

AR-Enhanced Indoor Construction Progress Monitoring Using Bim and Synthetic Data

Mathis Baubriaud^{1,2*}, Stephane Derrode¹, Rene Chalon¹ and Kevin Kernn²

¹Centrale Lyon, CNRS, Claude Bernard Lyon 1, INSA Lyon, Lumiere Lyon 2, LIRIS, UMR5205, 69130, Ecully, France

²SPIE Building Solutions, F-69320, Feyzin, France

*Corresponding author

Mathis Baubriaud, Centrale Lyon, CNRS, Claude Bernard Lyon 1, INSA Lyon, Lumiere Lyon 2, LIRIS, UMR5205, 69130, Ecully, France and SPIE Building Solutions, F-69320, Feyzin, France.

Received: September 08, 2025; **Accepted:** September 22, 2025; **Published:** October 02, 2025

ABSTRACT

Manual inspection of indoor construction sites for progress monitoring is time-consuming, error-prone, and inefficient. Automated solutions using Deep Learning (DL) and Augmented Reality (AR) offer significant potential, but are hampered by the scarcity of large labeled datasets, especially for complex indoor environments. This paper presents a novel and automated methodology for Indoor Construction Progress Monitoring (ICPM) that addresses this data bottleneck by leveraging Building Information Modeling (BIM) and synthetic data. Our approach uses a photorealistic graphics engine to generate a large, annotated synthetic dataset of Mechanical, Electrical, and Plumbing (MEP) components within BIM environments. A YOLOv8 instance segmentation model, enhanced with domain adaptation techniques, is trained on this synthetic data and integrated with an AR application on HoloLens 2 for real-time on-site progress monitoring. Experiments demonstrate that the proposed synthetic data-powered model achieved a substantial improvement in mAP50 compared to models trained on limited real-world data. A preliminary on-site validation further highlights the practical potential of the AR-integrated system for efficient and reliable ICPM, demonstrating a feasible path towards accessible and user-friendly automated inspection tools that can be readily adopted by construction professionals on real-world sites.

Keywords: Building Information Modeling (BIM), Synthetic Data, Deep Learning, Augmented Reality (AR), Progress Monitoring, Indoor Construction, User Study

Introduction

Background and Problem Statement

The Architecture, Engineering, and Construction (AEC) industry is under increasing pressure to improve efficiency accuracy, and safety amidst complex project demands [1]. Indoor Construction Progress Monitoring (ICPM) is vital for ensuring adherence to project schedules and budgets [2]. However, traditional ICPM methods, heavily reliant on manual inspections and documentation, pose significant challenges. These methods are inherently labor-intensive, time-consuming, and susceptible to human error [3]. This is particularly problematic in modern construction, where complex Mechanical, Electrical, and

Plumbing (MEP) systems are prevalent [4]. The dynamic nature of construction sites, marked by frequent changes and unforeseen events, further compounds these difficulties, often rendering manual reports obsolete and leading to inaccurate progress assessments [5].

Building Information Modeling (BIM) offers a centralized digital platform for project information [6]. However, realizing BIM's full potential for on-site ICPM requires bridging the gap between the digital model and the physical site. Augmented Reality (AR) shows promise for visualizing BIM data on-site, facilitating comparisons between as-designed and as-built conditions [7]. Critically, existing AR-based solutions often require manual alignment of the BIM model and rely on subjective visual inspection, limiting automation and accuracy, particularly for complex MEP components [8].

Citation: Mathis Baubriaud, Stephane Derrode, Rene Chalon, Kevin Kernn. AR-Enhanced Indoor Construction Progress Monitoring Using Bim and Synthetic Data. J Mat Sci Eng Technol. 2025. 3(4): 1-17. DOI: doi.org/10.61440/JMSET.2025.v3.75

Computer Vision (CV) and Deep Learning (DL), specifically instance segmentation, offer the potential to automate ICPM tasks [9]. Instance segmentation is particularly well-suited for ICPM of MEP systems due to its ability to detect and delineate individual components, even in cluttered and occluded environments. However, a major bottleneck for DL in construction, especially for indoor MEP systems, is the scarcity of large, labeled datasets [10]. Collecting and annotating such data is costly and time-consuming. Therefore, a need exists for a fully automated, data-efficient, and AR-integrated ICPM system that can accurately detect and segment MEP components in complex indoor environments, bridging the gap between digital design and physical construction.

Research Objectives and Contributions

To address the limitations of current ICPM methods, this paper proposes a novel approach for automated ICPM, leveraging the synergies of BIM, synthetic data, DL, and AR. The research objectives and primary contributions are:

- Develop a streamlined and automated pipeline for generating large-scale, photorealistic, and accurately annotated synthetic datasets of indoor construction scenes, specifically targeting MEP components. This pipeline addresses the data scarcity bottleneck and provides a scalable solution for future research. The generated dataset, MEP-SEG, is made publicly available [11].
- Adapt and optimize a state-of-the-art DL model, specifically YOLOv8, for accurate instance segmentation of MEP components in complex indoor construction environments. This objective focuses on achieving high precision and robustness in detecting and identifying critical building elements using a combination of synthetic pre-training and fine-tuning on real-world data.
- Integrate the trained DL model with a commercially available mobile AR platform for real-time, on-site construction progress monitoring and BIM-based comparison. This integration aims to create a practical and user-friendly tool for construction professionals.
- Conduct rigorous experimental validation using both synthetic (MEP-SEG) and real-world (MEP-REAL) datasets from diverse construction sites. This evaluation includes ablation studies to assess the impact of domain adaptation techniques and different ratios of synthetic and real data, ensuring the robustness, generalizability, and practical applicability of the methodology.
- Evaluate the system's usability, acceptability, and perceived effectiveness through a user study with 21 construction professionals, employing quantitative questionnaires and qualitative interviews. This provides practical insights into the real-world applicability and potential benefits of the proposed AR-integrated ICPM system.

This research advances the field of automated construction progress monitoring by providing a validated and innovative methodology that addresses key limitations of existing methods, offering significant improvements in accuracy, efficiency and automation. Furthermore, by integrating a commercially available AR solution and making the synthetic dataset public, it contributes to the adoption of advanced digital technologies in the AEC industry.

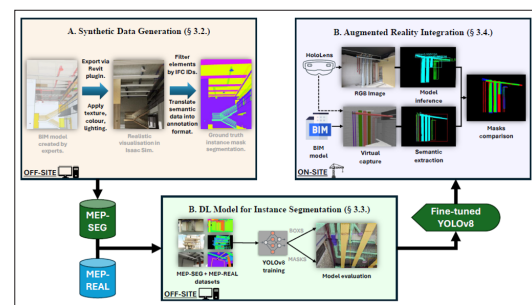


Figure 1: Overview of the proposed approach

Article Organization

The remainder of this paper is structured to detail the proposed methodology, experimental validation, and key findings. Following this introduction, Section 2 provides a comprehensive review of the relevant literature on automated ICPM, focusing on the application of CV, DL, AR, and synthetic data generation. Section 3 elaborates on the proposed methodology, detailing the BIM-based synthetic data generation pipeline, the architecture and training of the YOLOv8 instance segmentation model, and the integration of this model with an AR application for real-time ICPM. Section 4 presents a detailed account of the experimental setup, results (including evaluations on both synthetic and real-world datasets), ablation studies, mask alignment analysis, and user feedback from a study with construction professionals. Section 5 discusses the key findings, implications, limitations, and potential future directions of this research. Finally, Section 6 concludes the paper by summarizing the main contributions and highlighting the broader impact of this work on automated ICPM.

This paper significantly expands upon the foundational work presented in our previous conference paper [12], which introduced the BIM-based synthetic data generation pipeline. Specifically, this article provides a more detailed and comprehensive presentation of the methodology. Furthermore, this manuscript presents novel research outcomes, including the seamless integration of the trained DL model with an augmented reality application for on-site deployment, a thorough evaluation of mask alignment strategies, and a comprehensive assessment of the system's usability and user acceptance through detailed user studies and real-world experimentation.

Literature Review

Progress Monitoring in Construction

Ineffective progress monitoring is a barrier to successful project delivery because it prevents timely detection of deviations from planned schedules and budgets [13]. Traditional methods, including manual site inspections and paper-based documentation, are not only inefficient, but also introduce subjectivity and increase the likelihood of errors [14]. The complexity and dynamism of indoor construction environments, especially those with dense MEP installations, make comprehensive and accurate manual monitoring practically unsustainable [15,16]. As a result, delays in reporting, reactive project management and increased risk of cost overruns and schedule delays are common outcomes [17]. The need for automated and reliable solutions to address these challenges is therefore essential. To address these challenges, research has explored various automated and semi-automated technologies

[18]. Reality capture techniques, such as laser scanning and photogrammetry, offer accurate 3D representations of as-built conditions [19-21]. These methods enable Scan-vs-BIM comparisons for deviation analysis [22-24]. However, they often require specialized equipment, skilled operators, and extensive post processing, hindering real-time, on-site application and scalability [25]. Sensor-based tracking systems (RFID, UWB, BLE, GPS) provide real-time data on material flow, equipment utilization, and personnel location [26-30]. While valuable for resource management, these systems do not directly address the visual assessment of construction progress, particularly for complex MEP installations. Computer vision (CV) and Deep Learning (DL) are increasingly being applied in the AEC industry to automate tasks such as progress monitoring, safety inspection, and defect detection [31]. Specifically, for ICPM, DL-powered CV offers the potential to automate the visual assessment of construction progress, overcoming the subjectivity and labor-intensive nature of manual inspections [32]. Key CV techniques relevant to ICPM include object detection, semantic segmentation, and instance segmentation. Object detection has been used to identify and locate construction elements, such as MEP components, equipment, and workers [33,34]. Semantic segmentation provides a pixel-level classification of the scene, enabling the identification of different building materials and elements [35]. Instance segmentation combines the benefits of both, detecting and delineating individual object instances, even with occlusions, making it particularly suitable for tracking MEP components in complex indoor environments [36,37]. Various Convolutional Neural Network based (CNN) models have been explored for these tasks, including the You Only Look Once (YOLO) family of object detectors [38,39], Mask R-CNN [40,41], and, more recently, Vision Transformers (ViTs) [42,43]. While these models have shown promising results, their application to ICPM, particularly for indoor MEP systems, is often hampered by the lack of large, labeled datasets [44]. The complexity, clutter, and occlusions characteristic of indoor construction environments further complicate the task of accurate object recognition and segmentation. Existing research has applied CV and DL to structural element monitoring [45-47] and MEP system progress tracking [48]. However, many of these approaches rely on limited real-world data, hindering their generalizability and robustness in diverse construction scenarios. The need for extensive manual annotation of real-world images further limits the scalability of these methods. This data scarcity challenge motivates the exploration of synthetic data generation techniques, as discussed in the next section.

