YOSO (You Only Scan Once): A Large-Small Model Co-Adapter Framework for AI–Human Collaborative Scan-to-BIM Automation in Wastewater Infrastructure

Winson TC Leung[1], CH Wong[1], Senna CS Ng[1], Simon YO Lai[1], Carry PS Cheung[1], CY Lam[2], YM Qin[2] and LF Ren[2]
[1] Drainage Services Department, the HKSAR Government, Hong Kong, People's Republic of China
[2] Manifold Tech Limited, Hong Kong, People's Republic of China

ABSTRACT

Hong Kong's ageing wastewater infrastructure faces escalating climate risks, necessitating efficient asset management and swift upgrades through Modular Integrated Construction (MiC) and Multi-trade integrated Mechanical, Electrical and Plumbing (MiMEP) workflows. These efforts demand precise as-built BIM, yet traditional Scan-to-BIM methods remain hindered by labour-intensive segmentation, outdated semantic data and fragmented toolsets. This paper presents YOSO (You Only Scan Once), an AI-assisted framework built on a Large-Small Model Co-Adapter pipeline to automate Scan-to-BIM conversion. The Co-Adapter synergises a small model in a lightweight edge device, MindPalace Pocket, for real-time spatial data capture and a large model in a cloud-based DBSCAN-driven workflow for clustering and BIM generation. A confidence-driven interface routes low-certainty detections to engineers, reducing manual input by 80% while ensuring accuracy. Novel metrics—Intersection-to-Manual Ratio (IMR) for component prefabrication and Euclidean Distance (ED) for clash-free installation—are proposed for evaluation. Validated at Ting Kok Road Sewage Pumping Station No.8 using light tubes (representing standardised components) and gate valves (representing irregular geometries), YOSO achieves 99.9% and 84.6% average IMR, with minor average ED of 0.166 m and 0.116 m, respectively. By bridging edge-cloud AI with human expertise, YOSO directly supports digitalisation goals, offering a scalable blueprint for modernising ageing infrastructure globally.

KEYWORDS Adaptive segmentation, AI–human collaboration, Building Information Modelling (BIM), Edge-cloud computing, Infrastructure resilience, Scan-to-BIM
CONTACT Winson TC Leung tszchiuleung@dsd.gov.hk
Received 19 December 2025

1. Introduction

Hong Kong's wastewater infrastructure—comprising over 360 facilities and 2,400 km of pipelines—faces unprecedented challenges from climate change, with extreme rainfall events projected to increase by 51% by 2100 (Lee et al., 2012). These conditions accelerate corrosion in ageing pipes and mechanical systems (Smith et al., 2017), necessitating urgent upgrades through Modular Integrated Construction (MiC) and Multi-trade integrated Mechanical, Electrical and Plumbing (MiMEP) workflows. To meet the tight schedules of the dry season, a precise as-built Building Information Modelling (BIM) model is critical for component prefabrication and clash-free installation. Traditional Scan-to-BIM workflows, however, remain bottlenecked by labour-intensive manual segmentation and outdated semantic datasets that fail to capture the complexity of wastewater environments. Existing AI solutions, such as Scan2BIM-NET (Perez-Perez et al., 2021) and infrastructure-focused Scan-to-BIM frameworks (Justo Domínguez et al., 2021), underperform in dense, non-standardised environments due to the reliance on generic training data and fragmented workflows.

2. Literature Review

2.1. Evolution of Scan-to-BIM Automation

Early Scan-to-BIM workflows necessitated labour-intensive manual labelling of objects (e.g., pipes, valves) within point clouds, a process that, while precise, proved prohibitively time consuming for large-scale infrastructure projects (Jiang et al., 2021). Subsequent advancements aimed to automate segmentation through hybrid semantic–geometric methodologies. For example, Kim et al. (2022) combined PointNet++-based semantic segmentation with geometric reasoning to classify architectural elements (e.g., walls, floors) in standardised building layouts. However, such approaches remain constrained by their reliance on structured environments, failing to adapt to the non-standardised and spatially complex settings of wastewater infrastructure.

