

Adaptive object detection for construction sites via parameter-efficient fine-tuning with LoRA

Hyung-soo Kim ^a, Jaehwan Seong ^b, Yuree Choi ^c and Hyung-Jo Jung*

Department of Civil and Environmental Engineering, Korea Advanced Institute for Science and Technology,
291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

(Received September 11, 2025, Revised November 11, 2025, Accepted November 12, 2025)

Abstract. Construction sites present diverse and evolving visual conditions that challenge the generalizability of pre-trained object detection models. This study proposes a parameter-efficient fine-tuning approach based on Low-Rank Adaptation to enable adaptive object detection tailored to site-specific conditions. A general model was trained on a large-scale dataset and fine-tuned using both the proposed method and full fine-tuning across three real-world construction projects. Despite utilizing only 12% of the trainable parameters, the proposed approach achieved comparable or superior detection accuracy with 10% reduced training time and 30% lower GPU memory consumption. These results highlight its effectiveness in adapting object detection models to site-specific conditions under resource constraints. Furthermore, the approach can be extended with semi-supervised learning to support scalable adaptation in construction environments.

Keywords: adaptive object detection; computer vision; construction safety Low-Rank Adaptation (LoRA); parameter-efficient fine-tuning

1. Introduction

Recent advancements in computer vision have significantly enhanced the capabilities of construction site monitoring, particularly in productivity analysis and safety management (Fang *et al.* 2020, Alsakka *et al.* 2023, Alaloul *et al.* 2022, Liu *et al.* 2021, Ekanayake *et al.* 2021, Xu *et al.* 2021). Computer vision technologies offer the potential to automate the interpretation of visual data such as images and videos collected from construction sites, effectively addressing labor shortages and reducing operational costs. Among these technologies, object detection plays a pivotal role due to its ability to identify and localize key elements including workers, equipment, and materials in real time.

Object detection has already been implemented across various aspects of the construction industry. In safety management, it is used to detect potential fall hazards (Fang *et al.* 2018a, b), monitor compliance with safety regulations (e.g., hardhat usage) (Fang *et al.* 2018c, Wu *et al.* 2019, Nath *et al.* 2020, Kim *et al.* 2023a), and evaluate collision risks by tracking the proximity between workers and heavy equipment (Kim *et al.* 2019, Son *et al.* 2019, Seong *et al.* 2023, Gan *et al.* 2024). For productivity analysis, object detection aids in estimating equipment utilization rates and productivity levels, as well as supporting cost analysis

through the identification of the operational status of heavy equipment (Yang *et al.* 2023, Zhang and Zhang 2022, Kim and Chi 2020, Jeong *et al.* 2023).

However, despite these advancements, the real-world deployment of deep learning-based object detection models on construction sites is still limited. A major obstacle lies in the site-specific variability inherent to construction domains. Many existing models are trained on datasets that are either limited in size or insufficient in diversity to capture the highly variable and dynamic nature of actual job sites (Xuehui *et al.* 2021, Xiao and Kang 2021, Duan *et al.* 2022, Kim *et al.* 2023b). Visual characteristics on construction sites can vary significantly depending on the project phase, background context, equipment type, time of day, and weather conditions. These variations often lead to a decline in model performance when pre-trained systems are applied to new or independent sites, thereby reducing the reliability of downstream tasks such as safety monitoring and productivity evaluation (Kim and Chi 2024, Kim *et al.* 2024a). This underscores the importance of fine-tuning models on site-specific data to ensure robust and accurate performance in practical applications.

To address this issue, prior research has explored various fine-tuning strategies, including zero-shot learning (Kim *et al.* 2024b), few-shot learning (Kim and Chi 2021), active learning (Kim *et al.* 2020), semi-supervised domain adaptation (Hong *et al.* 2023, Tran *et al.* 2025, Kim *et al.* 2024a), and domain-specific data augmentation (Bang *et al.* 2020). While these approaches improve site-specific model performance, they often require additional fine-tuning, which presents practical challenges related to training time, computational demands, and deployment feasibility. One such challenge is that fine-tuning in centralized

*Corresponding author, Ph.D., Professor,
E-mail: hjung@kaist.ac.kr

^a Ph.D.

^b Ph.D. Student

^c Ph.D.

infrastructures, such as cloud servers or control centers, requires transmitting image data from each construction site. This process can be hindered by limited network bandwidth, data privacy and security concerns. Moreover, training and maintaining multiple site-specific models impose significant burdens, both in terms of time investment and the storage requirements associated with large model weight files, particularly in edge or resource-constrained environments.

To overcome these limitations, this study proposes a novel fine-tuning strategy for object detection in the construction domain based on Low-Rank Adaptation (LoRA) (Hu *et al.* 2022), a parameter-efficient fine-tuning technique initially developed for large language models. In contrast to conventional full fine-tuning methods, LoRA substantially reduces the number of trainable parameters by introducing pairs of low-rank matrices into the pre-trained

layers, enabling efficient adaptation with minimal resource consumption. Its design focuses on lightweight and incremental adaptation, making it particularly suitable for site- and task-specific fine-tuning rather than large-scale pretraining with substantial domain shifts. While LoRA has been successfully applied in natural language processing (NLP) (Hu *et al.* 2022) and generative models such as Stable Diffusion (Rombach *et al.* 2022), its application to object detection tasks, particularly within the construction domain, remains largely unexplored.

