CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0059

TRANSFORMING VISION-BASED INDOOR BUILT ENVIRONMENT MANAGEMENT WITH FEW-SHOT LEARNING

Gelare Taherian, Ehsan Rezazadeh Azar
Toronto Metropolitan University, Toronto, Canada

Abstract

Visual analysis of the built environment plays a critical role in the Architectural, Engineering, Construction, and Operation (AECO) sector. This type of analysis enables detection of visible discrepancies between the as-is and as-planned states of an asset to support proper decision making for different managerial and operational applications, such as construction progress monitoring and quality control. However, traditional approaches rely on manual inspections and expert knowledge, which can be time-consuming, error-prone, and labor-intensive. Therefore, automated detection of target objects within the visualizations is paramount. Recent advancements in computer vision and deep learning fields have enhanced automation of visual analysis, yet conventional deep learning approaches require large and well-annotated datasets, which may be challenging to develop. Alternatively, Few-Shot Learning (FSL) offers a promising solution by enabling the identification of regions of interest with minimal labeled data. This study investigates FSL, namely Prototypical Networks (PNs), to address the challenge of data scarcity in the indoor built environments, where the diversity of objects is high and data scarcity is a significant obstacle for effective training of conventional deep learning techniques. The method is evaluated on a custom dataset of AECO-specific indoor objects, achieving an average accuracy of 87.33% across three different tested folds of novel classes. The results indicate that the method enables the identification of unseen objects with rapid adaptation to new environments and tasks, while requiring only a limited number of labeled examples per class, thereby contributing to a significant reduction in time and effort required in real-world industry applications.

Keywords: Computer Vision, Few-shot Learning, Prototypical Networks, Object Detection, Indoor Built Environment Management.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

Construction jobsites are characterized by their dynamically changing environment, and this necessitates accurate and timely analysis of as-is against as-planned conditions [1] to support informed decision-making and resource allocation in construction projects [2]. Computer vision (CV)-based technologies have gained substantial attention in this field thanks to advances in data capturing, storage, and processing methods with lower cost and less time required [3]. Many research efforts proposed CV-based methods to detect physical changes in the jobsites for various purposes, such as construction progress monitoring [4], [5] and quality control [6], [7].

Recent advances in deep learning (DL)-based methods have enabled CV approaches to achieve high-level performances in various visual tasks, such as detection of object/regions of interest (ROIs) and classification [8]. Real-time progress monitoring has benefited from CV-based analysis, mostly by deep neural networks, providing information about workers' behaviour and equipment productivity as well as estimating work progress [9]. CV-based analysis has also been employed for post-construction analysis and management, where continuous monitoring is necessary to ensure the physical viability of the building components or condition assessment inspections [10], [11].

Despite these advances, DL-based methods face certain challenges to provide reliable results about the as-is state of built environments. For example, their performance heavily depends on the quantity and quality of the available data, which poses implementation challenges in environments where objects are unique and data is scarce, such as construction environments. Namely, indoor built environments

Corresponding author email address: gelare.taherian@torontomu.ca

(IBEs) contain diverse range of objects which often lack adequate representation in existing datasets [12]. Additionally, reliance on expert knowledge and high annotation costs limit the scalability of these methods. Therefore, other CV-based approaches, such as FSL, have been actively investigated for their potential to handle different ROIs with limited data [8]. Given the need for semantic identification of various unique ROIs for comparative analysis of the as-is and as-planned states of an IBE with limited data, FSL offers a promising solution by classification of novel classes using only a few labelled examples.

Since the research works on FSL applications in IBEs are limited [13], this paper investigates FSL-based techniques for classifying elements in IBEs based on datasets with limited samples. This approach can facilitate further semantic information extraction during and post construction phases with the final goal of detecting discrepancies in as-is and as-planned states. This approach can transform the vision-based analysis by reducing time and costs associated with data collection and labelling, thereby enhancing the practical usability of vision-based systems in IBEs with diverse classes. The remainder of the paper is structured as follows: Section 2 reviews relevant works on DL-based and FSL applications for managerial and operational tasks within the built environment management field. Section 3 describes the proposed workflow and utilized methods. Section 4 outlines the experiments and discusses their results. Finally, Section 5 concludes the study and suggests directions for future research.

