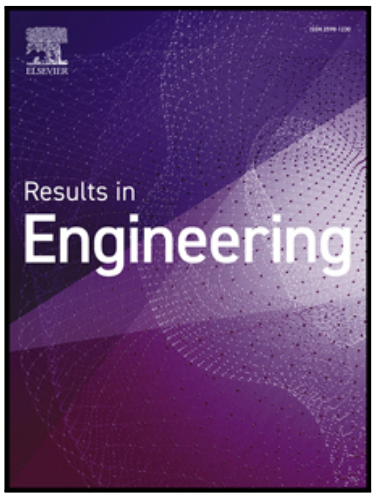


AR-based MEP Progress Monitoring using BIM and Synthetic Data

Mathis Baubriaud, Stéphane Derrode, René Chalon, Kevin Kernn

PII: S2590-1230(26)01414-3
DOI: https://doi.org/10.1016/j.rineng.2026.110380
Reference: RINENG 110380

To appear in: *Results in Engineering*

Received date: 27 August 2025
Revised date: 2 April 2026
Accepted date: 2 April 2026

Please cite this article as: Mathis Baubriaud, Stéphane Derrode, René Chalon, Kevin Kernn, AR-based MEP Progress Monitoring using BIM and Synthetic Data, *Results in Engineering* (2025), doi: https://doi.org/10.1016/j.rineng.2026.110380

Highlights

AR-Enhanced Indoor Construction Progress Monitoring using BIM and Synthetic Data

Mathis Baubriaud

- Proposed AR system enables real-time (<500ms) on-site MEP discrepancy detection.
- BIM pipeline generates 8.7k synthetic MEP images/masks to solve data scarcity.
- YOLOv8 achieves 79
- User study (N=21) confirms high usability and acceptance by site professionals.

AR-based MEP Progress Monitoring using BIM and Synthetic Data

Mathis Baubriaud[a,b], Stéphane Derrode[a], René Chalon[a], Kevin Kernn[b]

[a] *Centrale Lyon, CNRS, Claude Bernard Lyon 1, INSA Lyon, Lumiere Lyon 2, LIRIS, UMR5205, 69130, Ecully, France*
[b] *SPIE Building Solutions, F-69320, Feyzin, France*

Abstract

Manual inspection of indoor construction sites is labor-intensive and prone to human error. Currently, automating this process using deep learning is limited by the lack of large-scale, annotated datasets for complex mechanical, electrical, and plumbing (MEP) environments. This paper proposes an indoor construction progress monitoring framework that integrates a Building Information Modeling (BIM)-driven synthetic data pipeline with augmented reality (AR). To address the issue of limited data, we created an automated pipeline with NVIDIA Isaac Sim to generate the MEP-SEG dataset. This dataset contains 8,751 photorealistic synthetic images with automated pixel-level annotations. A YOLOv8 instance segmentation model trained via a domain adaptation strategy using this synthetic data and a 20% of available real-world data, achieved a Mean Average Precision (mAP_{50}) of 79.2% ± 1.4% on real site images. The model demonstrated resilience to varied occlusion and lighting; however, performance remains sensitive to low-contrast conditions below 300 lux. We deployed the trained model on a HoloLens 2 AR headset to provide a visual analysis of discrepancies between the as-built state and the BIM design at approximately two frames per second. A user study with 21 construction professionals indicates the system's potential for practical adoption, yielding positive feedback regarding usability and perceived efficiency. Our study shows that synthetic data can reduce the real-world annotation burden by 80% for certain construction tasks. This provides a framework to narrow the "reality gap" for automated on-site progress monitoring.

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

1. Introduction

1.1. Background and Problem Statement

The Architecture, Engineering, and Construction (AEC) industry faces escalating pressure to enhance efficiency, accuracy, and safety amidst increasingly complex project demands [1]. Within this context, Indoor Construction Progress Monitoring (ICPM) is critical for adhering to strict schedules and budgets [2]. However, traditional ICPM relies heavily on manual inspections and documentation. These methods are labor-intensive, time-consuming, and prone to human error, often yielding reports that are obsolete by the time they are filed [3, 4].

The challenge of monitoring is most acute in the installation of Mechanical, Electrical, and Plumbing (MEP) systems. MEP works represent 30 to 40% of total construction costs and frequently lie on the critical path [5]. Unlike planar structural elements such as walls or slabs, MEP monitoring presents unique and severe difficulties. First, components exhibit diverse geometric complexity, ranging from cylindrical pipes to rectangular ducts and compact junction boxes. These often have small cross-sections that challenge standard detection algorithms. Second, modern buildings contain thousands of interconnected components packed into confined ceiling plenums, creating heavy occlusion not found in structural monitoring. Finally, while datasets exist for structural defects [6], annotated imagery for MEP components remains scarce due to the specialized domain knowledge required for classification.

Building Information Modeling (BIM) provides a centralized digital platform to address these complexities [7]. To bridge the gap between the BIM model and the physical site, Augmented Reality (AR) has emerged as a promising visualization tool [8]. However, existing AR solutions typically rely on manual alignment and subjective visual inspection, limiting their utility for automated, quantitative progress tracking [9].

Concurrently, Computer Vision (CV) and Deep Learning (DL), specifically instance segmentation, offer a pathway to automate ICPM [10]. Recent studies have demonstrated the potential of these technologies: Li et al. [11] highlighted the shift toward data-driven quality control, while Xie et al. [12] utilized smartphone videogrammetry for piping reconstruction. In the realm of Mixed Reality (MR), Boan et al. [13] proposed frameworks integrating digital twins for discrepancy detection.

Despite these advancements, a critical bottleneck remains: the scarcity of large-scale, semantically annotated datasets for cluttered indoor environments prevents the robust deployment of DL models [14]. While techniques like semi-supervised domain adaptation have been explored [15], they still necessitate substantial real-world data collection. In contrast, purely synthetic training data often fail to generalize due to the "reality gap", which is the discrepancy between clean digital renders and the chaotic visual conditions of active construction sites involving variable lighting, dust and clutter.

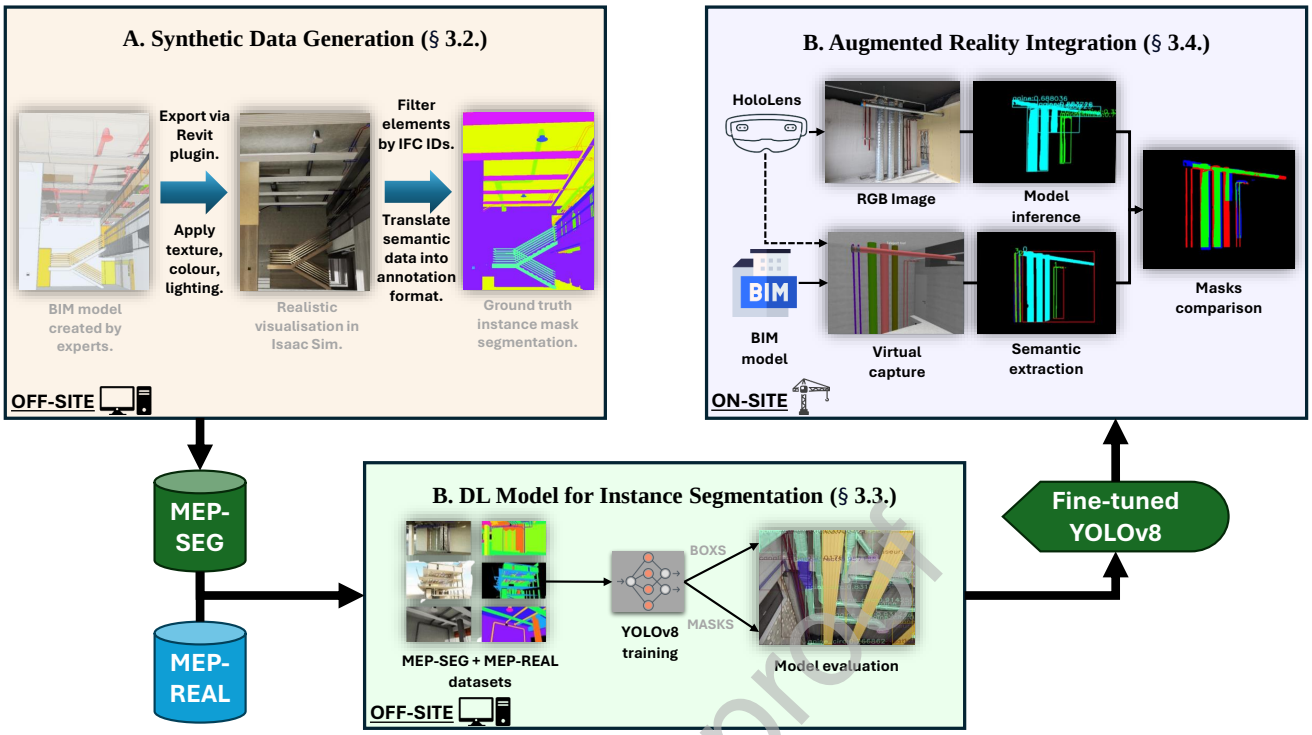


Figure 1: Overview of the proposed approach.

1.2. Research Objectives and Contributions

To address the limitations of current ICPM methods identified above, this research aims to validate the hypothesis that photorealistic synthetic data, generated entirely from BIM, can effectively enhance real-world training data for complex MEP segmentation tasks when paired with minimal domain adaptation.

In response to the identified gaps, the novelty of this work lies in the engineering of a closed-loop "BIM-to-Reality" pipeline that effectively mitigates the cold-start problem by reducing real-world annotation requirements by 80%. Unlike previous studies that focus on geometric reconstruction [12] or general inspection frameworks [13], we introduce a specific methodology for generating "infinite" labeled training data for instance segmentation. A key limitation of this approach, which we critically analyze, is the domain shift between synthetic renderings and the chaotic visual nature of active construction sites (e.g., lighting, dust, occlusion). We address this by quantifying the minimal amount of real-world data required to fine-tune the model for deployment.

The primary contributions of this paper are distinguished as follows:

Primary scientific contributions:

- Validation of a data-effective domain adaptation strategy: We demonstrate that a YOLOv8 model pre-trained on our synthetic dataset requires only 20% of the typical real-world data to achieve high-performance segmentation (79% mAP_{50}), effectively quantifying the value of synthetic pre-training for construction tasks.
- Development of the MEP-SEG dataset: We address the community-wide data scarcity bottleneck by creating and releasing a large-scale (8,751 images), auto-labeled synthetic dataset for indoor MEP components, enabling future research in semantic scene understanding.

Supporting technical implementations:

- Automated synthetic data pipeline: We developed a streamlined pipeline using NVIDIA Isaac Sim to convert BIM files into photorealistic training environments with pixel-perfect ground truth annotations.
- Real-Time AR integration: We engineered an end-to-end deployment architecture on HoloLens 2 that integrates the DL model with the NEXT-BIM application, enabling approximately 500ms latency for on-site progress monitoring.
- User-Centric evaluation: We conducted a usability study with 21 professionals to validate the system's practical acceptability and identify human-computer interaction barriers in field conditions.

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.

1.3. 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 [16], 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.

2. Literature Review

2.1. 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 [17]. Traditional methods, including manual site inspections and paper-based documentation, are not only inefficient, but also introduce subjectivity and increase the likelihood of errors [18]. The complexity and dynamism of indoor construction environments, especially those with dense MEP installations, make comprehensive and accurate manual monitoring practically unsustainable [19, 20]. As a result, delays in reporting, reactive project management and increased risk of cost overruns and schedule delays are common outcomes [21]. 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 [22]. Reality capture techniques, such as laser scanning and photogrammetry, offer accurate 3D representations of as-built conditions [23, 24, 25]. These methods enable Scan-vs-BIM comparisons for deviation analysis [26, 27, 28]. However, they often require specialized equipment, skilled operators, and extensive post-processing, hindering real-time, on-site application and scalability [29]. Sensor-based tracking systems (RFID, UWB, BLE, GPS) provide real-time data on material flow, equipment utilization, and personnel location [30, 31, 32, 33, 34]. 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 [35]. 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 [36]. 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 [37, 38]. Semantic segmentation provides a pixel-level classification of the scene, enabling the identification of different building materials and elements [39]. 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 [40]. Various Convolutional Neural Network-based (CNN) models have been explored for these tasks, including the You Only Look Once (YOLO) family of object detectors [41, 42], Mask R-CNN [43, 44], and, more recently, Vision Transformers (ViTs) [45, 46]. While these models have shown promising results, their application to ICPM, particularly for indoor MEP systems, is often slowed down by the lack of large, labeled datasets [47]. 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 [48, 49, 50] and MEP system progress tracking [51]. However, many of these approaches rely on limited real-world data, which restricts their applicability and effectiveness in various construction scenarios. The requirement for extensive manual annotation of real-world images also limits the scalability of these methods. The scarcity of this data motivates the exploration of synthetic data generation techniques, which will be discussed in the next section.