The proposed method enables fast and lightweight fine-tuning of object detection models tailored to specific construction conditions. Since LoRA-modified weight files are significantly smaller than full model weights, multiple site-specific models can be stored and deployed on a single edge device. This is particularly advantageous in edge computing environments, where computational and storage

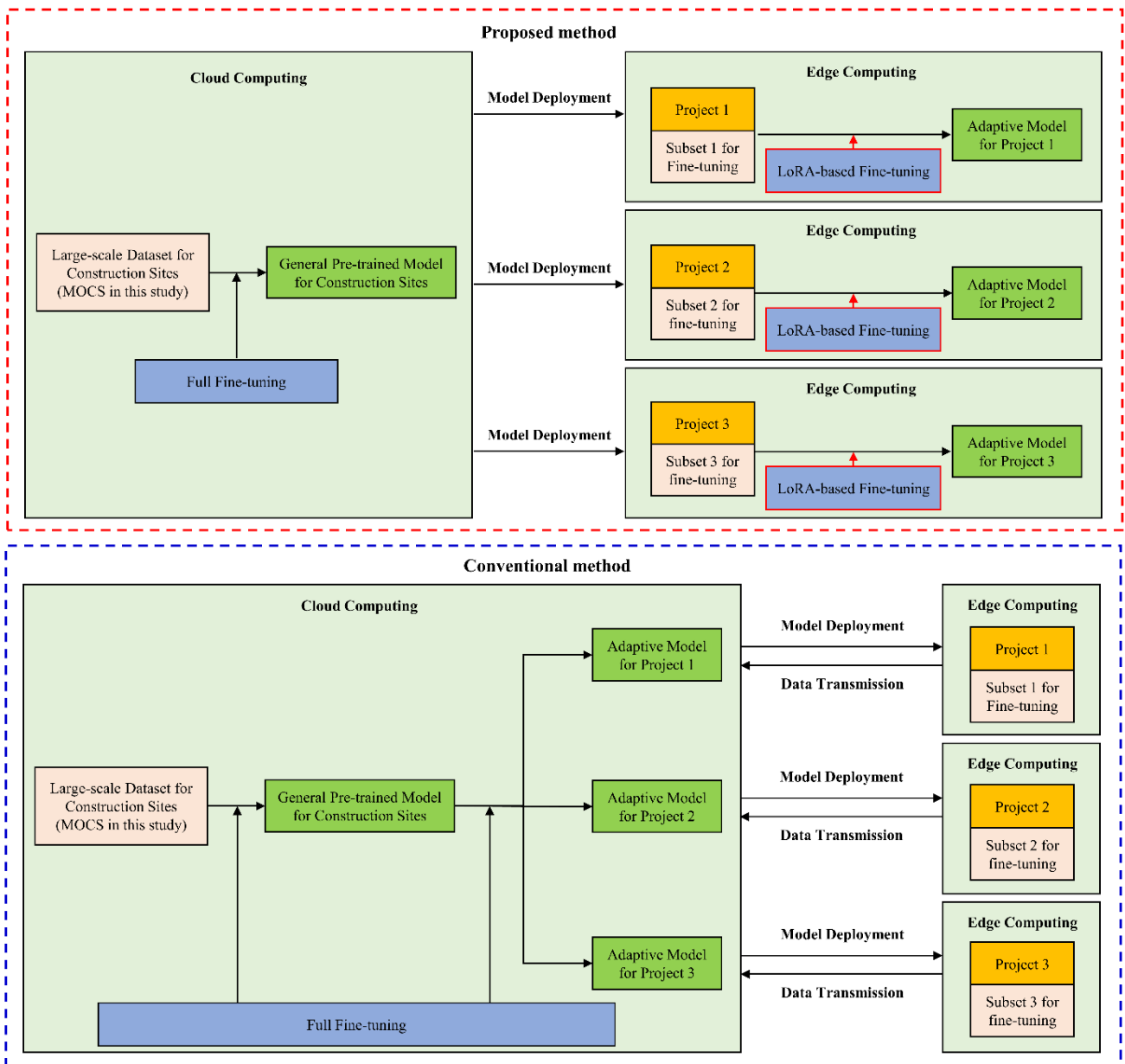


Fig. 1 Overall comparison between the proposed and conventional model adaptation for construction sites

resources are limited, and minimizing data transmission mitigates privacy and security concerns (See Fig. 1). Furthermore, the LoRA framework is compatible with other fine-tuning paradigms (e.g., few-shot and semi-supervised learning), making it a flexible tool for enhancing model adaptability.

The contributions of this study are as follows:

- 1) A LoRA-based fine-tuning framework is proposed for object detection models in construction environments, offering reduced training time and resource requirements compared to traditional methods.
- 2) The proposed method is compatible with other fine-tuning strategies such as few-shot learning and semi-supervised domain adaptation. This flexibility allows it to be integrated into existing pipelines to improve model performance on unseen site conditions.
- 3) The compact weights, low computational cost, and fast fine-tuning make the proposed method practical for on-device adaptation in edge computing environments.

The rest of this paper is organized as follows.

Section 2 reviews related work and establishes the necessity of a parameter-efficient fine-tuning approach in construction site object detection. Section 3 describes the proposed methodology, including the object detection architecture and the integration of LoRA into the model. Section 4 presents experimental results comparing the proposed LoRA-based fine-tuning with conventional full fine-tuning methods. Finally, Section 5 concludes the study by summarizing the main contributions, limitations, and future research directions.

2. Related works

2.1 Object detection in construction

The application of object detection in construction has become increasingly prevalent across a wide range of tasks, including safety management, productivity monitoring, and site operation optimization. Deep learning-based object detection models, particularly those using CNN or Transformer backbones, have enabled automated interpretation of site imagery, facilitating real-time understanding of complex and hazardous environments.

In safety-related applications, object detection has been used to identify workers, machinery, and personal protective equipment (PPE) compliance (Kim *et al.* 2023a). These models support real-time detection of fall hazards (Fang *et al.* 2018b), unsafe behaviors (Liu *et al.* 2021), and potential collisions between equipment and workers (Seong *et al.* 2023). For productivity analysis, object detection has helped monitor construction progress by tracking the activity and movement of heavy equipment such as excavators, cranes, and dump trucks, providing insight into task durations, equipment utilization rates, and idle times (Alaloul *et al.* 2022).

