study and ensures the consistency of the data collected. The functional unit, which is the measure to which all inputs and outputs are related, is also defined during this stage, providing a reference to which the performance of the product system is compared.



Figure 1. Overview of the proposed 3DP LCA framework.

2.2 Life Cycle Inventory

Life cycle inventory (LCI) is a critical stage of the LCA process. The objective of this stage is to collect and gather all relevant data including inputs and outputs of a system. In the context of this study, inputs may include building materials used and the energy consumed during the construction and operation of the construction project. Similarly, outputs can be referred to as the emissions from on-site construction processes and energy use during the operation and maintenance phase of a project. The aim is to create an inventory of every element and energy flow that goes into and out of the product's life cycle, from the extraction of raw materials through the production and use phases to end-of-life. Typically, this can be retrieved from the project documentation such as work breakdown structure (WBS) and bill of quantities (BOQ).

2.3 Life Cycle Impact Analysis (LCIA)

Life cycle impact analysis (LCIA) is the third and final stage in the LCA process, where the environmental impacts associated with the inputs and outputs identified during stage 2, namely LCI, are evaluated. The primary purpose of LCIA is to assess the magnitude and significance of potential environmental impacts using the data gathered in the LCI stage. This analysis involves several steps, including the selection of impact categories, classification, and characterization. The impacts associated with a product or process throughout its entire life cycle on the environment are categorized into different impact categories, each representing a specific aspect of environmental concern. The choice of impact categories depends on the goals and scope of the LCA study. Common LCA impact categories include global warming potential, ozone depletion, and eutrophication. Classification refers to the assignment of inventory data to the identified impact categories. Characterization refers to quantification of impact magnitude often resulting in a single score per impact category.

For example, in a building project, LCIA could assess the impacts of material extraction, energy use, waste generation, and emissions throughout the building's life span, from the construction phase to the demolition or end-of-life (EOL) phase. Moreover, it could investigate different construction methods and techniques. LCIA stage translates inventory data into a form that can be more easily understood and acted upon. This helps decision-makers identify the most significant environmental issues and the life cycle stages where improvements can be made for more sustainable product systems.

3 3DP LCA Framework Validation

The objective of this section is to evaluate the proposed general 3DP LCA framework through its implementation in a case study and provide a comprehensive evaluation of the environmental impacts of 3D contour crafting in residential construction. To achieve this, each of the subsections below follows a systematic approach, beginning with the selection of the 3D construction technique and moving through to the LCIA. This implementation ensures a thorough understanding of both environmental impacts and potential performance enhancements associated with the construction of a two-story residential villa using the contour crafting technique. This approach aligns with the standards set by ISO14044 and ISO14045, focusing on cradle-to-construction analysis. More specifically, sections 3.1 and 3.2 delve into the construction method selection and elaboration on the selected case study. Sections 3.3 to 3.5 follow the implementation and hence the validation of the threestage generic 3DP LCA framework developed and discussed in section 2.

3.1 Construction method selection

Based on the existing literature, contour crafting printing is one of the most widely adopted and demonstrated construction techniques [20]. Hence, the contour crafting technique is employed for this illustration. However, any 3DP technique can be selected to implement the specified steps within this framework, as the framework is designed to adapt AEC to various printing techniques.

After choosing the technique, the following needs to be done a) define the goal and scope, b) obtain the life cycle inventory, and d) perform impact assessment. The different materials and processes were quantified using the LCA systematic framework. To standardize the method of evaluating the burden on the environment, ISO14044 and ISO14045 were created by the International Organization for Standardization (ISO), addressing the associated environmental impacts and identifying possible performance enhancements during the lifespan of a system [21,22].

Two strategies are commonly used to evaluate the AEC industry: cradle-to-grave and cradle-to-construction. The first strategy evaluates all processes and materials comprehensively, whereas the second strategy concentrates on specific components of project elements, such as materials [16]. This research used the cradle-to-construction strategy, where DesignBuilderTM software [23–25] and OneClick[®] software [26–28] were used to perform the LCA analysis.

3.2 Description of Case Study

A two-story simulated residential villa was selected for the case study, as shown in Figure 2. Table 1 shows the different characteristics of the chosen villa. It was assumed that the villa was located in Sharjah, United Arab Emirates (UAE) following the typical dimensions and characteristics of the construction in the region. Based on the objective of this research, the selected structure was proposed to be built using the 3D contour crafting construction technique. The villa is mainly a concrete structure, with a plot area of 272 m² and a total built-up area of 394 m². For the 3D contour crafting, the dimensions of the elements were simply the length of the wall × the width of the wall, which was 30 cm for external walls and 20 cm for internal walls. The timeframe includes all building elements, such as heating, ventilation, and air conditioning (HVAC) systems, lighting, and finishes, to study their impacts.

Division	Description		
Building	Villa (G+1)		
Site	Sharjah, UAE		
Lifespan (years)	40		
Plot area (m ²)	272		
Total height (m)	8		
Ground floor (m ²)	197		
First floor (m ²)	197		
Total (m ²)	394		



Figure 2. A 3D model of the selected case study villa developed in DesignBuilderTM.

3.3 Goal and Scope

The first phase of any LCA is to specify the goal and scope of the proposed system. As discussed, the case study implementation evaluates the environmental impact of the concrete 3DP technique on a residential 2storey villa. The functional unit was chosen to be 1 m^3 to study the environmental impact of the proposed system. It was normalized to allow for a fair and meaningful comparison between the different materials. A cradle-toconstruction LCA was performed in this research, including material extraction, material production and manufacturing, and building construction. Figure 3 shows the general boundaries of the evaluated system in this research. Whereas Figure 4 illustrates the system boundaries for contour crafting, where the system includes different processes and materials such as material extraction, material production and manufacturing, and construction.



Figure 3. The general system boundary



Figure 4. System boundaries of contour crafting for the case study villa

3.4 Life Cycle Inventory

The input data for the contour crafting technique mentioned in the following sections was gathered from the literature [16,29]. The data will include the material mix, amount of each material, transportation, energy consumption, and equipment utilized. Only the production and construction phase elements are presented on the technique flowchart (cradle-toconstruction). In the context of the 3D printing construction methods, the concrete mix composition differs from traditional concrete mixes used in conventional construction methods [30,31]. In this scenario, the structure will be built using 3DP technology called contour crafting. In this technology, the structural elements will be built on-site. The functional unit combines all the building elements. The final materials and concrete mix were extracted after reviewing existing literature within the region and by examining a similar project facing comparable environmental conditions [11,29]. Through an in-depth review of their material choices, quantities, and a focus on their consistent watercement ratio of 0.4 across both studies, the concrete mix used in this scenario is high-performance concrete consisting of fly ash, micro-silica, plasticizer, perlite, microfiber, accelerator, and water. Detailed information amount of material obtained on the from DesignBuilder[™] for one functional unit is presented in Table 2. The concrete mix includes innovative materials that serve particular purposes in enhancing performance characteristics. For instance, perlite, which is a lightweight aggregate, was used to replace the sand in the mix. In addition, no steel will be added in the scenario because the mixed used was selected to be a selfreinforced mortar [24], includes components like microfiber and microsilica, which can significantly enhance the tensile strength and durability of the concrete. Moreover, the concrete mix was designed to be a lightweight mix by adding perlite.

 Table 2 Data Inventory of contour crafting for the studied villa per functional unit

Material	Quantity		
Fly ash (kg)	175		
Microsilica (kg)	88		
Plasticizer (kg)	9		
Perlite (kg)	68		
Microfiber (kg)	1		
Accelerator (kg)	9		
Cement (kg)	614		
Water (kg)	246		
Insulation (m)	0.08		

3.4.1 3DP system

A large-scale 3D printed construction requires an extrusion process, in which the structure is constructed by adding layers of the prepared mortar through a nozzle. For 3DP, the Putzmeister MP25 machine was considered to mix and pump the concrete. The ABB robot (IRB6700) was used to control and automate the nozzle movement. Table 3 summarizes the electric consumption required to print the desired structure based on the machine's characteristics [32,33].

Table 3 Energy consumption of the 3DP system

Equipment	Power required (kW)	Electricity consumption (kWh)		
Mixture and pump	7.38	1.55		
Robotic arm	3.4	0.71		
Total		2.26		

3.5 Life Cycle Impact Analysis (LCIA)

The environmental impacts of the proposed villa were evaluated using the integration between DesignBuilderTM and OneClick[®] (Figure 5). DesignBuilderTM was used to create the 3D model of the proposed villa, including all elements such as construction materials, structural systems, and energy systems. In addition, check the performance of the model, including energy, carbon, lighting, and comfort performance. The model extracted from DesignBuilderTM was exported to OneClick LCA[®], which is a comprehensive tool that easily integrates the DesignBuilderTM outputs to assess the environmental impacts. In this study and based on the location, the LCIA was based on the European Standard EN 15978 for conducting whole building LCA [34].



Figure 5. Input-Output diagram of software combinations

4 **Results and Discussion**

This section represents the LCA results of embodied carbon for global warming over the life cycle stages. LCA of CO_2 is performed in this research to assess the performance of the chosen villa. The goal of the LCA is to evaluate the embodied carbon in kilograms (kgs). The life cycle of CO_2 emission phases can be described as "cradle-to-construction", measuring the emission during material extraction, material production and manufacturing, and building construction.

The LCA results can be divided into different divisions: life cycle stages, contributing materials, annual impacts, elements classifications, and resource types. In this research, the concrete mix of contributing materials is analyzed. The estimated embodied carbon data shown below is based on bulk carbon data obtained from the Bath ICE and other data sources. These results do not cover the embodied carbon associated with building services such as lighting and HVAC equipment. Figure 6 shows the embodied carbon breakdown of contour crafting. The results show that the concrete mix contributed the most, with around 51.73%. Figure 7 shows the LCA for global warming in kg concrete over the life-cycle phases of the villa. Moreover, the total embodied carbon is about 103,135 kg.



Figure 6. Embodied carbon percentage breakdown for materials



Figure 7. Embodied carbon breakdown in kilogram CO₂ per unit area

The proposed model is designed to evaluate various villas' life cycle CO₂ emissions. Moreover, selecting the most sustainable material and evaluating the buildings' commitment to the CO₂ emission standards laid down in the UAE Green Building Certification Systems is helpful. Most of the data related to the life cycle phases of the DesignBuilderTM are provided by AEC industry professionals working on projects in the UAE or relevant literature. Moreover, the DesignBuilderTM model can be used to obtain additional data related to construction cost and energy consumption.

In order to validate the findings of our study, the total embodied carbon per unit function was compared to the existing literature within the same region constructing the same building type (Residential villa) and adopted the cradle-to-site system boundary, as demonstrated in Table 4.

Our study, employing the contour crafting 3DP technique as the construction method, resulted in an embodied carbon value of 261.76 kg CO_2 eq/Unit function. In contrast, Abdalla et al. (2021) reported values of 608.55 kg CO_2 eq/Unit function for 3D printing method and 1154.2 kg CO_2 eq/Unit function for conventional construction method.

	Embodied carbon	
Study	(kg CO ₂ eq/Unit	
	function)	
3D printing [Current study]	261.76	
3D printing [11]	608.55	
Conventional method [11]	1154.20	

Table 4 Comparison analysis

The comparison highlights that contour crafting resulted significantly lower embodied carbon emissions in comparison to conventional method and even when compared to similar 3D printing method documented in the literature.

5 Conclusion

This study introduces an innovative generic framework methodology designed for the comprehensive life cycle assessment (LCA) of concrete threedimensional printing (3DP) processes within the architecture, engineering, and construction (AEC) industry, with a specific focus on evaluating their environmental impacts. A noteworthy departure from previous research, which often concentrated on specific life cycle phases, materials, or methods, this study fills a crucial gap by developing a generic framework on 3DP LCA pertaining to the AEC industry. The framework's effectiveness is demonstrated through a meticulously chosen case study involving a 2-storey residential villa with a plot area of 272 m² in the emirate of Sharjah, United Arab Emirates (UAE), a locale synonymous with extensive concrete utilization. The contour crafting 3D printing method is employed in the case study, utilizing cradle-to-construction strategy to assess the а environmental impact exclusively up to the completion of the construction phase.

Embarking on the case study, the analysis of embodied carbon emissions reveals that the contour crafting technique emits approximately 103,135 kg CO2. Notably, concrete mix emerges as the predominant contributor, accounting for 52% of the total equivalent concrete mix over the life cycle stages under consideration. This information provides a nuanced understanding of the specific environmental implications associated with the chosen 3D printing method, aiding in the broader comprehension of the technology's ecological footprint.

While the research makes significant strides in addressing the environmental impact assessment of 3D printing in the AEC industry, it acknowledges a primary limitation, which is the scarcity of relevant literature and comparative studies. This scarcity poses a challenge to gathering the requisite data needed for a more comprehensive analysis. To address this, our future research agenda encompasses a commitment to expanding the framework's application to encompass various concrete 3DP technologies, such as D-shape, and diverse building types, including commercial structures. By broadening the scope, we aim to enhance the generalizability of our findings and contribute to a more nuanced understanding of the environmental impacts associated with different 3DP technologies and building typologies.

In addition to addressing the data limitations, our future work emphasizes the exploration of mitigation measures to curtail the environmental footprint associated with 3D printing in the AEC industry. By identifying strategies to minimize adverse environmental impacts, we hope to provide actionable insights that can guide decision-makers in adopting more sustainable practices. This forward-looking approach aligns with our overarching goal of not only identifying environmental challenges but also actively contributing to solutions that promote sustainability within the AEC sector.

References

- Mohammad M. Masad E. and Al-Ghamdi S.G.
 3D Concrete Printing Sustainability: A Comparative. *Buildings* 2020;10.
- [2] Andrew R.M. Global CO2 emissions from cement production. *Earth System Science Data* 2018;10:195–217. https://doi.org/10.5194/essd-10-195-2018.
- [3] Tinoco M.P. de Mendonça É.M. Fernandez L.I.C. Caldas L.R. Reales O.A.M. and Toledo Filho R.D. Life cycle assessment (LCA) and environmental sustainability of cementitious materials for 3D concrete printing: A systematic literature review. *Journal of Building Engineering* 2022;52. https://doi.org/10.1016/j.jobe.2022.104456.
- [4] Lu B. Weng Y. Li M. Qian Y. Leong K.F. Tan M.J. and Qian S. A systematical review of 3D printable cementitious materials. *Construction* and Building Materials 2019;207:477–90. https://doi.org/10.1016/j.conbuildmat.2019.02.1 44.
- [5] Saade M.R.M. Yahia A. and Amor B. How has LCA been applied to 3D printing? A systematic literature review and recommendations for future studies. *Journal of Cleaner Production* 2020;244:118803.

https://doi.org/10.1016/j.jclepro.2019.118803.

[6] Agustí-Juan I. and Habert G. An environmental perspective on digital fabrication in architecture and construction. *CAADRIA* 2016, 21st International Conference on Computer-Aided Architectural Design Research in Asia - Living Systems and Micro-Utopias: Towards *Continuous Designing* 2016:797–806. https://doi.org/10.52842/conf.caadria.2016.797.

- Kuzmenko K. Gaudillière N. Feraille A. Dirrenberger J. and Baverel O. Impact: Design With All Senses. Springer International Publishing; 2020. https://doi.org/10.1007/978-3-030-29829-6.
- [8] Han Y. Yang Z. Ding T. and Xiao J. Environmental and economic assessment on 3D printed buildings with recycled concrete. *Journal* of Cleaner Production 2021;278:123884. https://doi.org/10.1016/j.jclepro.2020.123884.
- [9] Long W.J. Lin C. Tao J.L. Ye T.H. and Fang Y. Printability and particle packing of 3D-printable limestone calcined clay cement composites. *Construction and Building Materials* 2021;282:122647. https://doi.org/10.1016/j.conbuildmat.2021.1226 47.
- [10] Muñoz I. Alonso-Madrid J. Menéndez-Muñiz M. Uhart M. Canou J. Martin C. Fabritius M. et al. Life cycle assessment of integrated additive– subtractive concrete 3D printing. *International Journal of Advanced Manufacturing Technology* 2021;112:2149–59.

https://doi.org/10.1007/s00170-020-06487-0.

- [11] Abdalla H. Fattah K.P. Abdallah M. and Tamimi A.K. Environmental footprint and economics of a full-scale 3d-printed house. *Sustainability (Switzerland)* 2021;13:1–19. https://doi.org/10.3390/su132111978.
- [12] Agustí-Juan I. Müller F. Hack N. Wangler T. and Habert G. Potential benefits of digital fabrication for complex structures: Environmental assessment of a robotically fabricated concrete wall. *Journal of Cleaner Production* 2017;154:330–40. https://doi.org/10.1016/j.jclepro.2017.04.002.
- [13] Agustí-Juan I. and Habert G. Environmental design guidelines for digital fabrication. *Journal* of Cleaner Production 2017;142:2780–91. https://doi.org/10.1016/j.jclepro.2016.10.190.
- [14] Yeon J. Rew Y. Kang J. and Kunhee C. Life Cycle Assessment-based Feasibility Study of Spall Damage Rehabilitation using 3D Printing Technology. 54th ASC Annual International Conference Proceedings 2018.
- [15] Long W.J. Tao J.L. Lin C. Gu Y. cun. Mei L. Duan H.B. and Xing F. Rheology and buildability of sustainable cement-based composites containing micro-crystalline cellulose for 3D-printing. *Journal of Cleaner Production* 2019;239:118054. https://doi.org/10.1016/j.jclepro.2019.118054.
- [16] Alhumayani H. Gomaa M. Soebarto V. and Jabi

W. Environmental assessment of large-scale 3D printing in construction: A comparative study between cob and concrete. *Journal of Cleaner Production* 2020;270:122463. https://doi.org/10.1016/j.jclepro.2020.122463.

- [17] Yao Y. Hu M. Di Maio F. and Cucurachi S. Life cycle assessment of 3D printing geo-polymer concrete: An ex-ante study. *Journal of Industrial Ecology* 2020;24:116–27. https://doi.org/10.1111/jiec.12930.
- [18] Weng Y. Li M. Ruan S. Wong T.N. Tan M.J. Ow Yeong K.L. and Qian S. Comparative economic, environmental and productivity assessment of a concrete bathroom unit fabricated through 3D printing and a precast approach. *Journal of Cleaner Production* 2020;261:121245. https://doi.org/10.1016/j.jclepro.2020.121245.
- [19] Muñiz M.M. Chantin M. Vintila C.R. Fabritius M. Martin C. Calvo L. Poudelet L. et al. Concrete hybrid manufacturing: A machine architecture. *Procedia* CIRP 2020;97:51–8. https://doi.org/10.1016/j.procir.2020.07.003.
- [20] Cabibihan J.-J. Gaballa A. Fadli F. Irshidat M. Mahdi E. Biloria N. Mansour Z. et al. A guided approach for utilizing concrete robotic 3D printing for the architecture, engineering, and construction industry. *Construction Robotics* 2023;7:265–78. https://doi.org/10.1007/s41693-023-00103-9.
- [21] ISO14044. Environmental Management—Life Cycle Assessment—Requirements and Guidelines. 2006.
- [22] ISO14045. Environmental Management -Ecoefficiency Assessment Of Product Systems -Principles, Requirements And Guidelines. 1st ed. 2012.
- [23] Abu Dabous S. Shanableh A. Al-Ruzouq R. Hosny F. and Khalil M.A. A spatio-temporal framework for sustainable planning of buildings based on carbon emissions at the city scale. *Sustainable Cities and Society* 2022;82:103890. https://doi.org/10.1016/j.scs.2022.103890.
- [24] Ghanbari M. Rusch R. and Skitmore M. BIMbased environmental assessment of residential renovation projects during the operational phase. *Architectural Engineering and Design Management* 2024. https://doi.org/10.1080/17452007.2024.2313026.
- [25] Sadati S.E. Rahbar N. and Kargarsharifabad H. Energy assessment, economic analysis, and environmental study of an Iranian building: The effect of wall materials and climatic conditions. *Sustainable Energy Technologies and Assessments* 2023;56. https://doi.org/10.1016/j.seta.2023.103093.

- [26] Sravani T. Prasanna Venkatesan R. and Madhumathi A. A comparative LCA study of passive cooling roof materials for a residential building: An Indian Case study. Mater. Today Proc., vol. 64, 2022, p. 1014–22. https://doi.org/10.1016/j.matpr.2022.05.089.
- [27] Grygierek K. and Ferdyn-Grygierek J. Analysis of the Environmental Impact in the Life Cycle of a Single-Family House in Poland. *Atmosphere* 2022;13.
 - https://doi.org/10.3390/atmos13020245.
- [28] Kurian R. Kulkarni K.S. Ramani P.V. Meena C.S. Kumar A. and Cozzolino R. Estimation of carbon footprint of residential building in warm humid climate of india through BIM. *Energies* 2021;14. https://doi.org/10.3390/en14144237.
- [29] Mohammad M. Masad E. and Al-Ghamdi S.G. 3d concrete printing sustainability: A comparative life cycle assessment of four construction method scenarios. *Buildings* 2020;10:1–20. https://doi.org/10.3390/buildings10120245.
- [30] Panda B. and Tan M.J. Experimental study on mix proportion and fresh properties of fly ash based geopolymer for 3D concrete printing. *Ceramics International* 2018;44:10258–65.

https://doi.org/10.1016/j.ceramint.2018.03.031.

- [31] Malaeb Z. AlSakka F. and Hamzeh F. 3D Concrete Printing: Machine Design, Mix Proportioning, and Mix Comparison Between Different Machine Setups. Elsevier Inc.; 2019. https://doi.org/10.1016/B978-0-12-815481-6.00006-3.
- [32] Agustí-Juan I. Hollberg A. and Habert G. Earlydesign integration of environmental criteria for digital fabrication. 2018.
- [33] Szabó L. Hidalgo I.J. Císcar J.C. Soria A. and Russ P. Energy consumption and CO2 emissions from the world cement industry. Brussels, Belgium: 2003.
- [34] Sustainability of construction works. Assessment of environmental performance of buildings. Calculation method,EN 15978:2011. European Committee for Standardization; 2012. https://doi.org/https://www.en-standard.eu/bsen-15978-2011-sustainability-of-constructionworks-assessment-of-environmentalperformance-of-buildings-calculationmethod/#:~:text=This%20European%20Standar d%20specifies%20the,the%20outcome%20of% 20the%20assessment.

Justification of Construction 4.0 maturity model with a case study of a data-driven façade company

Orsolya Heidenwolf¹ and Ildikó Szabó²

¹Doctoral School of Economics, Business and Informatics, Corvinus University of Budapest, Hungary orsolya.nagy@uni-corvinus.hu, ildiko.szabo2@uni-corvinus.hu

Abstract

The European Parliament released the 2030 Digital Compass, identifying the construction sector as one of the five key ecosystems in the digital transformation process. A Construction 4.0 maturity model (C4M) was created to lead companies through their digital transformation journey. This research aimed to justify the model categories of the developed C4M using a qualitative case study methodology. The case study highlighted the practical relevance of the model by identifying three areas within the C4M that were the main drivers of digital transformation: cloud-based systems, knowledge management and system integration to drive digitalisation further. It was revealed that the digital processes in the design, production, and construction phases are connected to each other through system integration.

Keywords -

construction 4.0, maturity model, digital transformation, data-driven organisation, AI

1 Introduction

In recent years, an emerging number of studies have been published on Construction 4.0 (C4) phenomena. In this paper C4 is defined as a new construction ecosystem that incorporates (1) integrated technologies and cyberphysical systems [1,2], (2) methods and processes [3,4], and (3) human resource competencies [3]. These three pillars are supported by automation and data analytics. In this ecosystem achieving sustainability is a crucial objective for industry players which can be attained by improving the efficiency and productivity of construction processes. However, only a few research efforts have addressed the systematic integration of C4 into the construction organisations' life-cycle and the tools that can support companies in this endeavour [5].

The European Parliament has released the 2030 Digital Compass, which identifies the construction sector as one of the five key ecosystems in the digital transformation process. The report emphasises that the Construction industry has had the lowest productivity of

all major sectors in the last 20 years. Executives believe that new production technologies and digitalisation can be the drivers of change [6].

The strategic implementation of C4 requires a shift in management mindset towards digitalisation [10] and to effectively use a toolset that supports digital transformation. One such tool is a maturity model, which helps assess a company's evolution stage-by-stage in a specific area [7]. Building Information Modelling (BIM) can also contribute to digitalization endeavours but researchers studied BIM Maturity Matrix from the perspective of designers rather than general contractors or subcontractors [8,9]. However, these matrixes often lack of important domains from the contractor's perspective such as technology management, business applications, innovation, or business processes. Contractors play a crucial role in the built environment as they are responsible for creating the physical product, such as buildings. Therefore, our research focuses on them.

This paper presents our Construction 4.0 Maturity Model (C4M) and how the qualitative case study of a Hungarian façade contractor revealed the relevance of each category element, and the connections between them. The selected company places significant emphasis on digitalisation, in particular, data-driven artificial intelligence (AI) solutions and its development. This case study helped us to justify the C4M developed from the literature review with a real-world example.

2 Research Methodology

The research applied Design Science (DS) framework to prepare the C4M. DS methodology provides steps to design artifacts such as algorithms, applications, information systems and so on. In DS research the design "focuses on the use of scientific principles, technical information and imagination in the definition of a structure"[10]. It illustrates the relevance underpinned by the business need arising from the environment and the rigour in the form of applicable and new knowledge.

In our research, problem space defines the environment which encompasses the construction firm's

existing and planned technological infrastructure that enhances its productivity. Businesses are inclined to lead this transformation process seamlessly. This induces the need for developing a C4M and justifying it. The knowledge base foundation includes the steps for developing the model (scope-design-populate-testdeploy-maintain) [11]. The DS process steps are introduced in the following paragraphs.

2.1 Problem identification

C4 presents numerous challenges at the organizational level [12]. The maturity model can be a useful tool in addressing these challenges and facilitating companies' digital transformation. A solution is required that considers companies' technological perspectives rather than solely focusing on the digital maturity of a construction project.

2.2 Objective and Solutions

The objective of the C4M is to conceptualise and quantify C4 in construction business activities and make them measurable over time. Our long-term purpose is to develop a rules-based information system. It will measure how ready a company is to adopt a particular technology. This C4M model serves as its basis, extended with a fuzzy inference engine built on an ontology[13]. DS recommends developing an artifact iteratively using a lot of feedback. This paper presents only one iteration. The case study will determine if the model accurately represents the company's digital maturity aspects and if it is practical to use from a market professional's perspective.

2.3 Development

During the model development phase, we aimed to identify distinct categories (level 1) underlying families (level 2), and elements (level 3). The developed model is based on the BIM Maturity matrix [9], which plays a central role in the digitalisation of the construction industry. In the second phase of model development, the main elements of the BIM Maturity Matrix model have been examined and expanded to include areas that can be important for the digitalisation of an organisation using Industry 4.0 and other construction-specific maturity models.

A literature review was conducted during the development process using the following keywords to identify relevant literature: Industry 4.0 maturity model, Industry 4.0 readiness, Construction maturity model, Construction 4.0 readiness, and Construction business model. The literature review identified 22 relevant articles in the field of Industry 4.0 and only 3 in the construction industry. A matrix was created during the

review to establish the main pillars of the C4M. These pillars supported the next phase of development where further literature was identified using the following keywords: pillar name + maturity model/readiness model, pillar name + construction + maturity model/readiness model. The maturity models identified in the literature were pruned and the model elements were categorised using ontology development methodology. Ontology is a "formal and explicit specification of a shared (shared) conceptualization"[14]. The steps of this methodology are described in detail in [13]. These steps led to the creation of the C4M families and elements.