2.2. 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 [52]. This is particularly crucial for training robust DL models for complex tasks like instance segmentation of MEP components in cluttered indoor environments [53]. Various methods exist for generating synthetic data, including graphics engine-based approaches and hybrid methods that combine synthetic and real-world data [54, 55]. However, for construction progress monitoring, BIM-based synthetic data

generation offers significant advantages [56]. 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 [57]. 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 [58]. 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 [59]. This automated annotation is a key advantage, enabling the creation of large-scale datasets with minimal effort [60]. Our approach builds upon this foundation, utilizing a streamlined pipeline and focusing specifically on the detailed representation of MEP components within complex indoor scenes [61]. 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 [62]. DL models trained solely on synthetic data may exhibit limited generalization performance when deployed in real-world scenarios. Therefore, domain adaptation techniques are crucial to bridge this gap and improve model transferability [63]. Recent work by Tran et al. [15] has demonstrated the efficacy of semi-supervised learning for domain adaptation in construction, specifically for long-tailed object detection in safety monitoring. Similarly, Gomaa et al. [64] utilized semi-automated dataset construction to enhance YOLOv8 performance through advanced domain adaptation. While these methods show promise, our research specifically investigates the "sim-to-real" transfer for instance segmentation of MEP components, quantifying the minimal real-world data required to fine-tune models pre-trained on photorealistic BIM-generated synthetic data. 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 [65]. Common approaches include Unsupervised Domain Adaptation (UDA) using adversarial training [66], and Semi-Supervised Domain Adaptation (SSDA) leveraging a small amount of labeled real-world data [67].

2.3. Augmented Reality in Construction

Augmented Reality (AR) offers significant potential for enhancing construction processes by overlaying digital information onto the real-world view [68, 69]. 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 [70]. 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 [71, 72]. The choice of device depends on factors such as user mobility, environmental conditions, and the required level of immersion [73]. Several studies have demonstrated the use of AR for on-site progress monitoring. For example, Martins et al. [74] proposed an AR-based framework for bridge inspection, while Kopsida and Brilakis [75] developed a system for real-time volume-to-plane comparisons. More recently, research has moved towards fully autonomous and integrated MR frameworks. Boan et al. [13] introduced an autonomous Mixed Reality framework that integrates digital twins for real-time construction inspection, enabling human-in-the-loop decision making. Additionally, Kuo et al. [76] proposed integrating BIM and AR with Visual Simultaneous Localization and Mapping (VSLAM) to improve the spatial accuracy of site inspections. In the domain of defect detection, Shojaei et al. [77] utilized Mixed Reality combined with DL to automate concrete crack detection, further illustrating the convergence of AI and AR. 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 [78]. This process can be time-consuming, subjective, and prone to errors [79]. 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 [5]. Our work addresses these limitations by integrating a DL-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 ICPM.

2.4. Comparative Analysis and Research Gaps

Table 1 systematically compares our approach with closely related works in MEP monitoring and synthetic data generation, revealing three critical gaps that our research addresses:

First, while BIM-based synthetic data generation has been explored [61, 60], these methods do not address MEP-specific challenges such as metallic reflectance and dense clutter. Most MEP-focused studies still rely on limited real-world datasets [25, 44]. Second, domain adaptation research [63, 15] demonstrates transferability but fails to quantify the minimal real-world data volume required, leaving practitioners without a clear cost-benefit metric. Finally, a gap exists between laboratory performance and field usability. While some frameworks integrate mixed reality for inspection [13], they lack automated segmentation, whereas segmentation studies frequently lack on-site validation via augmented reality.

3. Methodology

3.1. 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)

Table 1: Systematic comparison with related works on MEP progress monitoring and synthetic data approaches.

Study	MEP Focus	Synthetic Data	Domain Adapt.	AR Integration	Dataset Public	Quantified Efficiency
Bosché et al. [25]	✓(Pipes)	×	×	×	×	×
Kufuor et al. [44]	✓(Limited)	×	Transfer Learning	×	×	×
Ma et al. [61]	×(General Indoor)	✓	×	×	×	×
Ying et al. [60]	Partial	✓	×	×	×	×
Huang et al. [63]	×(Rebar)	✓	✓(UDA)	×	×	×
Tran et al. [15]	×(Safety)	×	✓(SSDA)	×	✓	×
Boan et al. [13]	✓	×	×	✓(MR)	×	×
Our Work	✓	✓	✓(Quantified)	✓	✓	✓

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 photoreal-istic 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-BIM[1], 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, on-site visualization of BIM models and instance segmentation results overlayed 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.

3.2. Methodological Positioning and Novel Contributions

To precisely delineate our contributions, Table 2 compares our pipeline against state-of-the-art approaches across key technical dimensions.

Our primary methodological novelty lies not in individual algorithmic components but in the engineering integration of several specialized techniques designed for real-world deployment. First, we utilize MEP-tailored synthetic generation; unlike general indoor scene synthesis, our pipeline incorporates domain-specific randomization, such as metallic PBR shaders and extreme lux ranges. These additions are specifically validated through ablation studies to ensure they accurately represent complex construction environments.

Furthermore, we provide a rigorous data efficiency quantification that is often missing from prior Unsupervised Domain Adaptation (UDA) research. By systematically varying real-world data ratios, from 10% to 100%, we establish the minimal sufficient labeled dataset required for performance. This allows for a pragmatic cost-benefit analysis, helping teams understand the threshold where additional data labeling yields diminishing returns.

Finally, we implement a geometric alignment strategy for AR that addresses the unique challenges of BIM-to-reality pose discrepancies. By utilizing a dual-alignment approach (centroid + affine ECC), our framework accounts for the SLAM drift and initial calibration errors common in active construction sites, moving beyond the limitations of lab-controlled Mixed Reality studies. Collectively, these contributions constitute a robust engineering framework for practical MEP monitoring rather than isolated algorithmic advances.

3.3. Synthetic Data Generation

Our methodology relies on the automated generation of a large, diverse synthetic dataset to train robust DL models for ICPM, thereby reducing dependence on scarce real-world labeled data. The pipeline leverages the geometric and semantic richness of BIM models alongside NVIDIA Isaac Sim's photo-realistic rendering capabilities.

3.3.1. BIM Model Provenance and Preparation

The process begins by selecting and preparing BIM models that represent indoor construction environments. The BIM models utilized in this study were obtained from industry partners and cover a diverse range of typologies, including office towers, scientific laboratories, and commercial retail spaces. This variety ensures the dataset encompasses a wide range of spatial configurations, architectural styles, and MEP system designs, enhancing the diversity of the training distribution.

To ensure relevance for MEP monitoring, a filtering process isolates essential components (*e.g.*, ducts, pipes, cable trays, HVAC units) and removes non-essential architectural elements. This optimization preserves computational resources while maintaining geometric fidelity. Using built-in BIM software functionalities (Revit), models are audited and exported to the Universal Scene Description (USD) format for full compatibility with the graphics engine.

3.3.2. Virtual Environment Setup and Randomization

The prepared BIM models are imported into NVIDIA Isaac Sim. To address the "sim-to-real" gap, we employ Domain Ran-

[1]https://next-bim.com/

Table 2: Technical comparison of proposed methodology with closely related approaches.

Dimension	Huang et al. [63]	Tran et al. [15]	Boan et al. [13]	Our Approach
Application	Rebar inspection	Safety monitoring	MEP presence check	MEP installation verification
Synthetic Pipeline	Blender + manual	N/A	N/A	Isaac Sim + automated BIM import
Domain Randomization	Lighting only	N/A	N/A	Lighting + materials + clutter + HDR
Segmentation Approach	Mask R-CNN	YOLOv8 (detection)	Rule-based	YOLOv8-seg
Domain Adaptation	UDA (unlabeled target)	Semi-supervised (pseudo-labels)	N/A	Quantified SSDA (20% labeled)
Evaluation Metric	AP improvement	Long-tail detection	Presence/absence	Data efficiency curve + mAP
Deployment	Offline analysis	Offline analysis	HoloLens (manual)	HoloLens (automated)
Public Dataset	No	No	No	**Yes (MEP-SEG)**

Figure 2: Snapshots of three BIM projects imported into the graphics engine. These environments provide the geometric ground truth for synthetic data generation.

domization (DR) techniques to systematically vary the visual properties of the scene.

- Texture and material: Materials are assigned using the NVIDIA Omniverse API. While MEP components retain realistic textures (*e.g.*, galvanized steel), background elements (walls, floors) are assigned randomized textures to prevent the model from learning spurious correlations.
- Lighting: Lighting is configured to simulate both natural and artificial illumination with varied intensity (300-1000 lux), color temperature (3000K-6500K), and position of light sources to mimic the harsh, dynamic lighting conditions often found on construction sites.
- Clutter: To replicate real construction environments, synthetic clutter objects (e.g., tools, ladders, debris boxes) are procedurally injected into the scene. These elements create realistic occlusions, challenging the DL model to accurately detect and segment MEP components even when partially hidden.

3.3.3. Virtual Camera Configuration

To capture diverse viewpoints, we configure a virtual camera within Isaac Sim that mimics the specifications of a typical mobile device used for on-site inspection. Key parameters include:

- Field of View (FOV): Calibrated to match a handheld camera, ensuring synthetic images capture a representative, realistic portion of the indoor scene.
- Resolution: Standardized to 640×640 pixels to balance image quality with computational efficiency during rendering and training.
- Distortion parameters: Radial and tangential distortion values are randomly sampled within a realistic range to simulate lens imperfections, enhancing the diversity and realism of the dataset.

Positioning is determined by defining routes that replicate the movement of a worker inspecting a site. While grid-based viewpoints are possible, we utilize the API to generate manually drawn routes, providing a more natural representation of human movement. This strategy ensures systematic coverage of the virtual environment while avoiding repetitive or unnatural viewpoints.

3.3.4. Automated Annotation Generation

A major advantage of synthetic data is the immediate availability of ground truth. The annotations for the synthetic dataset were generated using Isaac Sim's internal labeling system, which defines a unique semantic label for each BIM object ID. This allows for the generation of pixel-perfect instance segmentation masks.

Using BIM metadata and ray tracing, the pipeline produces a comprehensive set of labels alongside the visual data:

- Instance segmentation: Generated by identifying unique object IDs per pixel.
- Semantic segmentation: Classifies pixels based on predefined BIM object classes (*e.g.*, ducts, pipes).
- Depth maps: Provide geometric scene information to support depth-aware monitoring.
- Bounding boxes: Derived directly from the segmentation masks for object detection tasks.

3.3.5. Image Capture and Preprocessing

Following the environment and sensor configuration, an automated Python script within Isaac Sim triggers the rendering process. The engine captures photorealistic RGB images for each camera pose, incorporating the variations in lighting, materials, and clutter defined previously. To maximize diversity, camera angles are varied across different BIM spaces, and date-time settings are randomized to simulate shifting sun positions.

Preprocessing is kept minimal to allow the DL model to learn from raw data features. Images are resized to the fixed training resolution (640×640 pixels) using bilinear interpolation. No additional noise reduction or artificial enhancement is applied, preserving the realistic imperfections of the synthetic capture.

3.4. Deep Learning Model for Instance Segmentation

Accurate and robust instance segmentation of MEP components is crucial for effective ICPM. To achieve this, we systematically evaluated the segmentation performance of five architectures on a 500-image MEP validation subset (Table 3). To ensure a fair comparison, all models were initialized with weights pre-trained on the COCO dataset and subsequently fine-tuned on this MEP subset from scratch under identical hyperparameter configurations.