2. Background

2.1. Deep learning methods for object classification in AECO

Several CV-based methods have been developed to extract information from images and videos of construction jobsites. Early approaches utilized traditional image processing algorithms that necessitated feature engineering [14], such as Histogram of Oriented Gradients and background subtraction techniques [15]. While initial studies on vision-based progress tracking and monitoring with image processing were improved by employing machine learning techniques like Support Vector Machines [16], DL-based methods significantly expanded the applications of CV systems by their capability of processing substantial amount of unstructured data. Object/ROI recognition is required in many construction management tasks, such as progress monitoring, safety management, and quality control of the works/material [17]. DL-based methods have demonstrated superior performance over traditional methods in identifying objects/ROIs through classification, detection, semantic, and instance segmentation techniques. Various tasks have benefited from DL-based approaches, such as detecting no-hardhat workers for safety monitoring [18] or classifying construction resources to enable real-time monitoring of a construction site's status [19]. Additionally DL-based methods have been implemented to automate job-type classification for organizing visual records [20] and to detect discrepancies between as-built and planned structures using U-Net-based segmentation [21]. However, these methods were majorly constrained by limited data and viewpoint dependency, which were partially mitigated through data augmentation techniques.

The background studies highlight that the accuracy of deep learning models depends on the quantity and quality of the training data, and the accuracy of their annotations [22], which all can be challenging to attain in the context of IBEs. This is due to the diversity and specificity of the dynamic construction environments with unseen classes where there is not enough publicly available data for training [19]. Additionally, DL-based methods may encounter hardware limitations, including increased detection and training times, as well as high computational costs, limited memory, and system freezing. While certain techniques, such as image resizing and compression during preprocessing, can alleviate these computational constraints, they may limit the overall accuracy of the models [23].

2.2. FSL methods in AECO

Novel CV-based techniques, including FSL, have emerged as promising solutions to overcome the challenge of data scarcity and are getting attention in the AECO field. A recent work [24] explored the use of FSL for façade defect detection and its performance improvement by extensible classifier and contrastive learning, demonstrating its potential to classify novel defect types with limited training data with about 82% accuracy. Pozzer et al. [25] extended the application of FSL to identify ROIs in multimodal images, including thermographic and visible images used for detecting subsurface damages in concrete structures, which typically require specialized expert knowledge. With the aim of reducing false positives in detected delamination, the rate of false positive detections reduced, and the mean precision increased by 3.8% with 500 pairs of images. FSL methods have shown promise for recognizing dynamic and temporary objects in construction environments. Built upon Yolov2 model, Kim et al. [26] achieved promising performance in vision-based monitoring of construction sites using an FSL model

with 1-30 images per class, with a mean Average Precision of 73%. Similarly, Liang et al. [9] introduced a tailored FSL approach that utilized a multi-modal prototype technique to effectively classify temporary objects on construction sites with limited data.

Semantic identification of safety issues in construction jobsites is another field that has benefited from FSL methods. While conventional CV-based methods face implementation challenges in identifying diverse hazard scenarios, FSL techniques have shown promise in detection with imbalanced and limited data distributions. Wang et al. [27] proposed an FSL object detection method to deal with imbalanced distribution of objects and further developing an attribute recognition enabling semantic understanding of the safety measurements and regions of fall protection. Using a limited data size of 1098 images, they achieved 51.8%, 88.2% average precision on 10 shots and 50 shots, respectively.