2.4 Justification

To justify the integrity and applicability of the model to the environment, a qualitative case study was conducted in four steps: (1) designing the case study protocol, (2) conducting the case study, (3) analysing the evidence, and (4) developing the conclusion [15]. The data was collected through interviews. During the case study protocol phase, we developed a sampling strategy, to select a business that covers all areas of our model including planning and design, construction and production. We prioritised selecting a company that demonstrated a high level of maturity in terms of digitalisation compared to other market players.

During the second phase of the case study, two interviews were conducted. Firstly, the CEO of the company explained the main steps taken in their digital transformation journey. Secondly, the C4M elements were discussed in detail to assess the importance of each model element on a scale of 0 to 5. In the subsequent round of the case study, the Chief Technology Officer (CTO) was interviewed. During the analysis phase, we prepared the interview transcripts and created a matrix in Microsoft Excel to justify the model. The matrix had rows representing categories, families and elements (C4M elements) and the two columns (CEO and CTO) were filled with the data collected from the transcripts. During this phase, our main objective was to identify correlations between different processes and determine the drivers of digital transformation. Thus we created the process map of the case study and analysed the steps in detail. Based on the results, we drew conclusions, and considerations for future research.

3 Maturity Model

C4M has six main pillars: Technology Management and Business Applications [7,16], Culture and People Management [7,16–20], Collaboration and Communication [17], Technology for Automation [17,18], Innovation [17] and Change Management and Processes [7,16–18]. Technology for Automation category encompasses the use of industry-specific robotics [21–23], Building Information Modelling (BIM) and Digital Twin, Internet of Things (IoT) and sensing systems [21–23], and data-driven technologies [21,23] to enhance construction projects and production efficiency. The model is presented in more detail by [13]. Figure 1 illustrates the C4M and the Technology for Automation category in detail.



Figure 1 C4M Technology for Automation's families

4 Case study

The company operates with a team of 30 employees at its headquarter, with an additional 40 employees working in production and construction. The core team comprises individuals from various fields, including finance, logistics and procurement, sales, design, project management, CEO, ESG Manager (Innovation Manager), marketing, and software development. The company follows a matrix organizational structure. It is important to note that nearly one-third of the headquarters' employees work in software development.



Figure 2 The digital transformation process map

The strategic goal of the digital transformation journey was threefold: (1) automatic façade design, (2) production plan automation, and (3) predictive decisionmaking based on project data. The interviews revealed four development phases. Phase 1 involves developing a general IT infrastructure, Phase 2 includes creating a knowledge base which was used to educate employees about internal processes, Phase 3 focuses on system integration, and Phase 4 involves system scale-up and AI development based on data. The purpose of the case study was to test the relevance and existence of each category of the C4M, based on the company's digital transformation journey.

A process map in the context of C4M elaborated from the case study outlines the key steps involved in the digital transformation represented in Figure 2. This process map is introduced in the following paragraphs.

4.1 Technology Management and Business Applications

This category refers to the systematic coordination and application of technology to achieve its strategic goals and objectives [13]. This category examines how effectively integrated digital technologies can support the company's digital processes.

The journey towards digital transformation begins with defining a data-driven business strategy. During the system development process, the focus was on creating an IT framework, including both software and hardware, that enables data-driven operations. Specifically, the goal was to convert Excel sheets to an SQL database management system. The requirements for this can be defined by the IT governance policy, which also guides predictive data-driven decision-making. Cybersecurity should be included in this policy as it is becoming an increasingly important concern in the construction industry, particularly as core processes are digitised [24]. However, the case study did not reveal any particular focus on this area. Therefore, we did not include it in the process map.

"The biggest challenge was to build the system from the ground up to be able to apply data science "(CEO)

The importance of facilitating data-driven operations is underscored by the development of a cloud-based workflow management system that is compatible with existing software. These cloud-based workflows were later extended from core operational processes to projectrelated design and construction processes.

System integration and interoperability are significant challenges in digitisation, which require to rationalize the IT portfolio for efficiency. System integration refers to the process of combining individual component subsystems into a unified system to achieve its desired functionality effectively [25]. Defining system integration requirements is essential for the effective deployment of integrated systems. This task involves evaluating the compatibility and capacity of the hardware infrastructure, establishing testing criteria, selecting appropriate software, and validating integration to ensure system reliability. It is essential that functional requirements are clearly specified and that a cost-benefit analysis is carried out to assess the financial implications. Additionally, scrutinising the software data structure is necessary to enable seamless data flow. This systematic approach establishes clear integration requirements for successful operation.

During this process, nearly 100 construction-specific software programs were tested. Although individual software solutions have their advantages, the company decided to develop its own system called Dijkstra due to the lack of a comprehensive solution.

4.2 Culture and People Management

This category refers to maintaining a positive and productive work environment to drive organizational performance, with a focus on performance management, training and development, and employee engagement [13]. This category assesses the effectiveness of knowledge management within an organisation. It includes whether employees have the confidence and knowledge to use digital technologies effectively in the corporate environment.

The process map presents a thorough method for improving organisational capabilities through a systematic approach. It starts by establishing system integration requirements as guidelines for the knowledge base. Then, a knowledge base is developed, which includes technological instructions and case studies, to lay the foundation for digital expertise. Defining digital expertise fields and developing IT skills further empowers employees to navigate the digital landscape. Internal education initiatives, such as training sessions and onboarding programs, help ensure alignment with company processes and technologies.

The integration of operational processes streamlines workflow and enhances organizational efficiency. This integration aims to consolidate various activities within the organization, promoting cohesion and synergy. Subsequently, the implementation of a Key Performance Indicator (KPI) system enables the continuous measurement of employee efficiency. This system captures designers' hourly tasks, providing valuable insights for controlling and project pricing. The generated data facilitates pricing negotiations and automates invoicing processes, contributing to improved decision-making and resource allocation within the organization. A Business Intelligence (BI) system leverages this data for visualizing financial reports and project progress. Challenges were raised in defining KPIs for construction field workers, particularly in measuring their soft skills.

4.3 Technology for Automation

The Technology for Automation category assesses the use of several technologies, including data-driven technologies, robotics, BIM and Digital Twin, and IoT and sensing systems. The process flow diagram shows the close relationship between automation technologies and system integration requirements, which are defined in parallel.

4.3.1 BIM and Digital Twin

The design of the automation strategy is heavily influenced by the BIM environment, and the creation of an appropriate enterprise BIM environment is based on the system integration requirements. Special emphasis should be placed on ensuring interoperability among software. For instance, contractors often need to extract quantitative data from customer-provided IFC format and transfer it to Excel. To process the data, supplementary software is required, such as Revit, ArchiCAD, or even Inventor for certain manufacturers, which enables direct transmission of the data to production. Speckle's approach offers a viable way to address these data communication issues. Furthermore, the BIM ecosystem necessitates ongoing development of relevant IT expertise and internal training on software processes. If the data flow within the established ecosystem is continuous, it can provide real-time input data for data analytics. This data enables the company to track and measure project information.

The company utilises software such as Revit, ArchiCAD, Grasshopper, and Speckle to create project plans at Level of Development (LoD) 500, which represents the As-Built Stage of Modelling. This includes model elements with precise size, shape, location, quantity, and orientation for fabrication and assembly processes at the shop drawing level, with technical data connected to the modelled elements. Constant training and development have resulted in a high level of software usage, which has significantly improved its quality. The interviews identified barriers within the BIM and Digital Twin domain that are hindering the company's progress and the demonstration of BIM's added value. Firstly, Clients are increasingly requesting more extensive documentation from contractors, while not having the same expectations for their as-built BIM models. This is due to the contractors' far more advanced level of expertise in BIM compared to theirs. Secondly, there are data interchangeability issues due to challenges in software interoperability, especially between Revit and ArchiCAD. Finally, contractors may face additional challenges due to the large amount of data generated by designers.

These challenges resulted in the decision to integrate the Industry Foundation Class (IFC) model into the selfdeveloped Dijkstra platform using Speckle's solution. Figure 3 shows how this viewer enables the team to update and track the project through the model, visualising the state of each façade component.



Figure 3 Dijkstra platform and project tracking

4.3.2 Data-driven technologies

The development of a data-driven technology ecosystem is a sequential process. First, the data structure for analytical purposes must be designed, taking into account system integration requirements and guidelines from IT governance policies. It is crucial to consistently gather data by monitoring both project and organizational operations to preserve effectiveness.

To meet the requirements of data-driven technologies, the company focused on building the appropriate data structure. This structure would enable the system to be used for data analysis and BI reporting in the future. To achieve this, the company linked an SQL database to the SharePoint system and then to the Dijkstra interface.

The company is now developing an AI chatbot called Mici, which will integrate existing databases through the Dijkstra interface and extract information from background databases through human communication.

Recent research has identified three potential directions for data mining: data mining for intelligence, digitalization, and automation [26]. The company has identified two of these directions; and has developed three goals from an AI perspective:

- The purpose of this project is to use data mining techniques to access internal company SQL data and provide Project Managers with relevant information and knowledge.
- Additionally, the project includes the development of an AI chatbot to aid designers in documenting their design process and translating it into ArchiCAD.
- The project also involves developing a system for generating data to facilitate communication with the platform, recording data via mobile phone, and

managing the platform.

The study of how employees view these technologies highlighted the need for thorough training programs. If employees do not receive adequate feedback from AI, there is a possibility that they may return to using traditional job methods. The CEO highlighted that the internal IT staff has substantial concerns about future job loss due to AI.

4.3.3 Human-machine interfaces and Wearable devices

Cloud-based workflows and communication culture support the usage of human-machine interfaces. However, complete enterprise integration requires continuous monitoring of technology usage to track and record project data.

In the early stages of digital transformation, the company focused on prioritising mobile devices to ensure data availability across various workspaces and construction sites. Their mobile application was then updated to allow for microphone-based data entry, which was found to be more efficient than on-site typing. Furthermore, the company developed VR and AR areas to facilitate model visualisation. Continuous education is necessary to support the use of these devices for constant monitoring. Management control also influences the level of usage.

4.3.4 IoT and Sensing System

The integration of Internet of Things (IoT) and sensor systems is another key aspect that can facilitate real-time data collection in automated processes. This integration is a key aspect as it defines the ability of the robotic environment and its software to integrate with preexisting systems.

4.3.5 Robotics

Establishing a robotic manufacturing environment requires several preparatory stages. It is crucial to align the automation strategy, including the development of robotics, with the system integration requirements. The ongoing evolution of existing automated processes facilitates the development of a robotic environment.

In the field of robotics, the company prioritised the development and automation of production processes. They emphasized that the effectiveness of robotics depends on the user's expertise, thus the need for continuous training for welders operating Computerised Numerical Control (CNC) machines. These machines are used in computerized manufacturing, where preprogrammed software and code control the movement of production equipment. The CEO highlighted the growing importance of sustainability in automation and robotics. This includes the development of machines capable of managing waste produced within factories and on projects.

4.4 Change Management and Processes

This category refers to the degree of alignment between organizational and digital processes that enable the company to promptly and efficiently address customer requirements, as well as to integrate or introduce new processes [13].

Defining system integration requirements precedes the integration of corporate processes, enabling the extension of efficiently functioning operational processes to project-specific workflows. The development of automated processes emerges as a key element on the process map, and is driven by the automation strategy. IoT and sensor data contribute to seamless automated processes and continuous data flow. Additionally, support for development efforts across multiple domains is required, including internal training initiatives and continuous improvement of IT skills.

4.5 Collaboration and communication

Collaboration and communication cover how the organisation works, interacts, and cooperates within its physical, and digital environment and throughout its supply chain [13]. This category measures the effectiveness of workers in sharing information within and across organizations through digital technologies. Both the digital system and the staff are responsible for tracking project and operational information. Projectspecific digital processes should be properly integrated to support automated processes and the automatic flow of data. The cloud-based collaboration culture supports the use of a human-machine interface, allowing workers to capture the necessary data via mobile phones. Cloud collaboration also facilitates collaboration in the supply chain, which can contribute to the tracking of project information.

Internal communication was improved by eliminating WhatsApp and Viber-based communication, which has enhanced efficiency in cloud-based communication and allowed for better tracking of project information and changes. Supply chain collaboration aims to improve efficiency. The company implemented a supplier database to monitor pricing and effectiveness, enabling the Procurement and Project Manager teams to handle 70% of their work during offer preparation.

4.6 Innovation

This category assesses the corporate culture, leadership approach, and feasibility of innovation to promote the development and execution of new ideas, services, technologies, or procedures [13]. The Innovation Manager facilitates the ongoing implementation and maintenance of innovation. This is achieved through the triangle of knowledge base, technology, and process automation, which ensures a culture of innovation.

The Innovation Manager is responsible for maintaining a sustainable and innovative environment and ensuring ongoing reporting. With a comprehensive understanding of both manufacturing and IT processes, she is able to support the launch and implementation of innovation projects to keep employees motivated to innovate.

5 Findings

During the interview, the initial model element relevance importance was rated on a scale from 0 to 5 by the CEO. The mean value of the element's importance under the six main categories is shown in Figure 4.



Figure 4 The importance of the C4M categories

The results indicate that, among the items assumed in our model, several elements in the Collaboration and Communication category were not considered significant by the CEO. It is important to note that the company did not prioritize soft skills, as presented in the case study.

We aimed to use the case study to explore which processes and areas were more prominent within each category. Figure 5 constructed based on the process map presented in Chapter 4, shows the number of predecessors and successors for each process element, with the size of the bubbles representing the number of categories a process is linked to. The term 'predecessor' refers to the previous process step as defined in the process map. For example, the process element 'Measure efficiency by KPI' has one predecessor (Integrate Operational processes in Change Management and Processes category) and two successors (Track Project and Operational Information in Collaboration and Communication category, and BI System in Technology Management and Business Application category). The number of connected categories summarises the categories of the predecessors and successors, therefore it is three. The figure leads to the following conclusions: (1) the development of the automation process was

related to all categories, (2) Systems Integration, Knowledge Management, and Cloud-based Systems were related to four categories. From the location of these four processes on the grid, it can be concluded that (3) System integration was the most important initial element in the digital transformation, as most of the successors were related to it. (4) Knowledge management and cloud collaboration were both preceded and followed by a similar number of processes, indicating their importance in establishing and sustaining progress. (5) The development of automation processes had the majority of the preceding processes. Consequently, several preventive steps need to be taken before process automation can take place.



Figure 5 Visualisation of the process map results

6 Summary and conclusion

The purpose of this case study was to provide a justification for the developed C4M in terms of model categories with a qualitative case study. The company perspective demonstrated that all six categories of the model are equally important. However, our results indicate that the innovation category is the least important compared to the other five categories. This conclusion was drawn from the process map, which showed that innovation can be identified as the integration of industry-specific technology. The case study identified three main drivers of digital transformation for this company: cloud-based systems, knowledge management, and system integration leading to automated processes.

The objective of the C4M was to conceptualise and quantify C4 in construction business activities and make them measurable over time. The case study revealed that the model accurately represents the company's digital maturity aspects however further development is needed to measure these aspects over time. Our research will continue to verify the underlying elements of each category and to develop for each category indicator level.

In summary, this research has presented a detailed case study of the C4M that will significantly contribute

to guiding researchers and developing businesses. The results indicate that digital transformation was primarily driven by cloud-based systems, knowledge management, and system integration to drive further the digital processes (achieve higher level of automation) that are connected between design, production and construction phases through system integration with human resource capabilities.

Acknowledgements

This research is supported by the ÚNKP-23-3-II-CORVINUS-91 National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. The research was supported by ConTech Consulting Kft. Special thank you to the participating company and its CEO for sharing their experiences. Any opinions, findings, conclusions or recommendations expressed in this paper do not necessarily reflect the views of the companies mentioned in this article.

References

- Anil Sawhney, Mike Riley, Javier Irizarry, CONSTRUCTION 4.0 An Innovation Platform for the Built Environment, 2020. https://doi.org/https://doi.org/10.1201/9780429398 100.
- [2] A. Karmakar, V.S.K. Delhi, Construction 4.0: What we know and where we are headed?, Journal of Information Technology in Construction 26 (2021) 526–545. https://doi.org/10.36680/j.itcon.2021.028.
- K. Yang, R.Y. Sunindijo, C.C. Wang, Identifying Leadership Competencies for Construction 4.0, Buildings 12 (2022). https://doi.org/10.3390/buildings12091434.
- P. Schönbeck, M. Löfsjögård, A. Ansell, Quantitative review of construction 4.0 technology presence in construction project research, Buildings 10 (2020) 1–15. https://doi.org/10.3390/buildings10100173.
- [5] J. van der Heijden, Construction 4.0 in a narrow and broad sense: A systematic and comprehensive literature review, Build Environ 244 (2023). https://doi.org/10.1016/j.buildenv.2023.110788.
- [6] European Commission, 2030 Digital Compass: the European way for the Digital Decade, Brussels, 2021. https://eur-lex.europa.eu/legalcontent/EN/TXT/HTML/?uri=CELEX:52021DC0 118&rid=4 Accessed: 11/10/2023.
- [7] E. Gökalp, V. Martinez, Digital transformation maturity assessment: development of the digital transformation capability maturity model, Int J Prod Res 60 (2022) 6282–6302.

https://doi.org/10.1080/00207543.2021.1991020.

- [8] C. Ferraz, E.R. Loures, F. Deschamps, BIM maturity models evaluated by design principles, in: Advances in Transdisciplinary Engineering, IOS Press BV, 2020: pp. 504–513. https://doi.org/10.3233/ATDE200111.
- [9] Jason Underwood, Umit Isikdag, Handbook of Research on Building Information Modeling and Construction Informatics, IGI Global, 2010. https://doi.org/10.4018/978-1-60566-928-1.
- [10] A. van der Merwe, A. Gerber, H. Smuts, Guidelines for conducting design science research in information systems, in: Communications in Computer and Information Science, Springer, 2020: pp. 163–178. https://doi.org/10.1007/978-3-030-35629-3_11.
- T. De Bruin, Q. Health, U. Kulkarni, M. Rosemann, Understanding the Main Phases of Developing a Maturity Assessment Model, in: ACIS 2005 Proceedings. 109, 2005. http://www.efqm.org/Default.
- [12] O. Nagy, I. Papp, R.Z. Szabó, Construction 4.0 organisational level challenges and solutions, Sustainability (Switzerland) 13 (2021). https://doi.org/10.3390/su132112321.
- [13] O. Heidenwolf, I. Szabó, Construction 4.0 Maturity Tool development methodology for organisations, in: Proceedings of the Creative Construction Conference 2023, Budapest University of Technology and Economics, Online, 2023: pp. 246– 251. https://doi.org/10.3311/CCC2023-033.
- [14] Thomas R.Gruber, A translation approach to portable ontology specifications, Knowledge Acquisition 5 (1993) 199–220.
- [15] Winston M. Tellis, Application of a Case Study Methodology, The Qualitative Report 3 (1997) 1– 19.
- [16] C.A. Williams, D. Schallmo, K. Lang, C. Williams, L. Boardman, Digital Maturity Models for Small and Medium-sized Enterprises: A Systematic Literature Review Digital Transformation of Business Models in the Israeli AgTech Landscape View project Digital Maturity Models for Small and Medium-sized Enterprises: A Systematic Literature Review, 2019. www.ispim.org.
- [17] P. Das, S. Perera, S. Senaratne, R. Osei-Kyei, A smart modern construction enterprise maturity model for business scenarios leading to Industry 4.0, Smart and Sustainable Built Environment (2023). https://doi.org/10.1108/SASBE-09-2022-0205.
- [18] F. Pirola, C. Cimini, R. Pinto, Digital readiness assessment of Italian SMEs: a case-study research, Journal of Manufacturing Technology Management 31 (2020) 1045–1083. https://doi.org/10.1108/JMTM-09-2018-0305.

[19] B. Wernicke, L. Stehn, A.A. Sezer, M. Thunberg, Introduction of a digital maturity assessment framework for construction site operations, International Journal of Construction Management (2021).

https://doi.org/10.1080/15623599.2021.1943629.

- [20] M. Bou Hatoum, H. Nassereddine, F. Badurdeen, Towards a Canvas for Construction 4.0 Implementation in AECO Organizations, in: Proceedings of the Creative Construction E-Conference 2022, Budapest University of Technology and Economics, Online, 2022: pp. 214– 220. https://doi.org/10.3311/CCC2022-028.
- [21] A. Schumacher, T. Nemeth, W. Sihn, Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises, in: Procedia CIRP, Elsevier B.V., 2019: pp. 409–414. https://doi.org/10.1016/j.procir.2019.02.110.
- [22] A.P.T. Pacchini, W.C. Lucato, F. Facchini, G. Mummolo, The degree of readiness for the implementation of Industry 4.0, Comput Ind 113 (2019).

https://doi.org/10.1016/j.compind.2019.103125.

- [23] A.A. Wagire, R. Joshi, A.P.S. Rathore, R. Jain, Development of maturity model for assessing the implementation of Industry 4.0: learning from theory and practice, Production Planning and Control 32 (2021) 603–622. https://doi.org/10.1080/09537287.2020.1744763.
- [24] B.G. de Soto, A. Georgescu, B. Mantha, Ž. Turk, A. Maciel, M.S. Sonkor, Construction Cybersecurity and Critical Infrastructure protection: New Horizons for Construction 4.0, Journal of Information Technology in Construction 27 (2022) 571–594. https://doi.org/10.36680/j.itcon.2022.028.
- [25] M. Ghobakhloo, The future of manufacturing industry: a strategic roadmap toward Industry 4.0, Journal of Manufacturing Technology Management 29 (2018) 910–936. https://doi.org/10.1108/JMTM-02-2018-0057.
- [26] K. Wu, B. Garcí, D. Soto, Current State and Future Opportunities of Data Mining for Construction 4.0, 39th International Symposium on Automation and Robotics in Construction (ISARC 2022) (2022) 78– 85.

https://doi.org/https://doi.org/10.22260/ISARC202 2/0013.

A Bibliographic Multi Theme Review of Cybersecurity, AEC and 3D Printing: Learnings and Way Forward for the AEC Industry

Bharadwaj R. K. Mantha¹, Jannatul Tashfia¹, Jannatul Tabassum², and Saleh Abu Dabous¹

¹Civil and Environmental Engineering, College of Engineering, University of Sharjah, United Arab Emirates ²School of Information and Technology, Griffith University, Australia

rmantha@sharjah.ac.ae, U19200265@sharjah.ac.ae, jannatul.tabassum@griffithuni.edu.au,

sabudabous@sharjah.ac.ae

Abstract -

The exponential growth in both three-dimensional printing (3DP) and cybersecurity (CS) poses a significant challenge for the change-resistant architecture engineering and construction (AEC) industry. Despite growing research related to any of these fields combined, there is a research gap associated with their combination, particularly with themes and topics in conjunction with AEC and 3DP. To address this and the larger knowledge gap of a methodology for multi-theme review, this research a) develops a generic multi-theme bibliometric analysis and b) validates the developed methodology with the help of three themes mainly CS, AEC, and 3DP. To achieve this, data was collected from Scopus and analyzed using Biblioshiny. The results emphasize the need for further imminent research in combinations of CS and AEC and more particularly in CS, AEC, and 3DP, where the growth rate is null despite growing research in individual domains mentioned. It is also noted from the findings that interdisciplinary research within these domains call for increased collaboration. This should not only motivate researchers in these aspects but assist stakeholders in identifying the need, work towards addressing this and more importantly, enhancing collaborative efforts.

Keywords -

Cybersecurity; 3D Printing; Construction Industry; Text-mining; Biblioshiny; Threat Modeling

1 Introduction

Architecture, engineering, and construction (AEC) industry significantly contributes to the global economy accounting to 13% of the global GDP [1]. The resources required have also been exceptionally high at about 50% of the gross global resource usage [1]. Its adversity to

change, innovation, and lack of productivity [1] is the motivation for its growing connection with threedimensional (3D) printing (also referred to as 3DP), along with the reduction in labor [1][2], waste [2] and pollution [1], time [1][2] and costs [1], increased worksite safety and the opportunity to explore complex geometries [1][2] as they do not require formwork [3].

Furthermore, the United Arab Emirates (UAE) has initiated the "Dubai 3D Printing Strategy", aiming to have 25% 3D printed buildings in Dubai by 2030 [4]. Large-scale applications of 3DP are not novel, beginning from 1997 with contour crafting, freeform and additive manufacturing, rapidly progressing to whole residential buildings and bridges, making this target much more feasible [5]. Some of the existing structures constructed using 3DP, many of which were designed for a service life of 50 years or greater are as follows [2]: 1) Non-load bearing walls of a single-storey student house in Denmark in 2021; 2) A two-storey office building in UAE built in 2019; 3) A double-storey residence in Germany in 2021 constructed; 4) An army barracks hut in USA in 2019; 5) A one-storey residence in the Netherlands in 2021; 6) The Striatus footbridge in Italy in 2021.

Although research in these two fields namely AEC and 3DP has gained significant interest over the years, the cybersecurity (CS) aspect in either of them is still not given as much importance [6] owing to the limited studies conducted during the construction process and much of the focus taken up by the design stage and the post-construction stage [7]. This knowledge gap is what makes it crucial to understand how these three fields are connected and explore the future possibilities due to their assimilation.

2 Literature Review

AEC Industry has struggled to identify and protect against cyber-attacks. Studies suggest that among the

common targets of cyber-attacks, AEC industry is the third in the list [8]. Research also shows that only 25% of the construction firms have prioritized cyber-security issue [8]. The theft of Interserve construction firm's employees' personal information has cost around £11m due to a cyber-attack [9]. Moreover, according to an annual report by an insurance giant Hiscox, almost half of United Kingdom (UK) construction firms have suffered from cyber-attacks, ranking it as the fifth most vulnerable industry for the second consecutive year [10]. A more specific example of this in 3DP would be the attempt of hackers to steal data on Hadrian X, a onearmed robotic bricklayer in Australia [11]. [12] also cautions of increased risk of cyberattacks due to the exponential increase in the adoption of machine learning and robotics in the AEC industry. It is thus evident that more studies should focus on increasing the awareness, perception, implementation, framework, standard, threat modeling, attack maps, countermeasure of the cyber security related aspects to protection data, reputation, operations, and business at large.

3DP technology is still in its infancy in the AEC industry along with the investigation of CS aspects within the domain. The CS research within AEC and its susceptibility to cyberattacks can be studied with the help of existing standards, tools, frameworks, and case studies from other domains since the process is similar. The slow adoption of digitization and CS in AEC is arguably due to lack of awareness, reputation, and competition of the AEC industry, among many other reasons. Some specific reasons particularly for 3DP within AEC industry include requirement of special material for printing, lack of standards and regulations, limitations of 3D printers onsite, and operation and management of 3D printers [13][14]. Many characteristics of the AEC industry itself also contributes to this slow pace, such as its projectbased, uniqueness, conservative, change-resistant and fragmented nature, with various stakeholders' involvement at different phases of its life cycle [14][15][16]. Thus, it must be noted that these cannot be directly adopted given the inherent characteristic limitations as mentioned above.