2.2. AI-Driven Segmentation: Promise and Persistent Challenges

Early algorithmic approaches to automated segmentation, such as planar patch extraction and visibility reasoning by Xiong et al. (2013), demonstrated the feasibility of converting raw 3D point cloud data into

BIM elements. However, these methods lacked integration with real-time workflows and struggled with dynamic infrastructure environments. Subsequent advancements in deep learning, exemplified by Scan2BIM-NET (Perez-Perez et al., 2021), achieved 86.13% accuracy in geographic labelling but remained constrained by the reliance on generic training datasets, limiting their efficacy in non-standardised contexts like wastewater infrastructure. Maxime et al. (2024) further emphasised AI's potential in Scan-to-BIM but critiqued the fragmented implementation of workflows, which perpetuated delays.

Collectively, these studies highlight a critical limitation: while AI enhances segmentation accuracy, its isolation from domain-specific datasets and disjointed toolchains hinders practical deployment in infrastructure digitalisation.

2.3. Critical Research Gaps

Three unresolved challenges impede Scan-to-BIM adoption for wastewater infrastructure:

1. Dataset Scarcity: Existing datasets, including those underpinning Scan2BIM-NET (Perez-Perez et al., 2021), fail to capture the local wastewater infrastructure's unique spatial and semantic complexity.
2. Workflow Fragmentation: As highlighted by Kim et al. (2022) and Maxime et al. (2024), disjointed tools for scanning, segmentation and BIM generation create inefficiencies, particularly in MiC/MiMEP projects requiring rapid as-built modelling.
3. Non-adaptive Processing: Prior methodologies, such as Xiong et al.'s (2013) algorithmic workflows, lack mechanisms to dynamically balance automation with human oversight, resulting in either excessive manual intervention or error-prone fully automated outputs.

2.4. YOSO's Co-Adapter Paradigm

This study introduces YOSO, a Large-Small Model Co-Adapter framework that redefines Scan-to-BIM for complex infrastructure. YOSO addresses the identified gaps through its novel Large-Small Model Co-Adapter, a unified framework designed to harmonise edge-device agility with cloud-based precision. Unlike prior methodologies that handle scanning, segmentation, and BIM generation independently, YOSO integrates these stages into a cohesive pipeline. The Small Model leverages MindPalace Pocket—a handheld multi-sensor fusion system—to enable rapid, real-time spatial data capture in confined environments, overcoming the inefficiencies of traditional Terrestrial Laser Scanners (TLS). The Co-Adapter bridges edge and cloud workflows, introducing adaptive AI–human collaboration: high-confidence predictions are automated, while low-certainty cases are routed to engineers for verification, ensuring accuracy without excessive manual input. The Large Model refines outputs through advanced clustering and semantic enrichment, directly addressing dataset scarcity by iteratively building a localised wastewater infrastructure dataset. By unifying fragmented workflows and embedding adaptive processing mechanisms, YOSO resolves the critical limitations of prior Scan-to-BIM approaches, offering a scalable solution for infrastructure digitalisation.

This study's contributions are summarised as follows:

1. Unified Workflow: A cohesive pipeline integrating real-time scanning, AI segmentation, and BIM generation, eliminating workflow fragmentation.
2. Specialised Datasets: Hong Kong's first wastewater-specific semantic dataset (660 labelled images, continually update when more sites are applied) enables precise AI training for components like valves and ducts used in DSD's infrastructure.
3. Confidence-Driven Processing: Adaptive AI autonomously handles high-confidence predictions while flagging ambiguities for human verification.
4. MiC/MiMEP-tailored Metrics: IMR for component completeness (ensuring prefabrication accuracy) and ED for spatial fidelity (ensuring clash-free installation) are proposed to directly support MiC/MiMEP workflows.

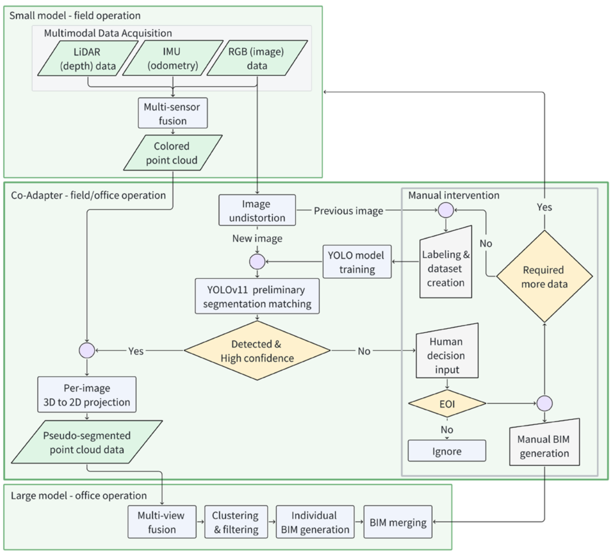


Figure 1. Pipeline of "YOSO".

