Manuscript 2115

Automated Detection for Alignment of Modular Integrated Construction Modules Using YOLOv9

Sulemana Fatoama Abdulai

Tarek Zayed Professor

Abdul-Mugis Yussif

Ridwan Taiwo

Mohamed Sherif AbdAllah EbnSallam Elrifaee

See next page for additional authors

Automated Detection for Alignment of Modular Integrated Construction Modules Using YOLOv9

Authors

Sulemana Fatoama Abdulai, Tarek Zayed Professor, Abdul-Mugis Yussif, Ridwan Taiwo, Mohamed Sherif AbdAllah EbnSallam Elrifaee, and Muhammad Muddassir Dr.

Automated Detection for Alignment of Modular Integrated Construction Modules Using YOLOv9

Sulemana Fatoama Abdulai*, sulemana.abdulai@connect.polyu.hk
Department of Building and Real Estate Department, The Hong Kong Polytechnic University, Hong Kong SAR

Tarek Zayed, tarek.zayed@polyu.edu.hk
Department of Building and Real Estate Department, The Hong Kong Polytechnic University, Hong Kong SAR

Abdul-Mugis Yussif, mugis.yussif@connect.polyu.hk
Department of Building and Real Estate Department, The Hong Kong Polytechnic University, Hong Kong SAR

Ridwan Taiwo, ridwan-a.taiwo@connect.polyu.hk
Department of Building and Real Estate Department, The Hong Kong Polytechnic University, Hong Kong SAR

Mohamed Elrifaee, mohamed.elrifaee@connect.polyu.hk
Department of Building and Real Estate Department, The Hong Kong Polytechnic University, Hong Kong SAR

Muhammad Muddassir, muhammad.muddassir@polyu.edu.hk
Department of Building and Real Estate Department, The Hong Kong Polytechnic University, Hong Kong SAR

Abstract
Modular Integrated Construction (MiC) represents an innovative approach in the construction industry, where building components are manufactured off-site and assembled on-site, offering enhanced quality control and reduced construction time. While computer vision technologies have advanced significantly in construction applications, the precise alignment of MiC modules during installation remains challenging, with traditional manual methods being time-consuming and error-prone. This study develops an automated detection system using a modified YOLOv9 architecture to enhance the precision and efficiency of MiC module alignment. The research employed a three-stage methodology comprising literature analysis, data acquisition of 5,000 images, and model development implemented on a GeForce RTX 3050 GPU. The system achieved exceptional performance with a mean Average Precision (mAP50) score of 0.995 and real-time processing capabilities of 3.1 ms per frame at 640×640 resolution. These results demonstrate the system's potential for transforming MiC practices through improved installation precision and reduced human error, contributing significantly to the advancement of construction automation.

Keywords
Computer vision, Modular Integrated Construction, Object detection.

1 Introduction

Modular Integrated Construction (MiC) represents a transformative approach in the construction industry, where building components are manufactured off-site in a controlled factory environment and subsequently transported to the construction site for assembly (Hussein and Zayed 2021). This innovative construction methodology offers numerous advantages, including enhanced quality control, reduced construction time, minimized environmental impact, and improved worker safety. The controlled manufacturing environment enables precise fabrication of modules, while the systematic assembly process on-site ensures efficient project delivery. As the construction industry continues to evolve toward more sustainable and efficient practices, MiC has emerged as a pivotal solution for addressing traditional construction challenges (Arshad and Zayed 2022).

However, one of the critical challenges in MiC implementation lies in the precise alignment of modules during the installation phase (Arshad and Zayed 2022). The accuracy of module alignment directly impacts the final building's structural integrity, functionality, and aesthetic quality. Traditional alignment methods rely heavily on manual measurements and visual inspections, which are time-consuming, labor-intensive, and susceptible to human error. Moreover, the substantial size and weight of modules make real-time adjustments during installation particularly challenging, emphasizing the need for accurate initial positioning and alignment (Shahtaheri et al. 2017).