2.3. Knowledge gap

DL methods require a significant amount of training data and are specifically used to classify unseen images of the same classes that were used during the training process. While FSL methods have been applied to address the challenges of limited data availability, complex data labelling, and imbalanced object distributions in previous AECO studies, their applications have mostly been limited to specific domains, such as safety monitoring or material defect detection. There is a lack of research focusing on using FSL methods to identify building elements to facilitate automated monitoring in indoor built environments during and post construction. This represents a gap in the literature, as the ability to classify diverse building elements with limited training data is beneficial for streamlining the discrepancy detection process in IBEs. Therefore, this research investigates the performance of the FSL approach and the extent of accuracy to classify the objects of interest with limited data to contribute toward comparative analysis of as-is and as-planned states.

3. Proposed method

Given the diversity and uneven distribution of object classes in IBEs, particularly pertinent to the construction monitoring and quality control applications, the proposed system implements FSL to classify the objects/ROIs with limited data available. Popular FSL methods include metric-based, optimization-based, and model-based techniques [28]. Considering the pros and cons of each approach as outlined by [28], this study initiated with metric-based methods, as they are easier to train and computationally efficient. Specifically, Prototypical Networks (PNs) [29] were selected due to their outperformance on general image classification tasks compared to other metric-based methods [30]. This research investigates the applicability and effectiveness of PNs for classifying AECO-specific objects within IBEs. Although future studies could explore alternative FSL approaches to further understand and expand application potentials in this domain.

3.1. Definition of Few-shot classification task and dataset

Meta-learning, often described as "learning to learn," equips a model with the ability to adapt rapidly to new tasks by using prior knowledge, optimizing a set of parameters for performance across episodes or tasks. In the context of FSL, this distribution comprises N-way K-shot tasks, where N represents the number of classes and K the number of support examples per class. Metric-learning complements this process by focusing on developing a robust sense of "distance" or "similarity" between examples, enabling the model to identify patterns and distinctions among new examples. Combining meta-learning with metric learning—referred to as "meta-metric learning" [28] —further empowers the model to classify novel instances during testing without altering its underlying parameters.

Given a dataset D = $\{(x_1, y_1), \ldots, (x_i, y_i)\}$, where x_i represents the features of an example, and y_i denotes the corresponding class label, the classes are divided into training set N_{train} and testing set N_{test} used during the meta-training and meta-testing phases, respectively. Each learning episode comprises a support set and a query set, both constructed from the same N classes. The support set consists of K labelled examples per class and serves as a reference for the model to learn class-specific features. The query set, containing additional unlabelled examples from the same classes, is used to assess the model's ability to classify based on the learned similarity space.

During meta-training, the model engages in tasks designed to refine its ability to measure similarities, positioning examples from the same class closer in feature space while separating those from different classes. This learned metric serves as the foundation for classification during meta-testing, where the model is presented with entirely new classes from N_{test}. In this phase, the model's generalization capability is evaluated by its performance on classifying unseen examples using only a limited number of support instances, without prior exposure.

3.2. Few-shot Learning model

PNs aim to learn a prototype representation for each class in each episode and use these prototypes to classify query samples based on their proximity to these representations in an embedding space. The definition of FSL task based on meta-metric learning setup and the dataset in this study are illustrated as an instance of 3-way 2-shot task in Figure 1**Hiba! A hivatkozási forrás nem található.**. In this instance, there are 3 classes with two supporting samples in each episode.

Figure 1: The definition of FSL and distribution of episodes for PNs in meta-metric learning; 3-way 2-shot task

During meta-training, PNs learn a robust embedding function that can map examples from different classes into an embedding space. This process involves training the model across many few-shot tasks, where each task has its own support and query sets. For each task, the model processes the support set examples to produce embeddings through a feature embedding function f_θ, resulting in the mean embedding for each class called the "prototype" P_c, computed as:

$$P_c = \frac{1}{K} \sum_{(x_s, y_s) \in S_c} f_\theta(x_s) \tag{1}$$

where S_c is the support set for class C. This prototype serves as a representative point for each class *C* within the task. Over many tasks, the embedding function is optimized to produce prototypes that are distinctive and separable, allowing the model to generalize better when encountering new classes.