YOLOv8m-seg achieved optimal accuracy-speed tradeoff for on-device deployment (52ms enables 2 FPS on HoloLens Snapdragon 850). 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) [45] and the Segment Anything Model (SAM) [81] 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. To address the reality gap between synthetic and real-world data, we employed a two-stage training strategy incorporating domain adaptation:

Table 3: Architecture selection on MEP validation set (500 images, 5 classes).

Architecture	mAP_{50} (%)	Inference (ms)	Params (M)
Faster R-CNN [80]	76.2	189	41.3
Mask R-CNN [43]	78.5	215	44.2
YOLOv5m-seg	75.8	45	26.4
YOLOv8m-seg	**79.1**	**52**	**27.3**
SAM [81]	72.3	412	93.7

1. 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 3.3. 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.
2. 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 [82] 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. Specific training details, hyperparameters, and evaluation metrics are presented in Section 4.

3.5. Augmented Reality Integration and On-site Comparison

To bridge the gap between the digital model and the physical construction site, we integrated the trained DL 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.

3.5.1. 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, 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.

3.5.2. 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.

3.5.3. BIM-Real World Alignment

Accurate registration between the BIM and the physical world is a prerequisite for discrepancy analysis. The alignment process is handled by the NEXT-BIM platform using a robust "3-Surface Calibration" method. Upon initialization, the user selects their approximate location in the building from a 3D external view. The view switches to a first-person immersive mode, and the user is prompted to manually align three orthogonal surfaces: the floor (elevation), a main wall (orientation), and a secondary element like a beam or perpendicular wall (translation). Following this initial calibration, the system employs a continuous alignment algorithm that constantly checks the correspondence between planar surfaces in the camera frustum and the BIM geometry to correct for SLAM drift in real-time.

3.5.4. 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:

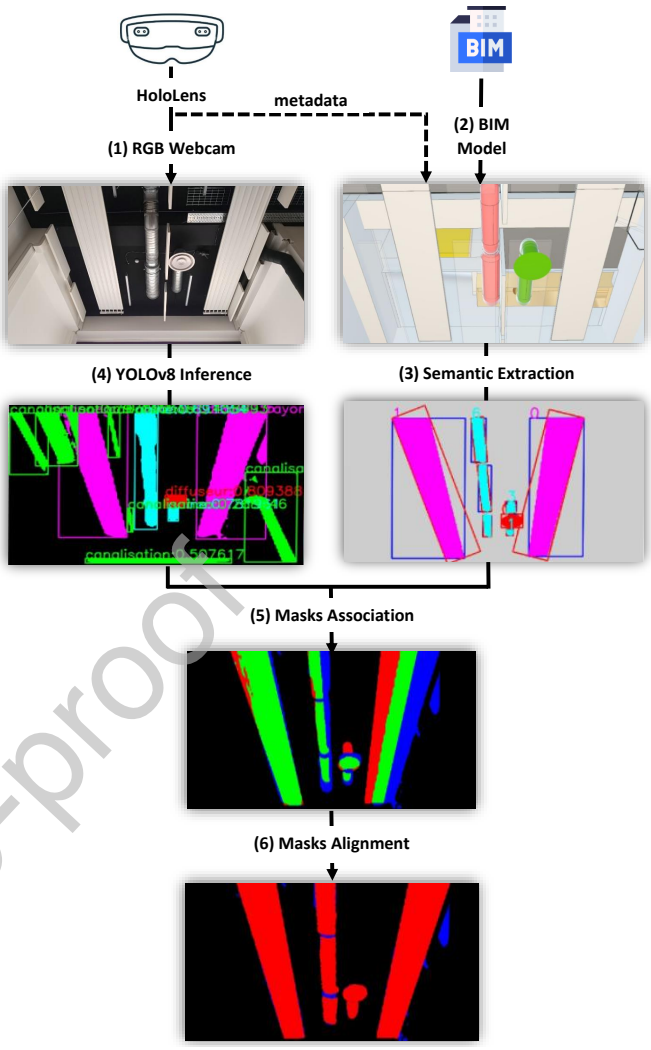


Figure 3: Process explaining the comparison of the BIM model with the prediction of the DL segmentation model in real-time.

(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).

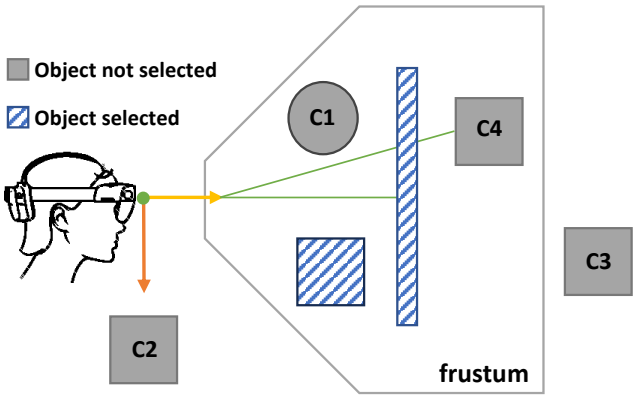


Figure 4: Illustration of the BIM model filtering process. Objects that meet all filtering criteria are shown with stripes; those that do not are shown in gray.

Objects are rigorously filtered based on four criteria, as illustrated in Figure 4:

- **C1** Object type: The object must belong to a predefined target category (*e.g.*, MEP equipment), specified by 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. This constraint is hardware-driven: according to Microsoft's specifications for the HoloLens 2 Time-of-Flight (ToF) depth sensor, and validated by spatial mapping studies [83], the spatial mesh accuracy significantly degrades beyond this range, leading to unreliable alignment between the BIM and the physical environment.
- **C4** Occlusion: The object must have a visibility score (V_s) above a threshold of 0.8. The visibility score is defined as the ratio of an object's visible pixels in the rendered frame (P_{vis}) to its total projected pixels if no occlusions were present (P_{total}):

$$V_s = \frac{P_{\text{vis}}}{P_{\text{total}}} \quad (1)$$

Objects with $V_s < 0.8$ are excluded to prevent matching against heavily occluded elements, where visual ambiguity could lead to false positives during the comparison process.

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

(3) Real-Time Ground Truth generation (virtual capture): To extract the GT masks for comparison, we implemented a real-time rendering pipeline on the HoloLens. When the user triggers an inspection, the application switches the rendering mode to "false color", where every filtered BIM object is rendered in a unique, solid RGB color corresponding to its ID. A screenshot of this view is captured instantly. This image is then processed using OpenCV to extract binary masks for each unique color. This method ensures that the ground truth masks geometrically align with the camera frustum based on available spatial metadata, occlusion, and optical characteristics of the user's current viewpoint.

(4) 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.

(5) 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 (1×8400×(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 (1×32×160×160) containing 32 prototype masks of 160×160 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.

(6) 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.

(7) 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:

$$C_x = \frac{\sum_{(x,y)\in M} x}{\text{Area}(M)}, \quad C_y = \frac{\sum_{(x,y)\in M} y}{\text{Area}(M)} \tag{2}$$

  where (x, y) are the pixel coordinates and $\text{Area}(M)$ 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 [84], which finds the optimal transformation matrix $\mathbf{A}$ and translation vector $\mathbf{t}$ by minimizing the difference between the warped predicted mask and the GT mask:

$$\arg\min_{\mathbf{A},\mathbf{t}} \| I_{\text{GT}}(x, y) - I_{\text{Pred}}(\mathbf{A}\begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{t}) \|^2 \tag{3}$$

  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:

$$\text{IoU} = \frac{\text{Area}(P \cap G)}{\text{Area}(P \cup G)} \tag{4}$$

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.

3.5.5. Progress Quantification Logic

To translate the instance segmentation results into actionable progress monitoring data, we define a binary status for each MEP component. An element E_{BIM} is considered "Installed" if the maximum IoU between its ground truth mask and any predicted mask M_{pred} exceeds the threshold $\tau_{IoU} = 0.5$. Elements with IoU < 0.5 or no corresponding detection are flagged as "Missing" or "Deviating".

The construction progress P for a given zone is quantified as the ratio of installed elements to the total expected elements within the verified camera frustums:

$$P = \frac{\sum_{i=1}^{N} (\text{IoU}_i > \tau_{IoU})}{N} \tag{5}$$

where N is the total number of visible BIM objects expected in the current view. This metric allows for the automatic generation of a "percent-complete" report for the inspected area.

Relationship to engineering accuracy criteria. It is important to note that IoU is a visual overlap proxy, not a dimensional tolerance in the sense of construction standards (e.g., ±10 mm pipe positioning per ISO 7268 or national equivalents). An IoU of 0.5 corresponds broadly to a gross presence/absence check: a correctly installed duct in the right plenum zone will satisfy this criterion even if its centreline deviates by several centimetres. For macro-level schedule progress reporting, this threshold is appropriate. For precision deviation measurement (e.g., verifying that a pipe is within tolerance of its BIM-designed centreline), the current system—limited by HoloLens 2 camera resolution and SLAM drift of typically 2–3 cm—is not a substitute for millimetric survey instruments. Future work integrating LiDAR point-cloud registration could complement the AR workflow with quantitative dimensional tolerances, narrowing the gap between visual inspection and macro-level progress assessment.

4. 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.

4.1. Hardware and Software Setup

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 API [2], and the AR application was developed using the NEXT-BIM C++ framework on the Microsoft HoloLens 2 platform.

[2] https://www.ultralytics.com/

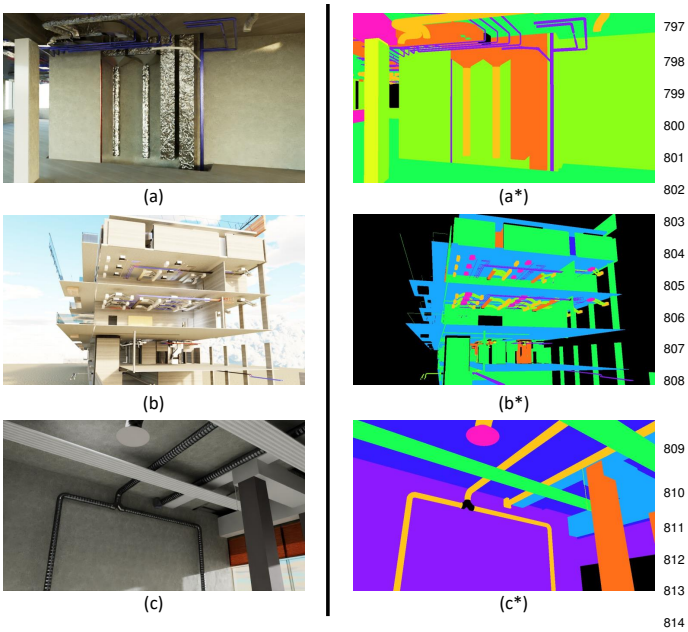


Figure 5: Synthetic image examples and instance segmentation masks. (a) RGB image. (a*) Mask. (b) RGB image (different lighting). (b*) Mask. (c) RGB image (partial occlusions). (c*) Mask.

4.2. Dataset Generation and Collection

4.2.1. Synthetic Data (MEP-SEG)

Our BIM-based pipeline generated the MEP-SEG dataset to address data scarcity. Derived from three diverse BIM models (an office tower, a laboratory, and a campus), the dataset contains 8,751 photorealistic 640×640 images. The pipeline, which ran for approximately 9 hours, produced pixel-perfect instance segmentation masks for 13 MEP classes. The distribution reflects real-world imbalances, with walls and pipes being the most frequent classes (Table 4). Figure 5 illustrates the high variance in lighting and clutter achieved through domain randomization. The dataset is publicly available [85].

Table 4: 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

4.2.2. Real-World Data (MEP-REAL)

To validate the model and perform domain adaptation, real-world images were collected from five construction sites (three matching the synthetic source BIMs and two new projects). Using both a smartphone and the HoloLens 2, we acquired and manually labeled 350 images (MEP-REAL). We focused on detecting five challenging classes: ducts, pipes, air vents, radiant panels, and fan coil units. While 350 images represent a limited dataset for standard deep learning paradigms, it serves here to effectively quantify the minimal real-world annotation effort required to bridge the “sim-to-real" gap when supported by a massive synthetic backbone.

4.3. Deep Learning Model Assessment

4.3.1. Training Procedure

Table 5 details the specific hyperparameters used for training. These were selected based on ablation studies and standard recommendations for the YOLO architecture to ensure optimal convergence and generalization.