Recent advancements in computer vision and deep learning have opened new possibilities for automating the module alignment process. Particularly, the YOLO (You Only Look Once) model has shown promising potential in real-time object detection applications. Its ability to process visual information rapidly while maintaining high accuracy makes it especially suitable for construction applications where real-time detection and precision are paramount (Xiong et al. 2024). The integration of such advanced detection systems in MiC processes could significantly enhance the efficiency and accuracy of module installation procedures.

Despite the growing adoption of computer vision technologies in construction, there remains a significant research gap in developing specialized detection systems for MiC module alignment. While existing studies have explored various aspects of construction automation (Cai et al. 2016;Muddassir et al. 2025; Zhang and Pan 2019, 2021), few have specifically addressed the unique challenges of real-time module detection and alignment in MiC applications. This research aims to bridge this gap by developing and implementing a YOLOv9-based detection system optimized for MiC module alignment.

The primary aim of this study is to develop an automated detection system using YOLOv9 for precise identification and alignment of MiC modules during the installation phase. The specific objectives include (1) implementing and training a YOLOv9 model for accurate module detection, (2) evaluating the model's performance in terms of detection accuracy and real-time processing capabilities, and (3) validating the system's effectiveness in supporting module alignment procedures. The significance of this research lies in its potential to enhance the precision and efficiency of module installation in MiC projects. Automating the detection process can significantly reduce installation time, serve as a step toward using computer vision module alignment in MiC, and improve overall construction quality. Furthermore, implementing such a system could contribute to the broader adoption of MiC by addressing one of its key technical challenges, thereby advancing the industrialization of the construction sector.

2 Literature Review

Computer vision technology represents a field of artificial intelligence that enables computers to derive meaningful information from digital images and video (Alsakka et al. 2023a). In construction applications, computer vision systems typically comprise image acquisition hardware, processing algorithms, and output interpretation mechanisms. These systems can automatically detect, classify, and track objects in real-time, making them particularly valuable for construction automation. Automated detection systems fundamentally operate through a pipeline of image preprocessing, feature extraction, and object recognition (Ekanayake et al. 2021). The integration of computer vision in construction automation has transformed traditional practices, offering enhanced precision, reduced human error, and improved efficiency in various construction processes. Particularly in MiC, these systems have become instrumental in addressing critical challenges such as component tracking, alignment verification, and quality assurance. Computer vision application in construction has seen significant advancement through various detection architectures, from traditional Convolutional Neural Networks (CNNs) to more sophisticated object detection frameworks (Alsakka et al. 2023b).

Early attempts at automation in MiC primarily focused on traditional sensor-based systems. With the advancement of computer vision and artificial intelligence, more sophisticated approaches have emerged to address the complex challenges of module alignment. Studies by Zhang and Pan (2019) documented the limitations of conventional methods, highlighting issues such as time consumption, labor intensity, and susceptibility to human error. Pan et al. (2022) demonstrated that computer vision systems could achieve alignment accuracies within 2mm, substantially surpassing manual methods. Zheng et al. (2020) implemented a Convolutional Neural Network (CNN) based approach, achieving 95% accuracy in module detection, though their system showed limitations in real-time processing capabilities. Gavrilas (2018) experimented with Faster R-CNN for module detection, reporting promising results but noting computational intensity as a significant drawback for on-site implementation.

Liu et al. (2018) utilized Faster R-CNN, achieving good accuracy but facing challenges with real-time processing. Thompson et al. (2021) demonstrated faster processing times with single-shot detectors (SSD) but lower accuracy than two-stage detectors. Wu et al. (2021) implemented YOLOv5, which showed promising results in construction applications. Previous implementations of YOLO variants (v3-v7) in construction contexts have shown progressive accuracy and processing speed improvements, though primarily focused on general construction element detection rather than specific module alignment applications.

Current systems face challenges, including environmental variations affecting detection accuracy, real-time processing requirements, balance between accuracy and computational efficiency, reliability under various lighting conditions, and handling of occlusions and partial visibility (Olaoye et al. 2024). While existing literature demonstrates significant progress in computer vision applications for construction, a notable gap exists in specialized systems for MiC module alignment. Most current systems are either not specifically optimized for MiC, unable to meet real-time processing requirements, or limited in their ability to handle complex construction site conditions. The emergence of YOLOv9, with its improved architecture and processing capabilities, presents an opportunity to address these limitations, though its application in MiC module detection and alignment remains largely unexplored, highlighting the need for this study.