3. Methodology

The proposed YOSO framework integrates a Large-Small Model Co-Adapter pipeline to automate Scan-to-BIM conversion. As illustrated in Figure 1, the workflow comprises three stages: (1) Small Model for real-time spatial data capture, (2) Co-Adapter for AI-driven segmentation and pseudo-segmentation, and (3) Large

Model for point cloud refinement and BIM generation. This cohesive architecture addresses the inefficiencies of fragmented workflows while balancing automation with human oversight.

3.1. Small Model: Robust Reality Capture in Non-exposed Spaces

The Small Model employs MindPalace Pocket, a handheld LiDAR-inertial-visual SLAM system, to enable rapid, drift-free 3D mapping in GPS-denied or low-light environments. The device integrates dual fisheye cameras (5 megapixel resolution, 5 frames per second), an Inertial Measurement Unit (IMU) and edge computing capabilities, capturing 200,000 LiDAR points per second alongside synchronised RGB data. A 360° non-repetitive scanning mode minimises occlusions while automatically filtering operator-generated noise within a 1-metre radius. Post-scan, the system outputs a dense, colourised point cloud without manual post-processing. This integration of LiDAR, real-time imaging, and automated processing delivers high-quality, precise and visually enriched 3D reconstructions.

3.2. Co-Adapter: Adaptive Segmentation via Edge-Cloud Synergy

Existing Scan-to-BIM methodologies predominantly rely on direct 3D point cloud segmentation, which depends on pre-labelled datasets or generic object libraries (Perez-Perez et al., 2021; Jiang et al., 2021). However, such approaches are ill suited for sewage pumping stations, where specialised mechanical and electrical components (e.g., gate valves, light tubes) lack standardised training data. To address this gap, the Co-Adapter introduces a 2D image segmentation-first strategy, leveraging the ubiquity of fisheye camera data captured during scanning to bypass computational and labelling bottlenecks inherent to 3D workflows.

At the core of the Co-Adapter lies its pseudo-segmentation process. As illustrated in Figure 2, YOLOv11, a domain-adapted iteration of the YOLO architecture (Redmon et al., 2016), generates precise 2D masks (Figure 2b) from undistorted fisheye images (Figure 2a). These masks are spatially mapped onto 3D point clouds using pose-aware transformation matrices derived from the scanning device's positional data, effectively segmenting target objects while relegating residual points to background data. This method circumvents the computational complexity of raw point cloud segmentation.

The selection of YOLOv11 over alternatives such as Meta AI's Segment Anything Model (SAM) (Kirillov et al., 2023) is driven by its suitability for automated workflows. While SAM excels in interactive 2D image segmentation, its dependency on manual prompts and reprocessing of entire image sequences renders it impractical for dynamic environments like wastewater facilities. In contrast, YOLOv11's custom training capability facilitates precise identification of irregular geometries and adaptability to new equipment types. The training process begins with geometric undistortion of fisheye images to correct lens distortions, followed by manual annotation of six critical infrastructure elements, namely air ducts, DI pipes, light tubes, gate valves, check valves, and flowmeters.

The framework incorporates a confidence-driven adaptive workflow: predictions with high confidence scores are automated, while ambiguous cases trigger human intervention via an interactive interface. This human-in-the-loop mechanism iteratively refines the model by incorporating verified results into the training dataset, progressively reducing manual verification as accuracy improves. By prioritising 2D segmentation and adaptive human–AI collaboration, the Co-Adapter not only generates a foundational dataset for wastewater infrastructure but also establishes a scalable framework for future digitalisation projects. This approach directly addresses the dual challenges of dataset scarcity and workflow fragmentation, positioning YOSO as a robust solution for MiC/MiMEP-driven infrastructure upgrades.