Synthetic Data Generation

Synthetic data generation offers a promising solution to the data scarcity challenge in construction, enabling the creation of large, labeled datasets without costly and time-consuming manual annotation [49]. This is particularly crucial for training robust DL models for complex tasks like instance segmentation of MEP components in cluttered indoor environments [50]. Various methods exist for generating synthetic data, including graphics engine-based approaches and hybrid methods that combine synthetic and real-world data [51-53]. However, for construction progress monitoring, BIM-based synthetic data generation offers significant advantages [54]. BIM models inherently contain rich geometric and semantic information about building elements, providing a readily available source

for creating realistic and accurately labeled virtual construction environments [55]. BIM-based approaches typically involve importing BIM models into graphics or game engines (e.g., Unreal Engine, Unity, Isaac Sim), configuring virtual environments with realistic materials and lighting, and rendering synthetic images from various viewpoints [56]. Crucially, the semantic information in BIM models allows for automated generation of ground truth annotations, including object classes, instance segmentation masks, and depth maps, eliminating the need for manual labeling [57]. This automated annotation is a key advantage, enabling the creation of large-scale datasets with minimal effort [58]. Our approach builds upon this foundation, utilizing a streamlined pipeline and focusing specifically on the detailed representation of MEP components within complex indoor scenes [59]. Despite the advantages of synthetic data, a key challenge is the "reality gap" - the difference in appearance and characteristics between synthetic and real-world images [60]. DL models trained solely on synthetic data may exhibit limited generalization performance when deployed in real-world scenarios. Therefore, domain adaptation techniques are crucial for bridging this gap and improving the transferability of models [61]. These techniques aim to reduce the discrepancy between the synthetic (source) and real-world (target) domains, enabling models trained on synthetic data to perform well on real-world images [62]. Common approaches include Unsupervised Domain Adaptation (UDA) using adversarial training [63], and Semi-Supervised Domain Adaptation (SSDA) leveraging a small amount of labeled real-world data [64].

Augmented Reality in Construction

Augmented Reality (AR) offers significant potential for enhancing construction processes by overlaying digital information onto the real-world view [65,66]. In the context of progress monitoring, AR enables direct visual comparisons between the as-planned BIM model and the as-built reality, facilitating the identification of discrepancies and deviations [67]. Various AR devices, including head-mounted displays (HMDs) like Microsoft HoloLens and Magic Leap, and handheld devices running ARKit or ARCore, have been explored for construction applications [68, 69]. The choice of device depends on factors such as user mobility, environmental conditions, and the required level of immersion [70]. Several studies have demonstrated the use of AR for on-site progress monitoring. For example, Martins et al. [71] proposed an AR-based framework for bridge inspection, while Kopsida and Brilakis [72] developed a system for real-time volume-to-plane comparisons. More recent work includes integrated systems combining AR with other technologies like 3D scanning and robotics for remote inspection and monitoring [73,74]. However, a key limitation of many existing AR-based progress monitoring systems is their reliance on manual alignment of the BIM model with the real-world scene and visual inspection for discrepancy detection [75]. This process can be time-consuming, subjective, and prone to errors [76]. Furthermore, many systems lack integration with automated object recognition and segmentation capabilities, limiting their ability to provide quantitative progress data and detailed analysis of specific building elements, especially complex MEP systems [77]. Our work addresses these limitations by integrating a deep learning-based instance segmentation model with an AR platform, enabling automated detection and segmentation of MEP components and facilitating

a more objective and efficient comparison between the BIM model and the as-built reality. This integration of DL-powered object recognition with AR visualization represents a significant step towards more automated and data-driven construction progress monitoring.

Methodology

Overview of the Proposed Approach

This research introduces a novel, integrated methodology for automated ICPM. The approach leverages the synergistic combination of BIM, synthetic data, DL, and AR to create a practical, efficient, and robust system for real-time, on-site progress assessment. (Figure 1).

The core of the methodology is a three-stage process. First, a large-scale, labeled synthetic dataset of indoor construction scenes, specifically focusing on MEP components, is automatically generated using existing BIM models and a photorealistic graphics engine (NVIDIA Isaac Sim). This addresses the critical data scarcity challenge hindering DL-based ICPM (Figure 1 A). Second, a state-of-the-art instance segmentation model, YOLOv8, is trained on this synthetic dataset and enhanced with a real images dataset to improve its robustness and generalizability to real-world construction sites (Figure 1 B). Third, the trained DL model is integrated into a commercially available mobile AR application, NEXT-BIM1, designed for the HoloLens 2. NEXT-BIM collaborated on this research project, providing essential support with their expertise in BIM and AR, with the goal of integrating the developed technology into their tools in the future. This AR application enables real-time, onsite visualization of BIM models and instance segmentation results overlaid onto the physical environment (Figure 1 C). This AR application also facilitates on-site progress comparison by superimposing the DL model's predictions onto the BIM model view, allowing inspectors to visually assess alignment and identify discrepancies. The integrated system is rigorously evaluated on both synthetic and real-world datasets, assessing accuracy, robustness, and user acceptance. The subsequent sections detail each stage of this methodology.

Synthetic Data Generation

Our methodology hinges on the automated generation of a large, diverse synthetic dataset to train robust DL models for ICPM, reducing dependence on scarce labeled real-world data. The pipeline leverages BIM models' geometric and semantic richness alongside NVIDIA Isaac Sim's photorealistic rendering capabilities. It consists of key steps including BIM model preparation and virtual environment setup.

BIM Model Preparation

The process begins with selecting and preparing BIM models representing indoor construction environments such as offices, laboratories, and commercial spaces (Figure 2). To ensure relevance for MEP system ICPM, a filtering process isolates essential MEP components (e.g., ducts, pipes, cable trays, HVAC units) while removing non-essential architectural elements. This optimization preserves computational resources and geometric fidelity.

Using BIM software's built-in functionalities, models are reviewed and adjusted to maintain compatibility with the

graphics engine. Simplifications are made where necessary to optimize rendering performance without compromising the accuracy of MEP components.

Virtual Environment Setup

The prepared BIM models are imported into NVIDIA Isaac Sim; a high-fidelity simulation platform built on NVIDIA Omniverse. Isaac Sim is selected for its advanced rendering capabilities, Python scripting automation, and support for BIM data manipulation.

To enhance photorealism, materials and textures are assigned using the NVIDIA Omniverse API (Application Programming Interface), employing physically-based rendering materials that accurately represent construction materials such as metal, concrete, plastic, and insulation. MEP components are given specific textures (e.g., galvanized steel for ducts, copper for pipes) to improve realism.

Lighting is configured to simulate both natural and artificial illumination. Sun and sky models replicate daylight conditions, while artificial lighting is adjusted to match fixtures present in BIM models. Variations in intensity and color temperature introduce diversity, mimicking real-world site conditions.

To replicate real construction environments, randomized scene clutter is introduced, including objects like tools, scaffolding, and debris. These elements create occlusions, challenging the DL model to accurately detect and segment MEP components under varying conditions, thereby bridging the gap between synthetic and real-world data.



Figure 2: Snapshots of three BIM projects imported into the graphics engine.

Virtual Camera Configuration

To capture diverse viewpoints of the virtual construction scenes, we configure a virtual camera within Isaac Sim and define a set of camera poses that mimic realistic on-site inspection paths. The virtual camera is configured to emulate the specifications of a typical mobile device camera, with parameters such as:

1. **Field of View (FOV):** A realistic FOV is set to mimic the field of view of a handheld camera, ensuring that the synthetic images capture a representative portion of the indoor scene.
2. **Resolution:** The image resolution is set to a standard resolution (e.g., 640x640 pixels) to balance image quality and computational efficiency during rendering and annotation.

- **Distortion parameters:** To further enhance realism and mimic real-world camera imperfections, we introduce lens distortion effects to the virtual camera model. Radial and tangential distortion parameters are randomly sampled within a realistic range to simulate lens imperfections and create more diverse synthetic images.

After configuring the virtual camera, the next step is to determine its possible positions within each scene. This involves defining routes that mimic the movement of a worker inspecting the construction site. The API provides a tool for manually creating these routes. While grid-based viewpoints could be used, manually drawn routes provide a more natural representation of an inspector's movement. This parameterized camera pose generation strategy ensures a systematic and comprehensive cover age of the virtual environment while maintaining realism and avoiding repetitive viewpoints.

Image Capture and Preprocessing

Once the virtual environment and camera configurations are set up, an automated image capture process is initiated using Python scripting within Isaac Sim. The graphics engine renders photorealistic RGB images for each defined camera pose, capturing variations in lighting, materials, and scene clutter. To enhance dataset diversity, camera angles are systematically varied across different BIM spaces, and date-time settings are randomized to simulate different sun positions and lighting conditions.

Minimal preprocessing is applied before training, primarily resizing images to a fixed resolution (e.g., 640x640 pixels) using bilinear interpolation to ensure consistency. Additional preprocessing, such as noise reduction or enhancement, is avoided to preserve realism and allow the DL model to learn from raw synthetic images.

Automated Annotation Generation

A major advantage of synthetic data is the automated generation of precise ground-truth annotations. Using BIM metadata and Isaac Sim's ray tracing, the pipeline produces detailed annotations, including instance segmentation masks, semantic segmentation labels, depth maps, and bounding box annotations.

