

DFBI 2026 aims to encourage the international exchange of innovative ideas between researchers from academia and industry. In addition to knowledge dissemination, the conference offers a valuable platform for professional networking, particularly benefiting university professors, graduate students, and postdoctoral researchers.

Research Article/ Review Article/ Perspective Article (Remove where relevant)

Smartphone-Based Indoor Localization and Object Detection for Construction Site Asset Management and Digital Twins

Bikash Lamsal¹, Masato Higo¹, Matteo Sardellitti², Bimal Kumar KC², Naofumi Matsumoto²

¹ Waseda University, Graduate School of Creative Science and Engineering, Tokyo, Japan

² NAIT Corporation, Tochigi, Japan

Correspondence: bikash@aoni.waseda.jp

Copyright: Copyright: © 2026 by the authors.

DFBI is an open-access proceedings distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). View this license's legal deed at <https://creativecommons.org/licenses/by/4.0/>



Abstract

This research addresses the challenge of keeping building asset information accurate and up to date for facility management and digital twin applications, especially in indoor environments where GPS signals are not available. The asset management is still largely carried out manually, making it time-consuming, difficult to update, and susceptible to information fragmentation. These limitations reduce on-site productivity and limit the practical use of digital twins throughout the building lifecycle. One of the main technical challenges in indoor environments is determining the exact location where images are captured without GPS. To address this issue, we introduce a smartphone-based indoor positioning system that uses the smartphone's built-in sensors and does not rely on expensive equipment such as beacons or LiDAR. The system can determine indoor locations with centimeter-level accuracy. This high level of accuracy allows photos to be placed correctly on indoor floor plans, making it easier to understand the current condition of buildings and the locations of newly installed assets. For asset identification, this study uses an image-based recognition approach based on a YOLOv8 object detection model. The model is trained on a mixed dataset that includes manually labeled images collected from real construction sites, along with selected asset categories from the COCO dataset. Once assets are detected, they are linked to their exact indoor locations and visualized within a digital twin environment. By combining accurate indoor localization with automated asset detection, the proposed approach can simplify asset surveys, reduce manual effort, and significantly improve productivity in construction and facility management.

Keywords: Asset management, Computer vision, Digital twin, Indoor Localization, Object detection

Highlights

- Solves indoor GPS limitations using smartphone-based localization.
- Reduces manual asset tracking with AI-powered object detection.
- Prevents fragmented data by linking photos to floor plans.

Cite this article:
Bikash Lamsal, Masato Higo, Matteo Sardellitti, Bimal Kumar KC, Naofumi Matsumoto (2026). Smartphone-Based Indoor Localization and Object Detection for Construction Site Asset Management and Digital Twins, 2026 (2),
<https://doi.org/10.66408/dfbi.2026.26329>

1 Introduction

The construction industry is undergoing rapid digital transformation (DX) [1], resulting in the generation of large volumes of digital data throughout project lifecycles. These data have strong potential for improving building operation and maintenance after project completion [2]. At the same time, many countries, including Japan, are experiencing a shortage of skilled construction workers. The declining number of construction workers highlights the need for automation and efficient information management strategies that can reduce manual workload and improve productivity.

Among the various types of data generated on construction sites, photographic records are one of the most frequently produced resources [3]. Images are captured for progress monitoring, inspection documentation, and safety management. However, these photographs are typically archived as independent visual records and are rarely reused as structured information for facility management. As a result, equipment installation locations and specifications are often difficult to retrieve during building operation and maintenance stages.

Reliable spatial positioning is essential for organizing construction photographs. While smartphones automatically attach GPS information to images, indoor positioning accuracy is significantly degraded due to signal obstruction [4]. To address this limitation, indoor self-localization technologies such as the Nav & Snap application have been developed. In this workflow, building floor plans are first converted into spatial datasets using the Map Creator tool [5], which enables visualization of shooting positions directly on floor plans, as illustrated in Figure 1. By embedding positional metadata into captured images, construction photographs can be systematically organized and retrieved based on spatial context.