Table 5: **Training hyperparameters for YOLOv8.**

Hyperparameter	Value
Model Architecture	YOLOv8m-seg
Optimizer	SGD
Initial Learning Rate (lr_0)	0.01
Momentum	0.937
Weight Decay	0.0005
Batch Size	6
Epochs	1000 (Patience: 50)
Input Resolution	640×640

The Stochastic Gradient Descent (SGD) optimizer and an initial learning rate of 0.01 were selected through a hyperparameter validation study conducted specifically for the MEP segmentation task. While the YOLOv8 architecture usually employs default values for general datasets like COCO, the unique visual characteristics of MEP components, including reflective metallic surfaces and intricate geometric patterns, require tailored adjustments to ensure stable convergence.

We compared three learning rate configurations ($lr \in \{0.1, 0.01, 0.001\}$) over 100 epochs on a subset of the Mixed-20% dataset. As shown in Table 6, $lr = 0.01$ provided the most stable reduction in validation loss and the highest mAP_{50} score. A higher learning rate (0.1) led to significant oscillations in the loss function, indicating gradient instability, while a lower learning rate (0.001) resulted in slow convergence and lower final precision, failing to escape local minima during early training phases.

Table 6: Comparison of initial learning rates (lr_0) on YOLOv8m-seg performance (100 epochs).

Learning Rate (lr_0)	mAP_{50} (m)	Val Loss	Status
0.1	0.62	1.45	Unstable
0.01	**0.79**	**0.88**	**Stable**
0.001	0.54	1.12	Underfit

Table 7: Performance Comparison (mean ± std over 5 runs): COCO vs. Synthetic Transfer Learning (TL) on Small (S) and Medium (M) Real Datasets.

Metrics	**COCO TL**		**Synthetic TL**	
S dataset	Box	Mask	Box	Mask
Precision (%)	43.2 ± 2.1	51.3 ± 2.3	**65.8 ± 2.4**	**64.1 ± 2.2**
Recall (%)	43.1 ± 1.8	34.2 ± 2.0	47.3 ± 2.1	46.1 ± 1.9
mAP_{50} (%)	42.0 ± 1.9	38.1 ± 1.7	53.2 ± 1.8	49.4 ± 1.7
$mAP_{50–95}$ (%)	26.1 ± 1.5	19.2 ± 1.3	37.0 ± 1.6	30.2 ± 1.4
M dataset	Box	Mask	Box	Mask
Precision (%)	52.1 ± 1.9	52.3 ± 2.0	**69.2 ± 2.1**	**63.1 ± 1.8**
Recall (%)	45.0 ± 1.7	42.1 ± 1.8	40.2 ± 1.9	38.3 ± 1.7
mAP_{50} (%)	43.2 ± 1.6	41.0 ± 1.5	47.1 ± 1.7	43.2 ± 1.5
$mAP_{50–95}$ (%)	29.0 ± 1.4	24.1 ± 1.2	30.3 ± 1.5	24.2 ± 1.3

Training utilized a linear warmup followed by cosine annealing decay. Early stopping (patience: 50 epochs) was employed to prevent overfitting. Standard data augmentations (rotation, scaling, mosaic) were applied to enhance robustness.

4.3.2. Quantitative Evaluation

The model performance was evaluated using standard computer vision metrics: Precision, Recall, and Mean Average Precision at different IoU thresholds (mAP_{50} and $mAP_{50–95}$). To ensure statistical robustness, all training configurations were repeated across five independent runs with different random seeds, and results are reported as mean ± standard deviation. The evaluation follows a two-stage analysis: first, assessing the effectiveness of synthetic transfer learning (TL) against standard benchmarks, and second, determining the optimal ratio of synthetic-to-real data for practical deployment.

Initial experiments compared pre-training on the COCO dataset versus our custom synthetic dataset. As detailed in Table 7, the Synthetic TL approach significantly outperformed COCO-based pre-training on the small (S) real-world dataset across all metrics and runs, particularly in mask precision (64.1% ± 2.2% vs. 51.3% ± 2.3%). The lower standard deviations observed in the Synthetic TL configurations further indicate greater training stability. While performance converged as the dataset size increased to medium (M), the synthetic pre-training consistently demonstrated superior boundary detection and feature extraction capabilities for specialized MEP components.

We further investigated the impact of mixing synthetic data with varying proportions of real-world imagery (Table 8). The "Mixed-20%" configuration (comprising 500 synthetic and 100 real images) achieved a peak mean precision of 80.1% ± 1.4% for box detection and 79.2% ± 1.5% for mask segmentation. The notably lower standard deviations for this configuration—compared to the Synthetic-only (±2.8%) and 10% Real (±2.3%) variants—confirm that the 20% real-data infusion stabilizes training convergence. This hybrid approach consistently outperformed the model trained exclusively on 142 real images across all five runs, suggesting that synthetic data acts as a powerful regulariser.

Table 8: YOLOv8 Performance (mean ± std over 5 runs) across different training data compositions.

Training Dataset	**Precision (%)**	
	Box	Mask
Real Only	77.1 ± 2.1	75.3 ± 2.0
Synthetic Only	30.2 ± 2.8	29.1 ± 2.6
Synthetic + 10% Real	71.2 ± 2.3	69.4 ± 2.2
Synthetic + 20% Real	**80.1 ± 1.4**	**79.2 ± 1.5**

Table 9: Comprehensive comparison of performance (mean ± std over 5 runs) and annotation effort across training strategies.

Training Strategy	**Precision (%)**	**mAP_{50} (%)**	**Effort**
Synthetic Only	29.1 ± 2.8	22.3 ± 2.4	None
Synthetic + 10% Real	69.4 ± 2.3	61.2 ± 2.0	Low
Synthetic + 20% Real	**79.1 ± 1.4**	**72.3 ± 1.4**	**Moderate**
Synthetic + 50% Real	81.2 ± 1.3	74.1 ± 1.3	High
Real Only (100%)	75.2 ± 2.4	68.1 ± 2.1	Very High

Training convergence and threshold optimisation for the Mixed-20% model are visualised in Figures 6 and 7. The F1-confidence curve identifies an optimal threshold of 0.365, yielding a maximum F1 score of 0.67. The Mixed-20% model maintained a significantly higher mean mAP_{50} (79.2% ± 1.4%) compared to the Real Only model (75.4% ± 2.1%), with non-overlapping 95% confidence intervals ([76.5%, 81.9%] vs. [72.8%, 78.0%]), confirming the statistical significance of the improvement under five independent runs.

A comprehensive granularity analysis (Table 9) highlights a critical "sweet spot" for engineering efficiency. The statistical results confirm that increasing real-world data beyond 20% yields only marginal *mAP* gains: the "Synthetic + 50% Real" strategy improves mean mAP_{50} by only ~2 points over the "Synthetic + 20% Real" configuration (74.1% ± 1.3% vs. 72.3% ± 1.4%), while requiring more than twice the annotation effort. Furthermore, the "Real Only (500 images)" model underperforms the "Synthetic + 20% Real" model (100 real images + synthetic) both in terms of mean precision and mAP_{50}, and exhibits greater run-to-run variability (±2.4% vs. ±1.4%), indicating that large-scale synthetic pre-training provides a more stable feature-extraction baseline that small real-world datasets cannot replicate independently. These results confirm that a 20% real-world data infusion is sufficient to bridge the sim-to-real gap for MEP components.

To assess whether the model generalises beyond the training sites, we evaluated it separately on the 70 images sourced exclusively from the two "unseen" construction projects (sites 4 and 5), which share no BIM geometry with the MEP-SEG synthetic data. Across five independent runs, the Mixed-20% model achieved a mean mAP_{50} of 73.6% ± 2.3% on this held-out cross-site subset, compared to 79.2% ± 1.4% on the full test set. The wider confidence interval on the cross-site subset ([69.1%, 78.1%] vs. [76.5%, 81.9%]) reflects the known domain shift between construction sites (different pipe diameters,

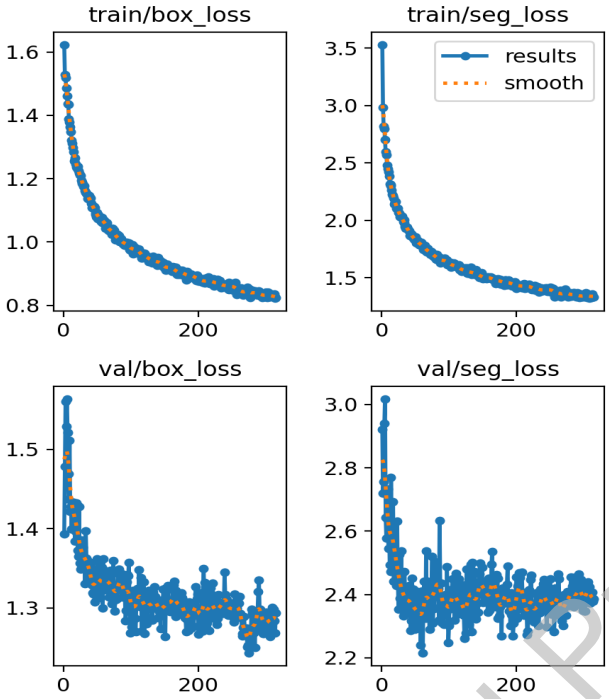


Figure 6: Training/validation loss curves (Mixed-20% dataset).

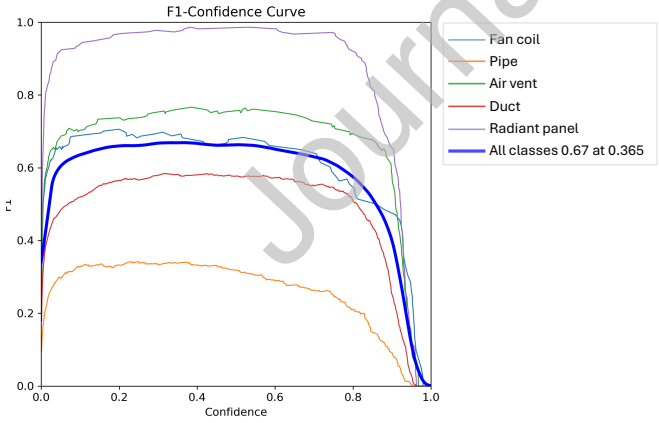


Figure 7: F1 score vs. confidence threshold.

duct brands, and installation styles), and the ~6-point drop confirms that while the framework generalises to unseen environments, site-specific fine-tuning on a small number of images is advisable for deployment on projects with highly distinct MEP configurations. Expanding the dataset to cover a broader range of building types, climatic zones, and national construction practices remains an important direction for future work to validate wider generalisation.

4.3.3. Methodological Comparison and Benchmarking Challenges

A direct quantitative comparison with recent state-of-the-art (SOTA) methods for MEP detection, such as [25] or [44], presents significant challenges due to the lack of standardized benchmark datasets in the construction domain. Most existing studies utilize proprietary datasets with varying object classes; for instance, while Kufuor et al. focus on small-scale components like sockets and switches, our work targets large-scale infrastructure such as HVAC ducts and radiant panels. Furthermore, traditional geometric approaches [25] are specifically tuned for cylindrical primitives and are not directly applicable to the diverse, non-primitive geometries addressed in this study.

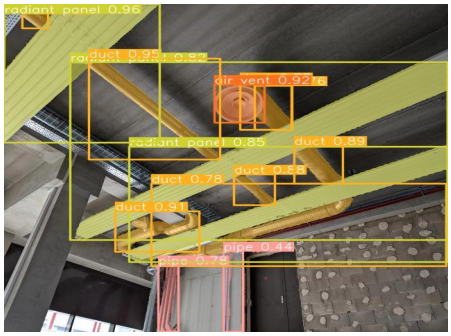
Consequently, we provide a systematic methodological benchmark in Table 1 to highlight our framework's original contributions, focusing on the quantified data efficiency and the public release of the MEP-SEG dataset to facilitate future comparative research.

4.3.4. Qualitative and Failure Mode Analysis

Qualitative results (Figure 8) confirm the model's ability to handle complex layouts, varying scales, and difficult lighting. However, while baseline performance is strong, a systematic analysis of 50 failure cases (where IoU < 0.3, with general examples shown in Figure 9) was conducted to identify persistent challenges in real-world construction environments. Three primary error modes were identified, as detailed below.