In the metric learning phase, PNs perform similarity measurement between prototypes (representing each class in the support set) and the embeddings of query samples. A distance metric is calculated between the embedding of each query sample x_q and each class prototype P_c, using a Euclidean distance metric, as has been shown to significantly enhance performance in PNs [29]. The query sample is then classified based on the nearest prototype, enabling fast classification. The distance d_θ between the embedded query $f_\theta(x_q)$ and the P_c is computed by:

$$d(f_\theta(x_q), P_c) = \left\| f_\theta(x_q) - P_c \right\|^2 \tag{2}$$

For each query sample x_q in class C, the probability of query sample belonging to class C is computed using the SoftMax function applied to the negative distance:

$$P(y = C | x_q) = \frac{exp\left(-d(f_{\theta_1}(x_q), P_c)\right)}{\sum_{c'} exp\left(-d(f_{\theta_1}(x_q), P_c)\right)} \tag{3}$$

The total loss for the episode is computed by maximizing $P(y = C|x_q)$ for matching class pairs and minimizing it for non-matching pairs, using the negative log-likelihood:

$$L = -\sum_{(x_q, y_q) \in Q} log\, P(y = y_q | x_q) \tag{4}$$

During testing, PNs sample episodes with classes from the test set N_{test} and the model's performance is assessed based on its classification accuracy on N-way K-shot tasks. The model relies on computing a prototype for each class in the embedding space based on the support set and query samples are classified by determining their proximity to these prototypes using the distance metric. The final output

is the average classification accuracy across all episodes, which serves as an indicator of the model's generalization to novel classes.

4. Experimental results and discussion

4.1. Dataset development

The dataset used in this study consists of indoor building elements collected from various sources pertinent to AECO industry. Real-world images were sourced from under-construction and post-construction sites captured by the authors and from publicly available resources such as Flickr and Google. This dataset contained 25 main object classes commonly found in IBEs, each with a varying number of instances, as provided in Figure 2. To make the sample images compatible with the FSL model, they underwent transformations including resizing to a standard input size (i.e., 224x224 pixels), colour jittering, and normalizing the pixel values. This preprocessing ensured consistency in the image dimensions and format, enabling the FSL model to work optimally with the inputs.

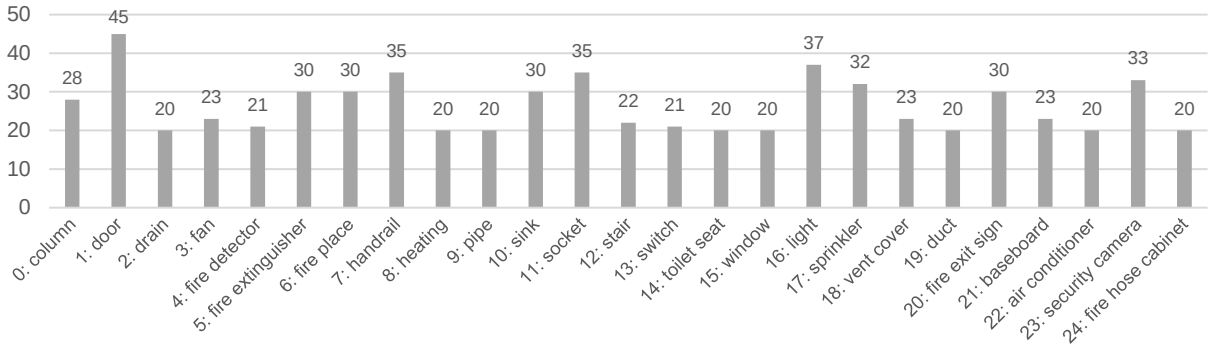

Figure 2: Distribution of images within the classes in the dataset

4.2. Implementation details

According to the theory of FSL, the model performance is related to the generation of episodes. Three different folds of base and novel classes were implemented, as performance may vary depending on novel class selection [27]. The 25 object classes were randomly divided into 20 base classes for training and 5 novel classes for testing in each fold. Across the three folds, a total of 15 classes were designated as novel classes, with no overlaps between the folds. The novel classes for Fold-1 were classes 0, 2, 9, 17, and 21; for Fold-2, the novel classes were 1, 11, 13, 18, and 20; and for Fold-3, the novel classes included 3, 10, 14, 15, and 23.