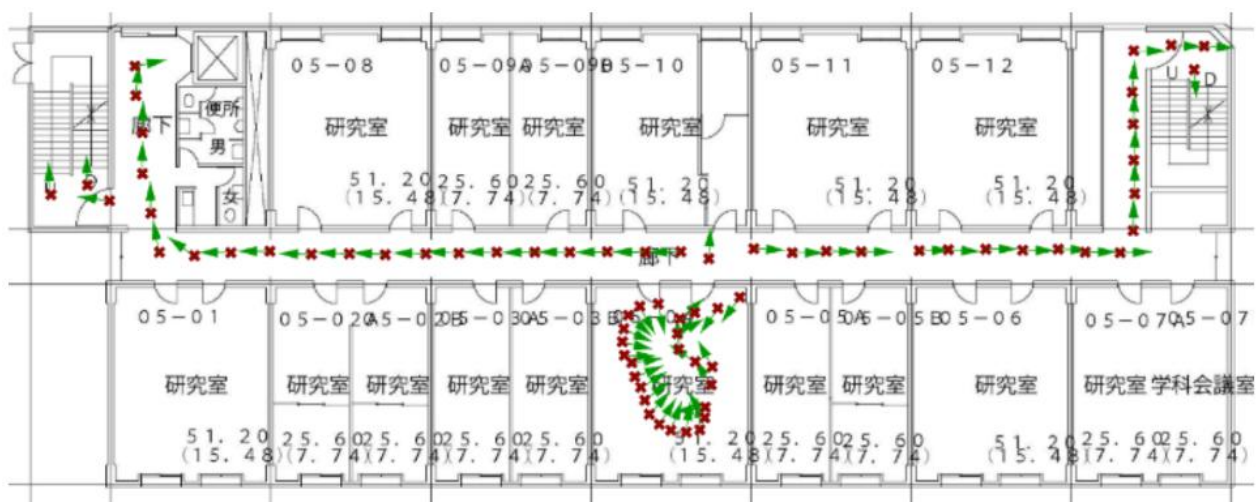


Figure 1. Visualizing the shooting position on floor plan

In addition to spatial organization, recent advances in deep learning enable automated interpretation of visual data [6]. Real-time object detection models such as YOLO can identify installed building equipment directly from construction photographs, reducing manual documentation effort. This approach enables visualization of building components that become concealed after construction, supporting spatial documentation and lifecycle facility management. Based on these developments, this research proposes a framework that transforms construction photographs into structured lifecycle information by combining indoor localization, AI-based equipment detection, and spatial reconstruction [7].

2 System Architecture and Conceptual Framework

The construction industry is increasingly adopting digital twin technologies to enable lifecycle-oriented management of building information [8]. However, practical implementation remains limited due to fragmented data collection processes and the lack of automated methods for organizing construction-stage information [9]. Construction photographs are widely captured during inspection, safety monitoring, and progress documentation, yet these images are typically stored as isolated records without meaningful spatial or semantic structure.

At the same time, modern buildings are becoming more complex due to the integration of advanced mechanical systems, smart controls, and safety infrastructure. This growing complexity increases maintenance workload and makes asset tracking more challenging when facility information is managed using conventional spreadsheet-based approaches [10]. Furthermore, indoor environments limit the effectiveness of GPS-based positioning, making it difficult to determine precise shooting locations of construction photographs [11].

To address these challenges, this study proposes a position-aware facility management architecture that transforms routine construction photographs into structured lifecycle information. As illustrated in Figure 2, the proposed framework bridges the gap between unstructured site photographs and digital twin-based facility management by integrating indoor localization, AI-driven object detection, and spatial asset database construction.