Instance segmentation masks are generated by casting rays to identify object instances, grouping pixels accordingly. Semantic segmentation labels classify each pixel based on predefined BIM object classes (e.g., ducts, pipes, walls). Depth maps provide geometric scene information, enhancing applications like depth-aware progress monitoring. Bounding boxes are derived from instance segmentation masks, enclosing each MEP component for object detection tasks.

This automated annotation process ensures pixel-perfect labels across all synthetic images, creating a comprehensive dataset for training DL models for ICPM tasks.

Deep Learning Model for Instance Segmentation

Accurate and robust instance segmentation of MEP components is crucial for effective ICPM. To achieve this, we evaluated several state-of-the-art DL architectures, including Faster R-CNN [78], Mask R-CNN [40], and various members of the

YOLO family [79,38]. While Faster R-CNN and Mask R-CNN offer high accuracy, their two-stage architectures result in significantly slower inference speeds compared to the single-stage YOLO models. Vision Transformers (ViTs) [42] and the Segment Anything Model (SAM) [80] were also considered; however, their higher computational demands made them less suitable for real-time deployment on a resource-constrained mobile AR device like the HoloLens 2.

Based on these considerations, we selected YOLOv8, a recent and efficient iteration of the YOLO family, as the most suitable model for our application. YOLOv8 provides an excellent balance of high accuracy and real-time inference speed, a critical requirement for on-site deployment.

To address the reality gap between synthetic and real-world data, we employed a two-stage training strategy incorporating domain adaptation:

- **Pre-training on synthetic data:** The YOLOv8 model was initially pre-trained on the large and diverse MEP-SEG synthetic dataset generated as described in Section 3.2. This pre-training provided a strong foundation for the model to learn robust features for MEP component detection and segmentation, leveraging the perfectly labeled synthetic data
- **Fine-tuning with mixed data:** The pre-trained model was then fine-tuned using a mixed dataset consisting of synthetic images from MEP-SEG and real-world images from the MEP-REAL dataset (detailed in Section 4). We experimented with different ratios of synthetic and real images to determine the optimal balance for achieving high performance on real-world data. This fine-tuning, a form of transfer learning, allows the model to adapt to the characteristics of real-world images while retaining the knowledge gained from the synthetic data. We also explored initializing the model with weights pre-trained on the large-scale COCO (Common Objects in Context) dataset [81] for comparison.

Data augmentation techniques, including random rotations, scaling, horizontal flips, color jittering, and mosaic augmentation (as implemented in the Ultralytics YOLOv8 framework), were applied during both pre-training and fine-tuning to further enhance the model's robustness and generalization capabilities. This two-stage training strategy, combining large-scale synthetic data pre-training with targeted fine-tuning, effectively bridges the reality gap and enables high accuracy and robustness for MEP component instance segmentation in real-world construction environments. Specific training details, hyperparameters, and evaluation metrics are presented in Section 4.

Augmented Reality Integration and On-site Comparison

To bridge the gap between the digital model and the physical construction site, we integrated the trained deep learning model within an AR application, leveraging the capabilities of the Microsoft HoloLens 2 HMD. This integration enables real-time, on-site visualization of the instance segmentation results superimposed onto the actual MEP components, facilitating intuitive progress monitoring and discrepancy detection.

AR Device and Application

The Microsoft HoloLens 2 was chosen as the AR platform for this research due to its advanced spatial mapping, object recognition, and gesture recognition capabilities, as well as its adaptation in construction environments. The HoloLens 2 is a self-contained, untethered holographic computer that allows users to interact with digital content and holograms in the real world. The device is equipped with a suite of sensors, including depth sensors, an Inertial Measurement Unit (IMU), and an RGB camera, which provide real-time data about the user's environment and position.

For this application, we utilized the NEXT-BIM application, which is specifically designed for on-site construction progress monitoring using AR. NEXT-BIM provides functionalities for visualizing BIM models in AR, aligning the virtual model with the physical environment, and interacting with the model through gestures and voice commands. We integrated our trained YOLOv8 model into the NEXT-BIM application to enable real-time instance segmentation of MEP components directly within the HoloLens 2's field of view.

Model Deployment

Deploying the trained YOLOv8 model on the HoloLens 2 required careful consideration of the device's computational resources and performance constraints. To achieve real-time inference, we optimized the model by converting it to the Open Neural Network Exchange (ONNX) format, which is suitable for efficient deployment on various hardware platforms, including mobile and embedded devices. The ONNX model was then integrated into the NEXT-BIM application, enabling on-device inference without the need for an external server or cloud connection. This on-device deployment ensures seamless user experience during field inspections without an internet connection. The specific details of the model conversion and integration process are beyond the scope of this paper, but we ensure it followed industry best practices for deploying DL models on resource-constrained devices.

Discrepancy Analysis

A crucial aspect of the AR-integrated system is accurately aligning and superimposing the YOLOv8 model's predictions (predicted masks) with the corresponding elements in the BIM model (ground truth masks). This enables a direct visual comparison between the as-designed and as-built states. The overall process is illustrated in Figure 3, and consists of the following steps:

- 1. RGB Image Capture and Metadata Acquisition:** The process begins by capturing the real-world scene using the HoloLens 2's built-in RGB camera. Simultaneously, we record crucial metadata associated with the captured image, including the camera's position, orientation, focal length, and resolution. This metadata is essential for accurately positioning the virtual camera within the BIM environment, enabling a direct comparison.
- 2. BIM model filtering for ground truth extraction:** To extract the Ground Truth (GT) masks, we filter the objects within the virtual BIM environment provided by the NEXT-BIM application. This filtering leverages the Industry Foundation Classes (IFC) specifications of the objects and the HoloLens 2's intrinsic camera data (acquired in step 1). Objects are rigorously filtered based on four criteria, as illustrated in Figure 4:

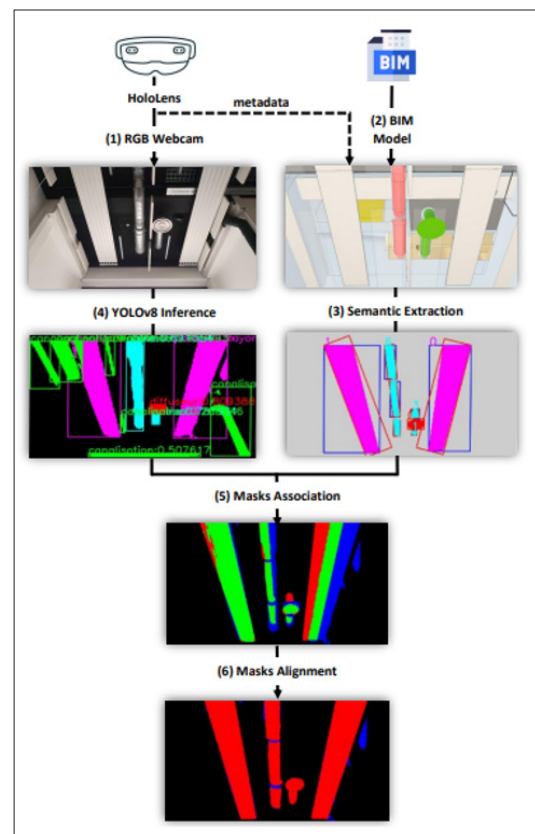


Figure 3: Process explaining the comparison of the BIM model with the prediction of the DL segmentation model in real-time.

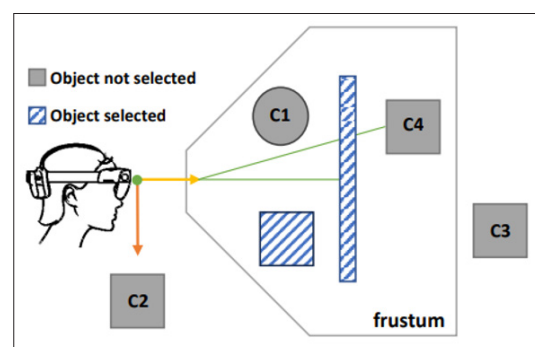


Figure 4: Illustration of the BIM model filtering process. Objects that meet all filtering criteria are shown with stripes; those that do not are shown in gray.

C1 Object type: The object must belong to a predefined target category (e.g., MEP equipment), specified by 523 the user and relevant to the inspection task.

C2 Camera frustum: The object must lie within the camera's frustum, the truncated pyramid defining the visible 3D space.

C3 Distance: The object must be within a 5-meter distance threshold, ensuring sufficient accuracy of the HoloLens 2's 3D mesh reconstruction.

C4 Occlusion: The object must have a visibility score above a threshold (e.g., 0.8), calculated based on raycasting to determine the percentage of the object's surface visible from the camera's viewpoint.

Only objects satisfying all four criteria are considered for GT mask generation.

3. Semantic extraction and GT mask generation: The filtered IFC objects are rendered within the virtual environment using distinct colors corresponding to their respective classes. This creates a visual representation of the expected visible objects. A screenshot of this rendered view is captured, providing a 2D projection of the relevant BIM elements. This screenshot then undergoes pixel-level processing to extract the GT masks. Pixels corresponding to the highlighted MEP components are identified and grouped based on their instance IDs, resulting in a set of binary masks representing the ground truth.
4. YOLOv8 inference and post-processing: The RGB image captured in step (1) is fed as input to the deployed YOLOv8 model. The model performs instance segmentation, generating a set of predicted masks. The YOLOv8 model outputs two key tensors: a detection tensor (1x8400x(4+5+32)) containing 8400 detection proposals (each with 4 bounding box coordinates, 5 confidence scores, and 32 segmentation weights), and a proto mask tensor (1x32x160x160) containing 32 prototype masks of 160x160 pixels. These are combined to generate the final instance masks. Post-processing includes extracting detection data, applying Non-Maximum Suppression (NMS) to remove redundant boxes, generating binary masks, and compiling results for comparison with ground truth.
5. Mask association: To ensure a robust comparison and handle potential false positives from the YOLOv8 model, we perform mask association before alignment. Each predicted mask is compared against all intersecting ground truth (GT) masks. This accounts for situations where a single predicted object might correspond to multiple objects in the BIM model, or vice-versa. Only predicted masks that have a non-zero intersection with at least one GT mask are considered for further processing. This step effectively filters out spurious detections that do not correspond to any element in the BIM model.
6. Mask alignment and discrepancy visualization: After the mask association step, the remaining predicted masks and their corresponding GT masks are aligned. This alignment is crucial to account for minor discrepancies between the as-designed positions in the BIM model and the actual as built positions of the MEP components, as well as potential inaccuracies in camera pose estimation. We evaluated two alignment methods.