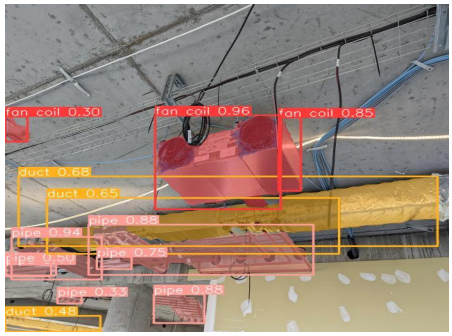
Error Mode 1: Scene Clutter and High Instance Density (38% of failures). Root Cause: While synthetic training scenes include clutter, the instance density on real active construction sites often exceeds that of the training distribution. When objects are tightly packed, standard Non-Maximum Suppression (NMS) thresholds may erroneously suppress valid overlapping detections, or the mask head resolution (160×160) may be insufficient to delineate distinct boundaries for small, adjacent components. Proposed Mitigation: Implement a "Tiling Inference" strategy, where high-resolution frames are split into overlapping patches to preserve small-scale features. Additionally, adopting Soft-NMS would allow the model to retain overlapping detections by decaying their confidence scores rather than completely suppressing them, which is critical for the parallel-heavy geometry of MEP networks.

Error Mode 2: Black Insulation in Low Light (28% of failures). Root Cause: The MEP-SEG synthetic pipeline was configured for 300–1000 lux. Real-world basements and utility corridors

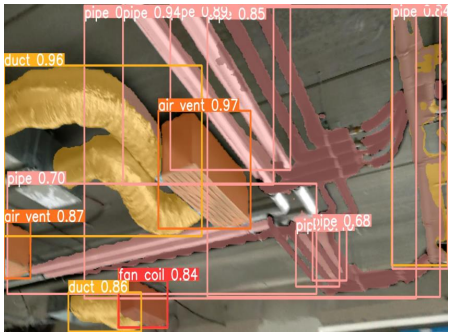


a) Scene with a high ceiling

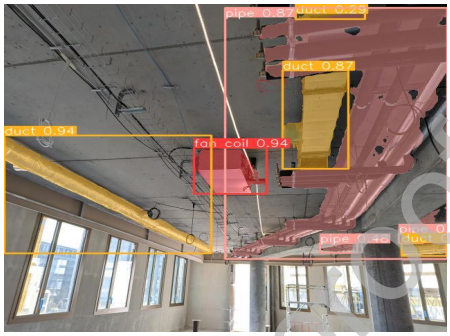
b) Scene with pipes and ducts



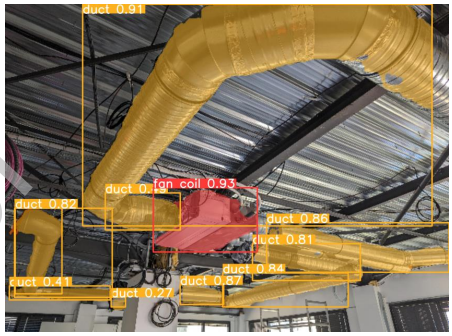
c) Scene with a fan coil



d) Scene captured on a lower end camera

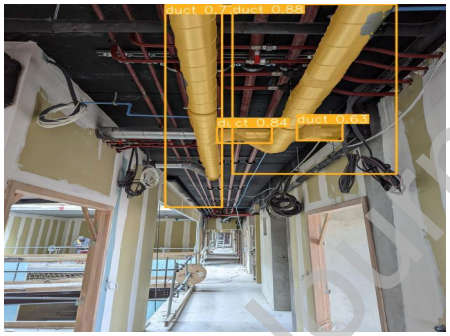


e) Long range scene with high contrast

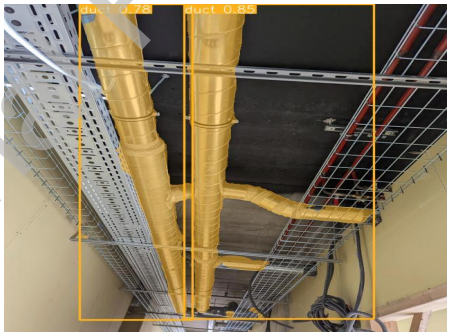


f) Robustness to different background

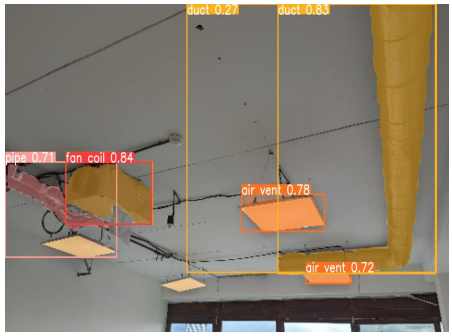
Figure 8: Qualitative examples of successful instance segmentation results.



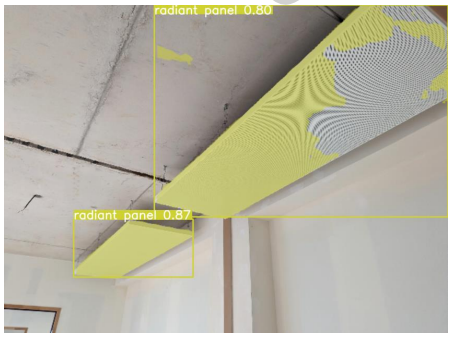
a) Undetected pipes



b) Cable routing frames are covering pipes

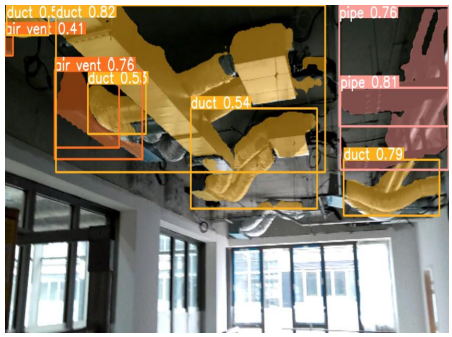


c) Light fixture mistaken for air vent



d) Segmentation incomplete

e) Black pipe insulation not detected



f) Challenging lighting conditions

Figure 9: Qualitative examples of failure cases and limitations.

Figure 10: Failure case: Overwhelming density of parallel pipes and conduits. High overlap leads to instances being merged by NMS or parts of elements being missed.

Figure 11: Failure case: Low lighting conditions (under 300 lux) and low contrast (σ = 15) prevents reliable edge detection.

often reach as low as 50 lux, where MEP elements lack sufficient contrast (σ) against dark concrete ceilings for standard RGB-based edge detection. Proposed Mitigation: Future work will extend Domain Randomization (DR) to a 50–2000 lux range using gamma-corrected rendering. Additionally, fusing RGB data with HoloLens 2 Time-of-Flight (ToF) depth data offers a robust alternative, as geometric cues and surface normals remain detectable independent of ambient lighting levels.

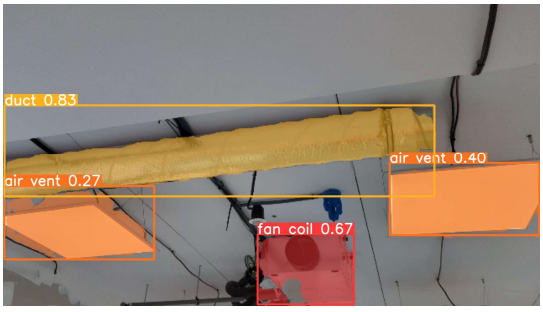


Figure 12: Failure case: Rectangular light fixture misclassified as an air vent due to similar geometry and color.

Error Mode 3: Semantic Confusion (34% of failures). Root Cause: The current MEP-SEG dataset lacks sufficient "negative samples", non-MEP cylindrical or rectangular objects that appear in similar contexts.

Proposed Mitigation: We intend to augment the synthetic dataset with "distractor" objects (e.g., light fixtures, sprinklers, security cameras) labeled as background classes. This forces the model to learn specific MEP textures (galvanized steel patterns, insulation seams) rather than relying solely on global geometric primitives.

Based on these findings, we estimate that implementing the proposed mitigations will yield a significant quantitative impact. Temporal filtering addresses 38% of failures with a 60% projected recovery rate, yielding a 22.8% reduction. Extended HDR lighting targets another 28% of failures, and with a 50% projected recovery, it contributes a further 14% reduction. Additionally, negative sample augmentation is expected to address the remaining 34% of failures with a 40% recovery rate, resulting in a 13.6% reduction. Combined, these strategies offer a total projected improvement of a 50.4% reduction in overall failures, which could potentially increase the mAP_{50} from 79.1% to approximately 84%.

4.4. Augmented Reality Implementation Results

4.4.1. Mask Alignment Analysis

We compared centroid-based and affine transformation alignment methods on the Avignon Archive Centre dataset (133 image pairs, 829 object occurrences).

Table 10: Average results of mask comparison using centroid-based and affine transformation alignment.

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. While efficient for detecting positional offsets (translation), this method is inherently limited as it cannot detect rotational deviations.

On the other hand, affine alignment significantly improves the alignment quality, with an IoU of 0.64 and a Dice coefficient of 0.76. This method theoretically enables the detection of rotational deviations in addition to translation and scaling. 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. Furthermore, our experiments indicated that the precision of the segmentation masks (often lacking sharp edges) makes the rotation estimate sensitive to noise, requiring further investigation into depth-enhanced segmentation for precise angular deviation measurement.

4.4.2. System Performance and Computational Load

Deploying DL models on wearable hardware requires careful resource management. The HoloLens 2 runs on a Qualcomm Snapdragon 850 Compute Platform (ARM64 architecture), which presents significant constraints compared to desktop GPUs. To characterize the on-site performance, we measured the end-to-end latency of the system. Centroid alignment is the default method used for the real-time on-device pipeline, while Affine was evaluated for offline/high-precision analysis. To characterize the on-site performance, we measured the end-to-end latency of the system. The total latency of approximately 500ms per inspection cycle is distributed across three primary stages:

- Preprocessing (12%, ~60ms): Includes RGB image capture from the HoloLens 2 sensor, resizing to 640×640, and normalization.
- DL Inference (76%, ~380ms): Execution of the YOLOv8m-seg ONNX model on the Qualcomm Snapdragon 850 DSP/GPU.
- Post-processing (12%, ~60ms): Comprises Non-Maximum Suppression (NMS), binary mask generation, and centroid-based alignment for discrepancy visualization.

Although a total latency of 500ms results in an operational frame rate of approximately 2 FPS, it is important to consider the nature of the target application. Construction progress monitoring, particularly for MEP installations, involves the assessment of static, cumulative assemblies rather than the tracking of high-speed dynamic objects. Therefore, a 2 FPS rate does not lead to "missed detections" of progress, as the physical state of the site remains constant during the inspection interval. Furthermore, to prevent visual desynchronization and user discomfort (cybersickness), the predicted instance masks are not anchored to the screen. Instead, they are locked to the spatial environment using World Anchors immediately after inference. This decoupling ensures that even while the inference loop runs at 2 FPS, the rendering loop maintains a steady 60 FPS, preserving holographic stability as the user moves their head. The system is designed for a "point-and-check" workflow where the inspector stabilizes their gaze on a specific area to verify installation status, rather than a continuous high-speed surveillance mode. For highly dynamic scenes (e.g., monitoring moving machinery), this latency would indeed be a bottleneck; however, for the verified engineering use case of as-built progress monitoring, it provides a sufficient balance between accuracy and on-device computational feasibility.

Future optimizations for deployment on resource-constrained devices could include:

- Model pruning: Reducing the depth of the YOLO backbone to create a specific "Nano" version tailored for MEP.
- Cloud offloading: Streaming video frames to an edge server for processing, although this introduces dependency on site connectivity.
- Asynchronous rendering: Decoupling the AR rendering loop (60 FPS) from the inference loop (2 FPS) to maintain hologram stability despite low inference frequency.

4.5. 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.

4.5.1. 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.