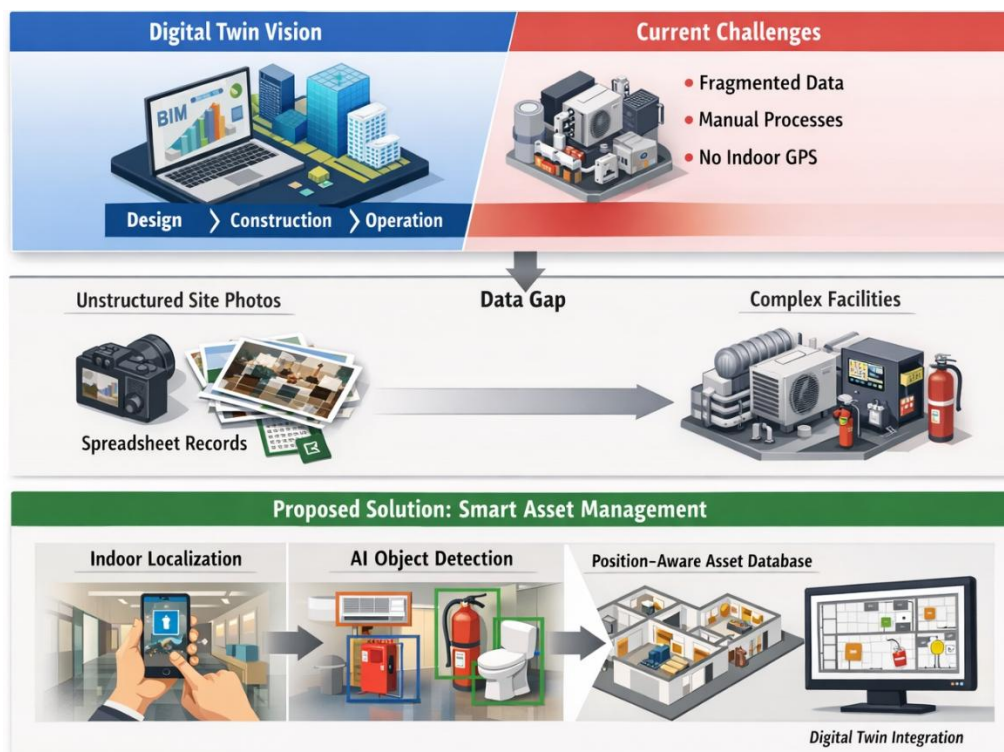


Figure 2. Conceptual illustration of the research motivation.

3 Methodology

This study develops and evaluates a position-aware equipment management framework that utilizes indoor self-localized construction photographs as structured information resources for facility management [12]. The framework aims to transform routine site documentation images into spatially

organized datasets that can support asset tracking, inspection planning, and lifecycle maintenance activities. By embedding reliable positional metadata into construction photographs, the proposed approach enables systematic integration of visual information with facility management workflows as shown in Figure 3.

The methodological procedure consists of three main components: indoor localization validation, development of an AI-based equipment detection system, and exploration of spatial reconstruction [13]. Through this multi-stage process, the study investigates the feasibility of combining spatial positioning technologies and computer vision techniques to enhance the usability of construction-stage data for long-term building operation and digital twin applications [14].