(a)

(b)

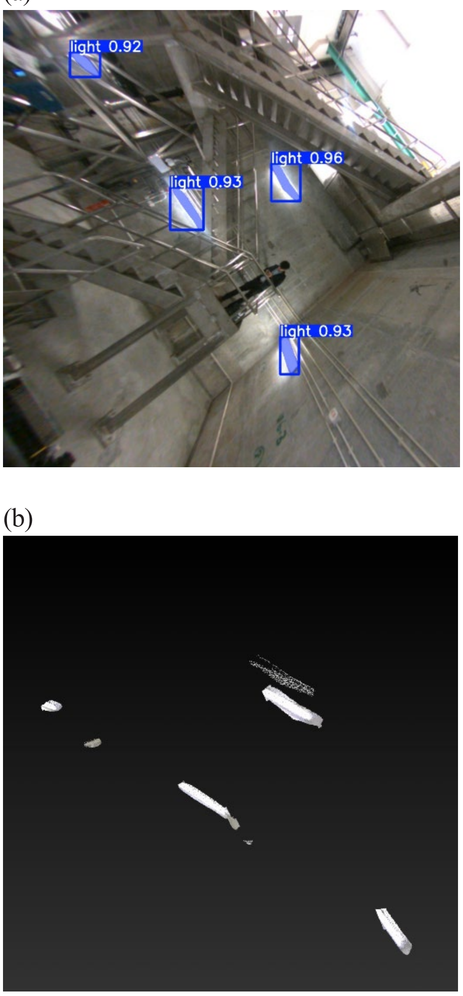


Figure 2. (a) Undistorted fisheye image captured during scanning; (b) Corresponding 2D segmentation mask generated by YOLOv11, highlighting light tube detection for pseudo-segmentation in the Co-Adapter workflow.

3.3. Large Model: Point Cloud Clustering and Automated BIM Generation

The Large Model refines pseudo-segmented point clouds from the Co-Adapter and automates BIM generation through a three-stage workflow. First, data consolidation aggregates point clouds classified under the same object category (e.g., light tubes) into unified datasets, mitigating fragmentation inherent in per-image processing. Next, the DBSCAN clustering algorithm groups spatially proximate points into discrete object clusters while segregating outliers. Clusters are filtered using geometric parameters, including point count, orientation and spatial distribution, to remove residual noise, preserving only valid objects. Finally, noise handling and completeness assurance merges unresolved noise into a consolidated background dataset. Objects undetected during prior stages (e.g., due to occlusions) are manually segmented from this background, ensuring comprehensive component detection.

Validated clusters are converted into 3D mesh models using Autodesk ReCap Pro's scan-to-mesh tool, which reconstructs surfaces with high geometric fidelity. These meshes, combined with contextual background data, are integrated into a unified BIM framework. The framework supports dynamic metadata embedding (e.g., specifications), enabling real-time updates and lifecycle management

3.4. Experimental Framework

The YOSO pipeline was tested through an end-to-end experiment at Ting Kok Road Sewage Pumping Station No. 8 (TKRSPS No. 8), a representative of Hong Kong's wastewater infrastructure. The YOLOv11 model was trained on 660 manually labelled fisheye images (9:1 train-validation split) from prior DSD scans, annotated with the six critical components. During testing, a 10-minute on-site scan using MindPalace Pocket captured new fisheye images and point clouds (~50 million points), processed through the Co-Adapter's pseudo-segmentation and Large Model's clustering workflow.

3.5. Evaluation Metrics

The evaluation of Scan-to-BIM workflows has historically relied on traditional metrics such as point-level segmentation accuracy and volumetric overlap ratios (e.g., Intersection over Union, IoU). While these metrics provide broad insights into spatial alignment, they exhibit critical limitations when applied to MiC/MiMEP. Point-level accuracy, which averages classification correctness across all points in a dataset, obscures localised errors, such as a partially omitted valve or misaligned pipe flange, by diluting critical inaccuracies within aggregate statistics. Similarly, IoU, calculated as the ratio of intersection to union volumes, penalises both omissions and over-segmentation equally. This normalisation risks misleading results: a component missing 10% of its true volume but over-segmented by 15% could yield a passable IoU score (~78%), masking deficiencies that jeopardise prefabrication.

To address these shortcomings, this study introduces two purpose-built metrics, Intersection-to-Manual Ratio (IMR) and Euclidean Distance (ED), designed to align evaluation with the operational demands of infrastructure digitalisation.

(1) Intersection-to-Manual Ratio (IMR)

IMR quantifies the completeness of a component's geometric capture in the BIM model, defined as:

$$IMR = \frac{(Volume_{YOSO} \cap Volume_{manual})}{Volume_{manual}}. \quad (1)$$

Unlike IoU, which normalises against the union of volumes, IMR focuses exclusively on ensuring that the automated model fully encompasses the ground truth. For instance, an IMR of 99% indicates near-total overlap with manually validated geometry, minimising the risk of prefabrication errors. This metric directly responds to the inadequacy of traditional approaches in penalising omissions, a critical flaw in modular workflows where incomplete components necessitate costly rework.