Beyond safety and productivity, object detection has been applied to other tasks including site access monitoring

(Wan *et al.* 2023), material tracking (Wang *et al.* 2021), and quality inspection (Lin *et al.* 2025). These applications demonstrate the wide applicability of object detection models in the construction domain.

However, despite promising outcomes in experimental or controlled testbeds, deploying these models in real-world construction sites remains a challenge. The high variability of site conditions, including changes in lighting, layout, background elements, and equipment types, makes it difficult to develop general pre-trained models that perform consistently across projects. As a result, pre-trained object detection models often exhibit poor generalization when applied to new sites.

To address this issue, recent studies have proposed fine-tuning strategies such as semi-supervised learning (Hong *et al.* 2023), few-shot learning (Kim and Chi 2021), and domain adaptation to train adaptive models to the specific sites (Kim *et al.* 2024a). These methods allow for model fine-tuning using a small set of site-specific images.

2.2 Fine-tuning for adaptive models in construction sites

To address the domain-specific performance degradation of object detection models in construction environments, various fine-tuning strategies have been proposed to adapt models to new site conditions. Kim and Chi (2021) proposed a few-shot learning method that enables models to rapidly adapt to new construction sites using only a limited number of labelled images from the target site. Their method demonstrated promising accuracy improvements in detecting site-specific objects. However, despite the reduced data requirement, they still demanded enough time and computational resources to cause significant delays between data acquisition and model deployment.

In parallel, efforts to reduce manual labelling labor have led to the development of semi-supervised learning techniques, which incorporate unlabelled data through methods such as pseudo-labelling and adversarial model architecture (Kim *et al.* 2022, 2024a). While these approaches alleviate the need for human annotation, they still involve iterative retraining and high computational cost. Such limitations underscore the need for lightweight fine-tuning strategies that are both time- and resource-efficient for real-world construction site applications.

Among such methods, LoRA stands out due to its computational efficiency and compatibility with other training strategies. LoRA can be integrated into complex training strategies such as semi-supervised learning or domain adaptation without significant architectural modifications. These characteristics make LoRA particularly suitable for construction scenarios, where new site conditions demand rapid and lightweight model updates.

In this study, a LoRA-based approach for adaptive object detection in construction environments was proposed. By minimizing training costs and enabling flexible model adaptation, LoRA offers a practical solution for deploying object detection models across diverse and evolving construction sites.

2.3 LoRA in object detection

While LoRA has demonstrated substantial performance gains and computational efficiency in fields such as NLP and image generation, its application in object detection remains relatively limited. This is largely due to the structural differences between model types: LoRA was originally designed for transformer architectures, where key components such as queries, keys, and values are implemented as linear projections, making them amenable to low-rank decomposition. In contrast, most object detection models are built on convolutional neural networks (CNNs), where the core operations involve convolutional kernels rather than linear layers.

To address this gap, Ding *et al.* (2024) introduced LoRA-C, a modified version of LoRA specifically adapted for CNN-based models. Their work explored which dimensions of convolutional kernels could be decomposed effectively without compromising the model structure. Although this represented one of the first serious attempts to bring LoRA into convolutional backbones, the overall performance improvements were modest, highlighting the inherent difficulty of applying LoRA to non-transformer architectures.

In another recent study, Talaoubrid *et al.* (2025) applied LoRA to DiffusionDet, a generative object detection framework based on diffusion models. Although conceptually interesting, DiffusionDet suffers from high computational complexity and slow inference, limiting its applicability to time-sensitive domains like construction monitoring. More promisingly, Pu and Xu (2025) explored LoRA in the context of Transformer-based object detectors, integrating LoRA into the attention layers and detection heads of a Swin Transformer-based architecture. Their results showed that LoRA can deliver performance comparable to full fine-tuning while using fewer parameters. However, the scope of their study was limited to satellite imagery, and primarily focused on architectural analysis rather than domain adaptability.

Building on this foundation, our study explores the application of LoRA to adaptive object detection in construction environments, focusing on its ability to generalize across diverse site conditions with minimal computational resources. Through this study, it has been demonstrated that the LoRA-based object detection fine-tuning approach effectively trains adaptive models by capturing cross-site variations in construction environments.

3. Methodology

3.1 Low-Rank Adaptation (LoRA)

Modern deep learning models, especially those used in vision and language tasks, often contain millions to billions of trainable parameters. With the rise of large-scale foundation models, the computational and storage demands for training and fine-tuning such models have increased dramatically. This has led to the development of parameter-efficient fine-tuning (PEFT) methods, which aim to optimize model performance on new tasks with minimal

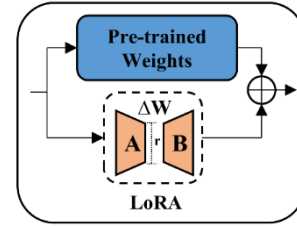


Fig. 2 The structure of LoRA module

resource consumption.

Traditional PEFT approaches include layer freezing, where only a subset of parameters (typically the final layers) are fine-tuned while the rest of the model is kept fixed. Another common method involves adapters, which insert lightweight trainable modules into a frozen pre-trained model. While these approaches offer efficiency benefits, they have notable limitations. Freezing early layers can limit model adaptability to new domains, and adapters may increase model complexity, require structural modifications, or introduce latency during inference.

LoRA takes a different approach by assuming a low-rank subspace for weight updates as shown in Fig. 2. The key idea is that during task-specific fine-tuning, the intrinsic rank of the required parameter updates is often much lower than the full rank of the weight matrices in the pre-trained model. LoRA exploits this by decomposing the weight updates into a pair of low-rank matrices, allowing the model to train only a small number of parameters during fine-tuning.