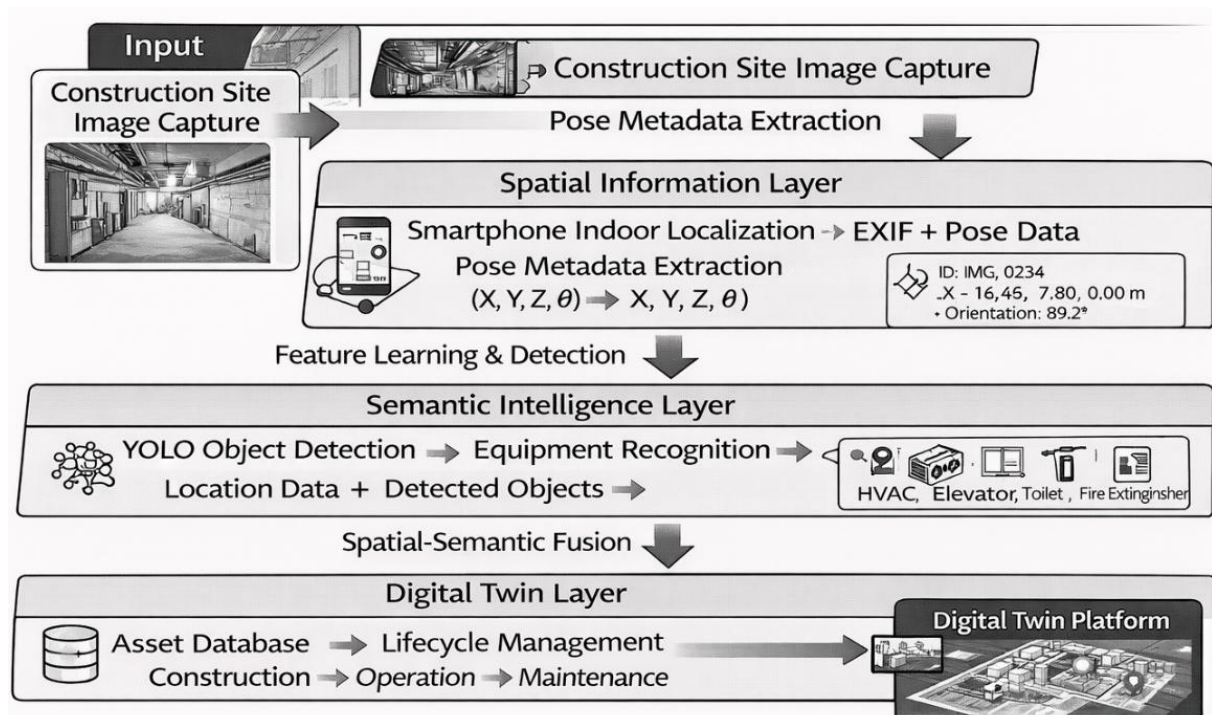


Figure 3. Methodology of the Conceptual framework

3.1 Indoor Localization Workflow

The first step of the methodology focuses on obtaining reliable spatial metadata from construction photographs. For this purpose, the Nav & Snap smartphone application is used to estimate camera position and orientation during image capture. Unlike GPS-based positioning, which is unreliable indoors, Nav & Snap determines location by detecting AR markers placed in the environment and tracking visual feature points while incorporating gravity sensor information [15].

Before image capture, building floor plans are processed using the Map Creator tool. Map Creator converts bitmap drawings into real-scale spatial datasets and assigns coordinate references to AR marker positions. This enables accurate visualization of shooting locations directly on floor plans.

During the experiment, photographs are captured at predefined intervals while moving through indoor spaces. The positional coordinates embedded in the EXIF metadata of each image are later extracted using Python tools. These coordinates are converted from pixel-based units into metric spatial values using known floor-plan scale information.

To evaluate positioning accuracy, the distances between consecutive shooting points calculated from localization outputs are compared with ground-truth measurements obtained using measuring tapes. Relative error values are then computed to assess the reliability of indoor localization results for subsequent facility management applications [16].

3.2 Object Detection and Management

After verifying the reliability of positional metadata, the study develops an automated equipment management workflow based on object detection technology. Spatially enriched construction photographs are first plotted on building floor plans using their estimated coordinates. This allows users to visually confirm camera coverage areas and understand equipment distribution within the building.

Next, the images are processed using a customized object detection model based on the YOLOv11 architecture [17]. Since general pre-trained datasets do not include sufficient indoor facility categories, a domain-specific dataset is prepared using publicly available images and manually annotated samples. Target detection categories include equipment commonly installed in buildings, such as air-conditioning units, sinks, toilets, fans, fire hydrants, piping components, and elevators.

During detection, the model outputs equipment labels together with bounding box coordinates and confidence scores. These results are automatically exported into an Excel-based equipment management platform. The platform records information such as image filename, shooting position, detected equipment type, and detection confidence.

Users can additionally input lifecycle attributes including model number, manufacturing year, physical size, installation cost, and maintenance records. By integrating spatial information, visual evidence, and equipment attributes, the platform enables centralized management of facility information and supports efficient inspection and maintenance planning [18].

4 Fundamental Experiment: Indoor Localization evaluation

To verify the reliability of positional metadata embedded in construction photographs, a series of indoor localization experiments were conducted. Since accurate spatial positioning is essential for mapping equipment locations on building floor plans, this experiment evaluates the measurement accuracy of the Nav & Snap self-localization system under controlled indoor conditions [19].

4.1 Experimental Setup

Localization experiments were carried out in indoor environments at Waseda University, including both classroom spaces and a large atrium area. Photographs were captured at regular intervals of approximately 0.5 m while moving along predefined paths. To reduce setup errors during image acquisition, smartphones were mounted on stabilization devices such as gimbals or holders.

The spatial configuration of shooting points and camera orientations used during the experiment is illustrated in Figure 4. Most images were captured using an iPhone 15, while one experimental trial was conducted using an iPhone SE to examine potential device-dependent differences in localization performance.

Positional coordinates stored in the EXIF metadata of captured images were extracted using Python-based tools. These coordinates were then converted from pixel-based units into metric spatial values using known scale information derived from building floor plans.