- Centroid-based alignment: This computationally efficient method aligns the masks by translating the predicted mask so that its centroid coincides with the centroid of the GT mask. The centroid of a mask M is calculated as:

$$C_x = \frac{\sum_{(x,y) \in M^x} (x,y)}{Area(M)}, C_y = \frac{\sum_{(x,y) \in M^y} (x,y)}{Area(M)} \quad (1)$$

where (x, y) are the pixel coordinates and $Area(M)$ is the number of pixels in the mask. This method is fast but less robust to significant shape variations and rotations.

- Affine transformation alignment: This method estimates an affine transformation (translation, rotation, and scaling) that best aligns the predicted mask to the GT mask. We use the Enhanced Correlation Coefficient (ECC) algorithm [82], which finds the optimal transformation matrix A and translation vector t by minimizing the difference between the warped predicted mask and the GT mask:

$$\arg \min \| I_{GT}(x, y) - I_{Pred}(A \begin{bmatrix} x \\ y \end{bmatrix} + t) \|^2 \quad (2)$$

where I_{GT} is the GT mask image, I_{Pred} is the predicted mask image, and (x, y) are pixel coordinates. This method is more robust to shape variations but is computationally more intensive.

After alignment, discrepancies between the predicted and GT masks are visualized within the AR view. Correctly identified and aligned MEP components (those with an Intersection over Union (IoU) score above a predefined threshold - typically 0.5) have their bounding boxes rendered in red, providing immediate visual feedback to the user. The IoU is calculated as.

$$IoU = \frac{Area(P \cap G)}{Area(P \cup G)} \quad (3)$$

where P is the predicted mask and G is the ground truth mask. Objects falling below the IoU threshold, or those not detected by the model, are not highlighted, indicating potential deviations from the BIM model.

This comprehensive alignment and superposition process, combining automated GT mask extraction (with rigorous filtering), real-time DL-based instance segmentation, and robust mask alignment, allows an accurate and efficient on-site comparison between the planned BIM model and the built reality.

Experiments and Results

To evaluate the effectiveness and generalizability of the proposed methodology, we conducted a series of experiments using both synthetic and real-world data. The experiments were designed to assess the performance of the trained YOLOv8 model for instance segmentation, the accuracy of the alignment algorithms, and the overall usability of the AR-integrated system for on-site progress monitoring.

The experiments were performed using a combination of hardware and software tools. Model training and synthetic data generation were conducted on a laptop equipped with an Intel Core i7-10750H processor, 32 GB of RAM, and an NVIDIA Quadro RTX 3000 GPU. The graphics engine used for synthetic data generation was NVIDIA Isaac Sim, leveraging its Omniverse platform and USD (Universal Scene Description) format for scene representation. The YOLOv8 model was implemented using the Ultralytics API2, and the AR application was developed using the NEXT-BIM C++ framework on the Microsoft HoloLens 2 platform. The HoloLens Development Mode enabled seamless communication between the device and a paired computer for data transfer and debugging.

Synthetic Data Generation Results

This section presents results from our BIM-based synthetic data generation pipeline, demonstrating its effectiveness in creating a large, diverse, photorealistic dataset (MEP-SEG) for training deep learning models for ICPM. The pipeline leverages BIM geometry and semantics, coupled with NVIDIA Isaac Sim's rendering capabilities, to realistically simulate real-world construction environments.

MEP-SEG was generated from three diverse BIM models: an eight-story office tower, a scientific university laboratory extension, and a business school campus. This variety ensures the dataset encompasses a wide range of spatial configurations, architectural styles, and MEP system designs. BIM models were pre-processed in Revit to remove irrelevant architectural elements, optimizing rendering performance.

The automated pipeline generated 8,751 640x640 pixel images with instance segmentation masks, covering 13 common MEP component classes (e.g., pipes, ducts, cable trays). The instance distribution (Table 1) reflects real-world imbalances.

Table 1: MEP-SEG instance distribution

Class	Instances
Wall	90,801
Pipe	44,998
Floor	44,266
Circular duct	34,973
Rectangular duct	26,227
Framework	11,627
Air vent	8,585
Pole	5,131
Fan coil	4,286
Radiant panel	3,031
Ceiling	2,431
Pipe accessory	1,449
Climatic equipment	1,309

Figure 5 demonstrates the photorealistic quality of the images, achieved through physically-based materials, realistic lighting (natural and artificial), and randomized clutter. Diverse viewpoints were captured via defined inspection paths within the BIM models.

Annotations (instance segmentation, semantic labels, depth maps) are pixel-perfect and consistent with the BIM data. The entire MEP-SEG dataset was generated in approximately 9 hours. The dataset is publicly available [11].

In summary, MEP-SEG provides a high-quality, diverse, accurately annotated dataset for ICPM deep learning, addressing data scarcity and advancing automated construction progress monitoring. Its photorealism and precise annotations make it a valuable resource for developing robust computer vision solutions for the construction industry.

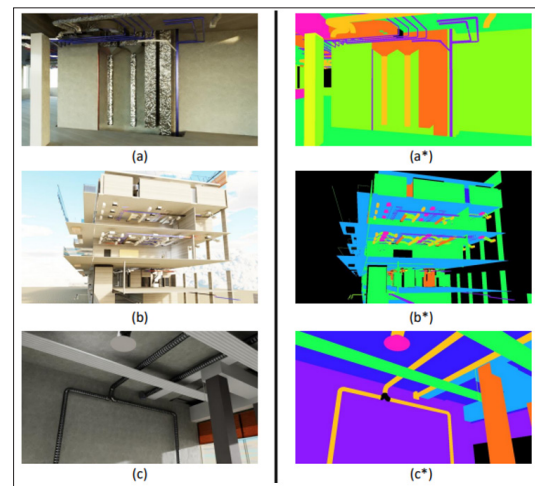


Figure 5: Synthetic image examples and instance segmentation masks. (a) RGB image. (a*) Mask. (b) RGB image (different lighting). (b*) Mask. (c) RGB image (partial occlusions). (c*) Mask.

Model Training and Evaluation

This section details the training procedure, evaluation methodology, and results for the YOLOv8 instance segmentation model on the synthetic and real-world datasets. The primary objective is to assess the effectiveness of synthetic data for training deep learning models for ICPM and to evaluate the impact of domain adaptation techniques. For evaluation purposes, real-world images were collected from inside five construction sites: three used for the MEP-SEG dataset and two new projects. Two distinct devices were used for data collection: a smartphone and the Microsoft HoloLens 2 AR glasses. A total of 350 images were acquired and manually labeled, resulting in the MEP-REAL dataset that will be used in the following experiments.

We focused on the detection of five classes: ducts, pipes, air vent, radiant panels, and fan coil units. These objects are among the most difficult to detect in MEP domain images, due to their varied shapes, textures and sizes.

Training Procedure

The YOLOv8 instance segmentation models were trained using a consistent set of hyperparameters across all experiments to ensure fair comparisons. The training configuration included:

- Optimizer: Stochastic Gradient Descent (SGD)
- Momentum: 0.937
- Weight Decay: 0.0005
- Initial Learning Rate: 0.01
- Learning Rate Schedule: Linear warmup followed by cosine annealing.
- Batch Size: 6
- Number of Epochs: 1000 (maximum)
- Input Image Size: 640x640 pixels

The learning rate schedule involved a linear warmup phase for the initial epochs to stabilize training, followed by a cosine annealing decay to gradually reduce the learning rate, aiding convergence.

To prevent overfitting and select the best performing model iteration, early stopping was employed with a patience of 50 epochs. Training was halted if the primary validation metric (typically a combination of losses or mAP) did not show improvement for 50 consecutive epochs. The model weights corresponding to the epoch with the best validation performance were saved and used for all subsequent evaluations and deployment.

Data augmentation techniques, including random rotations, scaling, horizontal flips, color jittering, and mosaic augmentation, were applied during training to enhance the model's robustness and generalization capabilities against variations encountered in real-world construction environments.

Evaluation Metrics

The performance of the trained YOLOv8 models was evaluated using standard instance segmentation metrics, including:

- Precision: The proportion of correctly predicted positive instances among all instances predicted as positive.
- Recall: The proportion of correctly predicted positive instances among all actual positive instances.
- mAP50: The mean Average Precision (mAP) calculated at an Intersection over Union (IoU) threshold of 0.5. These metrics measures the average precision across all classes and provides an overall assessment of the model's detection and segmentation accuracy.
- mAP50-95: The mAP calculated across IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05. This metric provides a more comprehensive evaluation of the model's performance across different levels of localization accuracy.

Model Evaluation and Comparison - First Evaluation

For this initial evaluation, we compared two Transfer Learning (TL) approaches using YOLOv8. The first model (COCO TL) was initialized with weights pre-trained on the COCO dataset [83]. The second model (Synthetic TL) used weights from a YOLOv8 model we pre-trained on the MEP-SEG synthetic dataset. These pre-trained models were then fine-tuned on real-world data extracted from MEP-REAL. We created two versions of the MEP-REAL dataset for fine-tuning: a small dataset (S, 45 training images, 19 validation images) and a medium dataset (M, 131 training images, 43 validation images). The fine-tuning process was identical for both models and datasets. Fine-tuning typically converged after around 200 epochs (20 minutes of training) using the COCO TL model and around 180 epochs for the Synthetic TL model. Performance was evaluated on a held-out test set of 40 unseen images.

Test results are summarized in Table 2.