3 Research Methodology

The overall methodology of the study is depicted in Figure 1. The first phase involves a comprehensive literature review on automated detection systems in MiC, particularly on computer vision applications in module alignment. This review identifies current technological gaps, methodological limitations, and potential areas for improvement in existing module detection approaches, which informed the development of our YOLOv9-based solution. The second phase focuses on data collection and preparation. A comprehensive dataset is created using images of MiC modules captured in a controlled laboratory environment at The Hong Kong Polytechnic University's Construction Technology Laboratory. The experimental setup simulated actual construction site conditions, including a scaled-down tower crane and modular components arranged to replicate real installation scenarios. The laboratory environment was designed to mirror authentic construction site conditions while allowing for controlled data collection. Images were captured during various simulated installation phases, with the setup incorporating typical site elements such as different module configurations, lighting conditions, and crane positions. This controlled environment enabled systematic data collection while ensuring the dataset's relevance to actual MiC applications. Phase 3 of the study highlighted the model development, and finally, the final phase presented the results, implications, and future direction of the study.

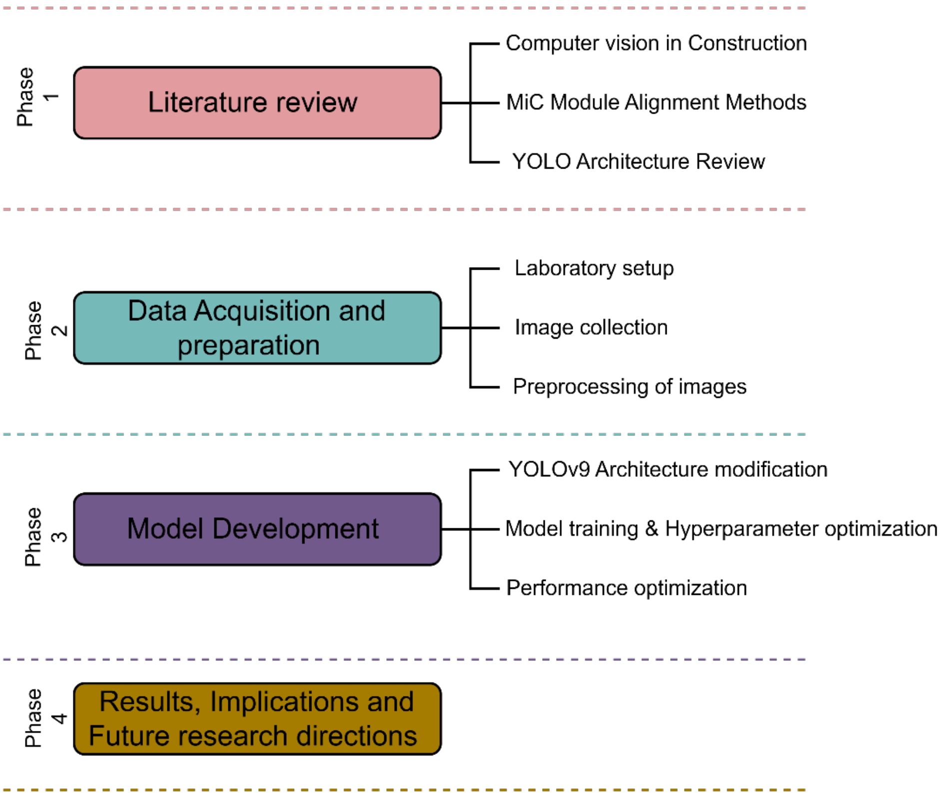

Figure 1: Framework of the study

To ensure robust model performance, images are collected under varying environmental conditions, lighting conditions, viewing angles, and weather situations. The dataset is enhanced through data augmentation techniques to increase its diversity and improve model generalization. Augmentation methods include horizontal and vertical flipping, rotation, brightness adjustment, and contrast variation. Additionally, the dataset is split into training, validation, and testing sets in the ratio of 70:20:10, respectively. The third stage centers on developing and implementing the YOLOv9-based detection model. YOLOv9 is selected for its superior real-time detection capabilities and proven performance in object detection tasks. The model training process involves several key steps: initial network configuration, hyperparameter optimization, and model training over 100 epochs. The training process utilizes three primary loss functions: bounding box loss for spatial accuracy, classification loss for module identification, and Distribution Focal Loss (DFL) for prediction confidence. The model's performance is evaluated using standard object detection metrics, including precision, recall, mean Average Precision (mAP), and F1-score.