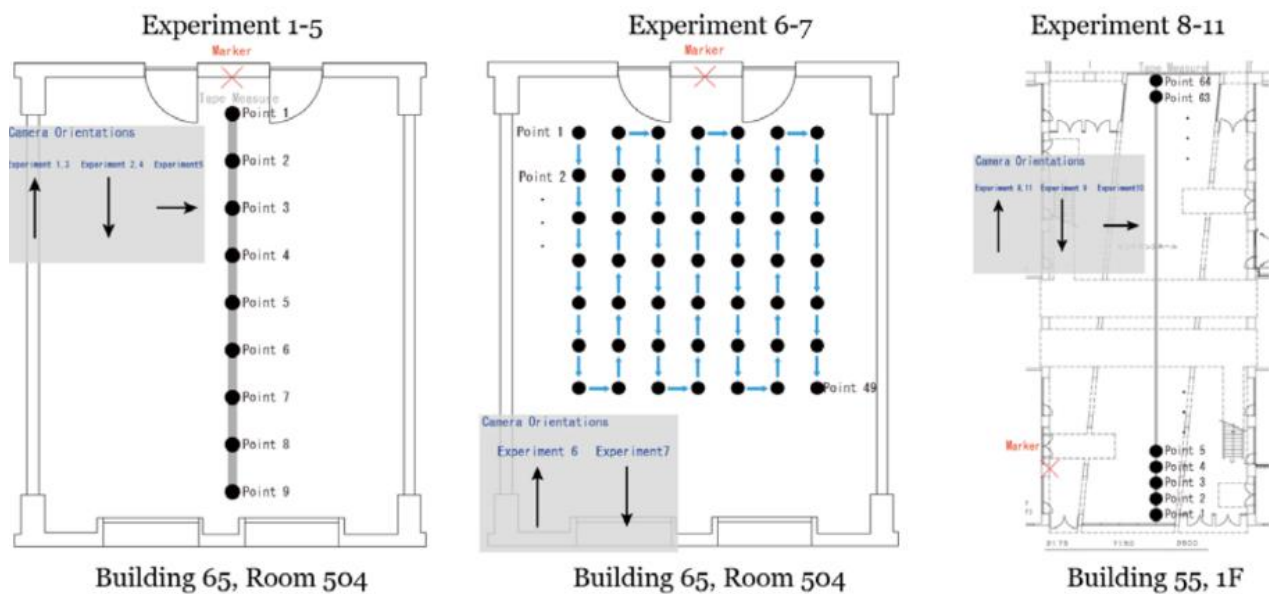


Figure 4. Experimental layouts for evaluating indoor localization accuracy.

4.2 Accuracy Evaluation Method

To evaluate localization accuracy, the distances between consecutive shooting points estimated from indoor localization outputs were compared with ground-truth distances measured using a measuring tape. The relative error was calculated for each movement segment to assess spatial consistency of localization results. This evaluation approach enables verification of whether position-encoded images can be reliably used for facility management tasks such as equipment mapping and inspection planning. By analysing error accumulation across different movement patterns, the study also investigates the robustness of localization performance under practical site conditions.

4.3 Experimental Results

Indoor localization experiments were conducted to evaluate the accuracy of positional metadata embedded in construction photographs captured using the Nav & Snap application. The comparison between estimated distances and ground-truth measurements is summarized in Figure 5, which shows the overall localization performance under different experimental conditions.

In short-range linear movement experiments (Experiments 1–5), the estimated distances between shooting points closely matched the ground-truth values. As shown in Figure 5, the relative error remained small and consistent, indicating stable positioning performance for simple movement scenarios.

In grid-based experiments (Experiments 6–7), a larger number of shooting points were used to evaluate positional drift over longer travel distances. Although cumulative error increased slightly as movement continued, the results in Figure 5 demonstrate that localization accuracy remained within a practical range for facility management applications.

Long-distance experiments conducted in a different indoor environment (Experiments 8–11) further confirmed the robustness of the localization method. While positional error was influenced by environmental factors such as visual feature availability and camera movement direction, the system maintained stable performance, particularly in open indoor spaces with limited occlusion.

Overall, the results indicate that indoor self-localized images provide sufficiently reliable spatial information for organizing construction photographs and supporting position-aware facility management workflows.