Table 2: Performances on COCO vs. synthetic TL on small (S) and medium (M) real datasets

Metrics	COCO TL		Synthetic TL	
S dataset	box	mask	box	mask
Precision (%)	43	51	66	64
Recall (%)	43	34	47	46

mAP50 (%)	42	38	53	49
mAP5095 (%)	26	19	37	30
M dataset	box	mask	box	mask
Precision (%)	52	52	69	63
Recall (%)	45	42	40	38
mAP50 (%)	43	41	47	43
mAP5095 (%)	29	24	30	24

As we can see on the S dataset, the model pre-trained on our synthetic data outperforms the one pre-trained on COCO in all metrics. Moreover, the gap between box and mask in terms of precision and recall is smaller with the synthetic TL method, suggesting that the model better detects the outline of objects. Looking at the results on the M dataset, we observe a less significant difference between the two methods, indicating that the more real data available, the less relevant it is to use a synthetic dataset for a pre-trained model. In conclusion, using a synthetic pre-trained model showed promising results for transferring knowledge to a small real dataset.

Model Evaluation and Comparison - Second Evaluation

For the second evaluation, we built four training datasets to explore the benefit of augmenting synthetic data with a small amount of real data:

- A purely real dataset comprising 142 real images from MEP-REAL.
- A purely synthetic dataset containing 500 carefully selected synthetic images from MEP-SEG to ensure they were the most representative.
- A mixed training dataset of 550 images, containing 500 synthetic images and 50 real images, i.e., 10% of the 500 synthetic images.
- A mixed training dataset of 600 images, containing 500 synthetic images and 100 real images, i.e., 20% of the 500 synthetic images.

In each training dataset, 80% of the images were used for training and the remaining 20% for validation. The YOLOv8 model pre-trained on the COCO image dataset served as the basis, and the same training strategy as in the previous evaluation was used. The evaluation of the precision measure on 50 new real test images from all training datasets is presented in Table 3.

Table 3: Performance of the YOLOv8 model trained on different datasets

Training dataset	Precision (%)	
Real	77	75
Synthetic	30	29
Synthetic +10% real	71	69
Synthetic +20% real	80	79

In each case, 80% of the images were used for training and 20% for validation. A YOLOv8 model pre-trained on the COCO

dataset served as the base, and the training strategy described in Section 4.2.1 was applied consistently. Performance, evaluated in terms of precision on 50 unseen real test images, is presented in Table 3. As expected, the model trained only on synthetic data showed poor generalization to real images (30% box, 29% mask precision). The model trained only on the 142 real images achieved respectable baseline performance (77% box, 75% mask). Most importantly, incorporating even a small amount of real data into the synthetic training set resulted in significant improvements. Adding just 50 real images (Mixed 10%) significantly increased accuracy (71% box, 69% mask), approaching the performance of the all-real model despite using fewer real images overall in its source pool. The Mixed-20% model, achieved the highest precision (80% box, 79% mask), surpassing the performance of the model trained on the larger dataset of 142 purely real images. This result strongly suggests the efficiency and effectiveness of the use of synthetic data. By strategically combining a large synthetic dataset with a relatively small amount of real data (in this case, 100 images), we can achieve superior performance compared to relying solely on a larger collection of manually annotated real-world images. In relation to the next section of this chapter, the model using the combination of 500 synthetic images and 100 real images is retained (Mixed-20%).

For this model, the training and validation losses for both bounding box regression (box loss) and segmentation (seg loss), are illustrated in Figure 6. The training losses consistently decreased and converged, indicating that the model was effectively learning from the training data. The validation losses also decreased initially but exhibited more fluctuation and plateaued after a certain number of epochs, suggesting the onset of diminishing returns or potential overfitting.

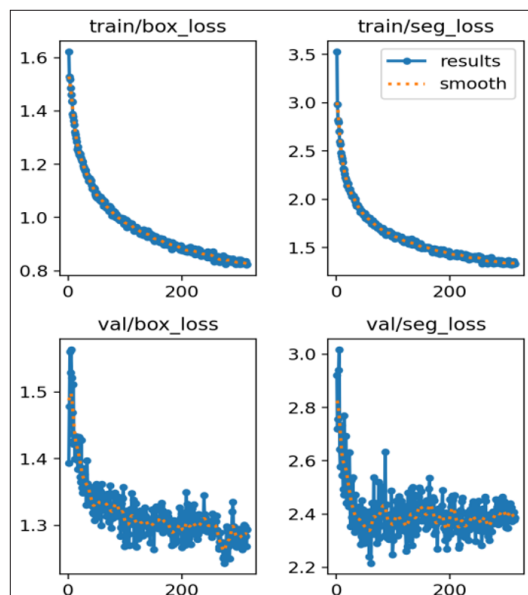


Figure 6: Training and validation loss curves for the YOLOv8 model (Mixed20% dataset). Top row: Training box loss and segmentation loss. Bottom row: Validation box loss and segmentation loss. The plots show convergence during training and the plateauing of validation loss used for early stopping.

Furthermore, after training, an analysis was performed to determine the optimal confidence threshold for balancing precision and recall, thereby maximizing the F1 score. Figure 7 shows the F1 score for each class and the average across all classes as a function of the confidence threshold, evaluated on the validation set. Based on this analysis, an optimal confidence threshold of 0.365 was identified, achieving the highest average F1 score of 0.67. This threshold was subsequently used during inference and evaluation to filter detections and ensure a balanced performance.

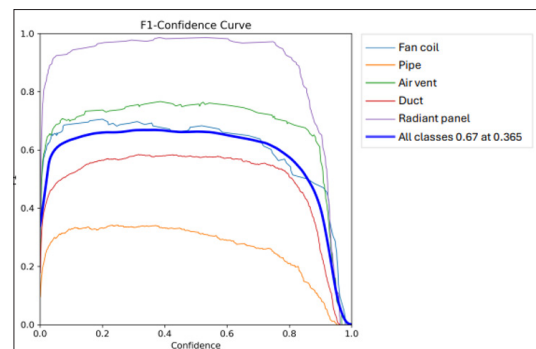


Figure 7: F1 score versus confidence threshold for each class and averaged across all classes (evaluated on the validation set). The plot indicates that the maximum average F1 score (0.67) is achieved at a confidence threshold of 0.365.

Qualitative Results

Beyond quantitative metrics, qualitative analysis of the model's predictions on unseen real-world images provides valuable insights into its practical performance. Figures 8 and 9 present representative examples from the best-performing model (Mixed-20%), illustrating both successful inferences and common failure modes, respectively.

Figure 8 demonstrates the model's capability to accurately detect and segment various MEP components across a range of challenging conditions often encountered on construction sites. As detailed in the caption, the model shows robustness to complex layouts, varying scales, visual clutter, different camera qualities, and difficult lighting scenarios (Fig. 8a-f). These successful examples validate the effectiveness of the mixed synthetic and real data training approach for achieving good generalization in practical settings.



Figure 8: Qualitative examples of successful instance segmentation results using the YOLOv8 model (Mixed-20%) on diverse real-world indoor construction scenes. (a) Scene with a high ceiling, showcasing detection of ducts and radiant panels. (b) Scene demonstrating differentiation between pipes and ducts. (c) Accurate detection of multiple fan coil units amidst clutter. (d) Performance on an image captured with a lower-end camera. (e) Long range scene with high contrast. (f) Robustness to different background.

- (e) Detection in a long-range scene with high contrast lighting.
 (f) Robustness to complex and varied background structures.

Conversely, Figure 9 highlights persistent challenges and typical failure modes. As described in the caption, these include difficulties with severe occlusion (Figure 9b), missed detections of smaller or less distinct objects (Figure 9a), misclassifications between visually similar items (Figure 9c), incomplete segmentations (Figure 9d), and sensitivity to specific material properties (e.g., black insulation, Figure 9e) or extreme lighting conditions (Figure 9f). These limitations point towards the need for further improvements, particularly in enhancing training data diversity and model robustness to handle edge cases and complex environmental factors.

Overall, the qualitative results provide a balanced perspective, confirming the model's strong potential for ICPM while clearly identifying areas for future refinement to increase reliability across the full spectrum of real-world construction scenarios.

While demonstrating strong performance in many typical construction scenarios, the identified failure modes pinpoint specific areas where targeted improvements in data and modeling can lead to enhanced reliability and broader applicability of the proposed ICPM system.

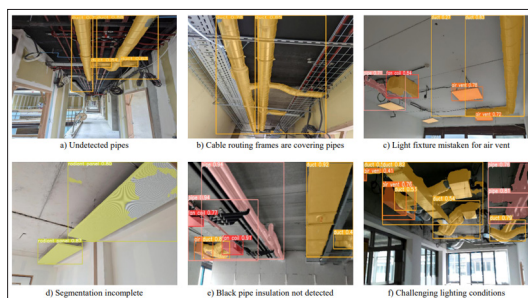


Figure 9: Qualitative examples of failure cases and limitations observed with the YOLOv8 model (Mixed-20%) on real-world indoor construction images. (a) Undetected pipes (false negatives) alongside correctly detected ducts. (b) Pipes heavily occluded by cable routing frames are missed. (c) Misclassification of a light fixture as an air vent. (d) Incomplete segmentation masks for detected radiant panels. (e) Failure to detect pipes covered in black insulation. (f) Reduced detection and segmentation accuracy under challenging lighting conditions (glare and shadow).

Mask Alignment Analysis

To quantify and visualize the discrepancies between predicted and ground truth masks, and to evaluate the effectiveness of different alignment strategies, we conducted a comparative analysis of centroid-based and affine transformation alignment methods. As described in Section 3.4.3, centroid-based alignment offers computational efficiency but is less precise, while affine transformation alignment (using the ECC algorithm) provides higher accuracy at the cost of increased computational time.

This analysis was carried out on a new dataset acquired as part of the Avignon Archive Centre project, which provides a controlled environment for evaluating alignment performance.

This database, consisting of 133 pairs of images, represents a total of 829 object occurrences divided into 5 classes: 79 fan coils, 175 ducts, 82 diffusers, 593 pipes and 40 radiant panels.