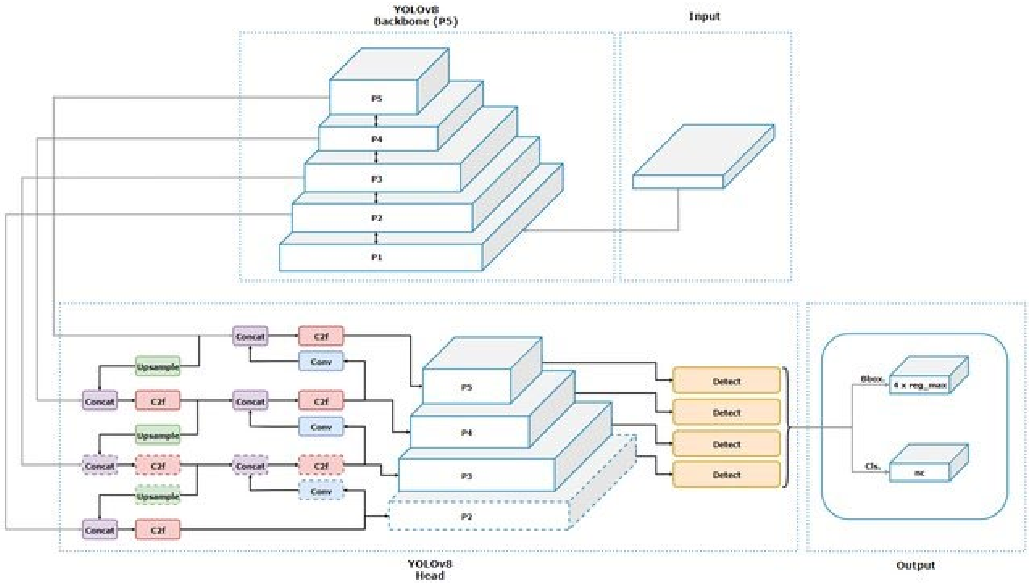

Figure 2: Improved YOLOv9 network architecture (Karna et al. 2023).

4 Model development

The proposed YOLOv9-based module detection model is developed through a systematic architectural design and training. The model architecture builds upon the standard YOLOv9 framework, consisting of a CSPDarknet backbone for feature extraction, a PANet neck for feature aggregation, and a detection head for object localization and classification. The backbone network utilizes Cross-Stage-Partial connections to enhance feature propagation and reduce computational complexity while maintaining detection accuracy. For module detection optimization, the model incorporates several key modifications to the detection head, customizing it to focus on single-class detection (MiC modules) rather than multi-class detection, allowing for more specialized feature learning (Yussif et al. 2024). The network employs a modified anchor-free detection approach, directly predicting module locations without predefined anchor boxes. It is more adaptable to varying module sizes and orientations commonly encountered in construction sites.

The model training process utilizes a custom loss function that combines bounding box regression loss (CIoU loss) for precise module localization, binary cross-entropy loss for module classification, and Distribution Focal Loss (DFL) for improved confidence prediction. Learning rate scheduling follows a cosine annealing approach with warm restarts, starting at 0.01 and gradually decreasing to 0.001, while batch size is set to 8 to balance computational efficiency and training stability. The model training runs for 100 epochs with an early stopping patience of 10 epochs to prevent overfitting.