(2) Euclidean Distance (ED)

ED complements IMR by measuring positional fidelity, calculated as the three-dimensional distance between the centroids of a component's bounding box in the YOSO-generated and manual models:

$$ED = \sqrt{(X_{YOSO} - X_{manual})^2 + (Y_{YOSO} - Y_{manual})^2 + (Z_{YOSO} - Z_{manual})^2} \ . \quad (2)$$

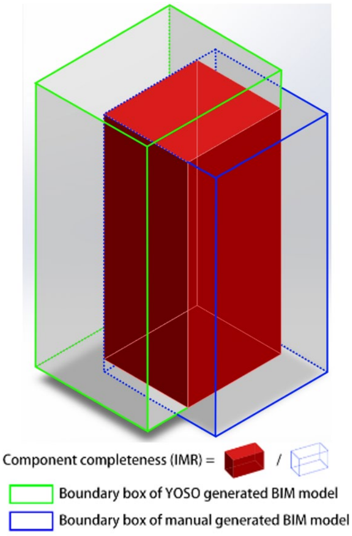
A low ED ensures minimal deviation from real-world coordinates, critical for seamless assembly. For example, a pump displaced by 0.5 m (high ED) might misalign with pre-cut floor penetrations, whereas traditional point-level metrics could overlook such localised inaccuracies.

3.5.1 Rationale for Metric Design

The development of IMR and ED stems from two unresolved challenges in Scan-to-BIM for infrastructure:

(1) Incompleteness in Prefabrication: Traditional metrics tolerate partial component capture, risking clashes during off-site manufacturing. IMR enforces rigorous completeness, ensuring that prefabricated units mirror the as-built conditions.
(2) Spatial Drift in Assembly: Positional errors, even for fully captured components, delay on-site installation. ED mitigates this by penalising centroid deviations, a factor ignored by volumetric averages.

(a)



(b)

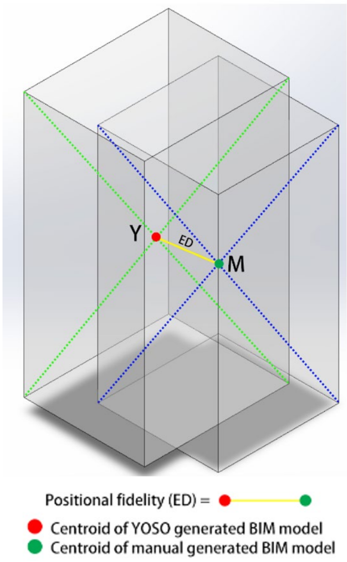


Figure 3. Visualisation of the novel evaluation metrics: (a) Intersection-to-Manual Ratio (IMR) demonstrates component completeness by the overlap (solid red) between automated (hollow green) and manual (hollow blue) segmentations; and (b) Euclidean Distance (ED) indicates positional accuracy through the centroid displacement (yellow line) between automated (red) and manual (green) models.

4. Experiments and Results

The experimental validation of the YOSO framework was conducted at TKRSPS No. 8, a representative site within Hong Kong's ageing wastewater infrastructure. The study targeted two critical components: light tubes (standardised geometries) and gate valves (irregular geometries), selected to evaluate YOSO's adaptability across diverse structural forms. The experiment encompassed four stages: (1) Small Model field operation, (2) Co-Adapter development, (3) Large Model refinement and (4) Metric-driven validation.

4.1. Small Model: Field Operation

The Small Model, implemented via the handheld MindPalace Pocket, executed rapid spatial and visual data acquisition at TKRSPS No. 8. The device captured around 50 million RGB point cloud points and more than 1,700 fisheye images per camera within 10 minutes, leveraging dual fisheye lenses for wide-angle coverage in congested spaces. An extension rod facilitated access to occluded areas (Figure 4a), while real-time feedback on a portable device verified scan completeness within 3 minutes, minimising downtime. Post-scan, fisheye images underwent geometric undistortion (Figure 4b) to correct lens distortions, ensuring alignment with spatial data. The LiDAR system automatically filtered operator-generated noise within a 1-metre radius, preserving accuracy. Despite challenges such as occlusions from large equipment, the Small Model's integration of real-time validation and adaptive hardware demonstrated robustness, seamlessly feeding undistorted data into the Co-Adapter for subsequent processing.