4.5.2. Participant Demographics and NEXT-BIM Usage

Table 11 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 11: 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%

4.5.3. Quantitative Results

Analysis of the Likert scale responses, presented in Table 12, 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. While a longitudinal field time-trial was not conducted, 90.5% of the 21 professionals surveyed agreed or strongly agreed that the system would significantly reduce inspection time. Specifically, qualitative interviews with site managers suggested that the “point-and-check" workflow could potentially replace the manual process of cross-referencing 2D paper drawings, which they estimated currently occupies up to 30% of their daily site walkthroughs. Furthermore, the system’s ability to automate the detection of small-diameter pipes suggests a theoretical reduction in the defect miss rate, particularly in high-density MEP zones.

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.

4.5.4. Qualitative Insights

A prominent theme was the substantial time savings afforded by the system. Participants reported a significant reduction in inspection times compared to manual methods; for instance, one construction manager estimated a decrease from half a day to approximately one hour.

Beyond speed, the AR visualization functioned as a shared, contextualized platform that improved communication and coordination. By allowing teams to visually compare “as-built" reality with the BIM model in real-time, the system facilitated clearer exchange of information regarding discrepancies and quality control than traditional 2D reporting methods.

Despite the positive reception, physical discomfort remains a significant barrier. Only 42.8% of users reported satisfactory comfort levels with the HoloLens 2, a finding that is consistent with the broader industry feedback on HMDs. Furthermore, operational longevity is currently limited by thermal management and battery constraints, restricting continuous use to approximately 90–120 minutes.

To address the identified ergonomic concerns, the proposed framework utilizes an ONNX-based model architecture, which is inherently device-agnostic. This flexibility allows the system to be ported from the HoloLens 2 to more lightweight AR glasses (e.g., Xreal) or high-performance handheld tablets like the iPad Pro with LiDAR, depending on user preference or environmental constraints.

To extend operational duration for full-day monitoring, future iterations will explore asynchronous processing or cloud-offloading to reduce the on-device thermal load. Overall, the high levels of agreement on usability and efficiency demonstrate the system’s potential for industry adoption, provided that hardware-specific limitations are addressed through these multi-platform and offloading strategies.

Table 12: User study results: Questionnaire responses (N=21)

Question	Disagree (1-2)	Neutral (3)	Agree (4)	Strongly Agree (5)	Median	IQR
1. Easy integration into workflow	3 (14.3%)	9 (42.9%)	8 (38.1%)	1 (4.8%)	3.0	1.0
2. HoloLens 2 comfort (long periods)	4 (19.0%)	8 (38.1%)	5 (23.8%)	4 (19.0%)	3.0	1.0
3. Clear & easy AR visualization	0 (0%)	5 (23.8%)	14 (66.7%)	2 (9.5%)	4.0	0.0
4. Increased work efficiency	0 (0%)	2 (9.5%)	8 (38.1%)	11 (52.4%)	4.0	1.0
5. Intuitive & easy user interface	0 (0%)	2 (9.5%)	13 (61.9%)	6 (28.6%)	4.0	0.0
6. Relevant app functionalities	0 (0%)	4 (19.0%)	10 (47.6%)	7 (33.3%)	4.0	1.0
7. Easy to use daily	2 (9.6%)	1 (4.8%)	11 (52.3%)	7 (33.3%)	4.0	1.0
8. Sufficient initial training	0 (0%)	4 (19.0%)	12 (57.1%)	5 (23.8%)	4.0	0.0

5. Discussion

5.1. Summary of Findings

The experimental evaluations and user study conducted in this research provide 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.

5.2. Limitations

While the proposed AR-integrated ICPM methodology demonstrates promising results, several limitations must be addressed to transition from a prototype to an industrial standard.

Generalization to diverse BIM models and construction environments. The current framework was validated on BIM models originating from three building types (office tower, laboratory, campus) in a single geographic and regulatory context. The MEP-REAL dataset spans five sites, including two projects unseen during synthetic data generation, and cross-site evaluation yielded a mAP_{50} of 73.6% ± 2.3%—a ~6-point gap compared to the full test set. This gap underscores that the model's performance is sensitive to variations in MEP product families (e.g., pipe diameters, duct manufacturers), installation conventions, and ambient conditions across regions and project types. Adaptation to a new BIM standard (e.g., US vs. French IFC conventions) or a new MEP component family currently requires re-running the synthetic data generation pipeline and collecting a small real-world fine-tuning set. While this process is automated and lightweight (as few as 100 labeled real images suffice), the effort is non-trivial for large-scale industrial deployment. Future work should expand the MEP-SEG dataset to cover a wider range of international BIM libraries and construction environments to quantify and reduce this generalization gap systematically.

User study constraints: The user study involved 21 participants, which provides preliminary validation but lacks statistical power for broad generalization. Furthermore, a significant selection bias must be acknowledged: all 21 professionals had prior experience with the NEXT-BIM application and the HoloLens 2 headset. While this ensured they could focus on evaluating the new AI-driven features rather than struggling with basic AR navigation, it inherently biases the acceptability and usability scores upwards. Future studies should include novices to evaluate the true learning curve of the system. Additionally, the inclusion of a standardized psychometric tool such as the System Usability Scale (SUS)...

Definition of deviation tolerances: A critical engineering limitation is the definition of "deviation". Currently, the system uses visual mask overlap (IoU) as a proxy for installation

correctness. However, construction standards often dictate millimetric tolerances (e.g., ±10mm for pipe positioning). The current resolution of the HoloLens cameras and the drift inherent in SLAM tracking (typically 2-3cm error in large open spaces) prevent millimetric verification. Consequently, the current system is best suited for macro-level inspection (presence/absence and gross positioning) rather than precision tolerance verification. Future work utilizing high-precision external trackers or LiDAR integration is necessary to achieve standard engineering tolerances.

Synthetic reality gap: While domain adaptation has narrowed the gap between synthetic and real-world performance, a persistent gap remains regarding extreme site conditions, such as atmospheric dust and High-Dynamic-Range (HDR) lighting. The current study used an illuminance range of 300 to 1000 lux, reflecting standard indoor BIM rendering practices. However, this range does not account for the extreme variability of unlit subterranean levels or the intense direct sunlight that is common near large glass façades [58]. These environmental omissions explain the observed performance degradation in low-contrast scenarios, particularly where black insulation material blends with dark ceilings. To bridge this gap, the synthetic data generation pipeline must evolve to include Physics-Based Rendering (PBR) of atmospheric effects and more aggressive Domain Randomization (DR) [62]. By extending the simulated illuminance spectrum from 50 to 5000 lux and incorporating ray-traced specular reflections to better represent metallic MEP components under harsh lighting, the model's robustness can be significantly enhanced. Furthermore, the integration of Active Infrared (AIR) or LiDAR-based depth sensors on the AR device could mitigate these purely visual limitations by providing ambient-light-independent geometric features to supplement RGB data [86].

5.3. Broader Engineering Applications and Cross-Domain Adaptation

While this framework was validated for indoor MEP installation monitoring, the underlying architecture holds significant potential for adaptation across other data-scarce engineering domains.

5.3.1. Digital Twins for Industrial Facility Management

The transition from construction to Operation and Maintenance (O&M) represents a critical lifecycle shift. As noted by Yildiz et al. [87], the integration of AR with Digital Twins (DT) is becoming essential for smart maintenance, yet it often suffers from data fragmentation. Our proposed pipeline can be extended to the O&M phase by replacing the "as-planned" BIM with a live Digital Twin. In this context, the system could be retrained on synthetic datasets of specific industrial components (pressure valves or HVAC actuators) to perform real-time predictive maintenance audits. By overlaying IoT sensor data onto the physical asset via the AR interface, as explored by Ghansah et al. [88], the system effectively bridges the gap between digital asset management and physical reality, reducing the cognitive load on maintenance technicians.

5.3.2. Aerospace Assembly and Surface Inspection

The aerospace industry shares distinctive challenges with MEP construction, specifically regarding the inspection of complex assemblies where real-world defect data is exceptionally rare. Recent work by Schmedemann et al. [89] on industrial surface inspection highlights that training data for specific defects (*e.g.*, micro-fractures or coating anomalies) is often unavailable. Our "Sim-to-Real" methodology parallels their findings, demonstrating that synthetic data generation is a viable pathway for training robust detection models for such restricted environments. By adapting our pipeline to ingest aerospace CAD models, the system could provide real-time AR feedback to technicians, flagging deviations in cable routing or surface integrity before the airframe is sealed, thereby mitigating costly rework.

5.3.3. Safety Engineering and Hazardous Scenario Simulation

Finally, the domain adaptation strategy employed in this study is particularly relevant to safety engineering, where collecting real-world data on accidents is ethically impossible. Yildiz [87] emphasize that AR-enhanced Digital Twins are critical for simulating hazardous scenarios without putting personnel at risk. Our synthetic data generation approach enables the simulation of rare and hazardous events, such as structural failures and high-voltage arc flashes, within a physics-enabled virtual environment. Training computer vision models on these synthetic "black swan" events would enable the deployment of AR systems capable of preemptively recognizing safety precursors in the field, moving the industry from reactive reporting to proactive hazard prevention.

6. Conclusion

This research presented an end-to-end framework for automating Indoor Construction Progress Monitoring (ICPM) by bridging Building Information Modeling (BIM) with Augmented Reality (AR) via synthetic data training. By addressing the critical bottleneck of data scarcity, we demonstrated that photorealistic simulation is a viable pathway for deploying Deep Learning (DL) in the built environment. The key findings of this study are summarized as follows:

- Synthetic data efficiency: A streamlined pipeline using NVIDIA Isaac Sim generated 8,751 labeled images in under 9 hours. We proved that pre-training on this data allows a model to achieve high performance (79% mAP_{50}) using only 20% of the typically required real-world data, significantly reducing deployment costs.

- Real-Time on-site inference: The YOLOv8 model was successfully optimized (ONNX/FP16) and deployed on HoloLens 2, achieving an operational latency of 500ms. This confirms the feasibility of edge-computing for immediate progress verification without reliance on cloud connectivity.

- Alignment and deviation: While centroid-based alignment proved fast (30ms) but limited, affine transformation provided higher accuracy for mask overlay. However, current hardware limitations restrict the system to detecting gross deviations (presence/absence) rather than millimetric engineering tolerances.
- User acceptance: A study with 21 construction professionals revealed strong acceptance of the AR visualization for communication and error reduction, though ergonomic concerns regarding the headset remains a barrier to continuous daily use.

Future work will focus on integrating depth sensor data to resolve lighting failures, implementing standardized usability metrics (SUS), and refining the geometric precision to meet strict construction tolerance standards.

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., Gemini) in order to improve language and refine phrasing. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the content of the published article.

CRediT authorship contribution statement

Mathis Baubriaud: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing – original draft. **Stéphane Derrode:** Supervision, Validation, Writing – review & editing. **René Chalon:** Supervision, Validation, Writing – Review & Editing. **Kevin Kernn:** Conceptualization, Supervision, Validation.

Declaration of interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Baubriaud Mathis reports a relationship with SPIE that includes: employment. Mathis Baubriaud reports financial support was provided by SPIE. Co-author is part of the company SPIE Building Solutions - Kevin Kernn If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The synthetic dataset generated during this study (MEP-SEG) is publicly available via [85]. 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.

References

[1] The Future: Building with BIM, in: BIM Handbook, John Wiley & Sons, Ltd, 2018, pp. 364–397, section: 9 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119287568.ch9. doi:10.1002/9781119287568.ch9.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119287568.ch9

[2] A. Aljohani, Construction Projects Cost Overrun: What Does the Literature Tell Us?, International Journal of Innovation, Management and Technology (2017) 137–143doi:10.18178/ijimt.2017.8.2.717.
URL http://www.ijimt.org/index.php?m=content&c=index&a=show&catid=83&id=1056

[3] B. Ekanayake, J. K.-W. Wong, A. A. F. Fini, P. Smith, Computer vision-based interior construction progress monitoring: A literature review and future research directions, Automation in Construction 127 (2021) 103705. doi:10.1016/j.autcon.2021.103705.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580521001564

[4] F. Bosché, A. Guillemet, Y. Turkan, C. T. Haas, R. Haas, Tracking the Built Status of MEP Works: Assessing the Value of a Scan-vs-BIM System, Journal of Computing in Civil Engineering 28 (4) (2014) 05014004. doi:10.1061/(ASCE)CP.1943-5487.0000343.
URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CP.1943-5487.0000343

[5] P. Dallasega, F. Schulze, A. Revolti, M. Martinelli, Augmented Reality to increase efficiency of MEP construction: a case study, Dubai, UAE, 2021. doi:10.22260/ISARC2021/0063.
URL http://www.iaarc.org/publications/2021_proceedings_of_the_38th_isarc/augmented_reality_to_increase_efficiency_of_mep_construction-a_case_study.html

[6] V. K. Reja, K. Varghese, Q. P. Ha, Computer vision-based construction progress monitoring, Automation in Construction 138 (2022) 104245. doi:10.1016/j.autcon.2022.104245.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580522001182

[7] R. Sacks, C. Eastman, G. Lee, P. Teicholz, BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers, 2018. doi:10.1002/9781119287568.