The analysis of these images by our algorithm identified 28% of the objects present in the scene (approximately 267 objects), confirming the presence of these elements on the construction site at the time of data acquisition. For each detected object, a comparison between the predicted mask and the corresponding ground truth mask was performed, including both alignment methods: centroid-based and affine transformation-based. This process generated a total of 450 compared alignments. The results, averaged across all samples, are presented below (Table 4):

Table 4: Average results of mask comparison using centroid-based and affine transformation alignment on the Avignon Archive Center Dataset

Metric	Centroid	Affine
Distance Centroids (px)	1	24
IoU	0.40	0.64
Dice Coefficient	0.54	0.76
Time (ms)	30	2110

Centroid-based alignment, which is very fast (30 ms on average), offers limited alignment accuracy, with an IoU of 0.40 and a Dice coefficient of 0.54. On the other hand, affine alignment significantly improves the alignment quality, with an IoU of 0.64 and a Dice coefficient of 0.76. However, this precision improvement comes at the cost of a significantly higher computation time (2.1 seconds on average), approximately 70 times slower than centroid-based alignment.

It is noteworthy that the centroid distance is significantly higher for affine alignment (24px versus 1px for centroid-based alignment). This can be explained by the fact that affine transformation, unlike the simple translation performed by centroid-based alignment, includes rotations and scale changes. These transformations can shift the centroid of the aligned predicted mask, even if the overlap with the ground truth mask is optimal. Consequently, centroid distance is not a relevant metric for evaluating the quality of affine alignment. IoU and the Dice coefficient, however, remain reliable indicators of the overlap quality between masks.

This experiment on a larger dataset highlights the performance and limitations of the proposed alignment methods. It confirms that affine alignment, although more computationally expensive, provides significantly higher accuracy. The choice of alignment method will thus depend on application constraints: prioritizing speed with centroid-based alignment or precision with affine alignment. The analysis of metrics across the entire dataset allows the identification of problematic cases and guides improvements in the segmentation model and alignment algorithms. It also validates the potential of the approach for work inspection and quality control, although optimizations are necessary for real-time use on devices such as the HoloLens 2.

User study and Acceptability Assessment

To comprehensively evaluate the practical usability, user acceptance, and perceived value of the AR-integrated ICPM system, a user study was conducted with a panel of 21 construction professionals. This section summarizes the key findings from this study, providing insights into the user experience, perceived benefits, and potential challenges associated with adopting the proposed technology in real-world construction settings.

Methodology

The user study employed a mixed-methods approach, combining quantitative and qualitative data collection techniques to provide a holistic understanding of user perspectives. The methodology comprised three main components:

- **Online questionnaire:** A quantitative questionnaire was administered to a panel of 21 construction professionals, all of whom had prior experience using the NEXT-BIM solution and the HoloLens 2 headset. The questionnaire utilized Likert scale questions (1-strongly disagree to 5-strongly agree) to assess usability, acceptability, comfort, and perceived effectiveness, complemented by open-ended questions for gathering qualitative feedback and suggestions.
- **Semi-structured interviews:** In-depth, semi-structured interviews were conducted with a subset of 10 questionnaire participants, selected to represent diverse roles and experiences within the construction industry. These interviews explored the user experience in more detail, looking at perceived advantages and disadvantages, integration challenges and recommendations for improvement.
- **On-site observations:** Throughout the research project and during on-site testing of the AR-integrated ICPM system, the researchers collected observational data, acting as both users and observers. These observations provided contextual insights into real-world usage scenarios, user interactions with the system, and the challenges and opportunities encountered in practical implementation.

This mixed-methods approach allowed for triangulation of data, enhancing the validity and reliability of the user study findings. The questionnaire provided quantitative measures of user perceptions, while the interviews and observations offered rich qualitative insights into the nuances of user experience and the practical implications of adopting the AR-integrated ICPM system.

Participant Demographics and NEXT-BIM Usage

Table 5 summarizes the demographics and NEXT-BIM usage patterns of the study participants. The majority of participants (47.6%) had used NEXT-BIM for over a year, indicating substantial experience with the platform. A further 28.6% had used it for between 6 months and a year. In terms of usage frequency (N=19), the most common response was "about once a month" (42.1%), followed by "about once a week" (36.8%). Participants represented a variety of roles, with BIM engineers (38.1%) being the largest group, followed by site managers (23.8%). This diverse range of roles and experience levels provides a representative sample of potential users.

Table 5: User Study Demographics and NEXT-BIM Usage (N=21, except where noted)

Characteristic	Percentage
NEXT-BIM Usage Duration	
Less than 3 months	9.5%
Between 3 and 6 months	14.3%
Between 6 months and 1 year	28.6%
More than 1 year	47.6%
NEXT-BIM Usage Frequency (N=19)	
Daily	15.8%
About once a week	36.8%
About once a month	42.1%
Less than once a month	5.3%
Primary Role on Construction Sites	
Site Manager	23.8%
Works Supervisor	9.5%
Quality Inspector	9.5%
BIM Engineer	38.1%
Design Office Manager	14.3%
Study Technician	0%
Quality Control Manager	4.8%

Quantitative Results

Analysis of the Likert scale responses, presented in Table 6, reveals a generally high level of user acceptance and perceived usefulness of the AR-integrated ICPM system. Regarding work efficiency (Question 4), a substantial majority of participants agreed (38.1%) or strongly agreed (52.4%) that the system enhanced their productivity. However, responses to the ease of integration into existing workflows (Question 1) were more divided. While 38.1% agreed and 4.8% strongly agreed with easy integration, a considerable 42.9% remained neutral, and 14.3% disagreed. This suggests that while the system is perceived as effective, further refinement may be necessary to optimize its integration with established construction processes.

Concerning usability, the system received overwhelmingly positive feedback. A significant portion of participants agreed (61.9%) or strongly agreed (28.6%) that the user interface was intuitive and easy to comprehend (Question 5). Similarly, a high proportion agreed (57.1%) or strongly agreed (33.3%) that the application was easy to utilize daily (Question 7). The initial training provided (Question 8) was deemed sufficient by a majority, with 57.1% agreeing and 23.8% strongly agreeing. The augmented reality visualization (Question 3) was also highly regarded, with 66.7% agreeing and 9.5% strongly agreeing on its clarity and ease of interpretation. This confirms the effectiveness of the AR component in presenting BIM models and instance segmentation results in a readily understandable manner. Concerning the application's functionalities (Question 6), 47.6% of participants agreed and 33.3% strongly agreed on their relevancy.

Conversely, the HoloLens 2 headset's comfort during extended use (Question 2) received a more mixed assessment. Only 23.8% agreed and 19% strongly agreed regarding comfort, while 38.1%

remained neutral, and a combined 19% disagreed. This indicates a potential area for improvement or consideration regarding prolonged use in field settings.

Table 6: User Study Results: Questionnaire Responses (N=21)

Question	Disagree (1-2)	Neutral (3)	Agree (4)	Strongly Agree (5)
1. Easy integration into workflow	3 (14.3%)	9 (42.9%)	8 (38.1%)	1 (4.8%)
2. HoloLens 2 comfort (long periods)	4 (19%)	8 (38.1%)	5 (23.8%)	4 (19%)
3. Clear & easy AR visualization	0 (0%)	5 (23.8%)	14 (66.7%)	2 (9.5%)
4. Increased work efficiency	0 (0%)	2 (9.5%)	8 (38.1%)	11 (52.4%)
5. Intuitive & easy user interface	0 (0%)	2 (9.5%)	13 (61.9%)	6 (28.6%)
6. Relevant app functionalities	0 (0%)	4 (19%)	10 (47.6%)	7 (33.3%)
7. Easy to use daily	2 (9.6%)	1 (4.8%)	12 (57.1%)	7 (33.3%)
8. Sufficient initial training	0 (0%)	4 (19%)	12 (57.1%)	5 (23.8%)

Qualitative Insights

The semi-structured interviews and on-site observations yielded rich qualitative data, providing nuanced perspectives that complement the quantitative findings. A prominent theme emerging from the interviews was the substantial time savings afforded by the system. Participants consistently reported a reduction in the time required for on-site inspections compared to traditional manual methods. One construction manager, for instance, estimated a decrease in inspection time from half a day to approximately one hour, a compelling illustration of the potential efficiency gains.

Beyond time savings, participants frequently highlighted the system's positive impact on communication and collaboration within the construction team. The AR visualization served as a shared, contextualized platform for discussing progress, identifying discrepancies, and coordinating corrective actions. This visual communication was perceived as a significant improvement over traditional reporting methods, facilitating clearer and more effective information exchange.

Furthermore, the integration of automated object detection and segmentation with the AR overlay was perceived to enhance the accuracy of progress monitoring and mitigate the risk of errors inherent in manual assessments. The ability to visually compare the as-built reality with the BIM model in real-time was considered a valuable asset for ensuring quality control and adherence to design specifications.

Despite the overwhelmingly positive feedback, participants also identified several challenges and limitations. Consistent with the quantitative findings, discomfort associated with prolonged use of the HoloLens 2 was a recurring concern. Additionally, the system's performance was acknowledged

to be contingent upon the quality and completeness of the underlying BIM model; incomplete or inaccurate models could limit the system's effectiveness. Environmental factors, such as complex geometries, cluttered environments, and suboptimal lighting conditions, were also noted as potential impediments to optimal performance. Finally, while generally perceived as user friendly, the seamless integration of the system into pre-existing workflows was identified as an area requiring further attention and potential adaptation.

The user study results indicate a strong positive reception of the AR-integrated ICPM system among construction professionals. The high levels of agreement on efficiency gains, usability, and the value of AR visualization demonstrate the system's potential to significantly improve progress monitoring practices. The mixed feedback on HoloLens 2 comfort, while a limitation, is consistent with broader user experiences with HMDs. The identified challenges, such as BIM model dependency and environmental factors, highlight areas for future development and refinement.

In conclusion, the user study provides valuable evidence for the practical usability, user acceptance, and perceived benefits of the AR-integrated ICPM system. The findings suggest that the proposed methodology has strong potential for adoption in the construction industry, offering significant improvements in efficiency, accuracy, and communication.