The integration of PNs with ResNet-18 has shown strong few-shot classification performance [31]. ResNet-18's lightweight architecture offers an optimal trade-off between computational efficiency and classification accuracy, making it well-suited for this study. Given that pre-trained deep neural networks for feature extraction improve the classification performance [30], this study investigated FSL with ResNet-18 as the feature extractor pre-trained on ImageNet [32]. The hyperparameters were tuned with respect to the easyFSL [33], a comprehensive open-source FSL library. The model was trained over 80 epochs and the SGD optimizer with the adaptive learning rate scheduling was used. The training process was implemented using the PyTorch library and was executed on Google Colab using an A100 GPU.

4.3. Few-shot performance on IBE classification

The performance of the proposed approach was evaluated using the standard classification accuracy metric on the test set with a 5-way 5-shot task. The classification accuracy is defined as the ratio of correctly classified query samples to the total number of query samples evaluated during the testing process, and the results of these three folds are presented in Figure 3. The overall accuracy of the PNs in the 5-way 5-shot scenario averages at 87.33%, indicating a promising baseline performance in handling few-shot learning tasks across various object classes. This metric highlights the model's general capability to understand and categorize new objects based on a limited set of examples, highlighting its potential utility in environments where quick adaptation to new data is essential. The accuracy variations across folds—with the highest in Fold-3 and the lowest in Fold-2— underscore the model's response to different sets of class combinations. The highest accuracy observed in Fold-3 at 89.55% suggests that the combination or type of classes in this fold may have been inherently more distinct or easier for the model to differentiate, potentially due to clearer or more characteristic features.

On the other hand, the lower accuracy of 84.47% in Fold-2 indicates a more challenging set of classes. This could be due to factors such as the presence of classes with similar features, less distinctive objects, or a combination of classes that do not contrast as sharply in the feature space defined by the model.

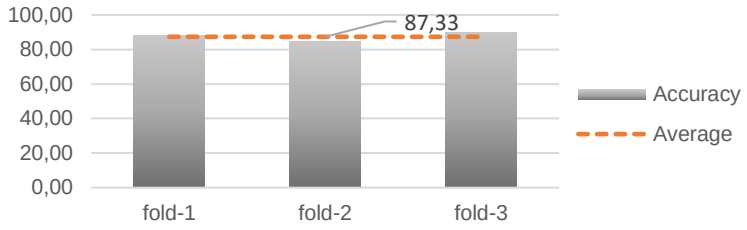

Figure 3: Classification accuracies of PNs on novel classes over the three different folds

The inherent challenges in each fold prompted further investigation. Confusion matrices were generated for the three data folds under the 5-way 5-shot experimental setup to further analyze the performance of PNs in classifying various AECO-specific objects in IBEs. Figure 4, provides a detailed view of the model's classification behavior by depicting how often each class was correctly or incorrectly predicted across the folds. Notably, Window (Class 15), Fire Exit Sign (Class 20), and Baseboard (Class 21) achieved the highest correct classification rates in their respective folds—96.24%, 98.46%, and 96.84%—indicating their strong visual separability. The confusion matrices show that these objects, likely due to their distinctive geometric features or unique positioning within the environment, were consistently recognized by the model with minimal misclassification. In contrast, certain classes such as Socket (Class 11) and Switch (Class 13) experienced higher rates of confusion, particularly in Fold-2, which indicate challenges in distinguishing between objects with visually similar features. Additionally, misclassification between Pipe (Class 9) and Column (Class 0) further underscores the limitations of relying solely on visual features, especially when dealing with elements that share similar shapes, textures, or spatial configurations. Figure 5 illustrates some instances with possible sources of failures. These findings suggest that, while PNs performance may decline when faced with visually ambiguous or structurally overlapping classes, incorporating contextual information or multimodal data may therefore be necessary to improve differentiation in such cases. However, PNs still demonstrated robust performance by achieving 87% accuracy matching or exceeding prior FSL studies reporting 62–80% accuracy in certain tasks such as temporary object classification on construction sites using few images per class [9] or façade defect classification [24]. Additionally, given that conventional deep learning models often require large and well-annotated datasets to surpass 64% accuracy in similar indoor settings [34], the results of this study underscore the effectiveness of PNs in delivering high performance with minimal data.