Formally, given a pre-trained weight matrix $W \in \mathbb{R}^{d \times k}$, LoRA introduces two low-rank matrices $A \in \mathbb{R}^{r \times k}$ and $B \in \mathbb{R}^{d \times r}$, such that the weight is updated as

$$W' = W + \Delta W = W + BA \quad (1)$$

Here, $r \ll \min(d, k)$, so the number of additional trainable parameters is drastically reduced. Importantly, the original weight matrix W is kept frozen, and only A and B are optimized during training.

LoRA offers several advantages that make it particularly suitable for practical deployment. It requires significantly fewer trainable parameters compared to the conventional full fine-tuning method. This results in reduced computational resource demands and shorter training times. Furthermore, the additional weight files generated through LoRA-based fine-tuning are considerably smaller in size than the original model weights, enabling storage of multiple task-specific weight files on a single device. During inference, these LoRA parameters are applied additively to the base weights, preserving the original model's pipeline and achieving the same inference speed without introducing bottlenecks.

3.2 Model architecture

To evaluate the performance of LoRA-based fine-tuning for object detection, a modified Faster R-CNN (Girshick 2015) architecture was adopted as a baseline detection model. Several improvements were applied to enable the

integration of LoRA.

Specifically, the model replaces the original RoI Pooling operation with RoI Align (He *et al.* 2017) to improve localization accuracy. For the backbone, Swin Transformer Tiny (Swin-T) (Liu *et al.* 2021), a vision Transformer known for its hierarchical attention structure and strong performance on detection tasks, is adopted. In particular, Swin-T was selected because its Transformer-based architecture is structurally compatible with parameter-efficient fine-tuning methods such as LoRA, whereas standardized and widely validated LoRA integration strategies for CNN-based backbones are still limited.

A Feature Pyramid Network (FPN) (Lin *et al.* 2017) was used as the neck component to support multi-scale feature extraction, enhancing the model’s ability to detect objects of varying sizes. The detection head of the proposed model consists of two stages. First, a Region Proposal Network (RPN) generates candidate object proposals, and RoI Align is applied to refine the spatial features extracted from these regions. Then, a classification and regression module, composed of fully connected linear layers, predicts the object categories and refines the bounding box coordinates.

In proposed LoRA-enhanced architecture, the Swin Transformer backbone was partially decomposed using LoRA. Specifically, LoRA modules were inserted only into the self-attention layers within the Swin blocks, while all other components were kept frozen to reduce the training cost. Additionally, LoRA decomposition was applied to the linear layers in the detection head, which constitute

a substantial portion of the trainable parameters in the base model. In contrast, the FPN and RPN consist primarily of convolution-based feature extraction layers, for which standardized and well-validated LoRA integration strategies are not yet established. Therefore, full fine-tuning was performed on the FPN, RPN, and the classification and regression blocks of the detection head to maintain overall model performance. The overall architecture with LoRA applied can be seen in Fig. 3.

In LoRA-based adaptation, the choice of rank is a critical factor, as it directly influences the model’s capacity while maintaining computational efficiency. The rank must be carefully balanced to avoid underfitting or excessive overhead. In this study, we follow the low-rank approximation-based selection strategy introduced by Pu and Xu (2025). Their method analyses the singular value distribution of weight matrices using Singular Value Decomposition (SVD) to assess the intrinsic dimensionality of each layer. Layers with slower singular value decay require higher ranks to avoid loss of representational capacity, while those with faster decay can be effectively approximated with smaller ranks.

Guided by this principle, a uniform rank of 48 was assigned to the LoRA modules applied to the Swin Transformer backbone, while the fully connected layers in the detection head (FC1 and FC2) were assigned ranks of 64 and 16, respectively. Table 1 presents the number of trainable parameters and compression ratios for each layer according to the specified rank values.

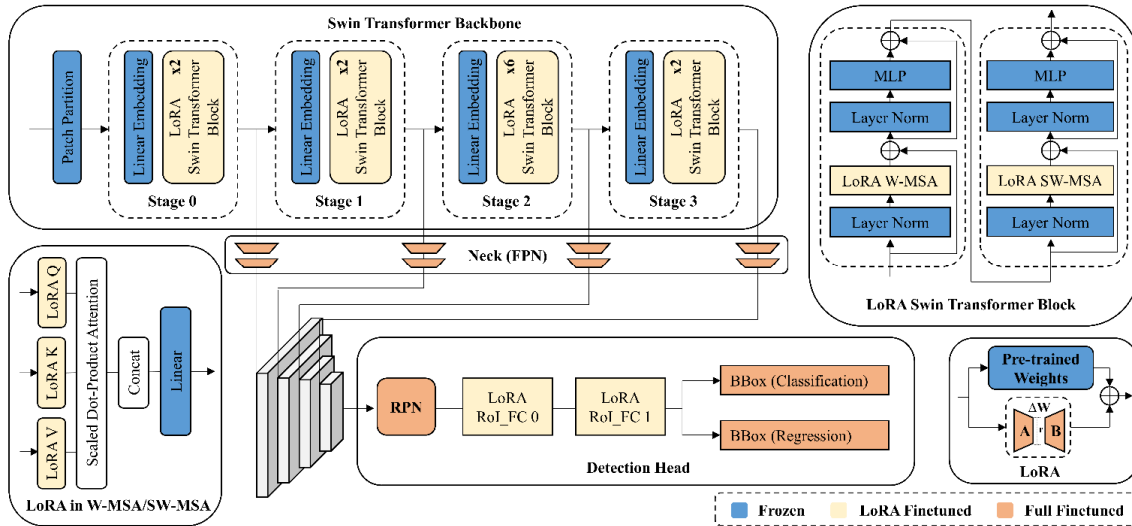


Fig. 3 Model architecture for LoRA-based fine-tuning

Table 1 Number of trainable parameters (K)

Methods	Backbone					Neck		Detection Head			Total
	Path partition	Stage 0	Stage 1	Stage 2	Stage 3	FPN	RPN	RoI_FC 0	RoI_FC 1	BBox	
Full fine-tune	8	299	1,188	11,842	14,184	2,730	594	12,846	1,050	11	44,752
LoRA	0	55	111	664	442	2,730	594	868	33	11	5,508
Low rank r	-	48	48	48	48	-	-	64	16	-	-
Compressed ratio p	0	0.184	0.093	0.056	0.031	1	1	0.068	0.031	1	0.123

In terms of parameter efficiency, conventional full fine-tuning involves approximately 44.75 million trainable parameters, requiring 171.02 MB of storage for the weight file. In contrast, the LoRA-based fine-tuning reduces the number of trainable parameters to just 5.51 million, which is 12.3% of the original model size, with a stored weight file size of only 21.04 MB. This substantial reduction in size allows for multiple fine-tuned model variants adapted to different conditions or tasks to be stored and accessed on a single device, which is particularly beneficial in construction implementation scenarios.