(a)

(b)

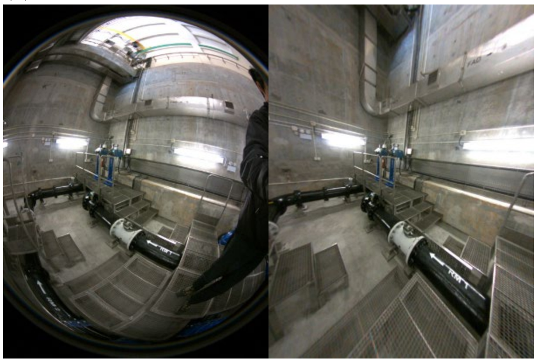


Figure 4. Small Model data acquisition workflow: (a) Scanning in congested spaces using an extension rod, with real-time point cloud visualisation transmitted to a portable device for quality assurance; and (b) Fisheye camera output during scanning: raw image (left) and geometrically undistorted result (right), correcting lens distortions for spatial alignment.

4.2. Co-Adapter Development and Pseudo-segmented Point Cloud Segmentation

The Co-Adapter constitutes the second stage of the YOSO pipeline, refining raw spatial and visual data from the Small Model into structured inputs for BIM generation. This stage integrates adaptive artificial intelligence with human oversight to address the challenges of segmenting irregular geometries and non-standardised components prevalent in wastewater infrastructure.

Fisheye images captured during field operations undergo geometric undistortion (Figure 4b) to correct lens-induced distortions, ensuring precise alignment with LiDAR-derived point clouds. A curated dataset of 660 representative images, manually annotated with six critical infrastructure elements, namely air ducts, DI pipes, light tubes, gate valves, check valves, and flowmeters, was partitioned into a 9:1 training–validation split to enable and optimise the YOLOv11 model for object detection in the specific context of wastewater infrastructure generalisability. Figure 5a illustrates the annotation process on undistorted images and subsequent validation of the model's iterative refinement.

Following training, new scans are processed through the YOLOv11 model to generate 2D segmentation masks for target objects. A confidence threshold of 0.5 is employed to automate high-certainty predictions, while low-confidence detections (e.g., a misclassified flow meter scoring 0.32, as shown in Figure 5b) are routed to human operators for verification. This hybrid approach can reduce manual effort, ensuring reliability in cluttered environments where algorithmic ambiguities arise. Undetected objects in the Co-Adapter phase are manually extracted from consolidated background data, preserving completeness for downstream BIM workflows.

The pseudo-segmentation process leverages pose-aware transformation matrices, derived from the scanning device's positional data, to project 2D segmentation masks (Figure 2b) onto 3D point clouds. Points intersecting with masks are classified as target objects (e.g., light tubes), while residual points are relegated to background data for subsequent refinement. This methodology circumvents the computational complexity of direct 3D point cloud segmentation, achieving scalable and efficient component isolation.

The Co-Adapter's innovation lies in its dual contribution: (1) the creation of Hong Kong's first localised wastewater infrastructure dataset, enabling continual model adaptation to new equipment types, and (2) the integration of human-in-the-loop validation, which ensures accuracy while minimising manual intervention.