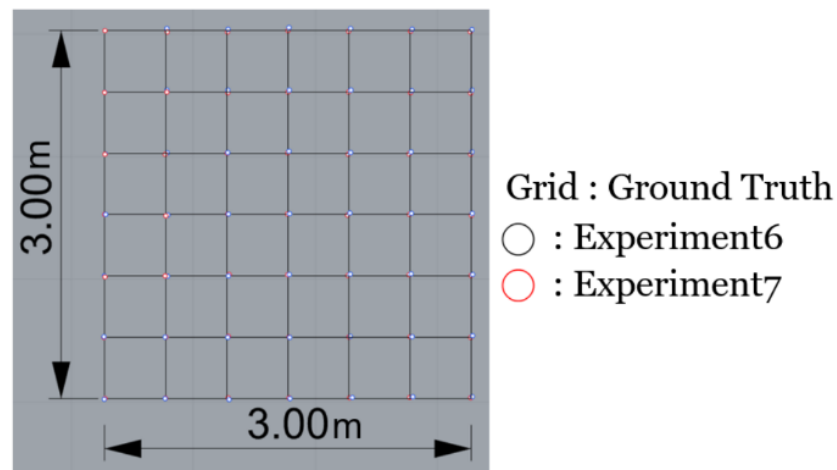


Figure 5. Comparison of ground-truth and measured positional distances in indoor localization experiments.

5 Facility Management System Using Object Detection

5.1 Object Detection Model Development

This study develops an automatic equipment identification method using a YOLOv11-based object detection model. Target categories include common building equipment such as various air conditioner types, fans, toilets, sinks, refrigerators, fire hydrants, pipes, and elevators selected because they frequently appear in commercial and office buildings and are difficult to manage manually.

Training images were collected from web, the COCO dataset, and publicly available single-class datasets. All images were annotated using COCO Annotator, and a Python script converted the COCO JSON format into YOLO text-format labels. The dataset was then split into training and validation sets.

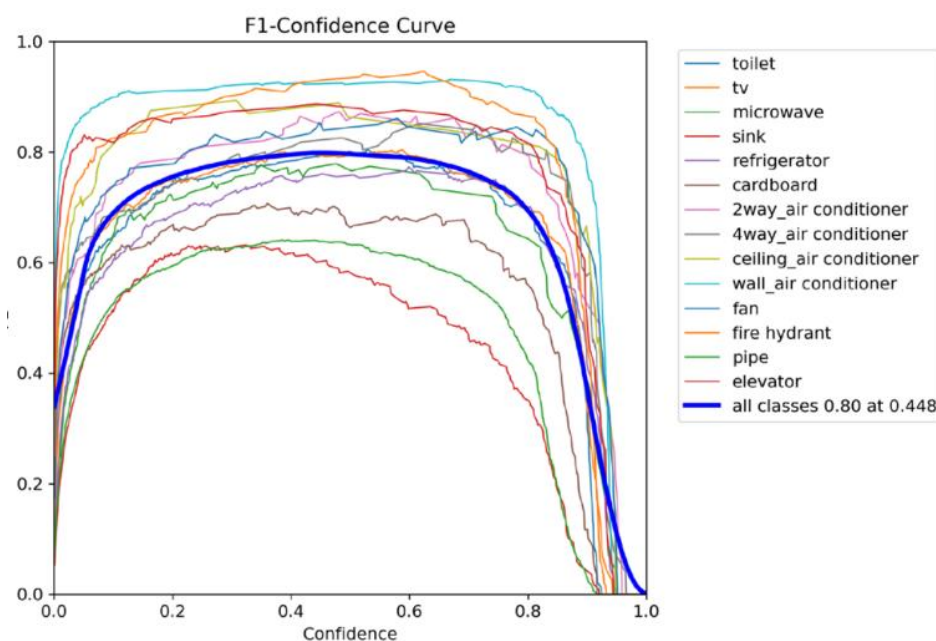


Figure 6. F1-Confidence Curve.

Fine-tuning was performed using YOLOv11n pretrained weights for 100 epochs at 640 × 640 resolution. Loss functions, including box loss, classification loss, and DFL loss, converged smoothly, and no clear overfitting was observed. As shown in Figure 6, the trained model achieved stable detection performance across confidence thresholds. The results reached approximately precision 0.85–0.9, recall 0.75–0.8, mAP@0.5 ≈ 0.85, and mAP@0.5:0.95 ≈ 0.70. F1-score analysis also indicated reliable recognition performance. These results confirm that the trained model can effectively detect major building equipment categories in indoor images.

5.2 Position-Aware Equipment Management Platform

Following the validation of indoor localization accuracy, this study develops an automated equipment management workflow using deep learning-based object detection techniques. The objective is to identify installed building equipment directly from construction photographs and integrate detection results with spatial metadata for facility management applications. As shown in Figure 7, the detected equipment locations are visualized on building floor plans based on the shooting positions of captured images. This spatial visualization allows users to understand equipment distribution within the building environment.