As an example, one of the methods of cyberattack during the 3DP process is acoustic side-channel attacks by using smartphones, which makes use of the Internet of Things (IoT) to steal confidential Intellectual Property (IP) information during 3DP. Using laboratory testing methods, some researchers were able to affirm IP leakage through the side channels of the 3DP process by using commercial off-the-shelf smartphones [17]. Sabotage attacks originating from the cyber domain can affect the physical domain and compromise systems' structural integrity potentially resulting in severe safety risks. Subtle variations in the sub-process can cause changes in the 3DP's parameters [18]. Other vulnerabilities include man-in-the-middle attacks to maliciously alter the ".STL" or the G-code file [19], which can compromise the quality of the product since the ".STL" file contains the design data, which is then sliced by a slicer software into layers as a G code to feed into the 3D printer [20]. [21] carried out a case study of cyberattack on the .STL file, wherein they altered the design to add a void, which impacted the specimen's strength. Other attributes that can be a target of such attacks are printing speed, the thickness of the layer, infill of the printing path [22]. Direct parallels can be drawn within the AEC industry and there can be several learnings for the same. Similarly, rapid digitalization incorporates several technologies such as the internet of things (IoT), building information modeling (BIM), machine learning (ML), robotics, and artificial intelligence (AI), which are also susceptible to cyberattacks. Examples of such attacks on HVAC (heating, ventilation, and air conditioning) and BMS (Building Management System) systems, unauthorized access, stealing construction plans, and breaching of sensitive and personal data are detailed in [7]. These can be seen mainly in OSC (Offsite Construction) and prefabrication [23], however, these are not separately detailed to consider AEC as the broad theme.

Table 1: Previous Studies Focus & Limitations

Study Focus	Limitations		
3DP/AM & AEC [24][25]	Focused on systematic mapping, cannot be directly adapted for bibliographic analysis.		
3DP/AM & AEC [26]	Focused on systematic literature review to identify gaps only.		
CS & AEC [7]	Restricted to two themes and does not identify multiple permutations and additional themes of interest.		
AEC & Industry 4.0 [27]	Focused on systematic mapping, cannot be directly adapted for bibliographic analysis.		
IoT & Smart Health Research [28]	Lack of clear framework in the methodology to conduct multi-theme analysis.		
Servitization & manufacturing [29]	Focused on systematic literature review to identify gaps only.		

To overcome such gaps and obtain learnings and a way forward for the future of the industry and research community, researchers resorted to summarizing and synthesizing existing literature on the respective subject. For example, the text mining method is utilized, and relevant bibliometric data is analysed. Table 1 lists some of the multi-theme reviews conducted both in the fields of interest, and others. Given their review, it is evident that no such analysis has been carried out for all three fields namely CS, AEC, and 3DP, which is the goal of this paper. Furthermore, there is no clear framework that can be directly adapted to our study, as eident in Table 1. Thus, the proposed methodology contributes to the existing body of knowledge by providing a general, systematic, easy to use multi-themed bibliometric analysis framework that can be applied for more than 2 themes irrespective of context and applications. The objectives of this study therefore are to a) develop a general research methodology for multi-theme bibliometric review, and b) validate the developed methodology with the help of three specific themes of interest in the current context namely CS, AEC, and 3DP.

3 Methodology

Figure 1 shows all the steps in the proposed methodology. Broadly categorized, the methodology employs a bibliometric approach to address the research objective of this study. Each of these steps is detailed in the sub-sections below.



Figure 1: Overview of Proposed Methodology

3.1 Identify Different Themes

The objective of this step is to identify different themes of interest pertaining to the context, application, and need. In the current context, three themes namely CS, AEC, and 3DP were chosen. Each of these terms is described briefly below. 3DP, also known as additive manufacturing or digital fabrication technology, is an emerging technology that creates physical objects from a geometrical representation through successive addition of materials. There is a mass increase of 3DP technology in the production of open-source design in agriculture, automotive and locative industries, aviation industries, including the construction industry. [30][31]. AEC, widely known as construction, refers to the creation of physical infrastructures as well as the maintenance and repair of the existing structures [32]. CS is a significant concern of every infrastructure and organization. It includes practical measures to protect information, data, and networks against any internal and external threats.

With the rise of 3DP technology in the AEC industry, CS has become a critical factor [33]. It has to be noted that the primary concern within CS is threat modelling (TM) and countermeasures. Although these are subtopics within the main theme of CS, in the latter stages of analyses in Section 4, it is also considered a theme of interest for further investigation and exploration. These themes will act as the foundation for further analyses and potentially pave the way for future research.

3.2 Outline Search Characteristics

Once the themes are identified, the immediate next step is to outline the search characteristics along with the selection of databases to conduct the search. In this study, Scopus was used as it is a vast database of abstracts and citations created by Elsevier, indexing more than 90% of journals and 70% more sources than WoS [34][35]. Since papers in combined fields are limited, Scopus is a better choice as it has 20% more coverage than WoS, including more journals despite their low impact [36].

Advanced search was done in Scopus to define the search characteristics due to the complexities of the search. Table 2 outlines the summary of the search characteristics and is described briefly below along with the rationale. The initial searches showed a significant of the papers published were in English, so the language was set to English, which also helps to verify the relevancy of the results. Since books and book chapters were both selected, the book sources from the first page results of the combination of 3DP and AEC, limited once only to books, and then only to book chapters, were used to compare them. Since there were no repeating sources, both document types were included. Mainly, the year range of 1950 to 2022 was chosen, as at this timeline, technology, and automation became of interest due to the third industrial revolution [37], and to keep the results consistent, the current year is excluded. The Boolean OR is used to group the words within the same domain, AND for combinations, AND NOT for excluding irrelevant words. It aims to find papers with at least one word from each domain. Wildcard is a tool in Scopus used to include

all similar words with different endings, for example, "3D print*" was used to include 3D printer, 3D printing, 3D printed, etc. All searches were carried out for selected keywords within the title, abstract or keywords of papers.

Table 2: Summary of Scopus Search Characteristics

Criteria	Option		
Search type	Advanced search		
Languages	English		
	articles, conference papers,		
Document type	reviews, book chapters,		
	conference reviews and book		
Timespan	1950 - 2022		
Booleans used	AND, OR, AND NOT		
Advanced search tool used	Wildcard (*)		
Searches within	TITLE-ABS-KEY (Article title, abstract and keywords)		

3.3 Determine Keywords and Combinations

In this step, relevant keywords and combinations need to be determined for the different themes identified in section 3.1. In addition, a list of exclusion words must be determined for irrelevant words that were found while manually checking for the search results. Different combinations of the chosen themes and topics can be determined subject to their relevance and applicability.

Table 3 outlines the summary of the keywords for the three main themes (CS, AEC, 3DP) along with the additional topic (TM) that was chosen for this study given its significance as discussed earlier. For example, regarding the exclusion words, the words hackneyed and hacksaw were taken from previously published papers and checked within these strings. Since they did bring up irrelevant papers, they were also added to the exclusion list. Section 3.4 discusses this further. Table 4 represents the 11 different combinations that were carried out. Although more were possible or done, due to the same results for those combinations, or irrelevancy, it was limited to only those shown in the table.

3.4 Define Different Search Strings

The objective of this step is to define the search string based on the identified based on the above steps (sections 3.1, 3.2, and 3.3. For example, the search for each combination from Table 4 was conducted using keywords from Table 3, the results of which were exported as a CSV file. The results from Scopus were last updated on 11th December 2023.

3.5 Update Keywords and Combinations

The immediate next step after conducting the search using the strings defined in Section 3.4 is to update the

keywords and combinations based on the results. The purpose of this step is to refine the keywords, exclusions, and combinations and improve the search results. The following sub-sections discuss significant results of the same.

Table 3: Keywords & Exclusions for Each Th	eme
--	-----

	Keywords	Exclusions
3DP	3D print, 3DP, additive manufacturing, additive fabrication, contour craft, extrusion-based printing, extrusion-based technology, rapid prototyping technology, rapid manufacturing, solid freeform fabrication, automated printing, three dimensional print	
AEC	(architecture, engineering, and construction), AEC, aecfm, (architecture, engineering, construction, and facility management), aeco, (architecture, engineering, construction, and operations), construction industry, construction sector	automatic exposure control, advanced eye center, array eddy current
CS	Cyberattack, Cyberbreach, Cybercrime, Cyberresilience, Cyberthreats, Cybervulnerability, Cyberspace, Cyberphysical, Cybersecurity, Sabotage attack, hack	Hackneyed , hacksaw, hackling
TM	Threat Models, digital model, Risk propagation, Vulnerability assessment, Counterattacks, Countermeasure analysis, Safety measures, Anti-hacking, Smart grid, Network security, Intrusion detection, Blockchain, Cloud-BIM	

3.5.1 3DP and AEC

Initially, more words under AEC, such as 4IR or Fourth Industrial Revolution, and under 3DP, such as automat, freeform were used, which resulted in papers under nursing, veterinary, and neuroscience. For example, "A study to determine the effects of industry 4.0 technology components on organizational performance" and "Freeform 3D printing of vascularized tissues: Challenges and strategies". Even the word construction, paired with 3D printing, resulted in several irrelevant papers, due to its usage as a verb. All these words were removed, and construction was replaced with words like construction industry and construction sector. The acronym AEC also brought up a few unwanted papers, but since they were fewer in number, those were added to the exclusion list.

No	3DP	AEC	CS	ТМ	Results
1					107,664
2					71,635
3					632
4					54,306
5					13,110
6					186
7					2
8					34
9					1
10					55
11					15

Table 4: Combinations & Results

3.5.2 3DP and CS

In this combination, papers like "A new technology to machine bimetal band saw" due to the use of Hack* could not be removed, so the keywords hacksaw, hackneyed, and hackling were excluded using AND NOT.

3.5.3 3DP and AEC and CS

This combination yielded 4 results, two of which are repeated and are simply the conference rather than an article within the conference. So, they were removed manually after exporting from Scopus.

3.5.4 3DP and CS and TM

Initially, TM was split into two domains, one for TM, and another for countermeasures. Due to the lack of papers, since the same results were obtained for both, the keywords in both groups were merged into one group, currently named TM. The threat model VAST was also a keyword, along with other types of threat models. However, since VAST also brought up results where it is used as an adjective, it was removed. The others were also removed since their presence made no difference in the results.

3.5.5 Other Combinations

The remaining combinations used the updated list from the previous searches, so no more modifications were done to them. Nonetheless, they were also checked for further improvement if required.

3.6 Conduct Bibliometric Analysis

Once the keywords and combinations are finalized, meaning no further refinement in search strings is required, updated results are extracted as per Section 3.4 and finally, bibliometric analysis of these results need to be conducted. In the current context, after all the searches were completed, the csv files exported from Scopus were used in Biblioshiny to analyse them. Significant parameters such as annual publication, annual growth rate, authorship, average citation per year, and frequent words from Biblioshiny were analysed and discussed.

4 Results and Discussion

This section discusses the results obtained from Scopus using the keywords and combinations as discussed in Section 3.3. It also discusses the results obtained from the analysis of Biblioshiny for specific themes and combinations as mentioned in Section 3.6.

4.1 Scopus Search Result

The Table 4 in Section 3.4 shows the number of total publications within the study timeframe for each combination. As can be observed, the publications within CS theme are approximately half of that in both the 3DP and AEC (i.e., construction) domains, with even fewer papers understandably when focused on the TM. A significant decrease was observed for combinations of these themes, with the least two-theme combination for 3DP and Construction, and only one when combining all domains. The one paper that comes up when all domains are combined is "Threat Modeling in Construction: An Example of a 3D Concrete Printing System", which discusses the use of a threat model in 3DCP.

4.2 Annual Scientific Production and Growth Rate

The total number of publications for each year is discussed in this section. Due to the large difference in the number of publications, they were divided into two parts, one for the main themes and another for the combinations. Although TM is a topic within CS, due to the high volume of papers and it being the primary focus within CS, it was also considered for further analysis and discussion.



Figure 2: Annual Production for Main Themes

Figure 2 shows the annual production for the main themes. A noteworthy observation is that production for

all of them began to pick up after the year 2000, with 3DP growing the fastest and TM the slowest. Both AEC and CS have a steady growth of approximately 10%, as seen in Figure 4. This suggests the growing prominence and relevance of 3DP further corroborating the theme selection for this study.

Since the numbers were smaller for combinations relative to the main themes, particularly in the earlier years and due to the exponential growth, logarithm (log) of the cumulative annual production for each was considered to better view and is shown in Figure 3. Additionally, the years that had only 1 publication cumulatively were removed as log (1) is equal to 0. It can be observed that all the graphs have a general increasing trend. Although the combinations of 3DP and CS with Construction (no 5 and 10) began in 2002, other combinations begin much later, especially those combinations that involve three domains namely no. 7, 8 and 11. This further emphasizes the motivation of this study.



Figure 3: Annual Production for Combinations

Figure 4 shows the annual growth rate (AGR) in percentage. As discussed, for no. 7 and 9, the number of publications were 2 and 1 respectively, hence the null growth rate. The AGR for the combinations is high despite the low number of publications because of the rapid growth of publications over the recent years. This shows that current research is largely focused on either 3DP and AEC or 3DP and CS and does not consider all three domains together.

4.3 Authorship

For each combination, the number of documents with single authors, multiple authors, and authors' collaboration internationally per article were analysed in this section and shown in Figure 5. The number of single-authored publications were huge for Construction and CS with 14,134 and 13,858 respectively, while for the combinations, it was between 0 and 43. It was in the midrange for 3DP and TM, at 4,949 and 1,368 respectively. The co-authored publications for all were in the range of

1 and 4.8. Since both single and co-author values use the number (left) axis, it was difficult to properly observe the co-author result in the same plot. Thus, the log of singleauthored publications was used instead. The international co-authorship is in percentage and uses the secondary (right) axis.



Figure 4: Annual Growth Rate (AGR)

Publications with single authors are relatively high for the main domains, but 0 for combinations of 3 domains. This may be due to the limited number of papers within these areas. As opposed to this, coauthored documents are published for all 11 areas of study. The international co-authorship, which refers to the authors of a co-authored publication being from different countries, is also 0 for two of the combinations, no (7) and (9). This is expected since there are only 1-2 papers from them. However, a positive trend is observed where combinations of domains encourage researchers' collaboration across these disciplines.



Figure 5: Single, Co-, & International Authorship

4.4 Average Citation per Year

The citations were also divided into two parts, one for the main themes in Figure 6 and another for the combinations in Figure 7. Figure 6 shows a steady increase for all of the themes, with the highest in 3DP. Similar to the annual production, the 3DP graph has a higher increase than the rest. TM, although beginning in
year 1994, increased rapidly and caught up with Construction and CS, both of which had a steadier rise since. The peaks at various years may be due to inconsistent research or lack of collaboration and interest [38]. These peaks have reduced in recent years as more research is conducted consistently every year.



Figure 6: Annual Citation for Main Themes



Figure 7: Annual Citation for Combinations

Figure 7 shows the citations for the combinations. Since there are not many papers for them, the citation graphs look randomized, with the exception of combination 6 which appears to have a more consistent increase than the others.

4.5 Most Frequent Words

A list of frequent words for each search were taken from Biblioshiny and displayed in the form of a Venn diagram in Figure 8. Although the combinations had a smaller result, any repeating keywords were placed in the intersections, according to the characteristics of a Venn diagram. These keywords not only highlight the current trends but will also help identify future research topics. As can be seen in Figure 8, a lot of the keywords that combine all three themes have to do with CS, such as risk assessment, security and unauthorized access. The extent of involvement desired from each domain is still missing, which requires combining the intersections of any two domains with others so more of those appear at the centre.



Figure 8: Frequent Words

5 Conclusions

This study developed and validated a multi-theme bibliometric analysis methodology addressing the critical knowledge gap in literature. Specifically, the three most significant and crucial themes of CS, AEC, and 3DP were investigated and analyzed. Results from the extensive, comprehensive and thorough bibliometric analysis suggest that a) CS-related topics such as security, privacy, and risk assessment are gaining momentum in the recent past but research within in AEC industry still lacks sufficient studies, b) combination themes have more coauthored papers than single-authored papers suggesting for collaborative approach on multi-disciplinary research endeavors. c) annual growth rate (AGR) was higher for combinations excluding AEC than otherwise, and zero for the 3DP-CS-AEC combination, emphasizing the need for further CS research within the AEC research community and more significantly for those including 3DP and AEC. A general roadmap for future research is discussed in Table 5 based on the observed recently published studies (now), ongoing research efforts (next), and what AEC community can aim at (future). The future roadmap in particular was derived from the summarized literature from the CS community.

One of the limitations of this study was that a) due to the large number of results for the original themes, it was practically not feasible to check for duplicates and remove them all b) due to the capacity of Biblioshiny, some of the papers were omitted during the analysis. Future research can be aimed at addressing these limitations and further comprehensively exploring all the parameters of the Biblioshiny and also other software and databases to compare the results.

Table 5: Roadmap to Future Research

Now	Next	Future
Currently, much of the focus in 3DP is within the manufacturin g industry, in terms of CS [17][18].	Although parallels can be drawn from manufacturing to AEC, it is required to have real case studies in the AEC to study its impact directly, due to the uniqueness of each project within AEC.	It is very likely that once focus is shifted to CS in AEC and 3DP jointly, with tested case studies, new issues of implementing 3DP into AEC will arise, which will require innovative solutions to be tackled.
3DP is generally vulnerable to hacking and tampering due to its digital reliance [19][20][22].	These vulnerabilities need to be tested in settings similar to the projects in AEC and holistically, rather than targeting only one aspect of it at a time.	Large-scale application of 3DP in AEC will require new research methodology and countermeasures to overcome.
Digitization of the AEC industry at present, such as adopting BIM, IoT, and ML already expose it to numerous threats of cyberattacks [7].	Evolution of cyberattack threats in digitized AEC that also integrated 3DP technology.	Countermeasures and threat models to combat these threats, which is only done in one paper (Garcia de Soto and Shibly, 2020). It is also the only paper that falls under 3DP- CS-AEC as seen from the results in Section 4.

References

- Valente, M., Sibai, A., & Sambucci, M. 2019. "Extrusion-Based Additive Manufacturing of Concrete Products: Revolutionizing and Remodeling the Construction Industry". *J of Comp Sc*, 3(3), 1-20. <u>https://doi.org/10.3390/jcs3030088</u>
- [2] Bos, F., Menna, C., Pradena, M., Kreiger, E., Da Silva, W. L., Rehman, A., Weger, D., Wolfs, R., Zhang, Y., Ferrara, L., & Mechtcherine, V. 2022. "The realities of additively manufactured concrete

structures in practice". Cem & Con Res, 156, 1-14. https://doi.org/10.1016/j.cemconres.2022.106746

- [3] Bedarf, P., Dutto, A., Zanini, M., & Dillenburger, B. 2021. "Foam 3D printing for construction: A review of applications, materials, and processes". *Autom in Const*, 130, 1-15. <u>https://shorturl.at/abG07</u>
- [4] Esmail, E. 2016. "Mohammed bin Rashid: 25% of Dubai's buildings will be 3D printed by 2030". *Emirates News Agency – WAM*. Accessed December 21, 2023. <u>https://shorturl.at/ryJV4</u>
- [5] Buswell, R., Leal de Silva, W., Jones, S., & Dirrenberger, J. 2018. "3D printing using concrete extrusion: A roadmap for research". *Cem & Con Res*, 112, 37-49. <u>https://shorturl.at/dhsuC</u>
- [6] Shibly, M., and Garcia de Soto, B. 2020. "Threat Modeling in Construction: An Example of a 3D Concrete Printing System" In 37th ISARC, 625-632, <u>http://dx.doi.org/10.22260/ISARC2020/0087</u>
- [7] Mantha, R. & García de Soto, B. 2021.
 "Cybersecurity in Construction: Where Do We Stand and How Do We Get Better Prepared". *Frontiers In Built Env.* 7, 1-13. <u>https://doi.org/10.3389/fbuil.2021.612668</u>
- [8] Baker, S. 2023. "Cybersecurity risks in the construction industry". *Lockton*. Accessed March 6, 2024. <u>https://rb.gy/tdgbsv</u>
- [9] Rogers, D. 2023. "Cyber attack costs Interserve more than £11m". *Building*. Accessed March 6, 2024. <u>https://rb.gy/ykatyz</u>
- [10] Building. 2023. "Construction among industries most at risk from cyber attacks, insurer warns". Accessed March 6, 2024. <u>https://rb.gy/u9yyo3</u>
- [11] Pash, C. 2018. "How hackers and spies tried to steal the secrets of Australia's one-armed robot bricklayer". *Yahoo Finance*. Accessed March 6, 2024. <u>https://rb.gy/xnq1w0</u>
- [12] Kunert, P. 2023. "US construction giant unearths concrete evidence of cyberattack". *The Register*. Accessed March 6, 2024. <u>https://rb.gy/t6d2fu</u>
- [13] El-Sayegh, S., Romdhane, L., & Manjikian, S. 2020.
 "A critical review of 3D printing in construction: benefits, challenges, and risks". *ArcCiv&MechEng*, 20. <u>https://doi.org/10.1007/s43452-020-00038-w</u>
- [14] Hossain, M.A., Zhumabekova, A., Paul, S. & Kim, J. 2020. "A Review of 3D Printing in Construction and its Impact on the Labor Market". *Sustainability*, 2 (20), 1-21. <u>https://doi.org/10.3390/su12208492</u>
- [15] Olsson, N., Arica, E., Woods, R. & Madrid, J. 2021.
 "Industry 4.0 in a project context: Introducing 3D printing in construction projects". *Proj Lead & Soc*, 2, 1-10. <u>https://doi.org/10.1016/j.plas.2021.100033</u>
- [16] Mantha, R, García de Soto, B., & Karri, R. 2020. "Cyber security threat modeling in the AEC industry: An example for the commissioning of the built environment". Sust Cit & Soc, 66, 1-15.

https://doi.org/10.1016/j.scs.2020.102682

- [17] Song, C., Lin, F., Ba, Z., Ren, K., Zhou, C., & Xu, W. 2016. "My Smartphone Knows What You Print: Exploring Smartphone-based Side-channel Attacks Against 3D Printers". In 23rd ACM Conf on Comp & Comm Sec, 895-907. Vienna: ACM. http://dx.doi.org/10.1145/2976749.2978300
- [18] Yu, S., Malawade, A., & Al Faruque, M. 2021. "Multi-Model Attack Detection for Cyber-physical Additive Manufacturing." *TC-CPS Newsletter*, 5(1), 2-6. <u>https://rb.gy/wpyzwi</u>
- [19] Wu, M., Song, Z., & Moon Y. 2019. "Detecting cyber-physical attacks in CyberManufacturing systems with machine learning methods". *J Intell Manuf*, 30, 1111-1123. <u>https://shorturl.at/mBZ19</u>
- [20] Agashe, K., Sachdeva, A., & Chavan, S. 2020. "3D Printing And Advance Material Technology". Int'l J of Grid & Dist Comp, 13 (2), 1899-1936. <u>https://rb.gy/stvd54</u>
- [21] Sturm, L., Williams, C., Camelio, J., White, J., & Parker, R. 2017. "Cyber-physical vulnerabilities in additive manufacturing systems: A case study attack on the .STL file with human subjects". *J of Man* Sys, 44 (1), 154-164. https://doi.org/10.1016/j.jmsy.2017.05.007
- [22] Gao, Y., Li, B., Wang, W., Xu, W., Zhou, C. & Jin, Z. 2018. "Watching and Safeguarding Your 3D Printer: Online Process Monitoring Against Cyber-Physical Attacks". Proc. of the ACM on Inter, Mob, Wear and Ubiq Tech, 2 (3). 1-27. http://dx.doi.org/10.1145/3264918
- [23] Wang, M., Wang, C., Sepasgozar, S., & Zlatanova, S. 2020. "A Systematic Review of Digital Technology Adoption in Off-Site Construction: Current Status and Future Direction towards Industry 4.0". Buildings, 10 (11), 1-29. https://doi.org/10.3390/buildings10110204
- [24] Labonnote, N., Rønnquist, A., Manum, B., & Rüther, P. 2016. "Additive construction: State-ofthe-art, challenges and opportunities". *Autom in Const*, 72 (3), 347-366. <u>https://doi.org/10.1016/j.autcon.2016.08.026</u>
- [25] Tay, Y., Panda, B., Paul, S., Mohamed, N., Tan, M., & Leong, K. 2017. "3D printing trends in building and construction industry: a review". *Virt & Phys Prot*, 12 (3). 261-276. http://dx.doi.org/10.1080/17452759.2017.1326724
- [26] Pessoa, S., Guimarães, A., Lucas, S., & Simões, N. 2021. "3D printing in the construction industry - A systematic review of the thermal performance in buildings". *Ren & Sust Ene Rev*, 141, 1-13. <u>https://doi.org/10.1016/j.rser.2021.110794</u>
- [27] Boton, C., Rivest, L., Ghnaya, O., & Chouchen, M. 2021. "What is at the Root of Construction 4.0: A Systematic Review of the Recent Research Effort".

Arc of Comp Meth in Eng, 28, 2331–2350. https://doi.org/10.1007/s11831-020-09457-7

- [28] Yang, X., Wang, X., Li, X., Gu, D., Liang, C., Li, K., Zhang, G., & Zhong, J. 2020. "Exploring emerging IoT technologies in smart health research: a knowledge graph analysis". *BMC Med. Inform. Decis. Mak.*, 20. <u>https://doi.org/10.1186/s12911-020-01278-9</u>
- [29] Raddats, C., Kowalkowski, C., Benedettini, O., Burton, J., & Gebauer, H. 2019. "Servitization: A contemporary thematic review of four major research streams". *Indus Mark Man*, 83, 207-223. <u>https://doi.org/10.1016/j.indmarman.2019.03.015</u>
- [30] Lee, T., Shahrubudin, N. & Ramlan, R. 2019. "An Overview on 3D Printing Technology: Technical, Materials, and Applications". *Proc Man*, 35, 1286-1296.<u>https://doi.org/10.1016/j.promfg.2019.06.089</u>
- [31] Chaturvedi, I., Jandyal, A., Irfan Ul Haq, M., Raina, A. & Wazir, I. 2022. "3D printing-A review of process, materials and applications in industry 4.0." *Sust Oper & Comp*, 3, 33-42. https://doi.org/10.1016/j.susoc.2021.09.004
- [32] Wells, J. 1984. "The construction industry in the context of development: A new Perspective." *Hab Int'l*, 8(3-4), 9-28. <u>https://doi.org/10.1016/0197-3975(84)90040-7</u>
- [33] Liu, Q. & Li, Y. 2021. "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments". *En Rep*, 7, 8176-8186. <u>https://doi.org/10.1016/j.egyr.2021.08.126</u>
- [34] Burnham J. 2006. "Scopus database: a review". Biomedical digital libraries, 3(1), 1-8. https://doi.org/10.1186/1742-5581-3-1
- [35] López-Illescas, C., De Moya-Anegón, F. & Moed, H. 2008. "Coverage and citation impact of oncological journals in the Web of Science and Scopus". J of Infor, 2(4), 304-316. <u>https://doi.org/10.1016/j.joi.2008.08.001</u>
- [36] Chadegani, A., Salehi, H., Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M. & Ebrahim, N. 2013. "A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases". Asi Soc Sci, 9(5). 18-26. http://dx.doi.org/10.5539/ass.v9n5p18
- [37] Hoque, A. 2019. "4th Industrial Revolution- Impact and Challenges". In *Nat Conf on Elec & Infor*, 1-9. Dhaka: Bang Elec Infor Soc & Bang Atom Ene Comm. <u>https://rb.gy/ye8aht</u>
- [38] OpenAI. (2023). ChatGPT [Large language model]. https://chat.openai.com
- [39] Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. 2021. "How to conduct a bibliometric analysis: An overview and guidelines". *J of Bus Res*, 133. <u>https://doi.org/10.1016/j.jbusres.2021.04.070</u>

Analogizing in Construction and Education Research: A Case Study

Ronie Navon¹ and Yoram M Kalman²

¹ Technion IIT [Past President of IAARC]. Currently a Ph.D. candidate at The Open University of Israel under the supervision of Prof. Y. M. Kalman

² The Research Center for Innovation in Learning Technologies, The Open University of Israel ronie@technion.ac.il, yoramka@openu.ac.il

Abstract

This paper is inspired by the concept of Analogizing. It explores the analogies between a concept from the Construction industry (Automated Project Performance Control - APPC) and a concept in Education (Organizational Learning Analytics -OLA). The APPC model deals with automating data collection and conversion from on-site construction projects in real-time to improve control and performance, while the OLA model expands the current application of learning analytics to managerial and strategic level decisions taken in higher education institutions (HEIs). By drawing connections between these two conceptual models, the paper demonstrates the innovative potential of Analogizing to identify novel solutions in one discipline using well established methods from an analogous, yet highly distinct, second discipline.