Data preprocessing involves resizing input images to 640×640 pixels while maintaining aspect ratio through padding. During training, mosaic augmentation is applied with a probability of 0.5, combining four training images to enhance the model's ability to detect modules at different scales and contexts. The training process incorporates random horizontal flipping, rotation (±15 degrees), and brightness adjustment (±25%) as additional augmentation techniques (Xiong et al. 2024). The model's inference pipeline is optimized for real-time performance through batch inference processing and non-maximum suppression with an IoU threshold of 0.45. A confidence score threshold is set at 0.25.

The main contribution of our work is centered on a module alignment verification step that computes detected modules' relative orientation and position, providing real-time feedback for alignment adjustments during installation. The final model outputs include bounding box coordinates, confidence scores, and alignment parameters for each detected module, enabling precise positioning guidance during installation. This comprehensive model development approach ensures robust module detection and alignment capabilities while maintaining real-time performance requirements for practical construction site implementation.

4.1 Model Implementation

The model implementation uses Python 3.10 with TensorFlow 2.15.0 and Keras 2.15.0 as the primary deep learning frameworks. The development environment operates on Windows 10, configured with CUDA 12.6 for GPU acceleration. The hardware setup consists of a GeForce RTX 3050 GPU, Intel Core i7-12700 CPU running at 2.10 GHz, and 32 GB of RAM, enabling efficient model training and inference. The training utilizes a comprehensive dataset of 5,000 annotated images, these images were systematically collected from our laboratory experimental setup, which was designed to replicate real construction site conditions while allowing for controlled data collection and precise annotation, with 4,000 (80%) allocated for training and the remaining 1,000 images evenly split between validation and testing sets. With this configuration, the model achieves an average inference time of 3.1ms per frame at 640×640 resolution. Implementation includes TensorFlow's built-in optimization techniques, such as automatic mixed precision and graph optimization, to enhance computational efficiency.

Table 1: Evaluation matrix for performance assessment.

mean average precision (mAP)	$\frac{1}{M}\sum_{k=1}^{M} AP_k$	M is the total number of classes AP_k is the average precision of the k-th class
Precision (P)	$\frac{TP}{TP+FN}$	TP = True Positives and FP = False Positives
Recall	$\frac{TP}{TP+FN}$	FN = False Negatives

5 Experimental Results and Discussion

The YOLOv9 architecture was customized and implemented for MiC module detection, and it was trained and validated on 5,000 images generated from a lab experiment. The model demonstrated exceptional detection and localization capabilities during validation, achieving a mean Average Precision (mAP50) score of 0.995 and mAP50-95 score of 0.755 at a 50% IoU threshold, with perfect precision (1.0) across varying recall values (see figure 3). The confusion matrix confirmed flawless discrimination between module and background classes (See Figure 4).

Training metrics showed significant improvements over 100 epochs. The bounding box regression loss decreased from 1.539 to 0.917, while the classification loss improved from 1.167 to 0.359. Validation metrics aligned closely with training performance, with final values of bounding box loss at 1.0631 and classification loss at 0.39507, indicating robust generalization without overfitting. The model exhibited remarkable stability, achieving optimal precision and recall values of 1.0 by epoch 15 and maintaining consistent performance thereafter. Spatial prediction analysis demonstrated accurate module dimension estimation, which is crucial for automated placement and alignment verification. Implementing a GeForce RTX 3050 GPU with CUDA 12.6 optimization achieved real-time processing with a 3.1ms per frame inference time at 640×640 resolution.

These results validate the proposed architecture's effectiveness for MiC module detection, combining high accuracy with real-time processing capabilities. The system's robust performance makes it suitable for automated construction monitoring and quality control, potentially improving efficiency and reducing errors in MiC operations.