Figure 7. Visualization of shooting positions and detected equipment distribution on the floor plan.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	filename	x_px	y_px	label	conf	maker	model no	year	price_jpy	size	maintenance cycle	last maintenance date	next maintenance due	notes
1	image_001.jpg	1093	353	elevator	0.9	日立	S-01	2010	¥1,000,000	2300×1985	1年	2024年2月	2025年2月	
2	image_002.jpg	1184	262	toilet	0.85	TOTO	US800CE	2016	¥80,000		5年	2021年5月	2026年5月	
3	image_002.jpg	1184	262	toilet	0.27	TOTO	TCF585	2016	¥100,000	600×300	3年	2025年5月	2028年5月	
4	image_003.jpg	1137	260	toilet	0.85	TOTO	TCF585	2016	¥100,000	600×300	3年	2025年5月	2028年5月	
5	image_004.jpg	961	375	4way air conditioner	0.34	DAIKIN	SZRC40BYNT	2010	¥800,000	800×600	1年	2025年8月	2026年8月	
6	image_005.jpg	1087	567	refrigerator	0.71									
7	image_006.jpg	552	437	fire hydrant	0.93	ヤマトプロテック	YA-10XIII	2021	¥12,000		6か月	2025年9月	2026年3月	耐用年2026年まで
8	image_007.jpg	370	430	elevator	0.86	日立	S-01	2010	¥100,000,000	2300×1985	1年	2024年2月	2025年2月	
9	image_008.jpg	295	299	fan	0.94	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	
10	image_009.jpg	297	97	4way air conditioner	0.9	DAIKIN	SZRC40BYNT	2010	¥800,000	800×600	1年	2025年8月	2026年8月	
11	image_009.jpg	297	97	fan	0.54	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	
12	image_009.jpg	297	97	fan	0.43	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	
13	image_010.jpg	343	186	fire hydrant	0.93	ヤマトプロテック	YA-10XIII	2021	¥12,000		6か月	2025年9月	2026年3月	耐用年2026年まで
14	image_011.jpg	1122	128	fan	0.89	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	
15	image_011.jpg	1122	128	fan	0.26	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	
16	image_011.jpg	1122	128	fan	0.25	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	
17	image_012.jpg	1223	375	sink	0.88	アイオ産業	KW-1000L	2020	¥80,000	1000×500	2年	2024年8月	2026年8月	
18	image_013.jpg	1217	362	fan	0.71	三菱電機	VD152VC7	2010	¥10,000	300×300	3年	2024年3月	2027年3月	

Figure 8. Interface of the position-aware facility management platform.

Furthermore, the detection results are integrated into a position-aware facility management system that supports lifecycle information management. The interface of the developed platform is presented in Figure 8, where users can review detected equipment information and input additional attributes such as model details, installation year, and maintenance records. This integrated workflow improves accessibility of facility information and supports efficient inspection and maintenance planning.

6 Discussion and Conclusion

This study presented a position-aware facility management framework that transforms routine construction photographs into structured spatial information for building lifecycle management. By integrating smartphone-based indoor localization and AI-driven object detection, the proposed approach enables construction-stage image data to be reused for asset tracking, inspection planning, and maintenance support in indoor environments where GPS positioning is unreliable.

The experimental results confirmed that indoor localization provides sufficiently accurate positional metadata for practical facility management applications. Average relative positioning errors remained approximately within 0.5% to 0.9%, demonstrating that construction photographs can be reliably mapped onto building floor plans. Although positional drift increased slightly during long-distance or grid-based movements, overall localization performance remained stable under different indoor conditions.

Furthermore, the customized object detection model showed reliable performance in recognizing commonly installed building equipment such as air-conditioning units, toilets, sinks, fire hydrants, piping components, and elevators. By integrating detection outputs with spatial metadata, the developed platform enabled intuitive visualization of equipment distribution and centralized management of lifecycle asset information. This workflow reduces manual documentation effort and improves accessibility of facility information, particularly in large and complex buildings.

Despite these promising outcomes, several limitations were identified. Localization accuracy may be affected by environmental factors such as visual feature availability, lighting conditions, and device differences. Similarly, detection performance depends on the diversity and quality of the training dataset. Future work should therefore focus on improving robustness in complex construction environments, expanding equipment datasets, and enabling real-time integration with BIM and digital twin platforms.