Keywords

Civil Engineering; Higher Education; Learning Analytics; Automation; Construction Project Performance Measurement

1 Introduction

Inspired by the concept of Analogizing, this paper presents a comparative case study of two conceptual models, one from Social Science and the second from Engineering.

The first model, known as Automated Project Performance Control (APPC), was developed to facilitate construction projects on-line control by automating the collection and conversion of field data from ongoing projects. The second model, currently under development, deals with learning analytics (LA). It aims to expand the application of LA beyond the classroom-level, which is its current focus. Thus, stakeholders outside the classroom-level will be able to make their decisions in alignment with HEI's overall organizational T&L. This model will be related to henceforth as organizational learning analytics (OLA).

The intention of this paper is to demonstrate how the concept of Analogizing can be used in a crossdisciplinary research, especially when researching new subjects. The authors use their experience in the development of the APPC and OLA models in a case study, which shows that the two, seemingly, distinct disciplines (the APPC and OLA models) share more commonalities than initially apparent. The concept of Analogizing, as well as the APPC and OLA models will be detailed in the following Sections.

2 Analogizing

Comparing dissimilar events, activities, or phenomena despite their differences characterizes analogical thinking, also known as Analogizing. This mental process is specifically designed to emphasize formal parallels across disparate contexts [1]. Zerubavel [1] illustrates this by pointing out that a female and an Afro-American professional in 1940s America are in parallel situations; both belong to low-status socio-ethnic groups while occupying high-status professional positions. He offers another cross-contextual example, highlighting the equivalence between gay individuals and those with disabilities, as both groups face societal "stigmatization." Sceptics often dismiss the possibility of finding equivalence among seemingly non-comparable items, asserting that it is like "comparing apples to oranges." However, Zerubavel [1] counters this argument by pointing out that even seemingly disparate entities, like apples and oranges, share commonalities, such as being fruits.

The Case Study will employ the Analogizing approach to underscore the shared characteristics between the OLA, utilized in the educational realm, and APPC in construction operations.

3 Automated Project Performance Control

This Section relates to the first of the two conceptual models – Automated Project Performance Control (APPC) from the construction industry.

Construction projects are complex and dynamic environments that prove challenging to monitor in realtime conditions [2]. Project Performance Control (PPC) broadly refers to the activities taken by the project management to ensure that the project's performance aligns as closely as possible with the initial plan. Performance is measured in terms of Project Performance Indicators (PPIs) such as cost, schedule, labor productivity, materials consumption, etc. [2], [3].

The reliance on manual data collection in traditional construction control methods inevitably leads to slow, inaccurate, and error-prone data collection processes [3], [4]. This, in turn, likely explains the prevalence of generic and infrequent control practices among construction managers. To achieve more timely and accurate control, project managers would need to dedicate an excessive amount of time to data collection, diverting their attention away from their primary responsibility of project management and supervision.

Automated Data Collection (ADC) technologies that can potentially measure performance indicators in realtime on construction sites are rapidly emerging with declining costs. However, the construction industry lags in adopting these technologies to measure performance indicators [2], [5], [6]. A key reason for this is that ADC technologies are unable to **directly** measure the required PPIs, they can only capture **indirect** metrics, raw data (RD) from the site [7], [8].

A conceptual model for APPC was developed to enable automated measurement of the indirect RD, which after conversion to PPIs, are used for controlling construction projects [9]. As illustrated in Figure 1, this model outlines how a typical control cycle begins by measuring RD, which will later be converted to the PPIs like cost, progress, and resource consumption (resources such as materials, manpower, etc.) as of the datacollection date. These values are compared against planned, or updated plans, performance levels. If deviations are found, an analysis is conducted to understand the factors causing them. Corrective actions, like adding resources, adding workers, or authorizing overtime work are then decided upon based on this analysis. The final phase of the control cycle involves implementing these corrective measures.



Figure 1. The Control cycle

The Achilles' heel of the automated construction control cycle lies in the real-time measurement of PPIs. The current absence of sensors capable of directly and automatically measuring PPIs creates a significant gap between the technologically capturable data on-site – RD – and the PPIs essential for real-time automated control. While automated data collection (ADC) technologies can measure various RD, project managers require, on the other hand, direct PPIs values to make informed decisions. This fundamental lack of direct PPI measurement sensors presents a major obstacle to achieving real-time automated control in construction projects.

To bridge the gap between RD collected from on-site construction projects in real-time and the necessary PPIs, Technion has developed conceptual conversion models. This approach aims to measure real-time values of indirect parameters (the RD) that can be captured on-site using existing technology. By converting these indirect parameters into the required PPIs, construction managers have actionable information to work with. The most important PPIs – cost and schedule, known as Key Performance Indicators (KPIs) – can now be compared in terms of actual performance against planned objectives. This enables timely course correction and improves project outcomes.

The specific indirect parameters likely differ across various types of activities (in building construction: skeleton activities, finishing activities, like flooring, in road construction: compacting, paving etc.). However, Technion models are often based on the fact that a "construction agent" – worker, earthmoving equipment or other – must be proximate to a building element, or the road section, to construct it. Thus, capturing the construction agent's location over time, combined with

data from the Building Project Model (BPM), enables determining what activity the agent is performing.

By collecting location data at regular intervals, the construction-activity-specific algorithms developed in Technion can determine what the construction agent is doing at any given time. The model links measured locations to activities (e.g. flooring, painting, paving etc.) based on proximity of the construction agent to building elements – this is the novelty of the concept.

This conceptual model was implemented in proof-ofconcept case studies and tested on-site with ADC, in various construction areas: manpower control [10], road construction [11], materials management [12], safety control [13], progress [14] and more. Subsequent efforts delved into enabling technologies and opportunities, reviewed the field's state-of-the-art, and proposed dashboards and concepts [2], [4]–[6], [15]–[18].

Like the OLA model that will be explored next Section, the APPC model demonstrates a data-driven approach to gaining insights for performance improvements.

4 Learning Analytics

Learning analytics (LA) is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs [19]. The data available in the HEI's repository are students' participating in online courses, submitting assignments, engaging with learning resources, accessing the university websites and databases, activities in the libraries, number of students per class, distribution of grades across courses and departments, courses where students need additional support, faculty-to-student ratios, and more.

LA is instrumental in identifying students who may be at-risk of dropping out or failing a course and allows for the implementation of targeted interventions to help these students succeed [20]–[22]. When combined with machine learning methods, LA serves as a tool for early identification of students who may not submit their upcoming assignments or face potential failure [23], [24].

Currently, LA primarily focuses on informing professors about their students for teaching and learning (T&L) improvement [22], [25]. However, this approach only utilizes a fraction of the valuable data available within HEIs' repositories. The activities and decisions of other stakeholders, including department heads,

administrative staff, presidents or provosts, and vice presidents, etc., all have a significant impact on the academic conduct and the quality of T&L. These stakeholders often focus on performance metrics specific to their individual roles, such as marketing, budgeting and finances, faculty recruitment, student enrolment, fundraising, etc. rather than metrics aligned with the institution's broader goals.

This demonstrates the gap between LA's potential and actual usage [26], [27]. In order to narrow this gap, The Open University of Israel is currently conducting a research study to develop a conceptual model which will empower a broader range of stakeholders, both within and outside HEIs, to effectively utilize LA to improve T&L – as mentioned in the Introduction, this conceptual model is called organizational learning analytics (OLA). The conceptual model adapts principles from management science models, demonstrating the Analogizing concept. An initial conceptual model was presented last year [28].

Ethical considerations still challenge LA researchers [22], [29], [30]. Expanding LA applications to more stakeholders could raise additional ethical issues. Therefore, the new OLA conceptual model will need to carefully account for ethical implications during its development.

As highlighted in the previous APPC model Section, the OLA approach also relies heavily on data analytics to derive decisions and outcomes.

Having outlined the key concepts within the APPC and OLA models, this paper now turns to analyze the disparities and commonalities between them.

5 Case Study

Construction project performance measurement and organizational learning analytics (OLA) operate in distinct contextual circumstances with unrelated objectives. Educational institutions function within structured environments defined by clear plans, such as curricula, and typically adhere closely to these plans. Conversely, construction projects, conducted in external environments, are subject to harsh weather conditions and other dynamic factors, making it challenging to adhere rigidly to initial plans.

In educational institutions, the key stakeholders – the students – exhibit stability by adhering to the curriculum, usually thorough the whole year or the entire program duration measured in terms of years. In contrast, construction workers rarely stay on a single project for an

extended period, spanning only days, weeks, or occasionally, months. The differences between these contexts are significant, and in Zerubavel's [1] terms, attempting to find cross-contextual equivalence between OLA and APPC would be like to comparing apples to oranges.

As in the case of apples and oranges, despite these apparent disparities, a more profound comparative analysis uncovers noteworthy commonalities between OLA and APPC. Both models rely on data to derive actionable insights. They both measure indirect, proxy, raw data interpreted by powerful algorithmic processes to generate Key Performance Indicators (KPIs).

OLA leverages raw data such as grades, student participation, course assignment submissions, engagement with learning resources, etc. These data are not direct indicators of the quality of teaching and learning (T&L). Direct KPIs might include, for example, the number of students successfully completing a course or graduating. Similarly, the APPC model tracks the locations of "construction agents" at regular intervals, associating them with building, or other construction elements based on schedule data and Building Information Model (BIM). However, the resulting information, such as the construction agent's activity, does not directly represent KPIs like cost and schedule.

Both OLA and APPC share a common limitation: the outputs of their algorithmic processes may not accurately reflect the actual activities they are intended to represent. In OLA, student participation metrics may not necessarily correlate with genuine engagement in the learning process. Similarly, in APPC, a worker's presence near a specific building element may not always indicate that work is being performed on that element. Conversely, a worker's absence from a particular area (despite having been there shortly before and returning shortly after) should not be interpreted as a lack of work on that element. The worker may still be engaged in tasks contributing to actual work being performed on that element, such as fetching materials or preparing tools, despite not being physically proximate to it.

Another cross-contextual commonality is the involvement of diverse stakeholders in both construction and HEIs, including internal and external parties. Project management stakeholders range from project managers, clients, owner/entrepreneurs, architects, engineers, mainand sub-contractors, and materials suppliers, to local authorities and regulators. This stakeholder network expands at the company level, emphasizing the complexity of project management. In HEIs, stakeholders span from students, professors, department heads, administrative staff, presidents, vice presidents etc., to external authorities like local/national authority and regulators.

While additional commonalities, such as the competitive environments in which both domains operate, exist, this paper focuses on the highlighted examples to underscore the strength of the Analogizing concept in this case study.

Having examined the concepts of the APPC and OLA models, along with their distinct features and their commonalities, this paper now transitions to its concluding remarks.

6 Concluding Remarks

This paper underscores the significance of crossdisciplinary analogies in the face of researching new subjects. Through the APPC and OLA case study, it becomes evident that these seemingly distinct disciplines share more commonalities than initially apparent. The case study effectively highlights these cross-contextual connections.

The Concept of Analogizing emphasizes the recognition of shared patterns despite variations in domains. Examining the case study reveals several common principles between OLA and APPC models. Both models rely on collecting data to derive insights for informed decision-making. They utilize indirect, raw data that requires sophisticated analysis methodologies. Additionally, both operate within complex stakeholder environments.

These shared elements highlight the potential of analogical thinking to support concept-driven developments. Even without directly implementing ideas from other contexts, this approach can facilitate brainstorming, promote the exchange of ideas, and enable the reuse of methodologies, building blocks, and approaches across disciplines.

The use of the concept of Analogizing in construction, as well as in education research, is novel. The benefits of using such an approach for future construction research are unmistakable, especially when dealing with advanced technologies, such as automation and robotics (A&R). The construction industry lags in adopting A&R technologies, although these technologies have been discussed in the construction realm for the last four decades and even longer. The idea of using Analogizing in these circumstances may help break through this deadlock. This paper advocates for the proactive extraction of lessons from broader contexts during the conceptualization phase. Further exploration of crossdisciplinary analogies holds promise for continual advancement.

References

- [1] Zerubavel E. *Generally speaking: An invitation to concept-driven sociology*. Oxford University Press, 2020.
- [2] Carmona et al. A. M. Instrumentation and data collection methodology to enhance productivity in construction sites using embedded systems and IoT technologies. In *Proceedings of the 35th CIB W78* 2018 Conference: IT in Design, Construction and Management, Springer, 2019: 637–644.
- [3] Navon R. Research in automated measurement of project performance indicators. *Automation in Construction*, 16(2): 176–188, 2007.
- [4] Cha H. and Kim J. A study on 3D/BIM-based onsite performance measurement system for building construction. *Journal of Asian Architecture and Building Engineering*, (19)6: 574–585, 2020.
- [5] Munawar H. S. Ullah Qayyum F. S. and Shahzad D. Big Data in Construction: Current Applications and Future Opportunities. *Big Data and Cognitive Computing*, (6)1. 2022.
- [6] Murguia D. Chen Q. Van Vuuren T. J. Rathnayake A. Vilde V. and Middleton C. Digital Measurement of Construction Performance: Data-to-dashboard strategy. In *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2022, p. 092009.
- [7] Griffin R. Navon R. Brecher A. Livingston D. Haas C. and Bullock D. Emerging technologies for transportation construction. *Transportation in the New Millennium*, 2000.
- [8] Navon R. and Goldschmidt E. Monitoring labor inputs: automated-data-collection model and enabling technologies. *Automation in construction*, (12)2: 185–199, 2003.
- [9] Navon R. Automated project performance control (APPC) of construction resources. In *ISARC 2003 The Future Site*, Eindhoven, The Netherlands, 2003, pages 79–82.
- [10] Navon R. and Goldschmidt E. Can labor Inputs be measured and controlled automatically?" *Journal of Construction Engineering and Management*, (129)4: 437–445, 2003.
- [11] Navon R. and Shpatnitsky Y. A model for automated monitoring of road construction. *Construction Management and Economics*, (23)9: 941–951, 2005.
- [12] Navon R. and Berkovich O. An automated model

for materials management and control. *Construction Management and Economics* (24)6: 635–646, 2006.

- [13] Navon R. and Kolton O. Model for automated monitoring of fall hazards in building construction. *Journal of Construction Engineering and Management*, (132)7: 733–740, 2006.
- [14] Navon R. and Haskaya I. Is detailed progress monitoring possible without designated manual data collection? *Construction Management and Economics*, (24)12: 1225–1229, 2007.
- [15] Alizadehsalehi S. and Yitmen I. A concept for automated construction progress monitoring: technologies adoption for benchmarking project performance control. *Arabian Journal for Science and Engineering*, (44): 4993–5008.
- [16] Bhagwat K. and Delhi V. S. K. Review of construction safety performance measurement methods and practices: a science mapping approach. *International Journal of Construction Management*, (23)4: 729–743, 2023.
- [17] Sardroud J. M. Perceptions of automated data collection technology use in the construction industry. *Journal of Civil Engineering and Management*, (21)1: 54–66, 2015.
- [18] Woo J. Shin S. A. Asutosh T. Li J. and Kibert C. J. An Overview of State-of-the-Art Technologies for Data-Driven Construction. in *Proceedings of the* 18th International Conference on Computing in Civil and Building Engineering, Toledo-Santos E. and Scheer S. Eds. pages 1323–1334 Springer International Publishing, 2021.
- [19] What is Learning Analytics? What is Learning Analytics? Society for Learning Analytics Research (SoLAR). [Online]. Available: <u>https://www.solaresearch.org/about/what-islearning-analytics/</u> Accessed: 09/11/2023.
- [20] Montuori L. Alcazar-Ortega M. Vargas-Salgado C. and Alfonso-Solar D. Learning Analytics as Data driven decision making in High Education: a case study. 2023.
- [21] Dawson S. Pardo Salehian-Kia A. F. and Panadero E. An Integrated Model of Feedback and Assessment: From fine grained to holistic programmatic review. in 13th International Learning Analytics and Knowledge Conference, 2023, pages 579–584.
- [22] Rets I. Herodotou C. and Gillespie A. Six Practical Recommendations Enabling Ethical Use of Predictive Learning Analytics in Distance Education. *Journal of Learning Analytics*, (10)1, 2023.
- [23] Herodotou C. Hlosta M. Boroowa A. Rienties B. Zdrahal Z. and Mangafa C. Empowering online teachers through predictive learning analytics.

British Journal of Educational Technology, (50)6: 3064–3079, 2019.

- [24] Lin, J. Dai, W. Lim, L.-A. Tsai, Y.-S. Mello, R. F. Khosravi, H. Gasevic, D. & Chen, G. Learnercentred analytics of feedback content in higher education. in 13th International Learning Analytics and Knowledge Conference, 2023, pages 100–110.
- [25] Ouyang F. Wu M. Zheng L. Zhang L. and Jiao P. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, 20)1: 1–23, 2023.
- [26] Kalman Y. M. A theoretical framework for organizational learning analytics. *Mini LAK Theory* Workshop in Learning Analytics and Knowledge Conference (LAK). (Online due to COVID-19). 2020.
- [27] Kalman Y. M. Learning analytics in higher education (in Hebrew)." *The Open University, Israel, Seminar*, 26/05/2022.
- [28] Navon R. and Kalman Y. Organizational Learning Analytics for Higher Education Institutions: A Novel Theoretical Framework (in Hebrew). In *The* 18th Chais Conference on the Study of Innovation and Learning Technologies: Learning in the Digital Era, Ra'anana, Israel, 02/2023.
- [29] Cerratto Pargman T. and McGrath C. Mapping the ethics of learning analytics in higher education: A systematic literature review of empirical research. *Journal of Learning Analytics*, (8)2: 123–139, 2021.
- [30] Tzimas D. and Demetriadis S. Ethical issues in learning analytics: a review of the field. *Educational Technology Research and Development*, (69): 1101–1133, 2021.

Enhanced Precision in Built Environment Measurement: Integrating AprilTags Detection with Machine Learning

Shengtao Tan¹, Aravind Srinivasaragavan¹, Kepa Iturralde^{1,2}, Christoph Holst¹

¹Chair of Engineering Geodesy, School of Engineering and Design,

Technical University of Munich, 80333 Munich, Germany

²Chair of Digital Transformation in Construction, Institute of Construction Management, Faculty of Civil and Environmental Engineering, University of Stuttgart, 70569 Stuttgart, Germany.

shengtao.tan@tum.de, aravind.srinivasaragavan@tum.de kepa.iturralde@tum.de christoph.holst@tum.de

Abstract -

In the field of building renovation with prefabricated modules, accurately locating and identifying connectors' positions and orientations is an essential technological challenge. For building renovation with prefabricated modules, traditional methods like total stations are not only time-consuming but also highly dependent on experienced technicians. However, previous research has proven that ApriTtag tags can be effectively used in building measurements. This paper proposes a refined AprilTag detection pipeline that integrates machine learning techniques, significantly improving detection accuracy. Moreover, this process can be easily used by non-experts making it more accessible and less timeconsuming.

Keywords -

AprilTag; Machine Learning; Neural Network; Building Measurement

1 Introduction

Existing building stock renovation is a critical aspect of construction automation. For instance, accurately determining the position of connectors is essential for the installation of prefabricated panels on building exteriors. However, traditional measurement methods, such as the use of total stations(see [1]), have several limitations, including being time-consuming, requiring skilled technicians, etc. To address these challenges and leverage the advantages of automation, the integration of computer vision techniques with a visual fiducial system is suggested as a viable solution. In this paper, we propose a refined Apriltag(see [2]) detection pipeline integrated with machine learning to tackle the above problem. Figure 1 shows its architecture. We will first discuss the research gaps in the accuracy of AprilTag localization. Then we will introduce the components of the pipeline in detail. The experiments in section 4 show that our refined pipeline has very good accuracy. This research is part of the ENSNARE project[3].

2 Research gaps

The AprilTag is widely applied across various domains, including robot navigation & localization, industry automation, and augmented reality. However, existing research puts limited focus on localization accuracy when employing AprilTags. López-Cerón et al.[4] conducted an analysis of AprilTag's accuracy, but their study was limited to camera-to-target distances of no more than 6 meters. Kallwies et al.[5] extended their testing range to 18 meters, yet their focus was on pixel-level errors, rather than millimeters. Similarly, Olson et al.[6] investigated distances up to 80 meters, but their error tolerance is in meters. Additionally, research on the large-scale layout of AprilTags is very scarce. Kallwies et al.[5] constructed a simulated 7 x 22 grid with 152 AprilTags, each 13 cm in size. Beyond this study, there is a lack of research exploring the potential and challenges of large-scale AprilTag layouts. To sum up, there is a notable gap in research and applications concerning the combination of high accuracy and the large-scale layout of AprilTags. But in our context, we mostly applied AprilTags in large-scale outdoor environments and want to achieve accuracy that is comparable to or even beyond traditional methods like total-station measurement. Therefore, this paper aims to bridge this gap.



Figure 1. A flowchart outlining the pipeline

3 AprilTag detection pipeline

As in previous experiences with OpenCV and AprilTags, we **calibrated a camera**, in our case a Sony A7R4 [7] with different lenses using a checkerboard in ambient lighting conditions. For each focal length, we took pictures in different camera positions and orientations while making sure that the checkerboard eventually covered the entire image frame. Finally, we loaded all the pictures to the MATLAB R2023b Camera Calibration toolbox and only left around 15 pictures that had the smallest pixel errors as our final calibration candidates.

Before using the AptilTag detector, we first **Preprocess the images** with the following steps:

- 1. **Removing outliers:** We filter out the pictures that have undetected tags (false negative) and non-existent tags (false positive).
- 2. **Minimizing image distortion:** cv2.undistort function is applied to undistort images using the camera calibration and distortion matrix from the calibration step. This step can increase the detection rate.
- 3. **Grayscale conversion:** We convert the images to the acceptable format of AprilTag Detector, which is grayscale.
- 4. **De-noising:** A Gaussian filter with kernel size 5 x 5 is applied to reduce the sensor noise of the camera.
- 5. **Sharpening:** We subtract the smoothed image of the last step from the original image

For the **AprilTag detection** algorithm, we used the AprilTag3(see [8]) library of Python. The motivation is that the Python bindings allow efficient development with powerful libraries, such as NumPy. We mostly use the default parameters of the AprilTag3 binder.

Using a **Coordinate transformation** one can express the pose of tags in the coordinate frame built up at the center of the bottom left tag. To do so, we first obtain the rotation matrix and translation vector for each tag from the tag frame to the camera frame using cv2.SOLVEPNP_IPPE_SQUARE function. With the bottom left tag as tag 1 and the other as tag 2, tag 2's position and orientation(in the new coordinate system) can be expressed using the formulas below:

$$P_{tag2} = R_1^T * (t_2 - t_1) \tag{1}$$

$$R_{2|1} = R_1^T * R_2 \tag{2}$$

where R_1 is the rotation matrix of tag 1. t_2 and t_1 are the translation vector of tag 2 and tag 1 respectively. $R_{2|1}$ is the rotation matrix of tag 2 w.r.t tag 1. The Euler angles of tag 2 can be further computed from $R_{2|1}$.

A **Postprocessing** was necessary because, we took 10 to 200 pictures of the tags, which means we have lots of measurements. In the end, we only need one final and

accurate measurement. Therefore, the postprocessing is a crucial step of the whole pipeline. We took two strategies for postprocessing:

- 1. **Mean values:** The mean value is used when there are not many measurements or DBSCAN can't find a valid dense area of the measured point cloud.
- DBSCAN clustering: By visualizing the positions of detected tags' center, there are usually some outliers and also a dense area of point cloud (See Figure 2). To filter out the outliers, we use DBSCAN(see [9]) to find the core measurements.



Figure 2. Point cloud visualization

We also implemented two other strategies based on the ranking of pose_err and blur scores.

- 1. **Pose_err ranking:** pose_err represents the objectspace error of the estimation. The idea is to only select the tag measurements that have lower pose_err.
- 2. **Blur score ranking:** Blurry images can contribute to poor detection accuracy. To avoid blurry images, we first mask the pictures and only keep AprilTags because those squares are our ROIs(Region of Interest). Then we compute the total variance of the laplacian of ROIs using cv2.Laplacian as the blur score for the image. Eventually, we only compute the mean of less blurry pictures.

The performance comparison of the above 4 strategies will be introduced later.

3.1 Machine learning correction

The AprilTag detection pipeline is highly susceptible to lighting conditions which can influence its performance and these "noises" are hard to determine and control. We have devised a neural network trained on a dataset of hundreds of images, such as the one shown in Figure 4c, taken in different lighting conditions and positions. Essentially, it is trained to predict the correct ground truth values based on the experimental observations. The idea is that the neural network can then improve on the values predicted by the AprilTag detection pipeline. The dataset consists of ground truth values of different AprilTags in different images such as their 3d coordinates as well as camera information. Our approach is inspired by [10].

The neural network is implemented in Python version 3.10.10 using Tensorflow(See [11]). We have used a standard fully connected feed-forward network with 2 hidden layers as shown in Figure 3.



Figure 3. Network Architecture

The dataset is first randomly shuffled to increase the uniformity of the dataset. We split the dataset randomly into a training set and a testing set. The size of the training set is 95% of the original dataset. We then initialize the neural network with the following parameters: we use Adamax for the optimization process along with 12 regularization with regularisation constant 0.01234. We use mean absolute error for the loss function and finally, we set the batch size to 40 and train the model for 200 epochs.

4 Experiments and results

The proposed pipeline was tested in simulation, indoor and outdoor environments respectively.





(b) indoor

(c) outdoor

Figure 4. Example pictures of real-world and simulation experiments

4.1 Simulation testing

For simulation, we use Blender(See [12]) to generate synthetic images (See Figure 4a). The first part of Table 1 shows that using mean values or DBSCAN outperforms the rest two postprocessing methods. In addition, without the influence of environmental noise, our pipeline reached millimeter-level accuracy.