[8] J. C. P. Cheng, K. Chen, W. Chen, State-of-the-Art Review on Mixed Reality Applications in the AECO Industry, Journal of Construction Engineering and Management 146 (2) (2020) 03119009. doi:10.1061/(ASCE)CO.1943-7862.0001749.
URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CO.1943-7862.0001749

[9] B. Schiavi, V. Havard, K. Beddiar, D. Baudry, BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction, Automation in Construction 134 (2022) 104054. doi:10.1016/j.autcon.2021.104054.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580521005057

[10] W. Fang, L. Ding, B. Zhong, P. E. D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Advanced Engineering Informatics 37 (2018) 139–149. doi:10.1016/j.aei.2018.05.003.
URL https://www.sciencedirect.com/science/article/pii/S1474034617304706

[11] C. Z. Li, J. Zeng, V. W. Tam, H. Wu, Advanced Information Technologies for High-precision Quality Control in Building Engineering, Journal of Building Engineering 101 (2025) 111918. doi:10.1016/j.jobe.2025.111918.
URL https://www.sciencedirect.com/science/article/pii/S2352710225001548

[12] Y. Xie, M. X. Teo, S. Li, L. Huang, N. Liang, Y. Cai, As-built BIM reconstruction of piping systems using smartphone videogrammetry and terrestrial laser scanning, Automation in Construction 156 (2023) 105120. doi:10.1016/j.autcon.2023.105120.
URL https://www.sciencedirect.com/science/article/pii/S0926580523003801

[13] T. Boan, L. Jiajun, F. Bosché, Autonomous Mixed Reality Framework for Real-Time Construction Inspection, Journal of Information Technology in Construction (ITcon) 30 (35) (2025) 852–874. doi:10.36680/j.itcon.2025.035.
URL http://www.itcon.org/paper/2025/35

[14] S. I. Nikolenko, Synthetic Data for Deep Learning, Vol. 174 of Springer Optimization and Its Applications, Springer International Publishing, Cham, 2021. doi:10.1007/978-3-030-75178-4.
URL https://link.springer.com/10.1007/978-3-030-75178-4

[15] D. Q. Tran, Y. Jeon, A. Aboah, J. Bak, M. Park, S. Park, Leveraging Semisupervised Learning for Domain Adaptation: Enhancing Safety at Construction Sites through Long-Tailed Object Detection, Journal of Construction Engineering and Management 151 (1) (2025) 04024190, publisher: American Society of Civil Engineers. doi:10.1061/JCEMD4.COENG-15259.
URL https://ascelibrary.org/doi/10.1061/JCEMD4.COENG-15259

[16] M. Baubriaud, S. Derrode, R. Chalon, K. Kernn, Accelerating Indoor Construction Progress Monitoring with Synthetic Data-Powered Deep Learning, IAARC (2024) 792–799Publisher: Automation in Construction. doi:10.22260/ISARC2024/0103.
URL https://hal.science/hal-04588885

[17] R. Pellerin, N. Perrier, A review of methods, techniques and tools for project planning and control, International Journal of Production Research 57 (7) (2019) 2160–2178, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00207543.2018.1524168. doi:10.1080/00207543.2018.1524168.
URL https://doi.org/10.1080/00207543.2018.1524168

[18] T. Omotayo, B. Awuzie, T. Egbelakin, I. R. Orimoloye, O. E. Ogunmakinde, A. Sojobi, The Construction Industry's Future: Systems, People and Projects, in: Innovations, Disruptions and Future Trends in the Global Construction Industry, Routledge, 2024, num Pages: 9.

[19] A. Warszawski, Industrialized and Automated Building Systems: A Managerial Approach, 2nd Edition, Routledge, London, 2003. doi:10.4324/9780203223697.

[20] T. D. Akinosho, L. O. Oyedele, M. Bilal, A. O. Ajayi, M. D. Delgado, O. O. Akinade, A. A. Ahmed, Deep learning in the construction industry: A review of present status and future innovations, Journal of Building Engineering 32 (2020) 101827. doi:10.1016/j.jobe.2020.101827.
URL https://www.sciencedirect.com/science/article/pii/S2352710220334604

[21] P. Kavaliauskas, J. B. Fernandez, K. McGuinness, A. Jurelionis, Automation of Construction Progress Monitoring by Integrating 3D Point Cloud Data with an IFC-Based BIM Model, Buildings 12 (10) (2022) 1754, number: 10 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/buildings12101754.
URL https://www.mdpi.com/2075-5309/12/10/1754

[22] N. Rane, Integrating Building Information Modelling (BIM) and Artificial Intelligence (AI) for Smart Construction Schedule, Cost, Quality, and Safety Management: Challenges and Opportunities (Sep. 2023). doi:10.2139/ssrn.4616055.
URL https://papers.ssrn.com/abstract=4616055

[23] W. S. Alaloul, A. H. Qureshi, M. A. Musarat, S. Saad, Evolution of close-range detection and data acquisition technologies towards automation in construction progress monitoring, Journal of Building Engineering 43 (2021) 102877. doi:10.1016/j.jobe.2021.102877.
URL https://www.sciencedirect.com/science/article/pii/S235271022100735X

[24] D. Rebolj, Z. Pucko, N. C. Babic, M. Bizjak, D. Mongus, Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring, Automation in Construction 84 (2017) 323–334. doi:10.1016/j.autcon.2017.09.021.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580517304107

[25] F. Bosché, M. Ahmed, Y. Turkan, C. T. Haas, R. Haas, 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 49 (2015) 201–213. doi:10.1016/j.autcon.2014.05.014.
URL https://www.sciencedirect.com/science/article/pii/S0926580514001319

[26] F. Amer, M. Golparvar-Fard, Decentralized Visual 3D Mapping of Scattered Work Locations for High-Frequency Tracking of Indoor Construction Activities, in: Construction Research Congress 2018, American Society of Civil Engineers, New Orleans, Louisiana, 2018, pp. 491–500. doi:10.1061/9780784481264.048.
URL http://ascelibrary.org/doi/10.1061/9780784481264.048

[27] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Automation in Construction 20 (8) (2011) 1143–1155. doi:10.1016/j.autcon.2011.04.016.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580511000707

[28] Y. Turkan, F. Bosche, C. T. Haas, R. Haas, Automated progress tracking using 4D schedule and 3D sensing technologies, Automation in Construction 22 (2012) 414–421. doi:10.1016/j.autcon.2011.10.003.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580511001956

[29] H. González-Jorge, B. Riveiro, P. Arias, J. Armesto, Photogrammetry and laser scanner technology applied to length measurements in car testing laboratories, Measurement 45 (3) (2012) 354–363. doi:10.1016/j.measurement.2011.11.010.
URL https://linkinghub.elsevier.com/retrieve/pii/S0263224111004179

[30] S. Hasan, R. Sacks, Integrating BIM and Multiple Construction Monitoring Technologies for Acquisition of Project Status Information, Journal of Construction Engineering and Management 149 (7) (2023) 04023051, publisher: American Society of Civil Engineers. doi:10.1061/JCEMD4.COENG-12826.
URL https://ascelibrary.org/doi/10.1061/JCEMD4.COENG-12826

[31] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, Automation in Construction 24 (2012) 89–99. doi:10.1016/j.autcon.2012.02.013.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580512000283

[32] D. Atherinis, B. Bakowski, M. Velcek, S. Moon, Developing and Laboratory Testing a Smart System for Automated Falsework Inspection in Con-

struction, Journal of Construction Engineering and Management 144 (3) (2018) 04017119. doi:10.1061/(ASCE)CO.1943-7862.0001439.
URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CO.1943-7862.0001439

[33] A. Shahi, M. Safa, C. T. Haas, J. S. West, Data Fusion Process Management for Automated Construction Progress Estimation, Journal of Computing in Civil Engineering 29 (6) (2015) 04014098. doi:10.1061/(ASCE)CP.1943-5487.0000436.
URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CP.1943-5487.0000436

[34] N. Pradhananga, J. Teizer, Automatic spatio-temporal analysis of construction site equipment operations using GPS data, Automation in Construction 29 (2013) 107–122. doi:10.1016/j.autcon.2012.09.004.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580512001537

[35] O. Moselhi, H. Bardareh, Z. Zhu, Automated Data Acquisition in Construction with Remote Sensing Technologies, Applied Sciences 10 (8) (2020) 2846, number: 8 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/app10082846.
URL https://www.mdpi.com/2076-3417/10/8/2846

[36] S. Paneru, I. Jeelani, Computer vision applications in construction: Current state, opportunities & challenges, Automation in Construction 132 (2021) 103940. doi:10.1016/j.autcon.2021.103940.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580521003915

[37] D. Roberts, M. Golparvar-Fard, End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level, Automation in Construction 105 (2019) 102811. doi:10.1016/j.autcon.2019.04.006.
URL https://www.sciencedirect.com/science/article/pii/S0926580518308525

[38] I. Jeelani, K. Asadi, H. Ramshankar, K. Han, A. Albert, Real-time vision-based worker localization & hazard detection for construction, Automation in Construction 121 (2021) 103448. doi:10.1016/j.autcon.2020.103448.
URL https://www.sciencedirect.com/science/article/pii/S0926580520310281

[39] F. Fooladgar, S. Kasaei, A survey on indoor RGB-D semantic segmentation: from hand-crafted features to deep convolutional neural networks, Multimedia Tools and Applications 79 (7) (2020) 4499–4524. doi:10.1007/s11042-019-7684-3.
URL https://doi.org/10.1007/s11042-019-7684-3

[40] X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, L. Kurach, A deep learning-based framework for an automated defect detection system for sewer pipes, Automation in Construction 109 (2020) 102967. doi:10.1016/j.autcon.2019.102967.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580519307411

[41] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv:2004.10934 [cs] (Apr. 2020). doi:10.48550/arXiv.2004.10934.
URL http://arxiv.org/abs/2004.10934

[42] K. Yang, Y. Bao, J. Li, T. Fan, C. Tang, Deep learning-based YOLO for crack segmentation and measurement in metro tunnels, Automation in Construction 168 (2024) 105818. doi:10.1016/j.autcon.2024.105818.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580524005545

[43] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, arXiv:1703.06870 [cs] (Jan. 2018). doi:10.48550/arXiv.1703.06870.
URL http://arxiv.org/abs/1703.06870

[44] J. Kufuor, D. Mohanty, E. Valero Rodriguez, F. Bosché, Automatic MEP Component Detection with Deep Learning, Pattern Recognition and Automation in Construction & the Built Environment 12667 (2021) 373–388, publisher: Springer. doi:10.1007/978-3-030-68787-8_28.