4. Experimental results

To evaluate the performance and applicability of the proposed LoRA-based fine-tuning method in real-world construction environments, a series of experiments was conducted. These experiments include training a general object detection model on a large-scale construction dataset, establishing site-specific testbeds, and comparing the effectiveness of LoRA against conventional full fine-tuning under resource and performance.

4.1 Experimental setup

4.1.1 General pre-training with MOCS datasets

To build a general object detection model for construction sites, a Faster R-CNN architecture was trained on the MOCS dataset (Xuehui *et al.* 2021), which comprises 41,668 images collected from 174 diverse construction sites. This dataset reflects a wide range of image acquisition conditions, equipment types, and environmental contexts. For consistency in evaluation and compatibility with site-specific datasets, the MOCS labels

were restructured into two high-level categories: ‘worker’ and ‘heavy equipment.’ All other object classes were excluded from these experiments.

A total of 19,404 images were used for training and 4,000 for validation. The model was initialized with weights pre-trained on the COCO dataset and fine-tuned for 12 epochs. Training was conducted using the AdamW optimizer with a learning rate of 0.0001, weight decay of 0.05, and β -values of (0.9, 0.999). A batch size of 4 was maintained throughout all training runs. This configuration served as the baseline for both full fine-tuning and LoRA-based adaptation experiments. The training and test results were obtained on a system equipped with an Intel i9-10980XE CPU, 256 GB of RAM, and an NVIDIA GEFORCE RTX 3090 GPU. For evaluation, mean Average Precision (mAP) at an IoU threshold of 0.5 (mAP@0.5) was used as the primary metric to assess detection performance.

Although the main objective of this study is to explore the use of LoRA for site-specific adaptation, its effectiveness during large-scale pretraining was also examined. LoRA modules were applied to the same model architecture described in Section 3.2, with all other training configurations held constant. However, the results (see Table 2) revealed that LoRA was not effective when used for full-scale pretraining on large-scale datasets with significant domain disparity. This is consistent with the core intuition of LoRA, which is designed for lightweight and incremental adaptation rather than addressing substantial domain shifts. Therefore, in scenarios involving large domain differences, full fine-tuning remains the more effective strategy for developing a general model in the construction domain. In contrast, for site-specific adaptation (see Table 3), LoRA achieves comparable performance to full fine-tuning, highlighting its suitability for efficient

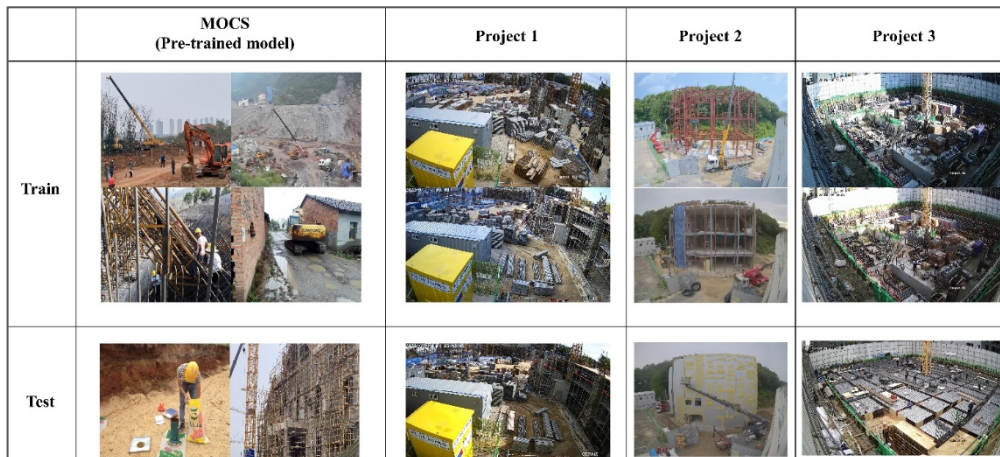


Fig. 4 Dataset configuration used for training and evaluation

Table 2 Comparison between full fine-tuning and LoRA on the MOCS dataset

Method	Dataset	Worker	Equipment	Time	Memory (MB)
Full fine-tune	MOCS	0.819	0.862	9h 37m 19s	12550
LoRA	MOCS	0.691	0.643	8h 44m 24s	9533

Table 3 Experiment results

Type	Dataset	Train				Test			Training time	Memory (MB)
		Banghwa		Kaist_animal		D2				
		Worker	Equipment	Worker	Equipment	Worker				
w/o Fine-tune	MOCS	0.819	0.862	0.689	0.236	0.606	0.485	0.769	9h 37m 19s	12550
	Banghwa	0.809	0.971	0.852	0.629				56m 58s	11339
Full fine-tune	Kaist_animal	0.834	0.952			0.82	0.941		23m 12s	9803
	D2	0.938						0.807	13m 19s	11340
LoRA	Banghwa	0.815	0.958	0.854	0.634				51m 16s	8423
	Kaist_animal	0.815	0.954			0.818	0.92		20m 29s	7290
	D2	0.929						0.833	11m 31s	8422
MOCS + subset full fine-tune	Banghwa	0.809	0.852	0.846	0.702				10h 46m 4s	12553
	Kaist_animal	0.805	0.85			0.775	0.919		10h 0m 53s	12559
	D2	0.813	0.849					0.797	9h 55m 27s	12553
MOCS + subset LoRA	Banghwa	0.812	0.863	0.837	0.643				9h 44m 15s	9527
	Kaist_animal	0.817	0.862			0.762	0.903		9h 5m 41s	9532
	D2	0.819	0.856					0.804	8h 56m 14s	9534

adaptation in resource-constrained environments.