Table 1. The results of proposed pipeline

Sinulation result					
		Mean	DBSCAN	Pose_err	Blur score
$\Delta x/m$	m	0.40	0.40	0.40	0.40
Δy/m	m	-0.45	-0.36	-0.45	-0.45
$\Delta z/m$	m	-0.10	0.10	-0.10	-0.10
Δxy/m	m	0.60	0.54	0.60	0.60
Δxyz/n	nm	0.61	0.55	0.61	0.61
$\Delta \alpha / \gamma$	D	0.04	0.07	0.04	0.04
$\Delta \beta / \gamma$	C	0.03	0.05	0.03	0.03
$\Delta \gamma / \gamma$	C	0.00	0.00	0.00	0.00
		Indoc	or experiment		
	Mean	DBSCAN	Pose_err	Blur score	Total station
$\Delta x/mm$	0.82	-0.43	1.50	1.71	0.60
Δy/mm	2.23	1.31	2.97	1.49	-0.20
$\Delta z/mm$	-5.59	-5.37	0.01	-8.29	-1.29
$\Delta xy/mm$	2.38	1.38	3.33	2.27	0.63
$\Delta x y z/mm$	6.07	5.54	3.33	8.59	1.44
Δά/°	-0.71	-0.38	-0.30	-1.07	-
$\Delta \beta / ^{\circ}$	-0.84	-1.07	-1.03	-1.14	-
$\Delta \gamma / ^{\circ}$	0.01	0.02	0.04	0.03	-
		Outdo	or experimen	t	
	Mean	DBSCAN	Pose_err	Blur score	Total station
$\Delta x/mm$	4.42	1.87	3.24	7.22	1.15
Δy/mm	-4.27	-4.20	-2.60	-11.72	1.00
$\Delta z/mm$	19.83	11.84	16.74	22.11	-1.29
Δxy/mm	6.15	4.60	4.15	13.77	1.52
$\Delta x y z/mm$	20.76	12.70	17.25	26.04	2.00
Δα/°	0.73	1.65	0.07	1.43	-
$\Delta \beta / ^{\circ}$	1.08	0.49	0.82	2.96	-
$\Delta \gamma / ^{\circ}$	0.10	-0.06	0.10	0.20	-
		Neural ne	etwork correct	tion	
		Indoor c	oorection	Outdoor	correction
		Mean	DBSCAN	Mean	DBSCAN
$\Delta x/m$	m	0.76	-0.42	4.00	-0.25
Δy/m	m	-3.08	-2.37	-7.33	-8.22
$\Delta z/m$	m	2.56	1.94	5.50	4.44
Δxy/m	ım	3.17	2.41	8.35	8.22
Δxyz/n	nm	4.08	3.09	10.00	9.35
$\Delta \alpha / \gamma$	D	0.37	0.48	1.52	-0.29
$\Delta \beta / ^{\circ}$	0	-2.85	-2.35	-5.24	-1.93
A A	C	3 7 3	3 05	5 3 8	2.06

4.2 Real-world testing

Building a large-scale testing environment is not only challenging, but it is also impractical to test the entire building's exteriors with exact ground truth. To mitigate this, we used a 2000 x 2000 millimeters calibrated wooden board as a substitution for placing our AprilTags. Each tag is 15 cm in length and is augmented with four reflective tapes at its corners, for measurements with a total station. The AprilTags were accurately placed with rulers.

Regarding the total station measurements, we used the Leica TC702. For better accuracy of the total station results, each layout was measured from 2 to 3 different positions in two phases.

The **indoor experiments** were carried out in a lab with shooting distances ranging from 4 to 15 meters. Figure 4b shows an example of the indoor picture.

The second part of Table 1 shows the accuracy of a representative indoor example layout. Bold numbers are the best results among 4 postprocessing methods.DBSCAN still performs the best in general, but pose_err also shows good results. Our proposed pipeline reaches the total-station-level accuracy.

The **outdoor experiments** are carried out on the campus with shooting distances ranging from 5 to 18 meters. The third part of Table 1 shows the accuracy of a representative outdoor example layout. One major reason that decreased the accuracy could be the poor illumination of the experiment environment. Among all the postprocessing methods, DBSCAN also achieved the best performance.

The **neural network** was tested on the test dataset and performed quite well with a mean absolute loss of 3.5547mm and an accuracy of 0.7551. The last part of the Table 1 shows a sample correction result. We can see there's an improvement in the depth measurement.

Further, we tracked the consumed time for our pipeline in the experiments. Table 2 shows that our method is much more efficient than the total station while achieving similar accuracy.

Table 2. Spent time comparison			
AprilTag detection total station			
setting up/min	3	13	
measuring/min	15	28	

5 Conclusion

The research described in this paper has successfully demonstrated a collection of techniques that allow for an accurate and faster process of building measurements, especially compared to conventional methods. Future research will include:

- 1. Improving the machine learning algorithm. We are currently exploring the possibility of using 3D rendering software to generate synthetic images through which we can generate more datasets.
- 2. Integrate and test the proposed pipeline with UAVs. We will work closely with another team on this project that is currently developing a UAV for sticking the AprilTags.

In the next steps of the research, the idea is to use this technique in real buildings.

Acknowledgements



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 958445.

References

- R. Volk, J. Stengel, and F. Schultmann. Building information modeling (bim) for existing buildings — literature review and future needs [autom. constr. 38 (march 2014) 109–127]. Automation in Construction, 38:109–127, 2014. doi:10.1016/j.autcon.2013.10.023.
- [2] K. Iturralde, J. Shen, and T. Bock. Apriltag detection for building measurement. In *Proceedings of the 40th ISARC*, pages 589–592, Chennai, India, 2023. IAARC. ISBN 978-0-6458322-0-4. doi:10.22260/ISARC2023/0079.
- [3] ENSNARE. https://www.ensnare.eu/, Accessed: Dec. 13, 2023.
- [4] A. López-Cerón and J. Cañas. Accuracy analysis of marker-based 3d visual localization. pages 1124–1131, 2022. doi:10.17979/spudc.9788497498081.1124.
- [5] J. Kallwies, B. Forkel, and H. Wuensche. Determining and improving the localization accuracy of apriltag detection. In 2020 IEEE International Conference on Robotics and Automation (ICRA), pages 8288–8294, 2020. doi:10.1109/ICRA40945.2020.9197427.
- [6] E. Olson. AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407. IEEE, 2011.
- [7] Sony Camera. https://www.sony.de/ electronics/wechselobjektivkameras/ ilce-7rm4a/specifications, Accessed: Dec. 19, 2023.
- [8] AprilTag Detector. https://github.com/ pupil-labs/apriltags, Accessed: Dec. 13, 2023.
- [9] scikit. https://scikit-learn.org/stable/ modules/generated/sklearn.cluster. DBSCAN.html, Accessed: Dec. 13, 2023.
- [10] B. Mariusz, D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. LaJackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba. End to end learning for self-driving cars, 2016.
- [11] Tensorflow. https://www.tensorflow.org/, Accessed: Dec. 16, 2023.
- [12] Blender. https://www.blender.org/, Accessed: Dec. 13, 2023.

Training and Generating Apartment Plan Graphic Images for Commercial Uses

Taesik Nam, Youngchae Kim, Sumin Chae, Youngjin Yoo, Jin-Kook Lee*

Dept. of Interior Architecture & Built Environment, Yonsei University, Seoul, Republic of Korea <u>nsmnts@yonsei.ac.kr</u>, <u>ca71232@yonsei.ac.kr</u>, <u>chaesumin17@gmail.com</u>, <u>yooyj@yonsei.ac.kr</u>, leejinkook@yonsei.ac.kr*

Abstract -

This paper introduces a methodology utilizing image generation AI to create 3D floor plan images for Korean-style apartments. The growing interest in spatial dynamics, driven by changes in residential preferences, has elevated the importance of widespread comprehension and the demand for floor plans. Specifically, apartment floor plans, crucial for commercial purposes, require visually compelling representations that are both diverse and detailed yet easily comprehensible. Traditional 2D floor plans face limitations in public understanding due to the use of architectural terminology and specialized symbols. As an alternative, 3D floor plan images offer a comprehensive understanding of three-dimensional structures and arrangements, with potential applications across various commercial sectors. The recent emergence of image generation AI presents opportunities for the automated creation and alteration of architectural images. Fine-tuning the generated results can be achieved through detailed prompt settings and additional training techniques. The methodology for constructing a 3D floor plan image generation model for Korean-style apartments involves: 1. Creating a 3D Building Information Modeling (BIM) model based on Korean-style apartment floor plans and extracting 3D floor plan images based on different areas. 2. Training the floor plan image generation AI model by combining the extracted images with text and utilizing the trained model to automatically generate 3D floor plan images. The resulting images can find applications not only in architecture but also in real estate, design, advertising, and other diverse commercial fields. This approach suggests the potential development of a versatile model applicable across various domains, especially when integrated with generative models.

Keywords -

Architectural Visualization, 3D Floor Plan, Image Generation AI

1 Introduction

This paper proposes a methodology for automatically generating 3D floor plan images of residential spaces. Due to the rapid changes in housing types and the increase in single-person households [1], there is a growing interest in the form and spatial usage of residential spaces [2]. In response to these changes, floor plans play a crucial role in satisfying the demands of residents. Among architectural drawings, floor plans, especially in 3D, are essential sources of information in the architectural and design processes, clearly depicting aspects such as spatial dimensions, gaps, and arrangements. Unlike traditional 2D floor plans, 3D floor plans convey the three-dimensional form and visual information of a space more vividly. 3D floor plans facilitate an intuitive understanding among users, enabling them to assess the layouts and specific details of the floor plans [3]. For example, in cases of selecting apartment blueprints, clients achieve a deeper understanding and expedite their decision-making process through 3D representations as opposed to 2D drawings [4]. Recent advancements in image generation AI models have significantly improved the representation of such visual information [5, 6, 7, 8]. These models can be utilized to create high-quality 3D floor plan images.

2 Background

2.1 Evolution of 3D Floor Plan Visualization

Architectural visualization is a crucial process in expressing architectural design and concepts visually. It is utilized from the early stages of design in architectural projects to effectively convey ideas and concepts to designers, stakeholders, and even construction personnel. Utilizing various forms and tools such as drawings, blueprints, photos, and 3D models architectural visualization is particularly exemplified by the use of floor plans, which accurately capture detailed spatial information, serving as a key element in conveying initial design requirements [9].

Among these, floor plans encompass essential spatial information, including precise dimensions, material specifications, and the arrangement of architectural elements, making them indispensable documents throughout the architectural process. Traditional 2D floor plans have the advantage of simplicity and clarity in conveying basic information. However. their representation of space in 2D has limitations in realistically visualizing both external and internal designs, especially posing challenges for non-professionals or the general public to comprehend [10]. On the other hand, 3D floor plans emphasize visual effects, providing a clear depiction of 3D information. For example, according to Kashlev(2008), 3D building models not only appear to be more aesthetic to ordinary people than architectural drawings, but also can benefit architects in detecting any inconsistencies in their designs [11]. Additionally, according to V Stojanovic et al.(2019), 3D floor plans can enhance stakeholder engagement by improving 3D visualizations or providing an initial assessment of the building's state [12].

They enable a detailed understanding of spatial form and layout, contributing to seamless collaboration and communication among architectural stakeholders [13, 14]. These advantages highlight the selective utilization of 2D and 3D floor plans based on various stages and purposes within architectural projects.

2.2 Floor plan and Generative AI

The evolution of image generation AI [15, 16, 17, 18] has progressed over time through continuous In the early stages, technological innovation. predominantly noise-based methods were employed using simple algorithms for image generation. Subsequently, with the advancement of deep learning technology and extensive research on generative models, the field of image generation has undergone revolutionary changes. Notably, the introduction of deep learning models such as Generative Adversarial Networks (GAN) has significantly enhanced the quality and diversity of image generation [19]. For example, there exists research on generating images by design style utilizing image generation AI model and large language model (LLM), employing these advanced technologies to foster innovation in the field of image generation [20, 21].

In the realm of image generation, two primary approaches are recognized: text-to-image (txt2img) and image-to-image (img2img). Txt2img involves generating images based on given text descriptions, utilizing semantic information extracted from the text [22]. On the other hand, Img2img is a technique that generates new images based on given input images. The image generation AI model understands the features and structure of the image from the input and produces an output image. Training image generation AI models for specific scenarios or types according to their intended purposes enhances the effectiveness of obtaining results. Such image generation AI finds applications in diverse fields, including education, architecture, art, virtual reality, and is anticipated to expand into even more varied domains in the future [23].

3 Model Fine-tuning for 3D Floor plan

3.1 Overall process

This paper employed the Stable Diffusion (SD) among various image generation AI models to fine-tune a model for 3D floor plan. SD maintains stable pixel values during the image generation process by utilizing the diffusion algorithm [1]. Consequently, it can generate high-quality, stable, and consistent images while minimizing noise. To train the base model, we utilize the Low-rank Adaptation (LoRA) method, which updates only a select few parameters for efficient model finetuning with minimal data. This approach allows for quick adaptation to new tasks by leveraging the knowledge from the base model without extensive retraining. Highquality, consistent training data, composed of image-text pairs, is crucial, as detailed and specific text descriptions enable the model to generate more accurate and expressive images [24]. Successful few-shot learning with LoRA necessitates high-quality training data with consistent representation [25].

3.2 Data Preparation

For high-quality dataset, we collected floor plans of Korean apartments from 'X' apartment and 'R' apartment which are reputational brands in Korea. The datasets are consisted of approximately 1,000 floor plans. We focused on the 2Bay and 3Bay structures characteristic of Korean apartments, generating for each floor plan, a single 3D Building Information Modeling (BIM) model. From these BIM models, 3D floor plans are extracted and used as training data. The 3D BIM models are created using 'R' software, and the 3D floor plans are rendered from them using 'E' software. Additionally, we resized the floor plans to the 500 pixels in width and height and prepared the text descriptions of the 3D floor plans for the LoRA training. We acquired the text datasets by observing and describing the 3D floor plans personally for raising the quality of training datasets, which are directly connected with the performance of the model. We produced the pairing datasets, matching the pairs of image and text data one on one. And we paired 100 pairing datasets in total. The example of the dataset for LoRA training is shown in figure 1.



Figure 1. The example of the dataset for 3D Floor plan

3.3 Training and Test

Based on the training datasets, we trained the base model with LoRA. The training was required about 30 minutes per one training session with RTX 3060 GPU. As a result, we build a 3D floor plan model with the extension ".safetensors" and a size of 144 megabytes. Inserting the 3D floor plan model file into the SD Web-UI's Model Management folder allows for the application of the model via text description. Consequently, this facilitates the generation of images based on the text data used for training when such data is inputted [24]. We proceeded performance test for evaluating the 3D floor plan model and comparing the base model and the trained model. The result of the floor plans generated with base model is on the left, and the other generated with trained model is on the middle and right. The result of comparison is presented in Table 1. The 3D floor plans of base models are not discriminated as bay structures unlike the plans of trained model.

Table 1. The result of comparison

model	Туре	Output
Base model	2 & 3Bay apartment	
Trained model (3D	2bay apartment	
floor plan model)	3Bay apartment	

The results from the base model were observed to be of low quality and resembled floor plans with errors. In contrast, the trained model applied to the LoRA model produced high-quality floor plans that are easy to understand. Furthermore, the trained model with LoRA created more accurate and well-discriminated floor plans compared to the base model. This performance test suggests that the importance of the LoRA trained model.

4 Demonstration

Table 2. The result of image generation from 3D floor plan model

Туре	Trained model (3D floor plan model)		
2 Bay Apartment			
3 Bay Apartment			

We found that the results of 3D floor plan are highquality and improved in performance. The layout of the floor plan is obvious and the factors such as doors are distinct. In addition, the furniture in the space is detailed. It is shown in table 2.

5 Conclusion

This paper proposes a process that leverages an image generation AI model for generating 3D floor plans. The research focuses on architectural visualization, with a particular emphasis on 3D floor plans that vividly represent spatial dimensions. We validated the potential effectiveness of utilizing image generation AI, specifically for Korean-style apartment spatial structures, to efficiently generate 3D floor plans. We discovered that this process can be utilized not only for the Korean-style apartment, but also for the layouts, such as those for other styles of apartments, houses, and offices. Moreover, we found that this model could be beneficial for architects in the field as well as for the public who struggle to understand traditional floor plans offering them intuitive understanding. We expect that this model to be useful not just for interior design but also for building construction, designing ships, and other architecture work.

6 Acknowledgement

This work is supported in 2024 by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269)

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MIST) (No. NRF-2022R1A2C1093310).

References

- Lim, Y., Hwang, M. (2023). Automatic Prompt Optimization Using ChatGPT: A Novel Approach for Tag Expansion and Image Generation with Stable Diffusion. Korea Artificial Intelligence Conference, 165-166.
- [2] Maeng, H. Y. and Hyeon, K. H. (2021). Deep Learning-Based Big Data Analysis of Apartment Unit Floor Plans: A Study Focused on Korean Apartments from 1970 to 2020. Journal of the Korean Society of Interior Design, 30(3), 65-76.
- [3] Choi, S. H., Nam, T. S., & Lee, J. K. (2023). Automated Generation of 3D Rendering Graphic Floor Plan based on Gen AI. Journal of the Korea Institute of the Spatial Design, 18(7), 263-271.
- [4] Gerstweiler, G., Furlan, L., Timofeev, M., & Kaufmann, H. (2018). Extraction of structural and semantic data from 2D floor plans for interactive and immersive VR real estate exploration. Technologies, 6(4), 1-27.
- [5] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674-10685.
- [6] Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3836-3847).
- [7] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125.
- [8] Betker, J., Goh, G., Jing, L., Brooks, T., ... & Ramesh, A. (2023). Improving image generation with better captions. Computer Science, 2(3), 1-19.
- [9] Huang, C. Y., Yien, H. W., Chen, Y. P., Su, Y. C., & Lin, Y. C. (2017). Developing a BIM-based visualization and interactive system for healthcare design. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 34), IAARC Publications, (pp.51-58).
- [10] Fikkert, F. (2019). *Linking indoor 3-dimensional visualizations with physical architectural floor plans*, Bachelor's thesis, University of Twente.
- [11] Kashlev, D. (2008). Efficient 3D building model generation from 2D floor plans. Master's thesis, Massachusetts institute of technology, MIT Libraries.
- [12] Stojanovic, V., Trapp, M., Richter, R., & Döllner, J. (2019, January). Generation of Approximate 2D and 3D Floor Plans from 3D Point Clouds. In VISIGRAPP (1: GRAPP), (pp. 177-184).

- [13] Clayton, M. J., Warden, R. B., & Parker, T. W. (2002). Virtual construction of architecture using 3D CAD and simulation. Automation in Construction, 11(2), 227-235.
- [14] Park, S., & Kim, H. (2021). 3DPlanNet: generating 3D models from 2D floor plan images using ensemble methods. Electronics, 10(22), 2729.
- [15] Bar-Tal, O., Yariv, L., Lipman, Y., & Dekel, T. (2023). Multidiffusion: Fusing diffusion paths for controlled image generation, Proceedings of Machine Learning Research, 202, 1737-1752.
- [16] Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G., & Cohen-Or, D. (2022). An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618.
- [17] Radhakrishnan, A. M. (2023). Is midjourney-AI the new anti-hero of architectural imagery & creativity?. GSJ, 11(1), 94-104.
- [18] Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., & Scarselli, F. (2020). Image generation by GAN and style transfer for agar plate image segmentation. Computer methods and programs in biomedicine, 184, 105268.
- [19] Goodfellow, I., Pouget-Abadie, J., Mirza, M., ... & Bengio, Y. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems, (pp. 2672-2680).
- [20] Jeong, H., & Lee, J.K. (2023). Study on the Applicability of Image Generation AI in Interior Architecture: Generating Images Based on Interior Design Styles. Journal of the Korea Institute of the Spatial Design, 247-250.
- [21] Ploennigs, J., & Berger, M. (2023). Ai art in architecture. AI in Civil Engineering, 2(8), 1-11.
- [22] Ramesh, A., Pavlov, M., Goh, G., Gray, S., ... & Sutskever, I. (2021). Zero-shot text-to-image generation. In International conference on machine learning (pp. 8821-8831).
- [23] He, X., & Deng, L. (2017). Deep learning for image-to-text generation: A technical overview. IEEE Signal Processing Magazine, 34(6), 109-116.
- [24] Zhang, C., Zou, Y., & Dimyadi, J. (2021). A systematic review of automated BIM modelling for existing buildings from 2D documentation. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 38, pp. 220-226). IAARC Publications.
- [25] Hu, J.E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. ArXiv, abs/2106.09685.

Research on an Open Source Physical Simulator for the Autonomous Construction Machinery Development

Daisuke Endo¹, Yosuke Matsusaka¹, Genki Yamauchi¹ and Takeshi Hashimoto¹

¹Public Works Research Institute, Japan

endou-d177cl@pwri.go.jp, matsusaka@mid-japan.com, yamauchi-g573bs@pwri.go.jp, t-hashimoto@pwri.go.jp

Abstract -

Facing a labor shortage, the civil construction industry is increasingly focused on automating machinery. The Public Works Research Institute in has introduced 'OPERA,' Janan an onen development platform, to encourage stakeholder participation in areas like IT, robotics, and AI. Key to **OPERA** is its open-source software for model-based autonomous construction. Currently, it offers two physical simulators with distinct physics engines, enabling user-specific applications. This paper covers their technical differences, and current states. Keywords -

Autonomous Construction; Dynamics Simulator; OPERA;

1 Introduction

To support technical research and development, the Public Works Research Institute is developing OPERA (Open Platform for Earthwork with Robotics and Autonomy), for autonomous construction development [1]. One key component of OPERA is its dynamics simulator, as shown in Figure 1, available on GitHub [2]. This simulator plays a crucial role in the development of automation software in construction machinery by integrating models of both machinery and terrain. It visually demonstrates how terrain reacts to machinery operations and generates numerical data about the conditions of both machinery and terrain. This paper elaborates on the dynamics simulator's technical aspects and discusses the two versions of simulators, each utilizing a different physics engine.

2 Overviews of the Simulator

This chapter introduces the dynamics simulator for research and development of autonomous construction technology provided by OPERA. First, the essential requirements for this simulator are described. In addition, the reasons for selecting the two types of dynamicsengine-based simulator platforms developed to meet these requirements are summarized.



Figure 1. Scope of the open platform OPERA

2.1 Simulator Requirements

The dynamics simulator is designed for developing software that automates construction machinery, aiming to replace real machines and environments with a virtual setup. Its primary objective is to be Open Source Software (OSS) to encourage widespread usage without mandating users to disclose their source code. Additionally, the simulator should be cost-effective and user-friendly[1]. To ensure these, it utilizes existing platforms with robust ecosystems instead of proprietary systems:

The simulator's technical specifications include:

- 1. Compatibility with ROS [3] and other middleware.
- 2. 3D rendering to display calculation results.
- 3. Physical-based behavior simulation of construction machinery.
- 4. Capability to handle terrain deformation due to interaction with construction machinery.

5. Real-time or faster calculation and rendering speeds. Concerning these specifications, it's crucial that the simulator is compatible with ROS (Robot Operating System) or ROS2, since these systems are widely utilized in the field of autonomous construction. The function to interact with other middleware is also important. For terrain interaction, the simulator is required to realistically mimic the way construction machinery alters the ground, as observed in earthwork operations, thereby accurately reflecting the physical transformations.

2.2 Selection for Simulator Platform

As a 3D physical simulator compatible with ROS middleware adopted by OPERA, Gazebo [4] is a widely recognized and freely accessible tool. Given its popularity among ROS users and the established ecosystem it offers, Gazebo is a general choice in this domain.

However, since Gazebo is designed for integration with ROS, it lacks compatibility with other middleware such as OROCOS [5], OpenRTM [6], etc. As previously mentioned, the policy for OPERA's simulator is to make it usable with middleware other than ROS, thus platforms other than Gazebo were also considered.

In recent years, there has been a trend to adopt game engines as platforms for physical simulators for software development to automate cars and robots. Compared to traditional dedicated simulators, they have the following characteristics and advantages, which are the main reasons for their increasing use:

Availability of abundant 3D rendering functions

and physical engines.

- Well-developed community and support.
- High versatility and extensibility.
- Potential for reduced development costs.

The major game engines adoptable for the simulator platform are Unity [7] and Unreal Engine [8]. Among these two, OPERA adopted Unity, which had a larger developer community and more abundant assets as of March 2021 when the selection was made. This decision was based on the aforementioned policy of making it easy for users to handle the simulator to increase the number of users as much as possible. Table 1 summarizes the results of the study on simulator selection. The criteria for each rating were determined by relative evaluation based on the results of our own preliminary research on each simulator. In conclusion, we decided to prepare two types of simulators with different physics engines, and let users choose which one to use depending on the cost-effectiveness they seek. This paper describes about these two types of simulators in following sections.

3. PhysX Version of the Simulator

This chapter describes the PhysX [11] version of the simulator, focusing on the technical content related to the construction machinery model, the soil ground model, and the interaction between construction machinery and soil ground.

Table 1. Results of platform selection for simulator.

					©: ∠:	○: ×:
	()	()	()
	()			()	()
1			O	O	0	0
			×	×	0	0
2			0	0	0	0
3			0	0	0	0
4			()	()	(()
5	_		×		$\odot \sim \odot$	O
			0:		0:	-
			0:	0:	0:	× :
			×	×	△~0	∆~0

3.1 Construction Machinery Model

The simulation model of construction machinery was crafted for realistic representation and user customization without infringing manufacturers' intellectual property. The process, shown in Figure 2, included:

- 1. Capturing the machine's (that is our property) exterior as a 3D point cloud using LiDAR.
- 2. Disassembling each machine into links to create a 3D mesh model from the point cloud. Joint connections, relative positions, and kinematic parameters were extracted, using both manufacturer catalogs and preliminary surveys.
- 3. Describing the machine in URDF (Unified Robot Description Format), including weight and other inertia parameters based on measurements, despite some errors.
- 4. The URDF file was imported into Unity, attaching components for operation as a game object. In the PhysX version, the ArticulationBody component was applied to all links, setting joint types, motion ranges, and dynamics. The tracks were modelled as a six-wheel skid steer utilizing WheelCollider Component, rather than precise tracks.



Figure 2. Making procedure for machinery model.

3.2 Soil Ground Model

In OPERA's simulator, the standard Unity Terrain object is enhanced to simulate terrain deformation due to construction machinery. To model moist soil's viscous behavior efficiently in real-time, a simplified method based on the Discrete Element Method (DEM), inspired by Daniel Holtz et al. [13], is implemented. This model treats soil as particles of radius R, where a cohesion force F acts between particles closer than a distance D. When particles are farther apart, no force acts between them (Figure 3). These parameters can be changed via GUI.



Figure 3. Adhesive force model for soil particles.

3.3 Physical Interaction

At December 2023, the PhysX version doesn't calculate the interaction force between construction machinery and soil ground during collapse or deformation. When parts of the machinery, excluding the hydraulic shovel's bucket, contact the ground, they are treated as rigid bodies in simulation. However, two distinct interference shapes are set for the bucket part:

1. Bucket-Terrain interference detection shape.

This is an interference shape that only detects interference and disables the reaction forces generation with other objects. Particles to represent soil behavior are generated when the bucket contacts the ground.

2. Particle retention interference shape.

This interference shape in the bucket prevents interaction with the ground (Terrain) while effectively managing particles and other objects.

When the bucket overlaps with the Terrain, an equivalent volume is removed, generating particles in the bucket. These interact as rigid bodies and are removed upon contact with the Terrain, increasing its height equivalent to their volume.

This method, a simplified representation of the construction machinery-soil-ground interaction, calculates only the force between particles and the bucket's retention shape. The force at particle generation is impulsive and unstable. Additionally, the feature for simulating terrain deformation through particle creation can be toggled on and off via the GUI.

4. AGX Dynamics Version of the Simulator

This chapter outlines the AGX [12] version of the simulator, highlighting the technical aspects of the construction machinery model, soil ground model, and their interaction. Due to space constraints, it primarily details the differences from PhysX version.

4.1 Construction Machinery Model

In the construction machinery creation process outlined in Section 3.1, the key distinction between the PhysX and AGX versions is in step 4. The AGX version, utilizing the AGXUnity [14] plugin for Unity by AGX Dynamics, attaches various components and sets parameters differently. Unlike the PhysX version, which uses Unity's standard ArticulationBody, the AGX version employs AGXUnity's RigidBody component to define physical properties, joint types, constraints, and dynamics model parameters. The tracks is simulated utilizing AGXUnity's TrackWheel and Track components, with sprockets and idlers modeled by TrackWheel and the track itself by the Track component. The AGX version's track comprises 44 evenly spaced flat plates revolving around the track's outer orbit, though further details are not included here.