[45] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv:2010.11929 [cs] (Jun. 2021). doi:10.48550/arXiv.2010.11929.
URL http://arxiv.org/abs/2010.11929

[46] J. D. Nunez-Morales, Y. Jung, M. Golparvar-Fard, Evaluation of Mapping Computer Vision Segmentation from Reality Capture to Schedule Activities for Construction Monitoring in the Absence of Detailed BIM, in: Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC), International Association for Automation and Robotics in Construction (IAARC), Lille, France, 2024, iSSN: 2413-5844. doi:10.22260/isarc2024/0110.
URL http://www.iaarc.org/publications/2024_proceedings_of_the_41st_isarc_lille_france/evaluation_of_mapping_computer_vision_segmentation_from_reality_capture_to_schedule_activities_for_construction_monitoring_in_the_absence_of_detailed_bim.html

[47] S. Xu, J. Wang, W. Shou, T. Ngo, A.-M. Sadick, X. Wang, Computer Vision Techniques in Construction: A Critical Review, Archives of Computational Methods in Engineering 28 (5) (2021) 3383–3397. doi:10.1007/s11831-020-09504-3.
URL https://link.springer.com/10.1007/s11831-020-09504-3

[48] Z. Kolar, H. Chen, X. Luo, Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images, Automation in Construction 89 (2018) 58–70. doi:10.1016/j.autcon.2018.01.003.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580517304314

[49] Arjmand,Mohsen; Jung,Jaehoon; Olsen, Michael J; Lassiter, H.Andrew; Jafari, Melika, Conceptual Design of Advanced Construction Progress Monitoring with Terrestrial and Robotic Laser Scanning SystemsPublisher: Zenodo (Mar. 2023). doi:10.5281/ZENODO.7807693.
URL https://zenodo.org/record/7807693

[50] R. Maalek, D. D. Lichti, J. Y. Ruwanpura, Automatic Recognition of Common Structural Elements from Point Clouds for Automated Progress Monitoring and Dimensional Quality Control in Reinforced Concrete Construction, Remote Sensing 11 (9) (2019) 1102. doi:10.3390/rs11091102.
URL https://www.mdpi.com/2072-4292/11/9/1102

[51] M. Choi, S. Kim, S. Kim, Semi-automated visualization method for visual inspection of buildings on BIM using 3D point cloud, Journal of Building Engineering 81 (2024) 108017. doi:10.1016/j.jobe.2023.108017.
URL https://www.sciencedirect.com/science/article/pii/S2352710223021976

[52] G. Wilson, D. J. Cook, A Survey of Unsupervised Deep Domain Adaptation, ACM Trans. Intell. Syst. Technol. 11 (5) (2020) 51:1–51:46. doi:10.1145/3400066.
URL https://dl.acm.org/doi/10.1145/3400066

[53] A. Y. Barrera-Animas, J. M. Davila Delgado, Generating real-world-like labelled synthetic datasets for construction site applications, Automation in Construction 151 (2023) 104850. doi:10.1016/j.autcon.2023.104850.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580523001103

[54] S. R. Richter, V. Vineet, S. Roth, V. Koltun, Playing for Data: Ground Truth from Computer Games, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 102–118. doi:10.1007/978-3-319-46475-6_7.

[55] S. Becker, R. Hug, W. Hübner, M. Arens, B. T. Morris, 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, pp. 13–21, arXiv:2107.00422 [cs]. doi:10.5220/0010621400003061.
URL http://arxiv.org/abs/2107.00422

[56] L. Xu, H. Liu, B. Xiao, X. Luo, DharmarajVeeramani, Z. Zhu, A systematic review and evaluation of synthetic simulated data generation strategies for deep learning applications in construction, Advanced Engineering Informatics 62 (2024) 102699. doi:10.1016/j.aei.2024.102699.
URL https://www.sciencedirect.com/science/article/pii/S1474034624003471

[57] Y. Hong, S. Park, H. Kim, Synthetic Data Generation for Indoor Scene Understanding Using BIM, International Symposium on Automation and Robotics in Construction (ISARC) Proceedings 2020 Proceedings of the

37th ISARC, Kitakyushu, Japan (2020) 334–338, iSBN: 9789529436347 Publisher: IAARC. doi:10.22260/ISARC2020/0048.
URL https://www.iaarc.org/publications/2020_proceedings_of_the_37th_isarc/synthetic_data_generation_for_indoor_scene_understanding_using_bim.html

[58] M. Alawadhi, W. Yan, BIM Hyperreality: Data Synthesis Using BIM and Hyperrealistic Rendering for Deep Learning, arXiv:2105.04103 [cs] (May 2021). doi:10.48550/arXiv.2105.04103.
URL http://arxiv.org/abs/2105.04103

[59] E. Frías, J. Pinto, R. Sousa, H. Lorenzo, L. Díaz-Vilariño, Exploiting BIM Objects for Synthetic Data Generation toward Indoor Point Cloud Classification Using Deep Learning, Journal of Computing in Civil Engineering 36 (6) (2022) 04022032, publisher: American Society of Civil Engineers. doi:10.1061/(ASCE)CP.1943-5487.0001039.
URL https://ascelibrary.org/doi/10.1061/%28ASCE%29CP.1943-5487.0001039

[60] H. Ying, R. Sacks, A. Degani, Synthetic image data generation using BIM and computer graphics for building scene understanding, Automation in Construction 154 (2023) 105016. doi:10.1016/j.autcon.2023.105016.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580523002765

[61] J. W. Ma, T. Czerniawski, F. Leite, Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds, Automation in Construction 113 (2020) 103144. doi:10.1016/j.autcon.2020.103144.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580519311884

[62] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, S. Birchfield, Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization, arXiv:1804.06516 [cs] (Apr. 2018). doi:10.48550/arXiv.1804.06516.
URL http://arxiv.org/abs/1804.06516

[63] T.-W. Huang, Y.-H. Chen, J. J. Lin, C.-S. Chen, Deep learning without human labeling for on-site rebar instance segmentation using synthetic BIM data and domain adaptation, Automation in Construction 171 (2025) 105953. doi:10.1016/j.autcon.2024.105953.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580524006897

[64] A. Gomaa, Advanced Domain Adaptation Technique for Object Detection Leveraging Semi-Automated Dataset Construction and Enhanced YOLOv8, in: 2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2024, pp. 211–214. doi:10.1109/NILES63360.2024.10753164.
URL https://ieeexplore.ieee.org/document/10753164/figures

[65] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, C. Raffel, FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, arXiv:2001.07685 [cs] (Nov. 2020). doi:10.48550/arXiv.2001.07685.
URL http://arxiv.org/abs/2001.07685

[66] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial Discriminative Domain Adaptation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2962–2971, iSSN: 1063-6919. doi:10.1109/CVPR.2017.316.
URL https://ieeexplore.ieee.org/document/8099799

[67] K. Saito, D. Kim, S. Sclaroff, T. Darrell, K. Saenko, Semi-Supervised Domain Adaptation via Minimax Entropy, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 8049–8057, iSSN: 2380-7504. doi:10.1109/ICCV.2019.00814.
URL https://ieeexplore.ieee.org/document/9010425

[68] J. O. Toyin, A. Sattineni, E. M. Wetzel, A. A. Fasoyinu, J. Kim, Augmented reality in U.S. Construction: Trends and future directions, Automation in Construction 170 (2025) 105895. doi:10.1016/j.autcon.2024.105895.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580524006319

[69] S. Ahmed, A Review on Using Opportunities of Augmented Reality and Virtual Reality in Construction Project Management, Organization, Technology and Management in Construction: an International Journal 11 (1) (2019) 1839–1852. doi:10.2478/otmcj-2018-0012.
URL https://www.sciendo.com/article/10.2478/otmcj-2018-0012

[70] C.-C. Hsieh, H.-M. Chen, S.-K. Wang, 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 27 (11) (2023) 4688–4707. doi:10.1007/s12205-023-1939-2.
URL https://www.sciencedirect.com/science/article/pii/S1226798824049705

[71] W. Fang, L. Chen, T. Zhang, C. Chen, Z. Teng, L. Wang, Head-mounted display augmented reality in manufacturing: A systematic review, Robotics and Computer-Integrated Manufacturing 83 (2023) 102567. doi:10.1016/j.rcim.2023.102567.
URL https://www.sciencedirect.com/science/article/pii/S0736584523000431

[72] Z. Oufqir, A. El Abderrahmani, K. Satori, ARKit and ARCore in serve to augmented reality, in: 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), 2020, pp. 1–7. doi:10.1109/ISCV49265.2020.9204243.
URL https://ieeexplore.ieee.org/abstract/document/9204243

[73] H.-L. Chi, S.-C. Kang, X. Wang, Research trends and opportunities of augmented reality applications in architecture, engineering, and construction, Automation in Construction 33 (2013) 116–122. doi:10.1016/j.autcon.2012.12.017.
URL https://www.sciencedirect.com/science/article/pii/S0926580513000022

[74] A. C. P. Martins, I. R. Castellano, K. M. Lenz César Júnior, J. M. Franco De Carvalho, F. G. Bellon, D. S. De Oliveira, J. C. L. Ribeiro, BIM-based mixed reality application for bridge inspection, Automation in Construction 168 (2024) 105775. doi:10.1016/j.autcon.2024.105775.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580524005119

[75] M. Kopsida, I. Brilakis, Real-Time Volume-to-Plane Comparison for Mixed Reality–Based Progress Monitoring, Journal of Computing in Civil Engineering 34 (4) (2020) 04020016. doi:10.1061/(ASCE)CP.1943-5487.0000896.
URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CP.1943-5487.0000896

[76] W.-L. Kuo, B.-K. Huang, S.-H. Hsieh, Y.-H. Tsai, Y.-T. Chang, Integration of BIM and ar with VSLAM to assist in construction site inspection, Journal of Civil Engineering and Management 31 (6) (2025) 646–669. doi:10.3846/jcem.2025.24360.
URL https://journals.vilniustech.lt/index.php/JCEM/article/view/24360

[77] D. Shojaei, P. Jafary, Z. Zhang, Mixed Reality-Based Concrete Crack Detection and Skeleton Extraction Using Deep Learning and Image Processing, Electronics 13 (22) (2024) 4426, publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/electronics13224426.
URL https://www.mdpi.com/2079-9292/13/22/4426

[78] B. Tao, F. Bosché, J. Li, Mixed Reality-based MEP construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison, Automation in Construction 168 (2024) 105852. doi:10.1016/j.autcon.2024.105852.
URL https://linkinghub.elsevier.com/retrieve/pii/S0926580524005880

[79] S. Rankohi, L. Waugh, Review and analysis of augmented reality literature for construction industry, Visualization in Engineering 1 (1) (2013) 9. doi:10.1186/2213-7459-1-9.
URL https://doi.org/10.1186/2213-7459-1-9

[80] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, issue: arXiv:1506.01497 arXiv: 1506.01497 [cs] (Jan. 2016). doi:10.48550/arXiv.1506.01497.
URL http://arxiv.org/abs/1506.01497

[81] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, R. Girshick, Segment Anything, arXiv:2304.02643 [cs] (Apr. 2023). doi:10.48550/arXiv.2304.02643.
URL http://arxiv.org/abs/2304.02643

[82] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: Common Objects in Context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 740–755. doi:10.1007/978-3-319-10602-1_48.

[83] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. Stühmer, T. J. Cashman, B. Tekin, J. L. Schönberger, P. Olszta, M. Pollefeys, HoloLens 2 Research Mode as a Tool for Computer Vision Research, arXiv:2008.11239 [cs] (Aug. 2020). doi:10.48550/arXiv.2008.11239.
URL http://arxiv.org/abs/2008.11239

[84] G. D. Evangelidis, E. Z. Psarakis, Parametric image alignment using enhanced correlation coefficient maximization, IEEE transactions on pattern analysis and machine intelligence 30 (10) (2008) 1858–1865. doi:10.1109/TPAMI.2008.113.

[85] M. Baubriaud, MEP-SEG DATASET : SYNTHETIC IMAGES GENERATED FROM BUILDING INFORMATION MODELING (BIM) (Dec. 2023).
URL https://ec-lyon.hal.science/hal-04488735

[86] A. Khairadeen Ali, O. J. Lee, D. Lee, C. Park, Remote Indoor Construction Progress Monitoring Using Extended Reality, Sustainability 13 (4) (2021) 2290. doi:10.3390/su13042290.
URL https://www.mdpi.com/2071-1050/13/4/2290

[87] E. Yildiz, A. Costa, J. Miranda, P. Faria, Smart maintenance solutions: Ar- and vr-enhanced digital twin powered by fiware, Sensors 25 (3) (2025) 845. doi:10.3390/s25030845.

[88] F. A. Ghansah, W. Lu, Optimizing facilities management through artificial intelligence and digital twin technology in mega-facilities, Sustainability 17 (5) (2025) 1826. doi:10.3390/su17051826.

[89] O. Schmedemann, D. Holst, J. Lund, T. Schuppstuhl, Introducing a tool for synthetic defect image data generation: enhancing industrial surface inspection, in: Automated Visual Inspection and Machine Learning 2025, Vol. 13459, SPIE, 2025, p. 134590E. doi:10.1117/12.3052620.