4.1.2 Construction site testbeds

To assess the adaptability of the LoRA-based fine-tuning method, three site-specific testbed datasets were constructed. In order to reflect realistic deployment conditions, images were split according to the actual construction timeline. Early-phase images were used for training, while visually distinct later-phase images were allocated to the test set. Training data for each project was further divided into training and validation subsets, with 80% allocated for training and 20% for validation.

Project 1 (Banghwa), located in Gangseo-gu, Seoul, included a total of 2,860 images, which were split into 1,716 for training, 572 for validation, and 572 for testing.

Project 2 (Kaist_animal), a structural frame project in Yuseong-gu, Daejeon, yielded 1,390 images, divided into 800 training, 200 validation, and 390 test images.

Project 3 (D2), also in Gangseo-gu, Seoul, produced 1,181 images, with 439 used for training, 110 for validation, and 632 for testing. Since this site did not include heavy equipment, evaluation was limited to the ‘worker’ class.

Fig. 4 shows representative samples from the MOCS dataset and each of the project datasets. The MOCS dataset covers a broad range of scenarios and captures diverse conditions, but lacks the temporal and spatial consistency found in the project datasets. In contrast, the project datasets exhibit distinct site-specific characteristics, such as consistent camera angles and background elements, which make them well-suited for evaluating the effects of site-specific fine-tuning on model performance. The temporal separation between training and test sets also introduces a moderate level of domain shift, reflecting the progression of a real construction project.

4.2 Site-specific fine-tuning with LoRA

In this section, the effectiveness of the proposed LoRA-based fine-tuning method was evaluated in adapting a general pre-trained model to site-specific conditions. The pre-trained model, trained on the MOCS dataset as described in Section 4.1.1, was evaluated on the test sets on the three construction sites introduced in Section 4.1.2. As shown in Table 3, the pre-trained model achieved an mAP of 0.557 on the test sets, significantly lower than the fine-tuned models which achieved 0.810, highlighting the need for site-specific adaptation. This highlights the difficulty of generalizing across diverse construction environments and reinforces the need for site-specific fine-tuning.

The model was subsequently fine-tuned for each project using both the conventional full fine-tuning approach and the proposed LoRA-based method. LoRA was configured as described in Section 3, with low-rank matrices applied to the attention layers in the backbone and the linear layers in the detection head. All other training settings remained unchanged.

The results of these experiments, summarized in Table 3, demonstrate that the proposed method achieved an average mAP of 0.812, compared to 0.810 for full fine-tuning, corresponding to a 0.2% difference across the projects. Moreover, the proposed method offers significant computational advantages: it reduced training time by approximately 10% and GPU memory usage by about 30%. These improvements are crucial in scenarios where models must be fine-tuned on-site or deployed in edge computing environments, where time and memory are limited.

In addition to site-specific fine-tuning, this study also investigates the effects of incorporating the MOCS dataset alongside each project’s subset during the fine-tuning process. These experiments were designed to evaluate whether the performance of the LoRA-based fine-tuning method is more sensitive to the degree of domain shift than

to the size of the dataset itself.

The results indicate a slight decrease in performance when fine-tuning is conducted using both the project subset and the MOCS dataset, compared to using the subset alone. In addition, the increased dataset size led to longer training times. However, despite the extended training time, the LoRA-based method still demonstrated a notable reduction in overall training duration compared to full fine-tuning.

These findings suggest that when the visual characteristics of the target site differ substantially from those represented in the MOCS datasets, introducing additional general data may dilute site-specific feature representations rather than reinforce them. Conversely, when the adaptation data closely matches the deployment environment, LoRA fine-tuning based solely on site-specific images can achieve better performance.

5. Conclusions

In this study, a LoRA-based fine-tuning framework was proposed for adapting object detection models to the dynamic and site-specific environments of construction sites. As construction sites exhibit a high degree of variability across location, phase, and time, pre-trained general models often suffer from performance degradation when applied directly to new environments. To address this challenge, a parameter-efficient fine-tuning strategy was adopted, which significantly reduces training time and computational demands.

Our experiments demonstrated that while the proposed method may not be suitable for full-scale model pre-training on large and diverse datasets, it proves highly effective when applied to site-specific adaptation tasks. Compared to traditional full fine-tuning, the LoRA-based approach achieved comparable or superior detection accuracy with a 10% reduction in training time and a 30% decrease in GPU memory consumption. These gains are especially important in construction domains, where model deployment often occurs under strict resource and time constraints, such as during live projects or on edge devices.

Through comprehensive testing across three real-world construction site datasets, the practical effectiveness of the proposed method was validated in handling both subtle and significant domain shifts. The results confirm that LoRA is a viable solution for scalable, adaptive object detection in the construction industry.

Future work could explore two promising directions. First, the compatibility of LoRA-based fine-tuning with semi-supervised learning frameworks could be investigated to reduce annotation costs during deployment. This integration would enable more efficient adaptation in real-world scenarios where labeled data is limited and manual annotation is labor-intensive.

Second, the practicality of the proposed approach may be further validated by evaluating its adaptability under diverse conditions within a single construction site, such as varying weather scenarios and distinct construction phases (e.g., earthwork, structural frame, and finishing). These evaluations would help determine whether multiple LoRA-adapted models can be selectively activated in response to

varying site conditions within the same project.