Discussion

Summary of Findings

The experimental evaluations and user study conducted in this research provide compelling evidence for the effectiveness and practical potential of the proposed AR-integrated ICPM methodology. The key findings demonstrate the successful integration of multiple technologies to address the challenges of traditional progress monitoring.

The MEP-SEG synthetic dataset proved to be a valuable resource for training high-performing DL models. YOLOv8 models trained on this synthetic data, especially when fine-tuned with a small amount of real-world data, achieved comparable or superior performance to models trained solely on limited real-world datasets. This highlights the potential of synthetic data to overcome the critical data scarcity bottleneck in construction applications.

Furthermore, the integration of the trained YOLOv8 model into the NEXT-BIM AR application on the HoloLens 2 successfully enabled real-time, on-site progress monitoring. The AR system provided users with an intuitive and immersive interface for comparing planned and built conditions, facilitating efficient and accurate progress assessment through the direct visualization of BIM models and instance segmentation results.

The implementation and evaluation of mask alignment methods (centroid-based and affine transformation) provided valuable tools for quantifying and visualizing discrepancies. Affine transformation alignment, while computationally more demanding, offered superior accuracy, enabling a more refined analysis of deviations from the BIM model. This highlights the

importance of choosing appropriate alignment strategies based on the specific application requirements.

Finally, the user study confirmed a generally positive perception of the AR-integrated system among construction professionals. Participants emphasized the system's usability, perceived usefulness, and its potential to improve efficiency, accuracy, and communication in progress monitoring. The AR visualization and real-time feedback were particularly well-received, demonstrating the practical value and user friendliness of the proposed solution.

These findings collectively demonstrate the successful development and validation of an innovative AR-integrated ICPM methodology, effectively leveraging BIM, synthetic data, and DL to address key challenges in construction progress monitoring.

Limitations

While the proposed AR-integrated ICPM methodology demonstrates promising results, it is crucial to acknowledge certain limitations inherent in the current study. These limitations define critical areas for future research and development, ultimately contributing to the enhanced robustness, generalizability, and practical applicability of the approach. A primary concern lies with the dataset realism and the persistent domain gap. Despite the photorealistic nature of the MEP-SEG synthetic dataset, a discernible difference remains between the visual characteristics of synthetic and real-world construction site imagery. Although domain adaptation techniques were employed, further investigation is required to fully bridge this reality gap and bolster the model's resilience to the complexities and variations encountered in authentic scenes. Factors such as sensor noise, fluctuating lighting conditions, and the unpredictable arrangement of objects in real-world environments may still present challenges to the model's generalization capabilities.

Furthermore, the accuracy and reliability of the proposed ICPM methodology are intrinsically linked to the quality and completeness of the BIM models utilized for both synthetic data generation and on-site comparison. Incomplete, inaccurate, or outdated BIM models can significantly limit the system's effectiveness and introduce discrepancies between the virtual and real-world representations. Consequently, further research should explore automated methods for validating and correcting BIM models to ensure data integrity and reliability for ICPM applications. The computational constraints of AR devices also pose a significant challenge. While the on-device deployment of the YOLOv8 model on HoloLens 2 enables real-time performance, it is inherently constrained by the limited computational resources of the mobile AR device. The complexity of the DL model, the resolution of the input images, and the object density within the scene can all impact the frame rate and responsiveness of the AR application, particularly within highly cluttered or complex indoor environments. Optimizing the model architecture, exploring model compression techniques, and leveraging hardware acceleration are therefore essential for ensuring seamless real-time performance on resource-constrained AR devices.

The scope of the current evaluation also warrants further attention. While the on-site evaluation of the AR-integrated ICPM system provided valuable user feedback and qualitative insights, it was limited to a preliminary study involving a single construction professional and specific types of construction sites. More extensive user studies, encompassing a larger and more diverse cohort of users and a broader range of construction projects and scenarios, are necessary to comprehensively assess the usability, acceptability, and practical impact of the proposed methodology in real-world settings. Finally, the study's focus on MEP components, while pertinent to indoor construction progress monitoring, represents only a subset of the building elements relevant to overall project management. Expanding the methodology to encompass a wider spectrum of construction elements, including structural components, architectural finishes, and temporary works, would enhance the comprehensiveness and applicability of the AR-integrated ICPM system for holistic construction progress assessment.

Conclusion

This research explored a novel AR-integrated methodology for indoor construction progress monitoring, combining BIM, synthetic data generation, Deep Learning (DL)-based instance segmentation, and AR visualization. Experimental validation and a user study demonstrated the initial feasibility and potential of the system to contribute to more automated progress monitoring. Key contributions include a scalable BIM-based synthetic data pipeline, the adaptation of a YOLOv8 model for MEP component segmentation, seamless integration with a HoloLens 2 AR application, and a preliminary evaluation.

Findings suggest the value of synthetic data for training robust DL models and the promise of the AR-integrated ICPM system for more efficient and reliable progress monitoring. However, it is important to acknowledge limitations related to synthetic dataset realism, reliance on accurate BIM models, and constraints inherent to AR devices. This research establishes a foundational step towards exploring automated, data-driven construction progress monitoring.

Future research will focus on addressing current limitations by improving synthetic data realism, exploring advanced domain adaptation techniques, optimizing the DL model for AR device performance, expanding the system's scope beyond MEP components, and conducting more extensive real-world deployments. This work represents an early step towards a more automated, data-driven approach to construction progress monitoring, with the potential to contribute to more efficient and accurate construction projects.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The work was conducted as part of a collaboration involving Centrale Lyon, CNRS, LIRIS (UMR5205), SPIE Building Solutions, and NEXT-BIM, who provided resources and support as detailed in the Acknowledgements.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the first author used Large Language Model technology (e.g., Llama) in order to improve language and refine phrasing. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Mathis Baubriaud: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing - Original Draft.

Stephane Derrode: Supervision, Conceptualization, Writing - Review & Editing.

Rene Chalon: Supervision, Conceptualization, Writing - Review & Editing.

Kevin Kernn: Supervision, Conceptualization.

Data Availability Statement

The synthetic dataset generated during this study (MEP SEG) is publicly available via [11]. A portion of the real-world dataset (MEP-REAL) analysed during the current study is available at <https://universe.roboflow.com/spie/gk-real>. Further data related to this study may be available from the corresponding author upon reasonable request.

Acknowledgements

The authors thank the LIRIS laboratory (UMR5205) for providing the research environment. We gratefully acknowledge SPIE Building Solutions for granting access to valuable BIM models and real-world construction sites, which were essential for data generation and validation. We also thank NEXT-BIM for their collaboration and technical support in developing the augmented reality application. The authors also thank the supervisors for their guidance throughout this research.

References

1. The future: Building with BIM, in: BIM Handbook, John Wiley & Sons, Ltd section: 9 print. 364-397.
2. Aljohani A. Construction projects cost overrun: What does the literature tell us. 2017. 137-143.
3. Ekanayake B, Wong JKW, Fini AAF, Smith P. Computer vision-based interior construction progress monitoring: A 1259 literature review and future research directions. 2021. 127: 103705.
4. Reja VK, Varghese K, Ha QP. Computer vision-based construction progress monitoring. 2022. 138: 104245.
5. Bosché F, Guillemet A, Turkan Y, Haas CT, Haas R. Tracking the built status of MEP works: Assessing the value of a Scan-vs-BIM system. Journal of computing in civil engineering. 2014. 28: 05014004.
6. Sacks R, Eastman C, Lee G, Teicholz P. BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers. John Wiley & Sons. 2018. 14.
7. Cheng JC, Chen K, Chen W. State-of-the-art review on mixed reality applications in the AECO industry. Journal of Construction Engineering and Management. 2020. 146: 03119009.
8. Schiavi B, Havard V, Beddiar K, Baudry D. BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction. Automation in Construction. 2022. 134: 104054.
9. Fang W, Ding L, Zhong B, Love PE, Luo H. Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. Advanced Engineering Informatics. 2018. 37: 139-149.
10. Nikolenko SI. Synthetic Data for Deep Learning of Springer Optimization and Its Applications, Springer International Publishing. 2021. 174.
11. Baubriaud M. MEP-SEG DATASET: SYNTHETIC IMAGES GENERATED FROM BUILDING INFORMATION MODELING (BIM).
12. Baubriaud M, Derrode S, Chalon R, Kernn K. Accelerating Indoor Construction Progress Monitoring with Synthetic Data-Powered Deep Learning. In International Symposium on Automation and Robotics in Construction. 2024. 3: 792-799.
13. Pellerin R, Perrier N. A review of methods, techniques and tools for project planning and control. International Journal of Production Research. 2019. 57: 2160-2178.
14. Omotayo T, Awuzie B, Egbelakin T, Orimoloye I, Ogunmakinde O, et al. The Construction Industry's Future, Routledge. 2024. 246-254.
15. Warszawski A. Industrialized and automated building systems: A managerial approach. Routledge. 2003. 2.
16. Akinosho TD, Oyedele LO, Bilal M, Ajayi AO, Delgado MD, et al. Deep learning in the construction industry: A review of present status and future innovations. Journal of Building Engineering. 2020. 2: 101827.
17. Kavaliauskas P, Fernandez JB, McGuinness K, Jurelionis A. Automation of construction progress monitoring by integrating 3D point cloud data with an IFC-based BIM model. Buildings. 2022. 12: 1754.
18. Rane N. Integrating building information modelling (BIM) and artificial intelligence (AI) for smart construction schedule, cost, quality, and safety management: challenges and opportunities. 2023. 16.
19. Alaloul WS, Qureshi AH, Musarat MA, Saad S. Evolution of close-range detection and data acquisition technologies towards automation in construction progress monitoring. Journal of Building Engineering. 2021. 43: 102877.
20. Rebolj D, Pučko Z, Babič NČ, Bizjak M, Mongus D. Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring. Automation in Construction. 2017. 84: 323-334.
21. Bosché F, Ahmed M, Turkan Y, Haas CT, Haas R. The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. Automation in construction. 2015. 49: 201-213.
22. Amer F, Golparvar-Fard M. Decentralized visual 3D mapping of scattered work locations for high-frequency tracking of indoor construction activities. In Construction research congress. 2018. 491-500.
23. Golparvar-Fard M, Bohn J, Teizer J, Savarese S, Peña-Mora F. Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques. Automation in construction. 2011. 20: 1143-1155.
24. Turkan Y, Bosché F, Haas CT, Haas R. Automated progress tracking using 4d schedule and 3d sensing technologies. 2011. 22: 414-421.