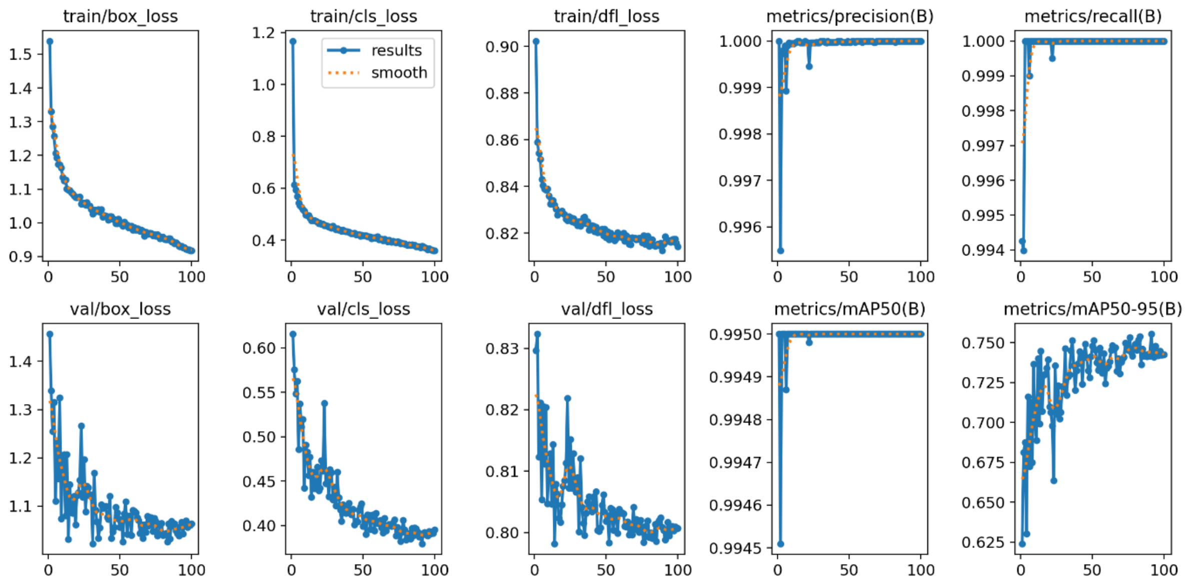

Figure 3: Training and validation performance metrics for YOLOv9 object detection model.

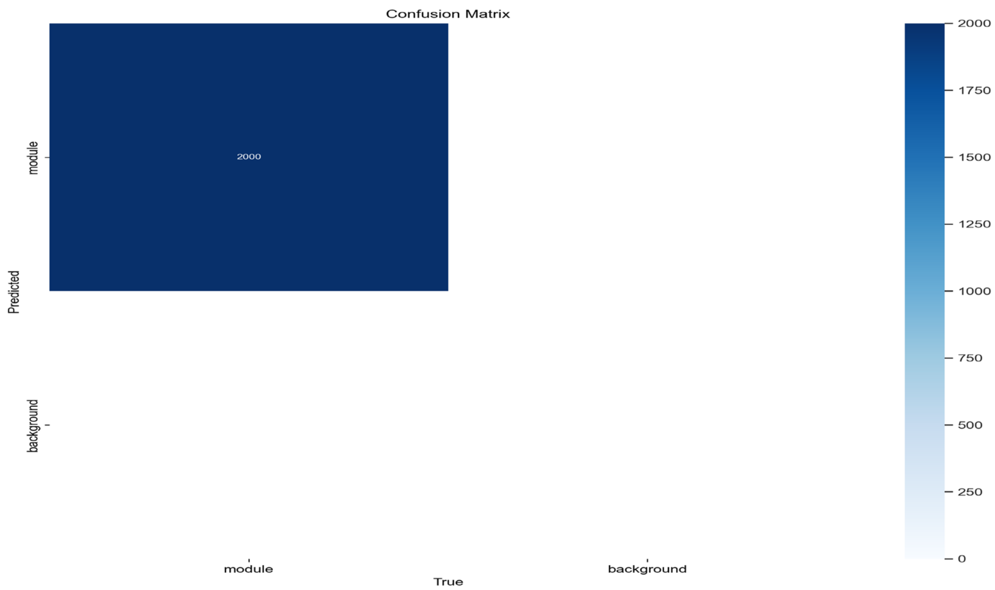

Figure 4: Confusion matrix

Figure 5: YOLOv9 detection test results

6 Implications of the study

This study offers significant theoretical and practical contributions while demonstrating important implications for the construction industry's advancement toward automation. Theoretically, our implementation of YOLOv9 achieved exceptional detection metrics, demonstrating a mean Average Precision (mAP50) of 0.995 and mAP50-95 of 0.755 at a 50% IoU threshold. The model maintained perfect precision (1.0) across various recall values, setting a new performance benchmark for module detection in construction applications. These results advance our understanding of deep learning optimization tailored to MiC applications.