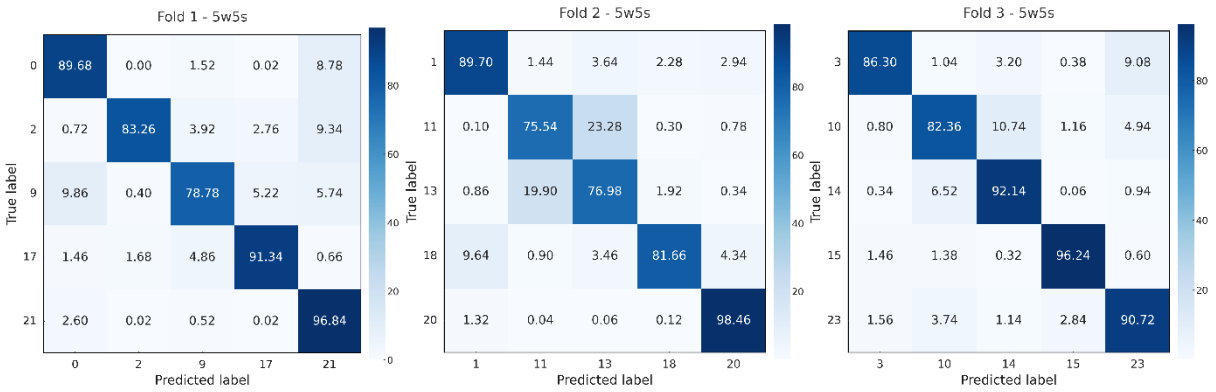

Figure 4: Confusion matrixes of PNs over the three folds from left to right; Fold-1, Fold-2, and Fold-3

Figure 5: Sample instances with possible influence on misclassifications

5. Conclusion and future work

This study implemented a few-shot classification approach using PNs within a meta-metric learning framework to classify AECO-specific object classes in IBEs, where data scarcity and class imbalance present significant challenges. Three different folds of novel classes were evaluated using a 5-way 5-shot configuration, with PNs achieving an average accuracy of 87.33%. This result demonstrates the model's potential for accurate classification using limited labelled data, offering a practical alternative to traditional deep learning methods that heavily rely on large-scale annotated datasets. The observed variability in accuracy across folds highlights both the strengths of the model in recognizing distinctive object classes and the need for improvement in handling visually ambiguous or visually similar instances. Given these findings, the proposed system is well-suited as a decision-support tool within broader construction management workflows. For tasks such as progress monitoring or quality controls, the model can automate object recognition and facilitate detection of discrepancies, thereby reducing manual effort and enabling timely decision-making. However, its direct application in safety-critical contexts, such as hazard identification or regulatory compliance, would require further enhancement, as such tasks demand extremely high accuracy with minimal tolerance for error. Enhancing the model's feature extraction capabilities, through more advanced feature extractor CNNs are worth investigating to check the improvement in recognition accuracy for classes with lower performance. Additionally, PNs performance on increased number of support images per class, though not as high as needed for classical DL-based models, could be investigated. While PNs was chosen as a metric-based FSL method for its computational efficiency, ease of training, and superior performance over other metric-based approaches, investigating optimization-based and model-based alternatives may further enhance adaptability in AECO contexts. A broader evaluation of FSL strategies could yield deeper insights into their suitability for diverse IBE tasks.

References

[1] S. Rankohi and L. Waugh, "Image-based modeling approaches for projects status comparison," in Proceedings Annual Conference Canadian Society for Civil Engineering, 2015.The referenced item does not yet have a DOI number.