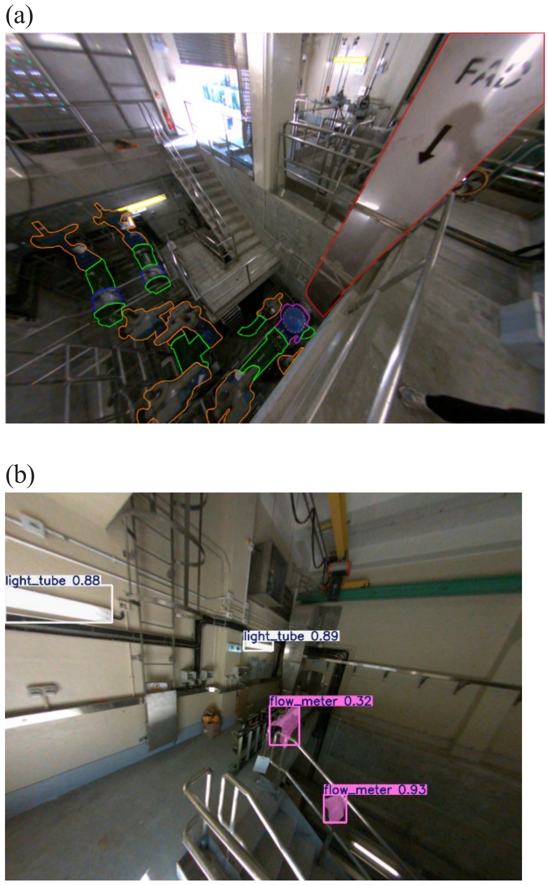


Figure 5. Co-Adapter training and validation: (a) Manual annotation of target objects on geometrically undistorted images; and (b) Validation of YOLOv11 segmentation accuracy, highlighting a misclassified flow meter (confidence score: 0.32) necessitating human intervention for correction.

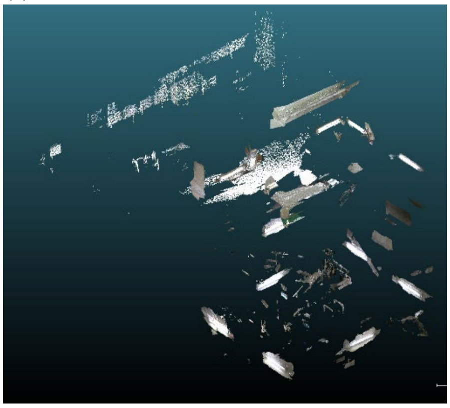
4.3. Large Model: Clustering, Filtering and BIM Integration

The Large Model constitutes the final stage of the YOSO framework, transforming pseudo-segmented point clouds into semantically enriched BIM through a structured two-phase workflow.

4.3.1. Clustering and Filtering

Pseudo-segmented point clouds, fragmented across individual images, are consolidated into unified datasets per object class. The DBSCAN algorithm clusters spatially proximate points into discrete objects while segregating outliers based on local density. This density-based approach ensures robustness in cluttered environments, overcoming the limitations of traditional distance-based clustering methods. Clusters are refined using geometric parameters, such as point count, orientation and spatial distribution, to eliminate noise, achieving clean per-object segmentation (Figure 6(a) vs. 6(b)). Residual noise is merged into a consolidated background dataset, enabling manual extraction of undetected components and ensuring comprehensive model completeness. This iterative refinement process addresses gaps in automated workflows, balancing efficiency with precision.

(a)



(b)

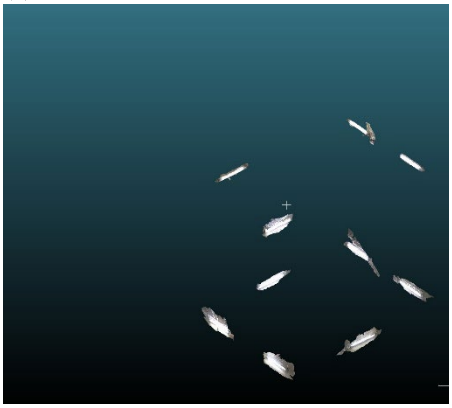


Figure 6. Large Model clustering and filtering outcomes: (a) Pseudo-segmented point cloud data of light tubes prior to refinement; and (b) Post-processing result after DBSCAN clustering and geometric filtering (point count, orientation), isolating 11 discrete light tube clusters and removing residual noise.

4.3.2. BIM Generation and Integration

Validated clusters are converted into 3D mesh models using Autodesk ReCap Pro's scan-to-mesh tool, preserving geometric fidelity (Figure 7). For instance, 11 light tubes were processed into individual meshes, while unclassified data formed a contextual background model. These meshes are integrated into a unified BIM framework, enriched with asset-specific metadata (e.g., maintenance schedules, valve specifications) to enable dynamic updates and lifecycle management.