Overall, this research demonstrates the feasibility of reusing construction photographs as structured digital resources for position-aware facility management. The proposed framework contributes to improving information continuity between construction and operation stages and supports the development of data-driven building lifecycle management systems [20].

References

- [1] Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114, 103179. <https://doi.org/10.1016/j.autcon.2020.103179>
- [2] Opoku, D.-G. J., Perera, S., Osei-Kyei, R., & Rashidi, M. (2021). Digital twin application in the construction industry: A literature review. *Journal of Building Engineering*, 40, 102726. <https://doi.org/10.1016/j.job.2021.102726>

- [3] Hamledari, H., & Fischer, M. (2020). Automated construction progress monitoring using computer vision. *Advanced Engineering Informatics*, 43, 101003. <https://doi.org/10.1016/j.aei.2019.101003>
- [4] Alarifi, A., Al-Salman, A., Alsaleh, M., Alnafessah, A., Al-Hadhrami, S., Al-Ammar, M., & Al-Khalifa, H. (2019). Ultra-wideband indoor positioning technologies: Analysis and recent advances. *Sensors*, 19(21), 4504. <https://doi.org/10.3390/s19214504>
- [5] Davidson, P., & Piché, R. (2017). A survey of selected indoor positioning methods for smartphones. *IEEE Communications Surveys & Tutorials*, 19(2), 1347–1370. <https://doi.org/10.1109/COMST.2016.2637663>
- [6] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- [7] Lamsal, B., Oka, M., Ryota, T., Sardellitti, M., KC, B. K., & Matsumoto, N. (2025). Development of a construction site photo management and inspection application using smartphone-based indoor localization in non-GPS environments. In *Proceedings of the Creative Construction Conference 2025*, Zadar, Croatia.
- [8] Sepasgozar, S. M. E. (2021). Digital twin and web-based virtual reality for construction management. *Automation in Construction*, 129, 103760. <https://doi.org/10.1016/j.autcon.2021.103760>
- [9] Pärn, E. A., Edwards, D. J., & Sing, M. C. P. (2019). The building information modelling trajectory in facility management: A review. *Automation in Construction*, 75, 45–55. <https://doi.org/10.1016/j.autcon.2016.12.003>
- [10] Motawa, I., & Carter, K. (2019). Sustainable BIM-based lifecycle asset management. *Journal of Facilities Management*, 17(3), 246–262. <https://doi.org/10.1108/JFM-09-2018-0057>
- [11] Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., & Furgale, P. (2015). Keyframe-based visual-inertial odometry using nonlinear optimization. *International Journal of Robotics Research*, 34(3), 314–334. <https://doi.org/10.1177/0278364914554813>
- [12] Deng, M., Menassa, C. C., & Kamat, V. R. (2021). Digital twin-enabled construction project management. *Journal of Construction Engineering and Management*, 147(12), 04021163.
- [13] Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin driven smart manufacturing. *Journal of Manufacturing Systems*, 48, 157–169. <https://doi.org/10.1016/j.jmsy.2018.06.012>
- [14] Liu, Z., Meyendorf, N., & Mrad, N. (2018). The role of data fusion in predictive maintenance. *Structural Health Monitoring*, 17(3), 558–575. <https://doi.org/10.1177/1475921717707357>
- [15] Volk, R., Stengel, J., & Schultmann, F. (2014). Building information modelling for existing buildings: Literature review and future needs. *Automation in Construction*, 38, 109–127. <https://doi.org/10.1016/j.autcon.2013.10.023>
- [16] Kiviniemi, A., & Codinhoto, R. (2014). Challenges in the implementation of BIM for facility management. In *Proceedings of the CIB World Building Congress*, Brisbane, Australia.
- [17] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8) [Open-source software]. Retrieved from <https://github.com/ultralytics/ultralytics>
- [18] Tang, P., Huber, D., Akinci, B., Lipman, R., & Lytle, A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds. *Automation in Construction*, 19(7), 829–843.
- [19] Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2009). D4AR: A 4-dimensional augmented reality model for automating construction progress monitoring. *Journal of Computing in Civil Engineering*, 23(1), 41–54.
- [20] Lamsal, B. (2025). Development of an indoor localization and positioning system in non-GPS environments using standalone smartphones and its implementation in construction photo management. In *Proceedings of the Digital Frontiers in Buildings and Infrastructure (DFBI 2025)*, Delft University of Technology, Netherlands.