[2] H. Deng, H. Hong, D. Luo, Y. Deng, and C. Su, "Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision," J Constr Eng Manag, vol. 146, no. 1, 2020,https://doi.org/10.1061/(ASCE)CO.1943-7862.0001744.

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

[4] H. Hamledari, B. McCabe, and S. Davari, "Automated computer vision-based detection of components of under-construction indoor partitions," Autom Constr, vol. 74, 2017,https://doi.org/10.1016/j.autcon.2016.11.009.

[5] B. Ekanayake, A. Ahmadian Fard Fini, J. K. W. Wong, and P. Smith, "A deep learning-based approach to facilitate the as-built state recognition of indoor construction works," Construction Innovation, 2022,https://doi.org/10.1108/CI-05-2022-0121.

[6] S. Vincke and M. Vergauwen, "Vision based metric for quality control by comparing built reality to BIM," Autom Constr, vol. 144, p. 104581, Dec. 2022,https://doi.org/10.1016/j.autcon.2022.104581.

[7] S. H. Hsu, H. T. Hung, Y. Q. Lin, and C. M. Chang, "Defect inspection of indoor components in buildings using deep learning object detection and augmented reality," Earthquake Engineering and Engineering Vibration, vol. 22, no. 1, 2023,https://doi.org/10.1007/s11803-023-2152-5.

[8] A. Nakamura and T. Harada, "Revisiting fine-tuning for few-shot learning," arXiv preprint arXiv:1910.00216, 2019,https://doi.org/10.48550/arXiv.1910.00216.

[9] Y. Liang, P. Vadakkepat, D. K. H. Chua, S. Wang, Z. Li, and S. Zhang, "Recognizing temporary construction site objects using CLIP-based few-shot learning and multi-modal prototypes," Autom Constr, vol. 165, p. 105542, Sep. 2024,https://doi.org/10.1016/j.autcon.2024.105542.

[10] T. Dawood, Z. Zhu, and T. Zayed, "Computer Vision–Based Model for Moisture Marks Detection and Recognition in Subway Networks," Journal of Computing in Civil Engineering, vol. 32, no. 2, 2018,https://doi.org/10.1061/(asce)cp.1943-5487.0000728.

[11] Z. Liu, Y. Cao, Y. Wang, and W. Wang, "Computer vision-based concrete crack detection using U-net fully convolutional networks," Autom Constr, vol. 104, 2019,https://doi.org/10.1016/j.autcon.2019.04.005.