The practical contributions directly translate into meaningful industry implications. The system's real-time processing capability of 3.1 ms per frame demonstrates immediate applicability to real-world construction scenarios and substantially enhances construction efficiency by reducing module alignment time and streamlining workflow processes. This processing speed enables real-time module detection that aligns with actual crane operation requirements, providing instantaneous feedback during the critical module placement phase. Furthermore, the system's high accuracy in module dimension estimation directly addresses one of the key challenges in MiC – precise alignment verification during installation. Beyond the technical achievements, the study's implications extend to construction safety and automation advancement. The automated detection system reduces workers' exposure to hazardous situations during module placement operations while enabling remote monitoring capabilities for enhanced site safety. These capabilities, combined with the system's practical performance metrics, support the broader adoption of MiC methodologies in the construction industry by offering tangible improvements to current module installation practices.

7 Conclusions and Further Research

This research successfully developed and implemented a YOLOv9-based detection system for MiC module alignment, addressing a critical challenge in MiC. The model achieved exceptional performance metrics, including a mAP50 score of 0.995 and perfect precision values, demonstrating its effectiveness in real-time module detection for alignment applications. The system's ability to process frames in 3.1ms while maintaining high accuracy represents a significant advancement in automated construction monitoring. The implementation demonstrates that deep learning-based detection systems can effectively address the challenges of precise module alignment in MiC projects, offering a viable alternative to traditional manual methods. The model's robust performance and consistent accuracy across various conditions suggest its potential for widespread adoption in construction automation. These results validate the feasibility of integrating advanced computer vision technologies in MiC processes, potentially transforming module installation procedures by reducing human error, improving efficiency, and enhancing overall construction quality.

Despite the significant contributions of this study, several limitations warrant acknowledgment. The primary limitation stems from our validation methodology, which was exclusively conducted in a controlled laboratory environment using a scaled-down tower crane setup, potentially not fully capturing actual construction site complexities. While our study demonstrated robust module detection capabilities, it did not address the complete module alignment process. However, this work establishes a foundation for future alignment systems that could incorporate pose estimation and additional sensing hardware. Integrating Inertial Measurement Units (IMU) and LiDAR systems could enable precise 3D mapping of construction sites and enhanced spatial awareness.

A notable technical limitation is that real-time performance metrics were not validated under actual construction site conditions and computational constraints. Future research should explore integrating multi-modal sensing approaches, combining IMU data with computer vision systems to enhance module tracking and positioning capabilities in real-time. This combined approach could provide more comprehensive monitoring of module movements, ultimately improving the efficiency and safety of MiC operations.

8 Acknowledgement

This research was supported by the Department of Building and Real Estate at The Hong Kong Polytechnic University. This study is part of a broader Ph.D research initiative focused on developing automated alignment systems for modular integrated construction. The authors express sincere appreciation to the Editor and Reviewers for their thorough feedback, which substantially enhanced the quality of this manuscript.

9 References

Alsakka, F., S. Assaf, I. El-Chami, and M. Al-Hussein. 2023a. "Computer vision applications in offsite construction." *Autom Constr*, 154. Elsevier B.V. https://doi.org/https://doi.org/10.1016/j.autcon.2023.104980.

Alsakka, F., I. El-Chami, H. Yu, and M. Al-Hussein. 2023b. "Computer vision-based process time data acquisition for offsite construction." *Autom Constr*, 149. Elsevier B.V. https://doi.org/http://dx.doi.org/10.1016/j.autcon.2023.104803.

Arshad, H., and T. Zayed. 2022. "Critical influencing factors of supply chain management for modular integrated construction." *Autom Constr*, 144. Elsevier B.V. https://doi.org/https://doi.org/10.1016/j.autcon.2022.104612.