4.2 Soil Ground Model

The Unity standard object Terrain has the DeformableTerrain component provided by AGXUnity attached. This synchronizes the shape of AGXTerrain with that of Terrain, and physical property parameters of the soil can be set using this GUI. In addition, some improvements have been made to AGXUnity's DeformableTerrainParticleRenderer to accommodate an increasing number of particles in the visible model. The physical parameters of soil that can be set in the TerrainMaterial within DeformableTerrain the component are diverse. AGXUnity provides three preset types. The AGX version simulator selects the dirt 1 parameter preset by default. It is recommended for simulating typical viscous mud [15] like our test site.

4.3 Physical Interaction

In the AGX version, it is possible to calculate the interaction force between construction machinery and soil ground when the ground is collapsing or deforming, which was not implemented in the PhysX version. This is achieved through a two-step process involving the DeformableTerrainShovel from AGXUnity. First, this componet is attached to the bucket link. Then, several key parameters are set: the Top Edge, Cutting Edge, and Cutting Direction. Finally, the bucket link is registered in the Shovels list within the DeformableTerrain component, as previous section.

5. Conclusion

This paper introduced two types of dynamic simulators provided by the open platform for Autonomous Construction OPERA. The OPERA platform offers two simulators: one based on the PhysX engine, available at no cost except for the Unity license, and the other using the AGX engine. While the PhysX version is freely accessible, it exhibits comparatively less refined soil deformation behavior in interactions with construction machinery than the AGX version. Conversely, the AGX version requires a paid license for its dynamics engine. Both versions are published on GitHub, making them accessible for any users. We highly welcome user feedback as it is instrumental in our ongoing process to enhance OPERA's future convenience and functionality.

References

- [1] Genki, Y., et al:"Proposal of an Open Platform for Autonomous Construction Machinery Development", Proceedings of the 40th ISARC, 2023.
- [2] GitHub: pwri-opera. On-line: https://github.com/pwri-opera, Accessed: 26/12/2023.
- [3] Open Robotics: ROS. On-line:, https://www.ros.org, Accessed: 26/12/2023 x
- [4] Open Robotics: GAZEBO. On-line: https://gazebosim.org, Accessed: 26/12/2023
- [5] OROCOS. On-line: https://orocos.org, Accessed: 11/3/2024
- [6] OpenRTM-aist. On-line: http://www.openrtm.org, Accessed: 11/3/2024
- [7] Unity Technologies: Unity. On-line: https://unity.com, Accessed: 26/12/2023
- [8] Epic Games, Inc.: Unreal Engine. On-line: https://www.unrealengine.com, Accessed: 26/12/2023
- [9] Russel Smith: Open Dynamics Engine. On-line: http://ode.org, Accessed: 26/12/2023
- [10] Bullet Real-Time Physics Simulation. On-line: https://pybullet.org/wordpress, Accessed: 26/12/2023
- [11] NVIDIA.DEVELOPER: PhysX. On-line: https://developer.nvidia.com/physx-sdk, Accessed: 26/12/2023
- [12] Algoryx Simulation AB: AGX Dynamics: Realtime multi-body simulation. On-line https://www.algoryx.se/agx-dynamics/, Accessed: 26/12/2023
- [13] Daniel, H., et al:"Real-Time Mud Simulation for Virtual Environments", ACM Siggraph Symposium on Interactive 3D Graphics and Games, i3D, 2018.
- [14] Algoryx: AGXUnity. On-line: https://github.com/Algoryx/AGXUnity, Accessed: 26/12/2023
- [15] Tomas, B., Mertin, S., Tech report V1.01agxTerrain, https://www.algoryx.se/download/agxTerrain_tech _report.pdf, Accessed: 26/12/2023

Development of an Educational Package for a Construction Robots Simulator

Ci-Jyun Liang¹, Cheng-Hsuan Yang², Liang-Ting Tsai³, and Bruce Alton³

¹Department of Civil Engineering, Stony Brook University, USA ²Department of Civil and Environmental Engineering, University of Alberta, Canada ²RoBIM Technologies Inc., Canada <u>ci-jyun.liang@stonybrook.edu</u>, <u>chenghsuan@ualberta.ca</u>, <u>liangtin@ualberta.ca</u>, <u>bruce@robimtech.com</u>

Abstract -

Despite construction robots being developed in the construction discipline, they are not widely deployed in the industry and field yet. One reason is complex robot control interfaces and the programming methods, which pose challenges for construction workers to learn and use. A robot simulator, RoBIM, has been designed to address this challenge by using a straightforward and userfriendly interface to simplify the robot control and programming process. This paper focuses on the development of an educational package tailored to train construction students and workers in utilizing the RoBIM simulator effectively. The package consists of three modules: column building, brick wall laying, and frame assembling. Through examples within the educational package, this paper illustrates its practical implementation and significance in overcoming the learning curve associated with construction robots, which promotes the integration of the RoBIM simulator within the construction industry.

Keywords -

Robot Simulator; Educational Package; Robot Arm; Tutorials

1 Introduction

The construction industry is at the nexus of technological innovation, with construction robots targeted to redefine the traditional construction process. These automated systems boast the potential to significantly enhance efficiency, precision, and overall project outcomes [1]. Although construction robots have potential in terms of efficiency and precision, their wide integration into construction workflows is still ongoing. This is mainly due to the complicated control interfaces and programming methods that present a formidable learning curve for construction workers [2].

To address this challenge, we developed a

transformative solution– the RoBIM robot simulator [3]. This simulator can directly import the BIM model, create the assembly procedure, plan the robot path, and generate the robot control program. The intuitive and straightforward user interface is the primary feature of this software, assisting construction students and workers in simplifying the construction robot assembly planning and controlling process. Figure 1 shows the interface of the RoBIM simulator.



Figure 1. RoBIM simulator interface

This paper focuses on developing a comprehensive educational package for the RoBIM simulator and showcasing its practical application. This educational package aims to provide a hands-on, accessible, and immersive training experience with the RoBIM simulator, which facilitates the seamless integration of construction robots into field practices. Robot simulator training plays a critical role in preparing workers for the complexities of robot control across industries. Existing literature emphasizes the significance of user-friendly interfaces to reduce the learning curve [4] or using emerging technologies, such as AR/VR [5, 6], in the training procedure. Therefore, it is important to develop an educational package to facilitate the RoBIM simulator training and smooth the learning process. ROS-based (Robot Operating System) or VR-based are two examples of robot training systems, but they are too difficult for workers without programming knowledge or not targeting construction robot programming [6-8].

2 Educational Package

The RoBIM simulator educational package includes three modules: column building module, brick wall laying module, and frame assembling module. Each module has several step-by-step tutorials to guide students to complete some tasks and a challenge task in the end to demonstrate the learning results. In the following subsections, we will introduce each module and some examples. To simplify the tutorial, we select Universal Robots UR5e as the industrial robot arm. Note that the UR5e robot arm in the tutorial can be replaced with any type of industrial robot arm to fulfill the project requirements and budget.

2.1 Column Building Module

The objective of the first module is to let students familiarize themselves with the robot hardware. Particularly, we design a tutorial to introduce the hardware interface of UR5e and demonstrate how to use their built-in functions to control the robot. Three tasks in this module are picking and placing bricks, building a column, and using the camera sensor.

In the pick and place bricks task, the students will write a short program to control the robot to pick up a small brick and place it at the target location. Figure 2 shows the pick and place task in the simulator. Next, the students will program the robot to repeat the pick and place task and build a column using the loop function. The location of each brick needs to be calculated in order to determine the robot placing location. In addition, each brick has to be placed in a different orientation, and thus, students have to consider different scenarios in the program.



Figure 2. Pick and place a brick using the robot arm and the simulator.

The final tutorial task in this module is to use the camera sensor to detect the position of the brick. The students will follow the procedure to calibrate the camera and set up the target object in the image. With the camera sensor, the robot can recognize the brick in the scene and pick it up directly.

The challenge task of this module is to build a brick column using the skills learned in the previous three tasks. Figure 3 illustrates the design of the brick column. The column has six layers with four bricks in each layer and different rotations. The camera sensor is required to automatically detect the position of the brick, as well as the loop function to avoid the redundant code.



Figure 3. The designed brick column.

2.2 Brick Wall Laying Module

The second module is to use the RoBIM simulator to lay a brick wall in both virtual and physical environments. This module starts with the introduction of the RoBIM simulator, including the background, the overall RoBIM simulator, and the user interface. Students will follow the steps to create a robot simulation scene that mimics the physical robot environment. Then, they will utilize the RoBIM programming function to generate the robot work plan for building a brick wall. Figure 4 demonstrates the planning procedure for the robot bricklaying.



Figure 4. The robot bricklaying planning.

Each brick's location has to be determined in the simulator, in particular, their transformation. With such information, the RoBIM simulator can automatically determine the robot motion by forward kinematics and inverse kinematics. Figure 5 shows the process of automatic generation of robot motions to pick and place the brick. The students only need to select the desired brick (Figure 5 up) and click the "pick and place button." The RoBIM simulator will then generate a series of robot motions in order to pick and place the desired brick (Figure 5 down). Finally, the robot control script can be generated directly from the RoBIM simulator and imported to the physical UR robot, as shown in Figure 6. With this control script, the UR robot can complete the task without the worker's intervention.



Figure 5. The automatic robot motion generation.



Figure 6. The robot control script generation.

The challenge task of this module is to lay a doublecurved brick wall (Figure 7). Students will design the brick layout that meets the design requirement and program the robot motions to complete the laying task. The outcome of this challenge is a scaled brick wall built by the physical robot.



Figure 7. The double-curved brick wall.

2.3 Frame Assembling Module

The third module is the frame assembling module. The objective is to program the robot to assemble a predesigned frame. Students can first design a model and then import it to the RoBIM simulator. For example, a timber or a brick wall layout can be designed using different software and imported to the simulator as a model. Figure 8 shows the imported brick wall model.



Figure 8. Import a brick wall layout model to the simulator.

Moreover, this module also introduces two additional sensors in the simulator: photoelectric sensors and force sensors. The photoelectric sensor can be used to detect the material to be picked up, similar to the camera sensor. The force sensor is used to measure the contact force. In this module, we want to use the force sensor to let the robot arm push the picked material to the corner before releasing the gripper.

Finally, to ensure the robot can safely complete the task, a collision avoidance algorithm is necessary [9]. Figure 9 demonstrates an example of the collision between the picked frame and the brick wall. Students will use the built-in collision avoidance function to generate a collision-free robot motion automatically.



Figure 9. Collision between the picked frame and the brick wall.

The final challenge of this module is to assemble a wood frame. Different types of studs are provided, and students need to design a wall frame. The photoelectric sensor and force sensors are required in the challenge to detect the stud and push it to the corner of the assembly table. A brick wall is built between the material and the target position as an obstacle such that the collision avoidance function in the RoBIM simulator is necessary to create collision-free paths for the challenge.

3 Discussion

The RoBIM simulator educational package represents a significant advancement in the field of construction robot training. The modular design of the package ensures a progressive learning experience, consisting of column building, brick laying, and frame assembling. The step-by-step tutorials and challenge offer a comprehensive understanding of tasks construction robot control and programming. With such an education package and robot simulator, students and workers do not require strong programming skills and expertise to control the construction robot. Future development directions include designing a variety of tutorials and challenge tasks, implementing different industrial robot arms such as ABB and KUKA robot arms, and conducting a user study to verify the effectiveness of the educational package.

4 Conclusion

In conclusion, this paper presents an educational package development for a construction robot simulator RoBIM. The target audience is construction students and workers who do not have strong background knowledge in robots and programming. Three modules in the package cover the basic workflow to design and program the robot motion in the simulator and to control the physical robot to follow the motion and complete the task. By completing the educational package, students will be able to design a wood frame or brick wall and program an industrial robot arm to assemble it.

References

- Liang, C. J., Wang, X., Kamat, V. R., and Menassa, C. C. Human-robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10): 03121006, 2021.
- [2] Yang, C. H., Wu, T. H., Xiao, B., and Kang, S. C. Design of a robotic software package for modular home builder. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 1217-1222, Banff, Canada, 2019.
- [3] RoBIM Technologies Inc. Robotics + Construction. Online: <u>https://www.robimtech.com</u>, Accessed: 26/12/2023.
- [4] Association for Advancing Automation. Simplifying the complex: the role of UI/UX in industrial automation. Online: <u>https://www. automate.org/blogs/simplifying-the-complex-therole-of-ui-ux-in-industrial-automation</u>, Accessed: 26/12/2023.
- [5] Wu, T. H., Wu, F., Liang, C. J., Li, Y. F., Tseng, C. M., and Kang, S. C. A virtual reality tool for training in global engineering collaboration. *Universal Access in the Information Society*, 18: 243-255, 2019.
- [6] Adami, P., Rodrigues, P. B., Woods, P. J., Becerik-Gerber, B., Soibelman, L., Copur-Gencturk, Y. and Lucas, G. Impact of VR-based training on human-robot interaction for remote operating construction robots. *Journal of Computing in Civil Engineering*, 36(3): 04022006, 2022.
- [7] Casañ, G. A., Cervera, E., Moughlbay, A. A., Alemany, J. and Martinet, P. ROS-based online robot programming for remote education and training. In *Proceedings of the International Conference on Robotics and Automation*, pages 6101- 6106, Seattle, WA, USA, 2015.
- [8] Sergeyev, A., Alaraje, N., Parmar, S., Kuhl, S., Druschke, V. and Hooker, J. Promoting industrial robotics education by curriculum, robotic simulation software, and advanced robotic workcell development and implementation. In *Proceedings of the Annual IEEE International Systems Conference*, pages 1-8, Montreal, QC, Canada, 2017.
- [9] Yang, C. H. and Kang, S. C. Collision avoidance method for robotic modular home prefabrication. *Automation in Construction*, 130: 103853, 2021.

Collaborative R&D and Mutual Utilization of Construction Robotics in the Construction RX Consortium

Hiroshi Tabai

Takenaka Corporation tabai.hiroshi@takenaka.co.jp

Abstract

The Construction RX Consortium was established to promote increased productivity and attractiveness of the Japanese construction industry as a whole. The Construction RX Consortium has established 12 subcommittees and stimulates various technology developments for mutual utilization. The automatic material delivery system subcommittee developed an automated transport system. The project is currently in the trial stage.

Keywords -

Construction robot; IoT tools; Collaborative R&D; Mutual utilization; Automation; Material Delivery; BIM

1 Introduction

In the construction industry in Japan, given the social background of the decreasing working population due to the aging of workers, securing future workers and realizing work style reforms have become urgent issues [1]. Under such circumstances, general contractors have promoted the development of construction support and construction management tools utilizing construction robots and IoT. However, it is inefficient for each company to conduct them individually because of the often-generated excessive development costs. Furthermore, the increased number of robot types and IoT tools is a factor that hinders the productivity improvement of cooperating companies who are actually utilizing the technology. The purpose of this research is to introduce overseas the outline of the joint development and mutual utilization efforts of the "Construction RX Consortium (hereafter RX Consortium)" established under this background, and to provide an overview of the development of an automatic material transport system.

2 Research method

This research adopted the case study research method. The target is an automatic material transportation system, which is one of the subcommittees of the Construction RX Consortium. The analysis method is to conduct trials at the five actual sites shown below.

Table 1. Trial list

No.	Building use	Trial year
1	Warehouse	2023
2	Hospital	2022
3	Office	2022
4	Office	2022
5	Office	2022

3 Overview of The Construction RX Consortium

The RX Consortium is a voluntary organization comprising general contractors and various companies supporting the construction industry. It was established with 16 companies in September 2021, and as of December 2023, the number of member companies is 242, of which 29 are regular members [2].

3.1 Aims and Objectives of the Construction RX Consortium

The RX Consortium aims to reduce the cost of technology development of construction robots and IoT tools, shorten the development period, accelerate their popularization by cost range reduction, and promote their introduction by cooperating companies. In addition, the goal is to improve the construction industry's productivity comprehensively and improve the work-life balance and workers' treatment. It is worth noting that the RX Consortium clearly distinguishes the areas of cooperation and competition so as not to hinder healthy competition. In other words, the members collaborate in the development of construction tools that everyone can mutually utilize. Nonetheless, they position which kind of construction can be made by employing them as a competition area.

3.2 The Construction RX Consortium Organization and Activities

The RX Consortium encompasses regular members (general contractors above a determined size with their own R&D organization) and cooperating members. As shown in Figure 1, the organization of the RX Consortium consists of a general meeting, a board of directors, a steering committee to promote and support daily activities, and subcommittees for actual joint development and mutual use. The themes addressed in the subcommittees are discussed internally and decided by the executive committee. In principle, each member may enter the subcommittee if a member wishes to participate based on the member's own free will. However, the subcommittee discusses the division of roles and the burden of costs, establishing a contract to conduct the R&D activities. The main role of the RX Consortium is the establishment of the subcommittee. The concrete activities are in charge of the participating members of the subcommittee. In the subcommittee, besides the members who carry out the technology developments at their expense, the members who contribute without bearing the development costs by feeding back the results of the field trials are also allowed to participate. The mechanism that enables participation in the subcommittee is relatively simple, in line with the purpose of the RX Consortium to contribute widely to the construction industry [3].



Figure 1. Organization

4 Subcommittee Activities

The RX Consortium currently has 12 subcommittees as listed in Table 2. The initiatives of one subcommittee in the table are reported in the next section.

Table 2. Subcommittee list

No	Subcommittee Name	Principal Company	Number of Regular Members	Number of Cooperative Members
1	Automatic material delivery system	Takenaka	18	14
2	tower crane remote control	Takenaka	10	6
3	Workplace Waste Management Technology	Takenaka	12	5
4	Concrete construction efficiency	Shimizu	14	7
5	Sumidashi Robot	Takenaka	16	9
6	Illuminance measuring robot	Kajima	7	17
7	Production BIM	Kalima	22	19
	Equipment Subcommittee	Kajima	13	14
8	interoperable technology	Shimizu	12	4
9	Use of commercially available tools WG1 drone WG2 vital sensor WG3 assist suit	Kajima	17 12 12	16 12
10	Airflow measuring robot	Kaiima	4	13
11	Safety belt non-use detection system using Al	Shimizu	10	10
12	Efficient reinforcement inspection using ICT technology	Obayashi	11	7

5 Subcommittee on Automatic Material Delivery Systems

5.1 Overview of the Subcommittee

In construction work, the delivery of various materials in the required time at the needed place without delay is very important to process control. The delivery of materials within a construction site is an ancillary task for construction workers, and by automating it, they can spend more time working on core tasks that require higher skills. Therefore, the RX Consortium has set up a subcommittee on "Automatic Material Delivery System" and is working on developing such a system.

As of March 2023, 18 construction companies and 14 cooperating companies have participated in the subcommittee. Each general contractor has started activities to realize the practical application of the automatic material delivery system by flexibly linking systems, robots, and delivery equipment developed so far, enabling robots to be used at any construction site of any construction company through repeated development, trial, and feedback among members.

5.2 Overview of the system

Figure 2 shows an overview of the Automatic Material Delivery System, which is positioned as the core system of this subcommittee. This system consists of an "Automatic delivery management system (JHS App)", which handles material delivery reservations and adjustments, delivery instructions, and collection of results, and a "Construction robot platform (RPF)", which receives instructions from the system and generates delivery routes for various robots in conjunction with BIM to perform operation control and condition management.

It works in conjunction with the construction elevator and elevator shutter to realize a series of automatic deliveries by sequentially instructing robot groups and construction elevators based on a planned delivery list. From partial to full automation, various systems and delivery robots can be flexibly combined depending on the purposes and needs. At the construction site, the material unloaded from the truck is horizontally transported to the construction elevator and loaded. Then, the construction elevator vertically transports them to the work floor. On the work floor, another robot unloads the materials from the elevator and distributes them to the workplace.



Figure 2. Automatic Material Delivery System.

5.2.1 Automatic delivery management system (JHS application)

The JHS App inputs delivery reservation information (i.e., material name, packing form, delivery destination floor yard, scheduled start and end dates, delivery means on a delivery route, etc.) for each material to be delivered. As shown in the left screen of Figure 3, the input information is listed and can be viewed for each day/delivery means, and the reservation can be adjusted. The JHS App links the construction autonomous driving elevator and the RPF described later by API. By sending the delivery reservation information from the JHS App, the work order is given to the construction elevator and the delivery robot, and the vehicle is automatically delivered. In addition, as shown in the right screen of



Figure 3. Example of the JHS App screen

Figure 3, the instructions to the worker are also possible through the smart device. When the delivery is completed, the actual information is automatically collected, and the status of the material can be checked on the screen in realtime.

5.2.2 Construction Robot Platform (RPF)

Traditionally, in the operation of construction robots in buildings under construction, it has been a problem that it is necessary to specify the range of operation by teaching in advance in the field, magnetic tape, twodimensional code, etc., requiring many man-hours for maintenance and operation management due to the increased number of robots in operation and the advancement of functions [4]. To solve this problem, we developed RPF. An overview of RPF is shown in Figure 4. As shown in Figure 5, RPF uses BIM data as map information to set the movement and operating range of the robot on the cloud, enabling the robot to run autonomously. In addition, it is also possible to remotely monitor the condition of the construction robot, such as the battery and abnormality, on the cloud. It is also possible to manage a wide variety of robots, such as marking robots, cleaning robots, and delivery robots.



Figure 4. Construction Robot Platform



Figure 5. Robot Right-of-Way in BIM

In the automatic delivery system, the route from the delivery source to the delivery destination is set beforehand, and it is simulated using BIM data for all materials to be delivered automatically. In the automatic delivery execution, the delivery route information is transmitted to the robot based on the conveyance reservation information from the JHS App to instruct execution. The RPF monitors the operation status of the delivery robot in real-time. The delivery performance information, such as the completion of delivery, can be fed back to the JHS App in real-time.

5.2.3 Delivery robot (Automatic hand pallet)

The specifications of the developed automatic hand pallet are shown in Table 1. The appearance is shown in Figure 6. A 2D LiDAR and stereo camera are added to the upper part of the rear housing of the main body. The 2D LiDAR detects objects in the rear range of 180 degrees and serves to grasp their position. The stereo camera is used to detect the position of the pallet. A PC for hand pallet control and a pocket WiFi are stored inside the housing and communicate with the RPF on the cloud.

The automatic hand pallet autonomously drives the route set by matching the self-driving map generated from the BIM with the point cloud data acquired by the 2D LiDAR and performing self-position estimation. When loading the pallet, it moves forward while grasping the shape of the pallet using a stereo camera, and drawl into the lower part of the pallet. When there is an obstacle that was not in the map acquired in advance, the 2D LiDAR detects it, detours around it, and runs autonomously to the destination [5].

Table 3.	Specification	of Automatic	hand palettes
----------	---------------	--------------	---------------

Dimensions	1,420mm×750mm×450mm
Self-weight	170kg
Carrying capacity	1,000kg
Running speed	Min42m/min, Max60m/min
Operating time	6h (battery)
Filling lifting dimensions	105-175mm



Figure 6. Automatic hand pallet

5.3 Trial at construction site

The developed system was tested at several construction sites. We confirmed that it is possible to automatically deliver materials to designated locations based on delivery instructions from the JHS App. In the future, we will conduct repeated trials at the general contractors participating in the subcommittee and assess the effects. We will also develop an operational system and aim to put it into practical use.



Figure 7. Trial at construction site

6 Conclusion

In addition to the subcommittee activities, the RX Consortium carries out various initiatives to promote information sharing and mutual technology utilization among members, such as holding exhibitions and introducing technologies and services by member companies. The number of members has also increased faster than initially expected, and a wide range of industries, such as insurance companies, trading companies, and consulting companies, are participating. The environment for incorporating technologies and services useful for the construction industry is becoming better. Through these activities, we hope to improve the productivity and attractiveness of the construction industry.

References

- White Paper on Land, Infrastructure, Transport and Tourism in Japan, 2023: <u>https://www.mlit.go.jp/statistics/file000004/pdfind</u> ex.html Accessed: 20/02/2024
- [2] Construction RX Consortium: <u>https://rxconsocom.dw365-ssl.jp/index.html</u>, Accessed: 18/12/2023.
- [3] Masahiro Sugata et al. Improvement the Attractiveness of the Construction Industry through Technical Alliances Such as Construction Robots. *Concrete Journal*, Vol. 60-5, p. 389, 2022.5
- [4] Eri Ohmoto et al, Development and application of autonomous route detection for the low-floor type AGV -Part 1. Summaries of Technical Papers of Annual Meeting Architectural Institute of Japan, p.1245-1246, 2019.9
- [5] Hiroshi Tabai et al, Collaborative R&D Mutual Utilization of Construction Robotics in RX Consortium Part4 Development of an Automation Material Delivery System. Summaries of Technical Papers of Annual Meeting Architectural Institute of Japan, p.1193, 2023.9

UAV for Sticking Markers in the Built Environement

Wen-Shuo Hsu¹, Soroush Fazeli¹, Kepa Iturralde^{1,2}, Christoph Holst¹

¹Chair of Engineering Geodesy, School of Engineering and Design Technical University of Munich, Germany ²Chair of Digital Transformation in Construction, Institute of Construction Management, Faculty of Civil and Environmental Engineering, University of Stuttgart, 70569 Stuttgart, Germany.

wenshuo.hsu@tum.de, soroush.fazeli@tum.de kepa.iturralde@br2.ar.tum.de christoph.holst@tum.de

Abstract -

UAVs have immense potential in construction applications. This paper gives a overview of a project that aims to expand the frontier of applications of UAVs from non-contact measurements to physically interact with buildings by sticking targets on the building. The goal of the project is to apply positioning tags onto a facade of an existing building with UAV. This paper looks into hardware, software and simulation aspects of the UAV and elaborates on advantages of the chosen solutions. Moreover, the UAV has been tested in laboratory environment.

Keywords -

UAV; Automation; ROS; Simulation; built environment

1 Introduction

Unmanned Aerial Vehicles (UAVs) are becoming popular as a solution to automation of a wide variety of tasks in different fields. For construction application, [1] utilized UAV to facilitate contact test for bridge inspection with a 1-DOF manipulator. [2] applied UAV for sensor installation and retrievement. [3] gives a thorough review on UAV with active multiple DOF manipulator and its development throughout the years.

The goal of this project is to apply AprilTags, a type of visual tag that provides localization with high accuracy and low overhead, onto facades of buildings with an automated UAV. These accurate positioning tags can later be used as reference for installation of insulated prefabricated modules with solar energy systems. UAVs are lightweight and low-cost, therefore it can be deployed fast and with multiple units simultaneously.

A quadcopter with an end effector was developed. The quadcopter will approach the target position with AprilT-ags loaded on the end effector. On the back side of the tags is adhesive applied. Once the tags have made contact with the target surface, the quadcopter will apply pressure on the tags to ensure a firm binding. The process is indicated in Fig. 1.

Compared to [1] and [2], our end effector does not require its own DOF. AprilTags are released mechanically. The contact position, force, angle is manipulated through



Figure 1. Sticking Process, check the video

controlling the vehicle directly. This reduced complexity and lighten its weight.

2 Design

As an experimental platform, the controllability, payload capacity and adjustability of dimension are the three priorities. Adjustability enables the UAV to operate with a wide variety of propellers for different payloads. X-shape quadcopter is chosen because of its symmetric maneuverability and controllability on X-Y plane.

Although multicopters with more motors have higher payload capacity than a quadcopter, the end effector would not have enough space between evenly spread motors. Stacking two motors vertically was also examined. However, in light of the efficiency loss of the second propeller [4], the additional thrust gain cannot compensate for the excessive complexity.

The size of the quadcopter, the diagonal shaft-to shaft distance between motors, is designed to be adjustable from 238 to 650 mm to be compatible with propellers up to 15 inches. This assures the quadcopter to be compatible with a heavy end effector. Compared to [5], the battery is moved from bottom to top. This brings 2 advantages. First, the battery is located closer to the Center of Mass (COM). As one of the heaviest parts on an UAV, moving the battery closer to the COM results in a smaller moment of inertia, which enhances the quadcopter's agility. Second, the battery on the top is more accessible for operators. Operators can install a battery much easier. It is also more

feasible to develop an automated battery exchange process with a battery on the top in the future.