[12] B. Zhong et al., “Mapping computer vision research in construction: Developments, knowledge gaps and implications for research,” Autom Constr, vol. 107, 2019,https://doi.org/10.1016/j.autcon.2019.102919.
[13] H. Hamledari and B. McCabe, “Automated Visual Recognition of Indoor Project-Related Objects: Challenges and Solutions,” in Proceedings of the Construction Research Congress, 2016.https://doi.org/10.1061/9780784479827.256.
[14] A. Pal, J. J. Lin, S. H. Hsieh, and M. Golparvar-Fard, “Automated vision-based construction progress monitoring in built environment through digital twin,” 2023.https://doi.org/10.1016/j.dibe.2023.100247.
[15] Z. Sun, E. Caetano, S. Pereira, and C. Moutinho, “Employing histogram of oriented gradient to enhance concrete crack detection performance with classification algorithm and Bayesian optimization,” Eng Fail Anal, vol. 150, 2023,https://doi.org/10.1016/j.engfailanal.2023.107351.
[16] N. D. Hoang, “Image Processing-Based Recognition of Wall Defects Using Machine Learning Approaches and Steerable Filters,” Comput Intell Neurosci, 2018,https://doi.org/10.1155/2018/7913952.
[17] K. Bacharidis, F. Sarri, and L. Ragia, “3D building façade reconstruction using deep learning,” ISPRS Int J Geoinf, vol. 9, no. 5, 2020,https://doi.org/10.3390/ijgi9050322.
[18] Q. Fang et al., “Detecting non-hardhat-use by a deep learning method from far-field surveillance videos,” Autom Constr, vol. 85, 2018,https://doi.org/10.1016/j.autcon.2017.09.018.
[19] H. Kim, H. Kim, Y. W. Hong, and H. Byun, “Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning,” Journal of Computing in Civil Engineering, vol. 32, no. 2, 2018,https://doi.org/10.1061/(asce)cp.1943-5487.0000731.
[20] D. Gil, G. Lee, and K. Jeon, “Classification of images from construction sites using a deep-learning algorithm,” in 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things, 2018.https://doi.org/10.22260/isarc2018/0024.
[21] J. Bae and S. U. Han, “Segmentation Approach to Detection of Discrepancy between As-Built and As-Planned Structure Images on a Construction Site,” in Computing in Civil Engineering: Data, Sensing, and Analytics, 2019.https://doi.org/10.1061/9780784482438.023.
[22] R. Khallaf and M. Khallaf, “Classification and analysis of deep learning applications in construction: A systematic literature review,” Autom Constr, vol. 129, 2021,https://doi.org/10.1016/j.autcon.2021.103760.
[23] Y. Hao et al., “Understanding the Impact of Image Quality and Distance of Objects to Object Detection Performance,” in IEEE International Conference on Intelligent Robots and Systems, 2023.https://doi.org/10.1109/IROS55552.2023.10342139.
[24] Z. Cui, Q. Wang, J. Guo, and N. Lu, “Few-shot classification of façade defects based on extensible classifier and contrastive learning,” Autom Constr, vol. 141, p. 104381, Sep. 2022,https://doi.org/10.1016/j.autcon.2022.104381.
[25] S. Pozzer et al., “A few-shot learning approach for the segmentation of subsurface defects in thermography images of concrete structures,” in Thermosense: Thermal Infrared Applications XLVI, SPIE, 2024.https://doi.org/10.1117/12.3013684.
[26] J. Kim and S. Chi, “A few-shot learning approach for database-free vision-based monitoring on construction sites,” Autom Constr, vol. 124, 2021,https://doi.org/10.1016/j.autcon.2021.103566.
[27] X. Wang and N. El-Gohary, “Few-shot object detection and attribute recognition from construction site images for improved field compliance,” Autom Constr, vol. 167, p. 105539, Nov. 2024,https://doi.org/10.1016/j.autcon.2024.105539.
[28] A. Parnami and M. Lee, “Learning from few examples: A summary of approaches to few-shot learning,” arXiv preprint arXiv:2203.04291, 2022,https://doi.org/10.48550/arXiv.2203.04291.
[29] J. Snell, K. Swersky, and R. Zemel, “Prototypical networks for few-shot learning,” in Advances in Neural Information Processing Systems, 2017.https://doi.org/10.48550/arXiv.1703.05175.
[30] S. Yang, F. Liu, N. Dong, and J. Wu, “Comparative Analysis on Classical Meta-Metric Models for Few-Shot Learning,” IEEE Access, vol. 8, 2020,https://doi.org/10.1109/ACCESS.2020.3008684.
[31] Dr. S. M. Kulkarni, S. S. Pawar, A. A. Dekhane, and S. L. Suryawanshi, “Enhancing Image Classification Using Few-Shot Learning Prototypical Networks with ResNet-18: Detection, Accuracy Enhancement, and Optimization,” International Journal of Scientific Research in Engineering and Management, vol. 08, no. 07, pp. 1–10, Jul. 2024,https://doi.org/10.55041/IJSREM36755.
[32] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2009.https://doi.org/10.1109/cvprw.2009.5206848.
[33] E. Bennequin, “easyfsl; Easy Few-shot Learning.” Accessed: Apr. 17, 2024. [Online]. Available: https://github.com/sicara/easy-few-shot-learning

[34] A. Nagarajan and G. M P, “Hybrid Optimization-Enabled Deep Learning for Indoor Object Detection and Distance Estimation to Assist Visually Impaired Persons,” Advances in Engineering Software, vol. 176, p. 103362, Feb. 2023,https://doi.org/10.1016/J.ADVENGSOFT.2022.103362.