Acknowledgments

This work was supported by a Korea Agency for Infrastructure Technology Advancement grant funded by the Ministry of Land, Infrastructure, and Transport (Grant RS-2020-KA156208).

References

- Alaloul, W.S., Alzubi, K.M., Malkawi, A.B., Al Salaheen, M. and Musarat, M.A. (2022), "Productivity monitoring in building construction projects: a systematic review", *Eng. Constr. Archit. Manag.*, **29**(7), 2760-2785.
<https://doi.org/10.1108/ECAM-03-2021-0211>
- Alsakka, F., Assaf, S., El-Chami, I. and Al-Hussein, M. (2023), "Computer vision applications in offsite construction", *Automat. Constr.*, **154**, p. 104980.
<https://doi.org/10.1016/j.autcon.2023.104980>
- Bang, S., Baek, F., Park, S., Kim, W. and Kim, H. (2020), "Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques", *Automat. Constr.*, **115**, p. 103198.
<https://doi.org/10.1016/j.autcon.2020.103198>
- Ding, C., Cao, X., Xie, J., Fan, L., Wang, S. and Lu, Z. (2024), "LoRA-C: Parameter-Efficient Fine-Tuning of Robust CNN for IoT Devices", arXiv preprint arXiv:2410.16954.
<https://doi.org/10.48550/arXiv.2410.16954>
- Duan, R., Deng, H., Tian, M., Deng, Y. and Lin, J. (2022), "SODA: A large-scale open site object detection dataset for deep learning in construction", *Automat. Constr.*, **142**, p. 104499.
<https://doi.org/10.1016/j.autcon.2022.104499>
- Ekanayake, B., Wong, J.K.W., Fini, A.A.F. and Smith, P. (2021), "Computer vision-based interior construction progress monitoring: A literature review and future research directions", *Automat. Constr.*, **127**, p. 103705.
<https://doi.org/10.1016/j.autcon.2021.103705>
- Fang, W., Ding, L., Luo, H. and Love, P.E. (2018a), "Falls from heights: A computer vision-based approach for safety harness detection", *Automat. Constr.*, **91**, 53-61.
<https://doi.org/10.1016/j.autcon.2018.02.018>
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H. and Li, C. (2018b), "Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment", *Automat. Constr.*, **93**, 148-164.
<https://doi.org/10.1016/j.autcon.2018.05.022>
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M. and An, W. (2018c), "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos", *Automat. Constr.*, **85**, 1-9. <https://doi.org/10.1016/j.autcon.2017.09.018>
- Fang, W., Ding, L., Love, P.E., Luo, H., Li, H., Peña-Mora, F., Zhong, B. and Zhou, C. (2020), "Computer vision applications in construction safety assurance", *Automat. Constr.*, **110**, p. 103013. <https://doi.org/10.1016/j.autcon.2019.103013>
- Gan, W., Gu, K., Geng, J., Qiu, C., Yang, R., Wang, H. and Hu, X. (2024), "A Novel Three-Stage Collision-Risk Pre-Warning Model for Construction Vehicles and Workers", *Buildings*, **14**(8), p. 2324. <https://doi.org/10.3390/buildings14082324>
- Girshick, R. (2015), "Fast R-CNN", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440-1448.
<https://doi.org/10.1109/ICCV.2015.169>
- He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017), "Mask R-CNN", *Proceedings of the IEEE international conference on*