The flight controller is located between upper and lower plate of the quadcopter. This change minimized the distance from COM to onboard Inertial Measurement Unit (IMU) to 23 mm, which can minimize the extra centripetal acceleration an IMU experienced when turning around COM and hence increase its accuracy.

3 Hardware

Hardware choices were made with max compatibility in mind. Within the budget limits, the power system is designed for a drone weighing 2.5kg.

Table 1. Hardware Overview		
Motor	T-Motor F100	
	2820 1100KV	
Propeller	HQ8037-3	
Battery	6s 2200mAh	
Flight controller	Raspberry pi 4B	
e	w/ Emild Navio 2	
Electronic speed controller	T-Motor F55A PRO II	
(ÉSC)		
RC receiver	FrSky X8R	
Laser range finder	GÝ-53	
IMŬ	LSM9DS1	



Figure 2. Hardware Position

Although propellers with higher pitch can generate more thrust, these propellers are also prone to being caught into Vortex Ring State (VRS), which results in lose of lift and control [6]. The adopted 8037 propellers have smaller pitch than the manufacturer's recommendation.

Although high kV motors on the old model [5] can generate more thrust with the same size, more current is drawn from batteries and more waste heat is generated. This results in less overall efficiency, which leads to a shorter flight time. By contrast, a low kV motor with high voltage has less side effects while having same output. Therefore, compared to previously 2550kV, the new motors' kV value is selected as 1100kV.

Low kV motors require higher voltage to operate. Thus, a 6-cell battery (rated at 22.2V) is used.

4 Material and Manufacture Process

For manufacturing, CNC machined Carbon Fiber Reinforced Polymers (CFRP) and Fused Deposition Modeling (FDM) with Polylactic Acid Plastic (PLA) are widely used.

CFRP is widely used in the aerospace industry because of its strength and stiffness with low density. In the case of this project, CFRP is used in critical parts such as arms which are cantilever beams with motors on the open end. The stiffness of beams can suppress vibration induced by motors and propellers. The main piece of the frame is also consists of CFRP, where this stiff material acts as a high pass filter against vibration, which is crucial to the precision of the vibration-sensitive onboard IMU.

FDM can shorten the time of manufacturing. Also, it can create sophisticated infill pattern inside work piece, which can be optimized and achieve high specific strength. Parts such as landing gears and brackets for sensors and controllers, which do not experience high shear stress, are made out of PLA with FDM.

5 End Effector

AprilTags is fixed in a square frame. After the end effector makes contact with the surface, the square frame will be pushed back by the surface and releases the AprilTag. At the end of the sticking process, the quadcopter is given the command to move forward to apply pressure on the AprilTag for a firm binding. In order to move forward, a quadcopter pitches downward. To compensate the pitch angle, a ball joint is added as shown in Fig. 3.



Figure 3. Cutaway View of End Effector

The ball joint can also absorb some misalignment of the quadcopter w.r.t. the target surface. To allow soft contact with surfaces, the end effector can slide along the rods stemming from the main frame. Springs and dampers between two parts can reduce the reaction force from the surface pushing the quadcopter away.

6 Simulation

To affix the tag to the wall, the drone must make contact, subjecting it to external forces that can introduce disturbances and nonlinear dynamics. Addressing this challenge involves measuring and integrating these forces into the system. This integration enables the controller to compensate for the forces, ensuring stability during flight and minimizing the impact of disturbances. One method to measure these forces is by incorporating force sensors on the drone. These sensors can gauge the forces during interaction, providing feedback to the controller. By assessing changes in the drone's attitude, the controller can adjust propeller speeds to maintain flight stability and ensure sufficient force is applied to affix the AprilTags to the wall. However, it's worth noting that this approach may be costly due to the price of force sensors and the additional weight they introduce, potentially impacting the drone's motors and battery requirements. Due to this drawback of using sensors, these forces can be measured in simulation. Hence the task of the drone can be tested in the simulation to assure a stable flight during the interaction and evaluate the performance of the controller. With the simulation, the drone would have certainly lower cost and probable crashes could be prevented. However, in the robotic industry majority of the simulations are implemented in well-known simulations like Gazebo or other ones which will not provide force measurements. Unity [7], a popular game engine, facilitates the simulation of drones by employing a physics engine to model dynamic behavior and measure applied forces. Objects, including the asymmetric drone, are treated as rigid bodies within Unity. To ensure accurate simulation, Solidworks models of the drone are utilized. Unity's robust rendering capabilities extend to testing computer vision algorithms, such as detecting a drone equipped with an AprilTag. The high rendering ability of Unity is crucial for accurate position estimation by the ground camera in detecting the drone and other tags on the building.



Figure 4. Simulation Diagram

On the other side what makes unity a perfect tool for us is that it can be easily integrated with ROS. Robot Operating System (ROS) [8] is an open-source robotics middle ware suite. ROS optimizes communication in robotics through hardware abstraction, low-level device control, common functionality implementation, and inter-process message passing for diverse computer clusters. Unity's simulation can seamlessly integrate with the controller and other components of the software stack via ROS, enabling effective communication between them. In the ROS network illustrated in Fig. 4, the ground camera within Unity detects tags and transmits desired waypoints. This information guides the trajectory generator to produce and publish desired states for the controller.

6.1 Controller Design

The controller node is a key part of the software stack, utilizing a geometric controller inspired by [9]. It processes desired states, computes errors in the drone's state space, and works to mitigate these errors for enhanced control. The control of the transnational dynamics of a UAV involves the management of the total thrust, denoted as $-fR_{e3}$. The magnitude of the total thrust, represented by f, is under direct control, while the direction of the total thrust, denoted as $-R_{e3}$, aligns along the third body-fixed axis b_3 .



Figure 5. Controller Structure

The UAV's interaction with the wall provides feedback for the controller to make informed decisions on velocity, ensuring stability and precise tag placement. The contact force applied on the drone will be in two direction, The contact force applied on the drone will be in two directions, one is tangential, and the other one is perpendicular to the moving direction. The drone needs to adjust its velocity and orientation in order to assure an approach normal to the wall. Thus the forces applied on the drone from the interaction would be minimized.

7 Results

The aerial system, controlled via a remote controller, combines autonomous algorithms to achieve precise and

dynamic maneuvers. The inclusion of sensors such as an Inertial Measurement Unit (IMU) ensures stable flight by providing real-time data on the drone's orientation. The integration of a laser range finder enhances the manipulator's spatial awareness, allowing it to accurately gauge distances from the walls. The drone was initially tested in a simulation environment to evaluate the controller's performance and accuracy. Subsequent testing on the physical drone assessed the controller's performance on hardware. The following section delves into the obtained results and overall performance.

7.1 Simulation

In simulation tests, the drone controller's performance was assessed based on its stability during wall interactions and accuracy in placing the target. Key metrics included contact force, drone velocity, stability, and target placement, measured by considering the collision area and the number of contact points. Results are summarized in the table below:

Table 2. Simulation Results

Case	Contact Force	Velocity	Accuracy
	(N)	(m/s)	(m)
1	2.9	1	0.01
2	7.7	2.4	0.05
3	12	4.8	0.11

7.2 Experimental Setup

In the experiments, the Ardupilot flight controller and geometric controller played key roles in drone operations based on RC commands and real-time sensor data. The Ardupilot ensured the drone's stable flight, while the geometric controller significantly improved maneuvering precision. To address differences between simulation and real-world experiments, the Extended Kalman Filter (EKF) was employed for state estimation, effectively managing uncertainties, especially in state determination using data from the IMU and laser range finder. Despite arguments surrounding certain factors, the comparison struggled to precisely evaluate the controllers' adaptability in real-world scenarios. However, the drone successfully completed its tasks, showcasing practical functionality despite the inherent difficulties in accurately comparing simulations with real experiments.

8 Conclusion

In conclusion, the project successfully demonstrated the quadrotor's ability to interact with and attach a target to a wall in both simulation and real-world tests, showcasing practical applications like surveillance and object manipulation. The integration of robust control mechanisms showed promise, and insights gained from improved simulation contact force modeling suggest potential for refining the system. Future developments could benefit from a better controller design, informed by insights gained through enhanced simulation, to further optimize the performance and adaptability of versatile quadrotor systems.

Acknowledgments



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No.958445.

References

- I. Takahiro et al. Wall contact by octo-rotor UAV with one dof manipulator for bridge inspection. In 2017 IEEE/RSJ IROS, pages 5122–5127, 2017. doi:10.1109/iros.2017.8206398.
- [2] H. Salua et al. Sensor installation and retrieval operations using an unmanned aerial manipulator. *IEEE RA-L*, 4:2793–2800, 2019. doi:10.1109/lra.2019.2918448.
- [3] A. Ollero et al. Past, present, and future of aerial robotic manipulators. *IEEE T-RO*, 38:626–645, 2021. doi:10.1109/tro.2021.3084395.
- [4] M. Ramasamy. Hover performance measurements toward understanding aerodynamic interference in coaxial, tandem, and tilt rotors. *Journal of the AHS*, 60(3): 1–17, 2015. doi:10.4050/jahs.60.032005.
- [5] K. Iturralde et al. UAV for target placement in buildings for retrofitting purposes. In *Proceedings of IS-ARC*, volume 40, pages 593–596. IAARC Publications, 2023. doi:10.22260/isarc2023/0080.
- [6] W. Johnson. Model for vortex ring state influence on rotorcraft flight dynamics. Technical report, 2005.
- [7] A. Juliani et al. Unity: A general platform for intelligent agents. ArXiv, abs/1809.02627, 2018.
- [8] M. Quigley et al. ROS: An open-source robot operating system. In Workshops at the IEEE ICRA, 2009.
- [9] T. Lee et al. Geometric tracking control of a quadrotor UAV on se(3). In 49th IEEE CDC, pages 5420–5425, 2010. doi:10.1109/cdc.2010.5717652.

Development of Mobile Inspection Robot for Concrete Wall Surface and Strength Estimation Considering Coarse Aggregate by Small Diameter Drill

Momoe Terata¹, Naoki Senno¹, Fumihiro Inoue², Satoshi Sato³ and Hiroshi Watase⁴

¹Graduate School of Shonan Institute of Technology, Japan ²Shonan Institute of Technology, Japan ³Total Service Planning, Japan ⁴Oriental-Shiraishi Corporation, Japan <u>23T1502@sit.shonan-it.ac.jp</u>, inoue@mech.shonan-it.ac.jp

Abstract

In order to easily estimate the strength of concrete members such as piers and building walls at actual sites, we have developed a small-diameter abrasive drill and proposed a method to estimate the strength of concrete from changes in drilling speed. In drilling with a 5 mm diameter drill, the drilling speed changes due to the influence of coarse aggregate mixed in the concrete. Therefore, the drilling speed due to coarse aggregate was determined from the analysis of drilling speed, and the relationship between the average drilling speed excluding the effect of coarse aggregate and concrete strength was obtained. Next, a mobile inspection robot equipped with an abrasive drill mechanism was developed to apply this inspection method to actual sites.

Keywords

Small-diameter drill; Concrete strength estimation; Inspection robot; Roller drive; Duct fan thrust

1 Introduction

A large number of civil engineering infrastructures (bridges, tunnels, highways, etc.) constructed mainly in urban areas during the high-growth era of the 1970s have begun to reach the end of their useful lives and need large-scale repair, renewal, or reconstruction. Concrete structures, in particular, have deteriorated over time, and in some cases, their strength has been significantly reduced due to salt damage and neutralization, forcing immediate action.

In recent years, with the development of robotics and information technology, there have been reports of automatic inspections of the deteriorated condition of concrete structures using wall-moving robots and drones. The main tasks are imaging of the wall surface and simple sounding inspection [1-3]. If the reaction force of the robot is large enough, it is possible to estimate the strength of the concrete by mounting the grinding drill developed above. On the other hand, if the reaction force of the inspection drill can be kept small, it can be mounted on the current inspection system [4].

In this study, the previously developed abrasive drill was modified to estimate concrete strength by drilling a small-diameter 5 mm hole. The small diameter of the drill also reduces the reaction force due to drilling. In the experiment, strength estimation by drilling was conducted using specimens of different strength. However, when the drill bore size was smaller, the coarse aggregate inside the concrete was drilled directly, and the results were different from the original strength estimation by drilling (relationship between average drilling speed and concrete strength). Therefore, we focused on the local perforation rate of drilling and devised a strength estimation method that removes the influence of coarse aggregate as much as possible.

On the other hand, it is necessary to check the deterioration and strength of concrete structures at the actual site to ensure the integrity of the structure and efforts to repair and maintain the structure. In this study, an inspection robot equipped with a drill system for strength estimation and capable of inspecting at actual sites was developed. The inspection robot is equipped with a high-power duct fan and can easily press the drill. The robot is roller-driven, can move up and down, and is lightweight, thus gaining potential for practical use.

2 Development of Small Diameter Abrasive Drills and Drilling Tests

2.1 Outline of Experimental Equipment

Fig.1 shows an overview of the abrasive drill used in the experiment. The system consists of (1) a smalldiameter drill, (2) a slide block (feeder), (3) a water circulation device, (4) a vacuum pump, and (5) an air compressor. The small-diameter abrasive drill is an abrasive drill with a diamond bit attached to its tip, which rotates at high speed to grind the mortar surface while drilling. During drilling, a water circulation device is used to cool and lubricate the bit and discharge the drilled noro water. The body of the polishing drill can be moved up and down by air pressure from an air compressor through a pressure reducing valve, which provides a (constant) pressing force for the drill during drilling. The tip of the drill consists of a diamond bit, and the drilling method is to drill while polishing the concrete. Experiments using an abrasive bit with a relatively large bit diameter have shown that the drilling depth is almost proportional to the drilling time, and the drilling speed is almost constant under the same conditions such as drill pushing force and rotation speed. In the experiments, five kinds of concrete specimens of different strength were prepared in order to understand the performance of the developed drilling system.

2.2 Changes in Drilling Gradient and Internal Observation of Drilling

When drilling with a small-diameter (5 mm) drill, the drilling speed is not always constant because the drill may directly drill through the coarse aggregate scattered inside the concrete. Figure 3 shows the relationship of drilling depth to drilling time when drilling a concrete specimen. In addition, images of the perforation taken continuously by an endoscope are also shown.

Observation of the state of the drilling depth shows that the area near the starting point of drilling (above: Fig.2) is white due to the finishing layer of concrete. From the start of drilling to around 60 seconds, the change is relatively linear, and the layer is a mixture of white (Fig.2) and black (coarse aggregate) (below: Photo 2). Around Fig.2, the proportion of mortar increases and the gradient of perforation is slightly higher.

In the range of 130 to 190 seconds of drilling time, the drilling depth is almost constant and the black coarse aggregate is drilled (Fig.2 (4)). It takes a considerable amount of time to drill through the coarse aggregate. After passing through the coarse aggregate, a layer of coarse aggregate and mortar is formed again, and the perforation gradient increases (Fig.2 ⑤) and then reaches the longest part of the perforation (Fig.2 ⑥). Thus, it can be seen that the 5 mm hole drilled through a mixed layer of coarse aggregate and mortar, and that the drilling gradient (drilling speed) changes constantly.

2.3 Coarse Aggregate Condition and Change Distribution of Drilling Speed

Since the drilling gradient varies with the internal condition of the concrete, the drilling speed per second



Fig.1 Overview of the drilling device and measurement system used in the concrete



Fig.2 Relationship of depth and drilling speed to time for concrete specimens containing coarse aggregate



Fig.3 Time distribution for each drilling speed range

(mm/s) was calculated. Figure-3 (bottom) shows the relationship of drilling speed for each drilling time. It can be clearly seen that the drilling speed changes in response to the gradient of the drilling depth. It can be seen that the

drilling speed is smaller than 1.0 mm/s in the area affected by the coarse aggregate. Figure 4 shows the distribution of the number of hours for each range of drilling speed. It can be seen that the number of hours is concentrated in the range where the drilling speed is smaller than 1.0 mm/s. This area is affected by the coarse aggregate. From this result, the average drilling speed $v_{mean,i}$ excluding the effect of coarse aggregate is obtained by excluding the area of small drilling speed and dividing the remaining total number of hours by the total drilling depth (the drilling depth is subtracted because drilling progresses according to the number of hours even if the drilling speed is small), using the Eq. (1). where L_{Total} is the total drilling depth (155 mm in this experiment), T_{Total} is the total drilling time, *i* is the number of the speed range in which the effect of the stone layer is removed, $v_{(c,i)}$ is the lower limit of the i_{-th} speed range, and t_i is the number of hours for the *i*-th speed range.

$$v_{mean,i} = \frac{\frac{L_{Total} - \sum_{k=1}^{i} \{(v_{c,k} + 0.125) \times t_k\}}{T_{Total} - \sum_{k=1}^{i} t_k}$$
(1)

2.4 Strength Estimation Considering Effect of Coarse Aggregate

Drilling tests were conducted on five types of concrete specimens with high coarse aggregate content. Figure-6 shows the relationship between average drilling speed and compressive strength. In the range where the drilling speed v is small ($v_{(c,i)} \sim 0.5$), the plots are scattered in different places, and it is difficult to see a relationship between the average drilling speed and the specimen strength. However, for $v_{(c,i)} = 0.75$ to 1.0, each blot is aggregated and the average drilling speed and specimen strength line up linearly (red dashed lines in the figure). Using this relationship, it may be possible to estimate the concrete strength from the average speed of drilling. On the other hand, for $v_{(c,i)} = 1.25$, the values of the plots are again scattered and the relationship becomes less relevant. This may be due to the fact that even the drilling speed for normal drilling (when there is no influence of coarse aggregate) is excluded.

Thus, when drilling with a 5mm abrasive drill, if $v_c = 0.75$, the concrete strength σ can be estimated from the slope of the straight line obtained in Fig 6 using Eq. (2).

$$\sigma = 83.9 - 59.3 * (v_{mean,i} - 1.5)$$
(2)

However, Equation (2) is a relational equation for the specimens used in this study, and may not necessarily be applicable to specimens with different coarse aggregate and other materials. By introducing the relationship of average drilling speed without the effect of coarse aggregate, it was found that concrete strength can be adequately estimated even in drilling tests using small-diameter drills. The results of this study are as follows.



Fig.4 Relationship between average drilling rate and compressive strength considering range of drilling rates

2.5 **Outline of Inspection Robot**

An inspection robot equipped with a drill system for strength estimation was developed for on-site inspection. As shown in Fig. 5, the inspection robot is equipped with an automatic vertical moving mechanism and various sensors to enable various inspections of vertical wall surfaces.

2.5.1 Vertical Moving Mechanism

When robots are used for inspection and repair of wall surfaces, the types of devices used for vertical movement are generally of the drone flying type or the rope winding type with the device suspended from above. However, conventional methods have disadvantages such as being affected by wind, difficulty in approaching walls, and limited lifting weight of the inspection device. In this study, the tether and roller drive principle were used for vertical movement, which is a simple mechanism, has good positional accuracy, and has excellent stopping performance. During inspection, the device can be fixed to the wall surface via casters, and lanes to be inspected are set up and inspected in sequence.

The climber can be raised and lowered under autonomous control, and by turning on the start switch, it can stop at each section according to a pre-set program. The system can be switched between wired and remote control, and can be checked visually or with a live camera
mounted on the inspection equipment. After completing the inspection to the top under autonomous control, it can be lowered by sensor detection. The power supply is powered by a battery, but for longer inspections, a wired connection is used to supply power. Safety ropes and mechanical brakes are installed to prevent falling.

2.5.2 Horizontal Push against Wall

Multiple casters are mounted on the wall side of the inspection system to push to maintain the verticality of the small-diameter drill and the wall surface. A highpower duct fan is located in the same line across the drill to press the main unit and the drill against the wall. As shown in Fig.6, when measuring the strength of the wall surface, the small-diameter drill can be moved via a slider, enabling drilling with a constant force of pressing. The maximum force to press the entire inspection system by two duct fans against the wall surface is about 30 N, which is sufficient for the pressing force (15 N) for the strength estimation experiment described in Chapter 2.

2.6 Operation of the Small Diameter Drill for Wall Surface Inspection

The drill is moved to a predetermined height using a climber and pressed against the wall surface using a ducted fan. After setting it in the position to be inspected, the small-diameter drill for wall surface inspection is operated. The drilling procedure is as follows (1) to (3). (1) The compressor is used to bring the tip of the drill into contact with the wall surface.

(2) A water circulation system is activated to eliminate concrete dust generated during drilling.

(3) Operate the motor connected to the drill to drill.

After $5 \sim 10$ cm of drilling is completed, the motor is stopped and the drill body is pulled up by a compressor. This series of operations is performed by remote control. The mobile inspection robot and strength estimation system have already been completed and will be tested at actual sites for practical use.

3 Conclusion

In order to simply estimate the strength of concrete members for vertical walls such as bridge piers and buildings at actual sites, we developed a small-diameter abrasive drill and proposed a method to estimate the strength of concrete from the change in its drilling speed. In drilling with a 5 mm-diameter drill, the drilling speed changes due to the influence of coarse aggregate mixed in the concrete. Therefore, the drilling speed due to coarse aggregate was found from the analysis of the drilling speed, and the relationship between the average drilling speed excluding the effect of coarse aggregate and concrete strength was found.



Fig.5 Overview of the inspection robot equipped with the drill system for strength estimation



Fig.6 Operation of a small-diameter drill for wall surface inspection mounted on an inspection robot

Next, a mobile inspection robot equipped with a drill mechanism was developed to simplify this inspection method. In the future, we would like to use these in actual sites to add the necessary systems and the possibility of calibrating them for each site.

References

- F. Inoue, S. Doi, T. Okada and Y. Ohta, "Development of Automated Inspection Robot and Diagnosis Method for Tile Wall Separation by Wavelet Analysis", Proc. of 26th on ISARC 2009, Austin, USA, pp.379-388, (2009).
- [2] Tokura, Y., Toba, K. and Takada, Y., Practical Applications of HORNET to Inspect Walls of Structures, Journal of Robotics and Mechatronics, Vol.28, No.3 (2016), pp.320-327.
- [3] F. Inoue, T Makino, "Inspection Robot System Using Ducted Fan and Deterioration Estimation of Building Wall that Can Be Applied Even in Disaster", Proc. of 18th on ICCAS 2018, Korea, pp. 331-334, (2018).
- [4] M. Terata, F. Inoue, "Development and Performance Experiment for Wall Moving Inspection Robot Using Ducted Fan Thrust", Proc. of the SICE 2022, Sep.t 6-9, 2022, Kumamoto, Japan, FrA04.5, PP.1-3.

Analysis of openBIM Adoptions and Implementations: Global Perspectives and Canadian Industry Adoption

Farzad Jalaei¹, Saman Davari², Yasir Sultan³, Claudia Cozzitorto⁴, Erik Poirier⁵, Arash Shahi⁶

¹National Research Council Canada, Construction Research Center, Ottawa, ON, Canada <u>Farzad.jalaei@nrc-cnrc.gc.ca</u>, <u>Saman.Davari@etsmtl.ca</u>, <u>Yasir.Sultan@nrc-cnrc.gc.ca</u>, <u>ccozzitorto@buildingSMARTcanada.ca</u>, <u>Erik.Poirier@etsmtl.ca</u>, <u>arash@aecoinnovationlab.com</u> **Abstract** Format (BCF) [3], building Smart Data

The transition towards openBIM solutions has been receiving an increasing amount of attention in the built asset industry. This shift is principally driven by the adoption of open standards and services, aiming to promote a more collaborative, interoperable, and transparent digital workflow. This paper highlights openBIM adoption across various jurisdictions, with focus on government a involvement, standardization, technological innovation, environmental integration, national implementation, and knowledge resources. Drawing from international best practices, the paper presents a comprehensive analysis of openBIM opportunities and interoperability challenges, offering kev takeaways and guidance for other nations such as Canadian construction industry to adopt and implement openBIM solutions efficiently.

Keywords -

openBIM; Built asset industry; Building Information Modeling; digitalization

1 Introduction

The past decade has seen a notable shift in the built asset industry regarding how projects are planned, designed, constructed, and operated. The traditional siloed approach to project delivery has given way to a more collaborative and integrated methodology, and at the forefront of this change is the concept of openBIM. According to buildingSMART International (bSI) [1], openBIM refers to: a universal approach to the collaborative design, realization and operation of buildings based on open standards and workflows. openBIM is an initiative of buildingSMART and several leading software vendors using the open buildingSMART Data Model. With the collaboration of international industry experts and stakeholders, bSI has developed specific standards and services to accelerate the implementation and adoption of openBIM. These open international standards and services - such as Industry Foundation Class (IFC) [2], BIM Collaboration

Format (BCF) [3], building Smart Data Dictionaries (bSDD) [4], etc. – are based on various types of Building Information Modeling (BIM) processes and information capabilities.

By utilizing open standards and services, openBIM enables the exchange of information across different software platforms and disciplines, facilitating a more efficient and transparent digital workflow [5]. At present, several leading jurisdictions have leveraged openBIM requirements into their building regulations and procurement processes, mandating its use in public projects. Professional organizations and industry alliances are actively promoting openBIM, developing guidelines, providing training, and advocating for its widespread adoption.

While the state of openBIM in the world today is promising, challenges still exist. Standardization efforts need to be strengthened further to ensure interoperable data exchange between different software platforms [6]. Education and training initiatives must be expanded to enhance the digital skills of professionals across the construction industry. Additionally, cultural barriers and resistance to change present ongoing obstacles that require continuous awareness campaigns and stakeholder engagement.

This paper aims to highlight the current state of openBIM at international level, with the primary goal of identifying and assessing the key trends, challenges, and potentials of openBIM within built asset industry. To do so, several leading jurisdictions, including the United Kingdom (UK), the United States (US), Norway, Finland, Denmark, France, Australia, Singapore, and Japan, were selected and analyzed. The selection of these primary countries was based on a diverse range of geographical locations, cultural contexts, and their contributions to advancing BIM workflow and processes. The key findings of reviewed international jurisdictions are outlined in the following sections.

2 Dimensions for Analysis

The adoption of openBIM across various jurisdictions is not a uniform process, owing to diverse regulatory

environments, technological capabilities, and strategic priorities. To critically evaluate the international state of openBIM, it is imperative to consider a multidimensional framework that reflects the complex interplay of factors influencing its adoption and implementation. To identify potential dimensions for this evaluation, the approach involved the deployment of surveys and direct engagements with international bSC chapters. These engagements were aimed at understanding the specific conditions and practices of openBIM implementation in their respective countries. The feedback received revealed commonalities in challenges and strategies across the board, enabling the research team to conduct a thorough analysis. The six dimensions identified for this analysis are as follows:

- Governmental role: the extent of governmental engagement in openBIM initiatives, examining how public policy, funding, and legislative frameworks shape the adoption landscape.
- Standardization: The adoption of universal standards is the cornerstone of openBIM's ethos. By evaluating the degree of standardization, this dimension pinpoints the efforts towards harmonizing procedures, data formats, and workflows, and the impact of these standards on cross-border and cross-industry collaborations.
- Technological innovation: the current state and evolution of technological advancements, spotlighting pioneering tools and platforms that enhance interoperability and facilitate openBIM implementation.
- Environmental integration: the integration of environmental considerations into openBIM reflects a commitment to sustainability. This analysis dimension gauges how openBIM interfaces with green building standards and eco-friendly practices, reinforcing its role in driving sustainable development within the built asset industry.
- Implementations at national level: the practical application of openBIM within national projects provides tangible evidence of its value and adaptability. This dimension scrutinizes the scope and scale at which openBIM is embedded into the public infrastructure, serving as a measure of its penetration and effectiveness in delivering complex projects.
- Knowledge resources: the generation and dissemination of openBIM knowledge through educational resources and training platforms are pivotal for its adoption. This

dimension evaluates the avenues through which expertise is improved and shared, reflecting the industry's capacity to nurture a proficient workforce well-versed in openBIM principles.