25. Gonzalez-Jorge H, Riveiro B, Arias P, Armesto J. Photogrammetry and laser scanner technology applied to length measurements in car testing laboratories. 2011. 45: 354-363.
26. Hasan S, Sacks R. Integrating BIM and multiple construction monitoring technologies for acquisition of project status information. *Journal of Construction Engineering and Management*. 2023. 149: 04023051.
27. Li N, Calis G, Becerik-Gerber B. Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations. 2012. 24: 89-99.
28. Atherinis D, Bakowski B, Velcek M, Moon S. Developing and laboratory testing a smart system for automated falsework inspection in construction. *Journal of Construction Engineering and Management*. 2018. 144: 04017119.
29. Shahi A, Safa M, Haas CT, West JS. Data fusion process management for automated construction progress estimation. *Journal of Computing in Civil Engineering*. 2015. 29: 04014098.
30. Pradhananga N, Teizer J. Automatic spatio-temporal analysis of construction site equipment operations using GPS data. 2012. 29: 107-122.
31. Moselhi O, Bardareh H, Zhu Z. Automated data acquisition in construction with remote sensing technologies. *Applied Sciences*. 2020. 10: 2846.
32. Paneru S, Jeelani I. Computer vision applications in construction: Current state, opportunities & challenges. 2021. 132: 103940.
33. Roberts D, Golparvar-Fard M. End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level. 2019. 105: 102811.
34. Jeelani I, Asadi K, Ramshankar H, Han K, Albert A. Real-time vision-based worker localization & hazard detection for construction. 2020. 121: 103448.
35. Fooladgar F, Kasaei S. A survey on indoor RGB-D semantic segmentation: from hand-crafted features to deep convolutional neural networks. *Multimedia Tools and Applications*. 2020. 79: 4499 - 4524.
36. Kirillov A, He K, Girshick R, Rother C, Dollar P. Panoptic segmentation. 2019. 9404-9413.
37. Yin X, Chen Y, Bouferguene A, Zaman H, Al-Hussein M, et al. A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in Construction*. 2020. 109: 102967.
38. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. *arXiv*. 2004.10934.
39. Yang K, Bao Y, Li J, Fan T, Tang C. Deep learning-based YOLO for crack segmentation and measurement in metro tunnels. *Automation in Construction*. 2024. 168: 105818.
40. He K, Gkioxari G, Dollar P, Girshick R. Mask r-CNN. *arXiv*. 1703: 06870.
41. Kufuor J, Mohanty DD, Valero E, Bosché F. Automatic MEP component detection with deep learning. In *International Conference on Pattern Recognition*. 2021. 10: 373-388.
42. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, et al. An image is worth 16x16 words. *arXiv preprint arXiv*. 2010.11929: 2020
43. Núñez-Morales JD, Jung Y, Golparvar-Fard M. Evaluation of Mapping Computer Vision Segmentation from Reality Capture to Schedule Activities for Construction Monitoring in the Absence of Detailed BIM. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. IAARC Publications. 2024. 41: 847-854.
44. Xu S, Wang J, Shou W, Ngo T, Sadick AM, et al. Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering*. 2021. 28: 3383-3397.
45. Kolar Z, Chen H, Luo X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2d images. 2018. 89: 58-70.
46. Arjmand M, Jung J, Olsen MJ, Lassiter HA, Jafari M. Conceptual Design of Advanced Construction Progress Monitoring with Terrestrial and Robotic Laser Scanning Systems.
47. Maalek R, Lichti DD, Ruwanpura JY. Automatic recognition of common structural elements from point clouds for automated progress monitoring and dimensional quality control in reinforced concrete construction. *Remote Sensing*. 2019. 11: 1102.
48. Choi M, Kim S, Kim S. Semi-automated visualization method for visual inspection of buildings on BIM using 3d point cloud. 2023. 81: 108017.
49. Wilson G, Cook DJ. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2020 Jul 5;11(5):1-46.
50. Barrera-Animas AJ, Davila Delgado JM. Generating Real-world-like labelled synthetic datasets for construction site applications. *Elsevier, Automation in Construction*. 151: 104850.
51. Richter SR, Vineet V, Roth S, Koltun V. Playing for data: Ground truth from computer games, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), *Computer Vision - ECCV*. Springer International Publishing. 2016. 102-118.
52. Becker S, Hug R, Hubner W, Arens M, Morris BT. Generating synthetic training data for deep learning-based UAV trajectory prediction, in: *Proceedings of the 2nd International Conference on Robotics, Computer Vision and Intelligent Systems*. 2021.1: 13-21.
53. Kuang Y. Deep reinforcement learning and transfer learning of robot in-hand dexterous manipulation. 2024.
54. Xu L, Liu H, Xiao B, Luo X. DharmarajVeeramani, Z. Zhu, A systemic review and evaluation of synthetic simulated data generation strategies for deep learning applications in construction. 2024 62: 102699.
55. Hong Y, Park S, Kim H. Synthetic data generation for indoor scene understanding using BIM 2020 *Proceedings of the 37th ISARC*, Kitakyushu, Japan ISBN: 9789529436347 Publisher: IAARC. 334-338.
56. Alawadhi M, Yan W. BIM hyperreality: Data synthesis using BIM and hyperrealistic rendering for deep learning. 2021. <https://arxiv.org/abs/2105.04103>.
57. Frías E, Pinto J, Sousa R, Lorenzo H, Díaz-Vilariño L. Exploiting BIM objects for synthetic data generation toward indoor point cloud classification using deep learning. *Journal of Computing in Civil Engineering*. 2022. 36: 04022032.
58. Ying H, Sacks R, Degani A. Synthetic image data generation using BIM and computer graphics for building scene understanding. *Automation in Construction*. 2023. 154: 105016.

59. Ma JW, Czerniawski T, Leite F. Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds. *Automation in construction*. 2020. 113: 103144.
60. Tremblay J, Prakash A, Acuna D, Brophy M, Jampani V, Anil C, To T, Cameracci E, Bochoon S, Birchfield S. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* 2018. 969-977.
61. Huang TW, Chen YH, Lin JJ, Chen CS. Deep learning without human labeling for on-site rebar instance segmentation using synthetic BIM data and domain adaptation. *Automation in Construction*. 2025. 171: 105953.
62. Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel CA, Cubuk ED, Kurakin A, Li CL. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*. 2020. 33: 596-608.
63. Tzeng E, Hoffman J, Saenko K, Darrell T. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 2017. 7167-7176.
64. Saito K, Kim D, Sclaroff S, Darrell T, Saenko K. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF international conference on computer vision* 2019. 8050-8058.
65. Toyin JO, Sattineni A, Wetzel EM, Fasoyinu AA, Kim J. Augmented reality in US Construction: Trends and future directions. *Automation in Construction*. 2025. 170: 105895.
66. Ahmed S. A review on using opportunities of augmented reality and virtual reality in construction project management. *Organization, technology & management in construction: an international journal*. 2018. 10: 1839-1852.
67. Hsieh CC, Chen HM, Wang SK. On-site visual construction management system based on the integration of SLAM-based AR and BIM on a handheld device. *KSCE Journal of Civil Engineering*. 2023. 27: 4688-707.
68. Fang W, Chen L, Zhang T, Chen C, Teng Z, Wang L. Head-mounted display augmented reality in manufacturing: A systematic review. *Robotics and Computer-Integrated Manufacturing*. 2023. 83: 102567.
69. Oufqir Z, El Abderrahmani A, Satori K. ARKit and ARCore in serve to augmented reality. In *2020 international conference on intelligent systems and computer vision (ISCV)*. IEEE. 2020. 9: 1-7.
70. Chi HL, Kang SC, Wang X. Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Automation in construction*. 2013. 33: 116-122.
71. Martins ACP, Castellano IR, Lenz Cesar Junior KM, Franco De Carvalho JM, Bellon FG, De Oliveira DS. BIM-based mixed reality application for bridge inspection. *Automation in Construction*. 2024. 168: 105852.
72. Kopsida M, Brilakis I. Real-time volume-to-plane comparison for mixed reality-based progress monitoring. *Journal of Computing in Civil Engineering*. 2020. 34: 04020016.
73. Khairadeen Ali A, Lee OJ, Lee D, Park C. Remote indoor construction progress monitoring using extended reality. *Sustainability*. 2021. 13: 2290.
74. Halder S, Afsari K, Serdakowski J, DeVito S, Ensafi M, Thabet W. Real-time and remote construction progress monitoring with a quadruped robot using augmented reality. *Buildings*. 2022. 12: 2027.
75. Tao B, Bosché F, Li J. Mixed Reality-based MEP construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison. *Automation in Construction*. 2024. 168: 105852.
76. Rankohi S, Waugh L. Review and analysis of augmented reality literature for construction industry. *Visualization in Engineering*. 2013. 1: 9.
77. Dallasega P, Revolti A, Schulze F, Martinelli M. Augmented reality to increase efficiency of MEP construction: a case study. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* IAARC Publications. 2021: 38: 459-466.
78. Ren S, He K, Girshick R, Sun J. Faster rcnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*. 2015. 28.
79. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 779-788.
80. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2023: 4015-4026.
81. Lin TY, Maire M, Belongie S, Hays J, Perona P, et al. Microsoft coco: Common objects in context. In *European conference on computer vision* Cham: Springer International Publishing. 2014. 740-755.
82. Evangelidis GD, Psarakis EZ. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE transactions on pattern analysis and machine intelligence*. 2008. 30: 1858-1865.