- computer vision, pp. 2961-2969.
<https://doi.org/10.1109/tpami.2018.2844175>
- Hong, Y., Chern, W.C., Nguyen, T.V., Cai, H. and Kim, H. (2023), "Semi-supervised domain adaptation for segmentation models on different monitoring settings", *Automat. Constr.*, **149**, p. 104773. <https://doi.org/10.1016/j.autcon.2023.104773>
- Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W. (2022), "Lora: Low-rank adaptation of large language models", *ICLR*, **1**(2), p. 3.
<https://doi.org/10.48550/arXiv.2106.09685>
- Jeong, I., Hwang, J., Kim, J., Chi, S., Hwang, B.G. and Kim, J. (2023), "Vision-based productivity monitoring of tower crane operations during curtain wall installation using a database-free approach", *J. Comput. Civil Eng.*, **37**(4), p. 04023015.
<https://doi.org/10.1061/JCCEE5.CPENG-5105>
- Kim, J. and Chi, S. (2020), "Multi-camera vision-based productivity monitoring of earthmoving operations", *Automat. Constr.*, **112**, p. 103121.
<https://doi.org/10.1016/j.autcon.2020.103121>
- Kim, J. and Chi, S. (2021), "A few-shot learning approach for database-free vision-based monitoring on construction sites", *Automat. Constr.*, **124**, p. 103566.
<https://doi.org/10.1016/j.autcon.2021.103566>
- Kim, J. and Chi, S. (2024), "Investigating the Effects of Training Image Dataset's Size and Specificity on Visual Scene Understanding AI in Construction", *Land Hous. Rev.*, **15**(4), 1-9. <https://doi.org/10.5804/LHR.2024.15.4.1>
- Kim, D., Liu, M., Lee, S. and Kamat, V.R. (2019), "Remote proximity monitoring between mobile construction resources using camera-mounted UAVs", *Automat. Constr.*, **99**, 168-182.
<https://doi.org/10.1016/j.autcon.2018.12.014>
- Kim, J., Hwang, J., Chi, S. and Seo, J. (2020), "Towards database-free vision-based monitoring on construction sites: A deep active learning approach", *Automat. Constr.*, **120**, p. 103376.
<https://doi.org/10.1016/j.autcon.2020.103376>
- Kim, T., Shin, G., Kim, S. and Kim, H. (2022), "Generating Pseudo Label of Object Detector for Construction Site Monitoring", In: *International Conference on Computing in Civil and Building Engineering*, pp. 111-119.
https://doi.org/10.1007/978-3-031-35399-4_9
- Kim, H.S., Seong, J. and Jung, H.J. (2023a), "Real-time struck-by hazards detection system for small-and medium-sized construction sites based on computer vision using far-field surveillance videos", *J. Comput. Civil Eng.*, **37**(6), p. 04023028.
<https://doi.org/10.1061/JCCEE5.CPENG-5238>
- Kim, S., Hong, S.H., Kim, H., Lee, M. and Hwang, S. (2023b), "Small object detection (SOD) system for comprehensive construction site safety monitoring", *Automat. Constr.*, **156**, p. 105103. <https://doi.org/10.1016/j.autcon.2023.105103>
- Kim, H.S., Seong, J. and Jung, H.J. (2024a), "Optimal domain adaptive object detection with self-training and adversarial-based approach for construction site monitoring", *Automat. Constr.*, **158**, p. 105244.
<https://doi.org/10.1016/j.autcon.2023.105244>
- Kim, T., Koo, M., Hyeon, J. and Kim, H. (2024b), "Zero-shot Learning-based Polygon Mask Generation for Construction Objects", In: *ISARC, Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 41, pp. 81-88.
- Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017), "Feature pyramid networks for object detection", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117-2125.
<https://doi.org/10.1109/CVPR.2017.106>
- Lin, P.H., Lin, J.J. and Hsieh, S.H. (2025), "Construction Site Scaffolding Completeness Detection Based on Mask R-CNN and Hough Transform", arXiv preprint arXiv:2503.14716.
<https://doi.org/10.48550/arXiv.2503.14716>
- Liu, W., Meng, Q., Li, Z. and Hu, X. (2021), "Applications of computer vision in monitoring the unsafe behavior of construction workers: Current status and challenges", *Buildings*, **11**(9), p. 409. <https://doi.org/10.3390/buildings11090409>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021), "Swin transformer: Hierarchical vision transformer using shifted windows", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Nath, N.D., Behzadan, A.H. and Paal, S.G. (2020), "Deep learning for site safety: Real-time detection of personal protective equipment", *Automat. Constr.*, **112**, p. 103085.
<https://doi.org/10.1016/j.autcon.2020.103085>
- Pu, X. and Xu, F. (2025), "Low-rank adaption on transformer-based oriented object detector for satellite onboard processing of remote sensing images", *IEEE Transactions on Geoscience and Remote Sensing*.
<https://doi.org/10.48550/arXiv.2406.02385>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022), "High-resolution image synthesis with latent diffusion models", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684-10695.
<https://doi.org/10.48550/arXiv.2112.10752>
- Seong, J., Kim, H.S. and Jung, H.J. (2023), "The detection system for a danger state of a collision between construction equipment and workers using fixed CCTV on construction sites", *Sensors*, **23**(20), p. 8371. <https://doi.org/10.3390/s23208371>
- Son, H., Seong, H., Choi, H. and Kim, C. (2019), "Real-time vision-based warning system for prevention of collisions between workers and heavy equipment", *J. Comput. Civil Eng.*, **33**(5), p. 04019029.
[https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000845](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000845)
- Talaoubrid, H., Mokraoui, A., Ayed, I.B., Prouvost, A., Hang, S., Korn, M. and Harvey, R. (2025), "Analyzing the Impact of Low-Rank Adaptation for Cross-Domain Few-Shot Object Detection in Aerial Images", arXiv preprint arXiv:2504.06330.
<https://doi.org/10.48550/arXiv.2504.06330>
- Tran, D.Q., Jeon, Y., Aboah, A., Bak, J., Park, M. and Park, S. (2025), "Leveraging semisupervised learning for domain adaptation: Enhancing safety at construction sites through long-tailed object detection", *J. Constr. Eng. Manag.*, **151**(1), p. 04024190. <https://doi.org/10.1061/JCEMD4.COENG-15259>
- Wan, H.P., Zhang, W.J., Ge, H.B., Luo, Y. and Todd, M.D. (2023), "Improved vision-based method for detection of unauthorized intrusion by construction sites workers", *J. Constr. Eng. Manag.*, **149**(7), p. 04023040.
<https://doi.org/10.1061/JCEMD4.COENG-13294>
- Wu, J., Cai, N., Chen, W., Wang, H. and Wang, G. (2019), "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset", *Automat. Constr.*, **106**, p. 102894.
<https://doi.org/10.1016/j.autcon.2019.102894>
- Wang, Z., Zhang, Q., Yang, B., Wu, T., Lei, K., Zhang, B. and Fang, T. (2021), "Vision-based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase", *J. Comput. Civil Eng.*, **35**(1), p. 04020056.
[https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000933](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000933)
- Xiao, B. and Kang, S.C. (2021), "Development of an image data set of construction machines for deep learning object detection", *J. Comput. Civil Eng.*, **35**(2), p. 05020005.
[https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000945](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945)
- Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.M. and Wang, X. (2021), "Computer vision techniques in construction: a critical review", *Arch. Computat. Methods Eng.*, **28**(5).
<https://doi.org/10.1007/s11831-020-09504-3>

- Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L. and Zhiwei, L. (2021), "Dataset and benchmark for detecting moving objects in construction sites", *Automat. Constr.*, **122**, p. 103482. <https://doi.org/10.1016/j.autcon.2020.103482>
- Yang, J., Wilde, A., Menzel, K., Sheikh, M.Z. and Kuznetsov, B. (2023), "Computer vision for construction progress monitoring: A real-time object detection approach", In: *Working Conference on Virtual Enterprises*, pp. 660-672. https://doi.org/10.1007/978-3-031-42622-3_47
- Zhang, S. and Zhang, L. (2022), "Construction site safety monitoring and excavator activity analysis system", *Constr. Robot.*, **6**(2), 151-161. <https://doi.org/10.1007/s41693-022-00077-0>