Figure 1 summarizes the analytical findings pertinent to the identified dimensions, emphasizing the most recent advancements in the adoption and implementation of openBIM across prominent jurisdictions. This depiction is current as of the date of this publication, acknowledging that these details may evolve over time.

3 openBIM implications and opportunities

Review of the selected jurisdictions through identified dimensions demonstrated potential implications and opportunities for other countries worldwide. The following sub-sections describe such findings.

3.1 Advocacy for open standards

Standardization serves as the foundation for the efficient and interoperable practice of digital construction [7]. The reviewed jurisdictions are currently advocating openBIM standardization, recognizing for the importance of standardized procedures in harnessing the true collaborative potential of the construction industry. In the UK, NIMA has been a front-runner in its support for openBIM through initiatives like the UK BIM Framework. Statsbygg, as a governmental body in Norway, has actively promoted openBIM, ensuring that projects adhere to these standards from the onset, ensuring quality and cost-efficiency. Singapore, a bustling hub of modern construction, is yet another shining example. Building Construction Authority (BCA) in Singapore has published different versions of BIM Guides, which provide detailed insights into the adoption of open standards and services in the city-state.

Moreover, advocating the standards is just one aspect; the true essence lies in ensuring their implementation and providing the community with access to guidelines and case studies, fostering an ecosystem where openBIM becomes ingrained in the industry's DNA. By prioritizing the generation and dissemination of openBIM knowledge through educational resources and training platforms, jurisdictions can effectively equip their workforce with the necessary skills and understanding to implement and adhere to these standards, thereby amplifying the impact of standardization efforts and fostering a more collaborative and efficient construction ecosystem.

	Governmental Role	Standardization	Technological Innovation	Environmental Integration	Implementation at National Level	Knowledge Sources
×	 Regulatory (very low) Mandate renews Leadership as client & owner 	 ISO19650 series bSI standards 	 Information Management Platform (IMP) GIIG platform 	 GenZero for design of zero-carbon schools Guides and policies of UK Green Building Council (UKGBC) 	 GIIG projects GenZero project Sellafield (CDE) Babcock Major Infrastructure Projects (MIP) 	 NIMA guides and services UK BIM Framework
	 Mandates through federal agencies (GSA). Moderate support for openBIM in infrastructures 	 ISO19650 forwards & annex US-IDS CSI IBD (level of accuracy) bSI standards NBIMS-US 	 Central facility repository (GOTS) platform 	NIBS-BETEC programs	 TPF-5(372) BIM roadmap Implementation of MALD for DoT in various US states 	 bSI-US openBIM trainings GSA's BIM hub NIBS guides
	 Strong support for standards, mandates, regulatory framework 	 NS 8360-1 NS 8360-2 NS 8360-3 NS 8360-4 NS 8360-5 bS1 standards NS/EN ISO19650 	 Simplebim and Pillr platform BIMQ platform 	 Handbook V770 & R760-model basis, to meet UN sustainability goals. 	 Bane NOR: Project KIM National Digital building permit Statens vegvesen infra projects 	 Statsbygg's online database Resource Bank (by bSN)
+-	 Strong support for standards, mandates, roadmaps 	 bSI standards ISO19650 series ISO 191XX (WMS, WFS, GML) ISO 16739 ISO 12006 	 IFCtoolbox (by Eurostep) COVE (by YIT) Model Checkers (e.g.Solibri) 	 RYTV program for sustainable use of resources 	 RASTI projects RYTV projects Open InfraBIM Vera projects 	bSFinland's WIKI
C	 Strong support for standards, mandates, roadmaps 	 bSI standards ISO19650 series IFC-SG 	• Corenet X	 BCA-SkyLab sustainability program Singapore Green Building Masterplan (SGBMP) Super Low Energy Programme (SLE) 	 BCA BIM implementation roadmap CORENET X for regulatory checks 	 BCA community for education & knowledge sharing
	 Moderate support for certifications Sustainability regulations 	 bSI standards ISO19650 series 	Semantics4BIM	 Manifesto for openBIM at the service of ecological transition CSTB 2030 projects 	 Projects of Digital Transition Plan (PTNB) BIM 2022 Plan 	 bSFrance online database for guides and policies
:=	 Moderate support for certifications Sustainability regulations 	 bSI standards DS ISO19650 series DS/EN ISO 23386 DS/EN ISO 23387 DS/EN ISO 17412 	 Molio's product series LeksiCON Molio Huset 	• ConTech Lab's SME initiatives	 Molio House ICT program for openBIM 	Molio's BIM services and education
*	 Moderate support for national BIM guides and rules 	 bSI standards AS ISO19650 series OBOS AS ISO 16739 BIM-MEPAUS 	 NATSPEC BIM Properties Generator BIM Value tool BIM benchmarking tool 	 Policies & guides of Green Building Council Australia (GBCA) AS sustainable built environment programs 	 Sydney Opera House BIM implementation Brisbane's Queens Wharf Development 	 NATSPEC BIM knowledge hub and trainings
•	 Moderate support for national BIM guides and rules 	 bSI standards AS ISO19650 series OBOS AS ISO 16739 BIM-MEPAUS 	 National automated regulatory check platform (by 2025) 	 National BIM program to mitigate CO2 emissions. 	 National DX for building and city planning 	 bSJ online database for guides and regulations

Figure 1. Summary of openBIM adoption and implementation across the globe

3.2 Government-led initiatives and collaboration

The active involvement of a country's government can be an important force in the widespread adoption and success of openBIM. Governments possess the regulatory and financial clout to incentivize, mandate, and guide the direction of industry practices [8]. Considering the government role as one of the key dimensions of this study, Nordic countries and Singapore stand out in this regard due to the foresight of their government entities. Their push for openBIM isn't just about ticking a box; it's a larger vision for standardizing practices, ensuring sustainability across projects, and ultimately uplifting the nation's entire construction ecosystem. Singapore's National BIM program is anchored in openBIM principles, and represent the nation's vision for disaster resilience, integration of openGIS and openBIM, and reduction of carbon footprint. This government-led integration of openBIM into broader developmental goals showcases how Singapore can use it as a tool to address larger challenges, ranging from urbanization to environmental conservation.

Collaboration at both national and international levels has proven beneficial. The US and Nordic countries, for instance, have initiated joint ventures to bolster openBIM's potential. Such collaborations not only pool resources but also bring together diverse expertise, leading to refined solutions and strategies. Moreover, challenges faced in one region can be addressed by solutions tried and tested in another. This reciprocal sharing of experiences and resolutions fast-tracks problem-solving and fosters an environment of continuous improvement. The results are refined strategies, enhanced openBIM standards, and improved tools and methodologies that are the by-products of shared expertise.

3.3 Standardized terminology and semantic consistency

In a fragmented and complex domain like the built asset industry, the use of common semantics, definitions, and terminologies allow for effective implementation and collaboration across various stakeholders. Drawing inspiration from countries that have successfully implemented openBIM workflows, establishing common semantics and terminologies are the initial steps towards the effective exchange of digital assets across lifecycle stages. Denmark provides a stellar example in this regard, through the development of the LeksiCON, Molio has focused on achieving consistency in the way construction-related information is shared and goal is understood. The primary to reduce misinterpretations, which could lead to costly errors. By making LeksiCON available openly, they ensure that every participant in the construction process is on the same page, figuratively and literally. In France, buildingSMART France is promoting Semantic4BIM to ensure that when a project transitions from one phase to another, or one discipline to another, the terminology remains consistent, avoiding potential misunderstandings or errors. Such a resource wouldn't just be a repository but a reflection of the country's commitment to ushering its construction sector into a new digital age while maintaining clarity, consistency, and cohesion.

3.4 openBIM and sustainability

Beyond the immediate benefits of open standards in the built asset industry, openBIM offers avenues to address larger sustainability challenges. The move towards sustainable and smart cities necessitates a departure from siloed decision-making. Urban planners, environmentalists, transportation experts, and technologists must collectively envision urban spaces. openBIM can serve as the unifying ecosystem where diverse stakeholders converge to share insights, analyze data, and measure sustainability of asset and products. This collective approach can lead to more informed decisions, ensuring that urban infrastructures are functional, sustainable, adaptable, and in tune with the environmental condition, as well as needs and aspirations of their residents. Having a centralized and open

ecosystem provides numerous opportunities to integrate openBIM and openGIS digital assets. This enhances the integration of environmental data with as-built assets and enables users to trace the circularity of information across multiple resources. It also increases the longevity and efficiency of the materials, systems, and strategies implemented.

4 Conclusion

The research presented openBIM's transformative impact on the built asset industry, marking a shift toward collaborative, efficient, and sustainable practices. Despite momentum in adoption propelled by government initiatives and technological advancements, challenges persist demanding greater standardization, enhanced digital competency, and the navigation of cultural resistance. Notably, openBIM's role in fostering sustainability and its integration into smarter, sustainable cities is gaining recognition. Canada stands on the cusp of a transformative era, leveraging global best practices and openBIM as a blueprint for a sustainable digital future. Embracing open standards fosters collaboration and consistency across projects, while proactive governmental roles, as seen in Norway, Japan, UK, and Singapore, illustrate the impact of regulatory mandates. The potential for openBIM in enhancing sustainability is pivotal, offering opportunities for circular information flow and efficiency in materials and systems. A comprehensive roadmap, blending technology, collaboration, and sustainability, positions Canada for a promising future in the built environment. Collaboration, both domestically and internationally, can accelerate Canada's digital journey, promoting shared expertise and mutual growth. Aligning with a global data dictionary enables harmonization of classifications, definitions, and terminologies, fostering enhanced collaboration and understanding among diverse professionals. Initiatives to enhance awareness, conduct cost benefit analyses, and ensure access to standardized documentation are vital for owners and government bodies. Implementation should be guided by a culture of accountability and a recognition that BIM's utility extends well beyond the design and construction phases.

Acknowledgment

This paper is prepared from a project funded by National Research Council Canada in collaboration with the team at buildingSMART Canada.

References

[1] buildingSMART International Ltd., What is openBIM; available online:

https://www.buildingsmart.org/about/openbim/ope nbim-definition/

- [2] buildingSMART International Ltd. Industry Foundation Classes; available online: <u>https://www.buildingsmart.org/standards/bsi-</u> <u>standards/industry-foundation-classes/</u>
- [3] buildingSMART International Ltd. BIM Collaboration Format (BCF); available online: <u>https://www.buildingsmart.org/standards/bsi-</u> <u>standards/bim-collaboration-format-bcf/</u>
- [4] buildingSMART International Ltd. buildingSMART Data Dictionary; available online: <u>https://www.buildingsmart.org/users/services/build</u> ingsmart-data-dictionary/
- [5] Jiang, S.; Jiang, L.; Han, Y.; Wu, Z.; Wang, N. OpenBIM: An Enabling Solution for Information Interoperability. *Applied Sciences* **2019**, *9*, 5358, doi:10.3390/app9245358.
- [6] Bui, N.; Merschbrock, C.; Munkvold, B.E.; Hjelseth, E. ROLE OF AN INNOVATION COMMUNITY IN SUPPORTING BIM DEPLOYMENT: THE CASE OF BUILDINGSMART NORWAY.; Seville, Spain, December 18 2019; pp. 329–342.
- [7] Howard, R.; Björk, B.-C. Building Information Modelling – Experts' Views on Standardisation and Industry Deployment. *Advanced Engineering Informatics* 2008, 22, 271–280, doi:10.1016/j.aei.2007.03.001.
- [8] Aksenova, G.; Kiviniemi, A.; Kocaturk, T.; Lejeune, A. From Finnish AEC Knowledge Ecosystem to Business Ecosystem: Lessons Learned from the National Deployment of BIM. *Construction Management and Economics* 2019, 37, 317–335, doi:10.1080/01446193.2018.1481985.

Evaluating Site Scanning Methods: An Assessment System for Quantitative Comparison and Selection

Maria Mortazavi¹, Bin Wu¹, Sofie Lore Steinbrenner¹ and Shervin Haghsheno¹

¹Institute of Technology and Management in Construction, KIT, Germany <u>maria.mortazavi@kit.edu</u>, <u>bin.wu@kit.edu</u>, <u>sofie.steinbrenner@student.kit.edu</u>, <u>shervin.haghsheno@kit.edu</u>

Abstract –

With the rapid advancement of digitization, a multitude of tools for as-built modeling has emerged, posing a challenge in choosing the most suitable product. This article presents an evaluation system for identifying and selecting a site scanning method to create a digital twin of existing buildings. The evaluation criteria developed for this purpose were integrated into an evaluation scheme that forms the basis for a quantitative assessment. A case study was conducted to validate the developed evaluation system. Two Operating Systems (OS) applications, a Faro laser scanner, and a handheld LumoScanner were selected to generate a 3D model for a floor with a 500 m² area. The presented method supports construction and design companies in their decisionmaking process when selecting scanning methods for practical building applications.

Keywords -

Evaluation System, Laser Scanner, OS Scanning Application, As-built Modeling

1 Introduction

This work is part of the research project NaiS (Nachhaltige intelligente Sanierungsmaßnahmen). It focuses on digitizing information that is hard to reach and on optimizing it through human collaboration. One of the goals of NaiS is to employ advanced technologies in the field of site scanning to generate the digital twin.

The AEC sector is embracing an increasing availability of site scanning technologies [1], revolutionizing construction processes. However, adopting these technologies faces technical, economic and awareness-related difficulties. Previous studies [2], have explored the challenges associated with adapting laser scanners. Our paper addresses these challenges by employing an assessment system, allowing a quantitative comparison of site scanning technologies. This approach contributes to overcoming barriers and promoting the effective integration of site scanning technologies in the AEC sector. Therefore, the primary aim of this work is to develop a concept for objectively evaluating site scanning technologies and hence supporting decision making.

For the development of the conceptual framework, qualitative criteria were defined based on existing literature. This decision resulted from the realization in [3] & [4], that the exclusive use of quantitative measures is not sufficient for a comprehensive assessment of the multiple dimensions that characterize the performance of technologies. As qualitative criteria cannot be measured directly, a new instrument is needed to evaluate them. A utility value analysis is proposed, as recommended in [5] & [6]. This is particularly suitable for complex decisionmaking problems, as it offers the possibility of quantifying all evaluation criteria and therefore subjecting them to a final assessment. Consequently, this approach enables the objectification of qualitative criteria.

2 Methodology

This paper presents a comprehensive assessment framework designed for the evaluation of building scanners. The framework covers all essential aspects and enables a thorough analysis. It aims to make the framework to be both accessible and practical for users. This framework is divided into three components: evaluation criteria, evaluation scheme, and utility analysis. Each component will be discussed in the following discussion.

2.1 Evaluation Criteria

Kühnapfel [6] addresses essential details to consider in selecting criteria. To ensure the practicability, particular emphasis was placed on formulating the criteria in a manner that makes them applicable to a broad spectrum of building scanning methods. Similar criteria were avoided to prevent collinearity, which could result in disproportionate weight of certain categories. Furthermore, attention was given to selecting criteria that are relevant to the decision-making problem.

Table 1 presents the content of the evaluation criteria which have been selected based on existing literature. The weights are determined based on a subjective weighting method known as point allocation. Decision makers assign 100 points to the criteria, with a criterion's importance increasing with the number of points it receives. The total weighting for all main criteria must equal 100, and the sum of all sub-criteria within each main criterion must also equal 100 [7]. The weighting of the criteria in the study presented in this paper were assigned based on the author's expertise.

Now, the criteria need to be made quantifiable and comparable. To achieve this, an evaluation scale must be defined, ensuring a shared understanding among decision-makers. Therefore, we developed an evaluation scheme, which is described in the following chapter.

Table 1: Evaluation Criteria for building scanner

Main Criteria	Weighting (%)	Sub-Criteria	Weighting (%)
Functionality	36	·Technology [8]	8
		·Data acquisition [9]	9
		·Data analysis [10]	25
		·Result [8]	25
		·Costs [8]	25
		·Internet connection [11]	8
Maintainability	8	·Community [12]	34
and		·Maintenance readiness [12]	33
Sustainability		·Evolvability [12]	33
Performance	10	·Duration [9], [10]	80
		·Performance [11]	20
Compatibility	4	·Compatibility [12]	100
Usability	28	·Comprehensibility [12]	15
		·Documentation [12]	11
		·Installability [12]	11
		·Learnability [12]	11
		·Self-descriptiveness [13]	15
		·Controllability [13]	15
		·Findability [14]	11
		·Support [12]	11
Reliability	3	·Fault tolerance [13]	100
Security	3	·Data security [11]	100
Portability	8	·Operating System [12]	50
		·Mobility [9]	50

2.2 Evaluation scheme

In the presented Assessment System, the 5-point scale was selected to meet the requirements described below, including the practicability of the assignment to a scale value, the representation of the number of scaling levels, the uniformity of the scale, and the directional equality of the value development, as discussed in [6].

To guarantee the uniformity of the assessment system, a rating scale is created, where each sub-criterion is rated from 1 (very negative performance) to 5 (very positive performance). Therefore, all criteria can be evaluated on a common basis. The assessment of the criteria depends mainly on the hands-on experiences and knowledge of the evaluator.

A common understanding of the scale concerning criteria among all assessors is crucial to achieving comparable and reliable results. Therefore, a scoring guideline is provided in an evaluation scheme. Positive and negative characteristics are compared. More positive characteristics result in a higher score, while a higher degree of negative impact leads to a lower score. Table 2 illustrates an example of the evaluation scheme.

Sub-	Target Value				
criteria	Positive Impact	Negative Impact			
Compa-	There are various interfaces to other software programs.	There are no interfaces to other software programs.			
tibility	Various data formats can be generated for export	Only one data format can be generated for export.			

2.3 Utility Analysis

Zangemeister [15] declines utility analysis as an evaluation method designed to support rational decisionmaking. In this paper, it is used both to evaluate a building scanner and to compare multiple options. The evaluation criteria created above are integrated into the utility analysis. By applying weighted evaluation criteria, different options can be compared to determine an optimal result from a holistic perspective. Each subcriterion weighted according to its relevance is added together to give the total value. According to [6], the criteria weights are determined individually. During the evaluation process, the target values can be determined by using the evaluation scheme and incorporated into the utility analysis. A maximum of 5 points can be achieved. The higher the score, the more capable the potential building scanner is.

Table 3 illustrates an example of the mathematical expression of the utility analysis.

Table 3: Example of an Evaluation

Main Criteria	Weighting (%)	Sub- Criteria	Weighting (%)	Unweighted evaluation (Points 1-5)	Weighted evaluation	Sum of Main Criteria
Security	3%	Data	100%	4	$4^{(-1, y, 4)}$	0.12
		security			(=1X4)	(=0.03X4)

3 Case Study and Results

3.1 Description

To validate the assessment system presented above, two mobile-phone-based Light Detecting and Ranging (LiDAR) OS applications (MagicPlan and PolyCam), a photogrammetry-based scanning device (LumoScanner), and a laser scanner (Faro Focus S Plus 150) were tested. The OS applications have been chosen as they fulfill the requirements for our case study. For example, the user license is free during a trial period, the acquisition settings can be customized, and a 3D model is generated as an Industry Foundation Class (IFC)-Model. The test occurred on an approximately 500 m² floor in an office building (Figure 1), located in Karlsruhe, Germany.

T_ HIMES	15	14	13		12	<u>t</u> t	4	******
5-900 J				e Ka				1
	1 2	3 4	PC 5	6	7 8	9410		

Figure 1: Floorplan of the study object

To guarantee the comparability of the recorded scans, all scans were performed, under consistent lighting conditions. The detailed procedure in the case study is presented below.

Two OS applications are separately installed on two iPad Pro 11-inch (3rd generation), both equipped with a LiDAR sensor with a detection range of up to 5 meters. PolyCam offers four scanning modes (e.g. LiDAR, Room, Foto, and 360). LiDAR and Room modes were chosen to generate point clouds along with 3D models. The results can be exported as point clouds or 3D models in various formats. In LiDAR mode, each room is scanned separately, creating individual point clouds (exportable as .las files). In CloudCompare, those point clouds emerged to create a complete point cloud. The emerged point cloud was processed in ReCap and Revit for asbuilt modeling.

MagicPlan captures the spatial data and supports various export file formats. It allows users to edit the layout and add various elements such as windows and doors. A Bluetooth-enabled laser range finder can be lined for increased accuracy within a five-meter range. Besides, an internet connection is not required during recording.

LumoScanner creates 2D point clouds from images. Each scan takes around 2 seconds and requires 1 to 2 scans per room. This highly depends on the room size and layout. Due to its 2-dimensional nature, a handheld laser rangefinder supplements the LumoScanner for more detailed measurements, such as the position and size of windows or doors. The floor plan was created in advance and saved in the LumoApp. The final 3D modeling process is conducted in Revit.

The Faro laser scanner offers up to 150 meters of scanning range and rapid data capture. In the case study, its settings were adjusted for a high scanning speed and a suitable quality, including a 10 m indoor range, a resolution of 1/32, and no color. One scan took approximately 40 seconds. The recorded point clouds were saved on an SD card, merged in CloudCompare, and used for 3D modeling in Revit.

3.2 Results

The case study was conducted in two separate days. On day one, MagicPlan and PolyCam were tested to scan 15 rooms and three staircases. Due to highly complex circumstances in room 15, the scanning result was not used in the as-built modeling.



Figure 2: (a) 3D model in MagicPlan; (b) 3D point cloud model from PolyCam; (c) 3D point cloud model from Faro scanner; (d) 2D point cloud model from LumoScanner (illustration)

The data acquired with MagicPlan can be exported as an IFC model directly (Figure 2a), comparatively, the data from PolyCam was first exported as a set of 3D point clouds, each one of them represents a room or staircase (Figure 2b). On day two, the Faro laser scanner and LumoScanner were utilized to capture 14 rooms and three staircases. A set of 3D point clouds was exported from the Faro laser scanner (Figure 2c). The Data from LumoScanner were uploaded to Lumo Cloud, and a set of 2D point clouds was generated in the platform (Figure 2d) just for illustration, not original data).

Table 4 presents the data processing speed of the four scanning methods. Due to high integration of MagicPlan, it reaches the highest speed of $9.02 \text{ m}^2/\text{min}$.

Table 4: The processing speed of the four methods

	MagicPlan	PolyCam	Faro	Lumoview
Area/Time [m²/min]	9.02	1.27	1.05	1.42

Table 5 presents the final evaluation results derived from the assessment system. The results and detailed explanations for the assigned ratings are recorded in a separate table. However, not all sub-criteria are qualitative, such as the sub-criterion duration. Therefore, the duration for each processing step is recorded also in a table. The approach streamlines the assessment process and provides a transparent documentation review and analysis. MagicPlan is evaluated as the best method with a score of 4.26. The scores of the remaining three methods range between 3.50 and 3.60 points.

	Weigh- ting (%)	Magic- Plan	PolyCam	Lumo- view	Faro
Functionality	36	1.44	1.23	1.37	1.22
Maintainability & Sustainability	8	0.29	0.29	0.27	0.28
Performance	10	0.48	0.37	0.21	0.45
Compatibility	4	0.16	0.2	0.20	0.13
Usability	28	1.31	1.15	1.01	1.09
Reliability	3	0.06	0.06	0.06	0.03
Security	3	0.12	0.12	0.09	0.09
Portability	8	0.4	0.16	0.32	0.24
Total	100	4.26	3.58	3.53	3.53

Table 5: The final evaluation results

3.3 Discussion

MagicPlan results as the most efficient method, particularly in functionality, performance, and usability: (1) low cost (99.99€/year subscription); (2) high processing speed (9.02m²/min); (3) user-friendly interface (editable room layout). It is noticeable that almost all the tested objects are at a similar level. It can therefore be assumed that mobile laser scanners can be used just as effectively as terrestrial laser scanners. This finding is also supported in studies [1] and [16]. The high ratings of MagicPlan in the Google Play Store and the frequency of downloads further validate our findings [17].

4 Conclusion

This work developed a comprehensive evaluation system for site-scanning methods to enhance decisionmaking for users. A case study was conducted to verify the evaluation system and assess the performance of various scanning methods. The results highlight that the system can deliver a reliable evaluation with limited accessible information in practice. It should be noted that its application is not limited to the field of site-scanning. Additionally, the weighting rate for each criterion can be adjusted based on the users' need and a particular use case. However, a wider range of scanning methods should be tested in a more diverse building environment to verify this system. In conclusion, the paper lays a solid foundation for developing an evaluation system for the applications in BIM and offers a practical tool for users.

References

- Teppati Losè L, et. al. Apple LiDAR Sensor for 3D Surveying - Tests and Results in the Cultural Heritage Domain. *Remote Sens*, 14(17):4157, 2022.
- [2] Waqar, A., et. al. Complexities for adopting 3D laser scanners in the AEC industry Structural equation modeling. *Applications in Engineering*

Science, 16:100160, 2023

- [3] Radujković, M., et al. Application of key performance indicators in South-Eastern European construction. *Journal of Civil Engineering and Management*, 16(4):521-530, 2011
- [4] Chan, A.-P.-C., Chan, A.-P.-L. Key performance indicators for measuring construction success. *Benchmarking - An International Journal*, 11(2):203–221, 2004
- [5] Gelbmann, U., et al., Innovationsleitfaden: Der Weg zu neuen Produkten. Graz: Eigenverlag, 2003
- [6] Kühnapfel, J.-B., Scoring und Nutzwertanalysen -Ein Leitfaden für die Praxis. Wiesbaden: Springer Gabler, 2021
- [7] Zardari, N. H., et. al., Weighting Methods and their Effects on Multi-Criteria Decision Making Model Outcomes in Water Resources Management. SpringerBriefs in Water Science and Technology, 2015
- [8] Hudli, A., et al. An evaluation framework for selection of mobile app development platform. *Proceedings of the 3rd International Workshop on Mobile Development Lifecycle*, pages 13-16, 2015
- [9] Catalan R. B., et. al. Evaluation of 3D Scanners to Develop Virtual Reality Applications. *Electronics, Robotics and Automotive Mechanics Conference,* pages 551-556, Cuernavaca, Mexico, 2007.
- [10] Jang, A., et. al. Structural Stability Evaluation of Existing Buildings by Reverse Engineering with 3D Laser Scanner. *Remote Sens*, 14(10): 2325, 2022.
- [11] Huy, N. P. and van Than, D. Selecting the right mobile app paradigms. *Fifth IEEE International Conference on Service-Oriented Computing and Applications*, pages 1-6, Taipei, Taiwan, 2012
- [12] Jackson, M., et. al. Software evaluation: criteriabased assessment, *Software Sustainability Institute*, 2011.
- [13] ISO 9241-110:2010-03, Ergonomics of humansystem interaction, part 110: Dialogue principles
- [14] Morville, P. Experience design unplugged. Association for Computing Machinery SIGGRAPH 2005 Web program, pages: 10-es, New York, USA, 2005.
- [15] Zangemeister, C. Nutzwertanalyse in der Systemtechnik - Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen. 5., Hamburg: BoD – Books on Demand, 2014
- [16] Di Stefano F. et. al. Mobile 3D scan LiDAR a literature review. *Geomatics, Natural Hazards and Risk*, 12(1): 2387-2429, 2021
- [17] GooglePlay-Magicplan. Online: https://play.google.com/store/apps/details?id=com. sensopia.magicplan&hl=de&gl=US, Accessed: 13/03/2024.