Cai, P., Y. Cai, I. Chandrasekaran, and J. Zheng. 2016. "Parallel genetic algorithm based automatic path planning for crane lifting in complex environments." *Autom Constr*, 62: 133–147. Elsevier. https://doi.org/10.1016/j.autcon.2015.09.007.

Ekanayake, B., J. K. W. Wong, A. A. F. Fini, and P. Smith. 2021. "Computer vision-based interior construction progress monitoring: A literature review and future research directions." *Autom Constr*, 127. Elsevier B.V. https://doi.org/10.1016/j.autcon.2021.103705.

Gavrilas, Mihai. 2018. *EPE 2018 : proceedings of the 2018 International Conference and Expositions on Electrical and Power Engineering*. IEEE.

Hussein, M., and T. Zayed. 2021. "Crane operations and planning in modular integrated construction: Mixed review of literature." *Autom Constr*, 122. Elsevier B.V. https://doi.org/10.1016/j.autcon.2020.103466.

Karna, N., A. P. Putra, S. M. Rachmawati, M. Abisado, and G. A. Sampedro. 2023. "Towards Accurate Fused Deposition Modeling 3D Printer Fault Detection using Improved YOLOv8 with Hyperparameter Optimization." *IEEE access*. https://doi.org/10.1109/ACCESS.2017.DOI.

Liu, B., W. Zhao, and Q. Sun. 2018. "Study Of Object Detection Based On Faster R-CNN." *018 25th IEEE International Conference on Image Processing (ICIP)*.

Olaoye, F., K. Potter, and L. Doris. 2024. *Computer Vision and Image Recognition Techniques*.

Muddassir, M., T. Zayed, A. H. Ali, M. Elrifaee, S. F. Abdulai, T. Yang, and A. Eldemiry. 2025. "Automation in tower cranes over the past two decades (2003–2024)." *Autom Constr*, 170. Elsevier B.V. https://doi.org/10.1016/j.autcon.2024.105889.

Pan, M., Y. Yang, Z. Zheng, and W. Pan. 2022. "Artificial Intelligence and Robotics for Prefabricated and Modular Construction: A Systematic Literature Review." *J Constr Eng Manag*, 148 (9). American Society of Civil Engineers (ASCE). https://doi.org/10.1061/(asce)co.1943-7862.0002324.

Shahtaheri, Y., C. Rausch, J. West, C. Haas, and M. Nahangi. 2017. “Managing risk in modular construction using dimensional and geometric tolerance strategies.” *Autom Constr*, 83: 303–315. Elsevier B.V. https://doi.org/10.1016/j.autcon.2017.03.011.

Wu, T. H., T. W. Wang, and Y. Q. Liu. 2021. “Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network.” *2021 3rd World Symposium on Artificial Intelligence, WSAI 2021*, 24–28. Institute of Electrical and Electronics Engineers Inc.

Xiong, C., T. Zayed, and E. M. Abdelkader. 2024. “A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks.” *Constr Build Mater*, 414. Elsevier Ltd. https://doi.org/10.1016/j.conbuildmat.2024.135025.

Yussif, A. M., T. Zayed, R. Taiwo, and A. Fares. 2024. “Promoting sustainable urban mobility via automated sidewalk defect detection.” *Sustainable Development*. John Wiley and Sons Ltd. https://doi.org/10.1002/sd.2999.

Zhang, Z., and W. Pan. 2019. “Virtual reality (Vr) supported lift planning for modular integrated construction (mic) of high-rise buildings.” *HKIE Transactions Hong Kong Institution of Engineers*, 26 (3): 136–143. Hong Kong Institution of Engineers. https://doi.org/10.33430/V26N3THIE-2019-0015.

Zhang, Z., and W. Pan. 2021. “Virtual reality supported interactive tower crane layout planning for high-rise modular integrated construction.” *Autom Constr*, 130. Elsevier B.V. https://doi.org/https://doi.org/10.1016/j.autcon.2021.103854.

Zheng, Z., Z. Zhang, and W. Pan. 2020. “Virtual prototyping- and transfer learning-enabled module detection for modular integrated construction.” *Autom Constr*, 120. Elsevier B.V. https://doi.org/10.1016/j.autcon.2020.103387.