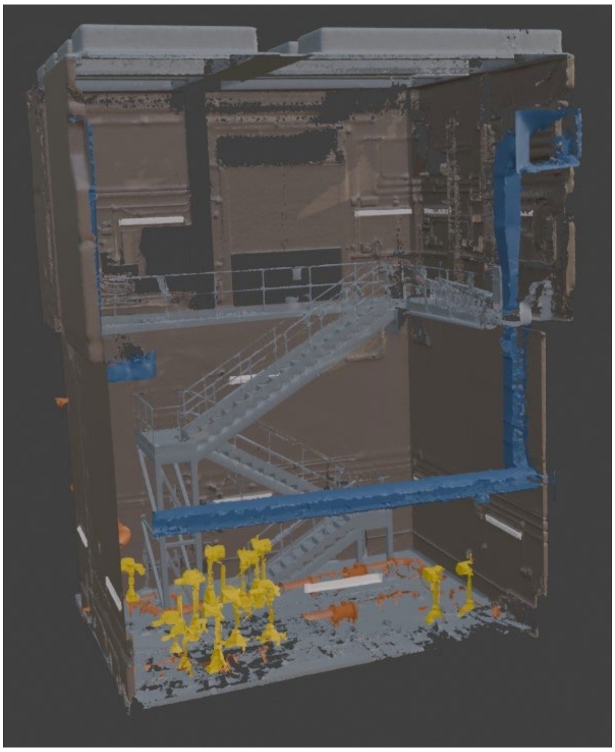


Figure 7. Finalised BIM model of TKRSPS No.8 generated through YOSO workflow.

5. Results and Discussion

5.1. Experimental Results

The YOSO framework was validated through an end-to-end experiment at Ting Kok Road Sewage Pumping Station No. 8. The semi-automated workflow achieved a total processing time of five days, which included 10 minutes of on-site scanning with the MindPalace Pocket, four days for initial manual annotation of the training dataset (a one-time investment), and one day of semi-automated office processing managed by a single operator. In contrast, a traditional manual Scan-to-BIM workflow for a comparable 90 m² pumping station typically requires one week. This represents an 80% reduction in total manual effort.

5.1.1. Component-wise Accuracy Analysis

The framework successfully segmented six types of critical infrastructure components: light tubes (n = 11), gate valves (n = 12), check valves (n = 3), air ducts (n = 2), ductile iron (DI) pipes (n = 2) and flow meters (n=2). Performance was evaluated using the proposed metrics, IMR for geometric completeness and ED for positional fidelity. The component-wise accuracy results are summarised in Table 1.

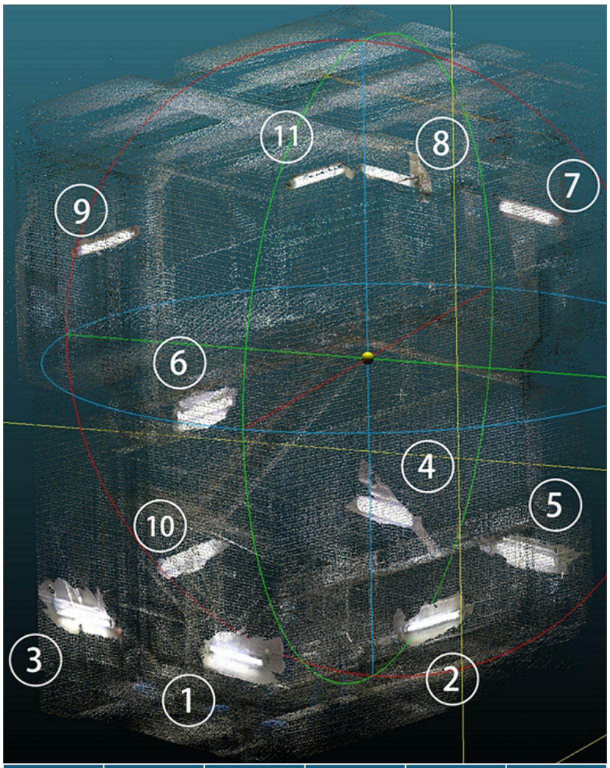
Table 1. Component-wise accuracy of the YOSO pipeline. Manual modelling results (IMR = 100%, ED = 0 m) serve as the baseline for each component.

Component	Sample Size (n)	YOSO IMR (mean ± SD)	YOSO ED (mean ± SD)
Light Tubes	11	99.9% ± 0.2%	0.166 m ± 0.093 m
Gate Valves	12	84.6% ± 12.7%	0.116 m ± 0.072 m
Check Valves	3	49.1% ± 26.4%	0.107 m ± 0.045 m
Air Ducts	2	64.0% ± 29.0%	0.466 m ± 0.036 m
DI Pipes	2	77.1% ± 26.3%	0.309 m ± 0.099 m
Flow Meters	2	87.4% ± 4.08%	0.025 m ± 0.013 m

For a robust comparative analysis, this study focuses on light tubes and gate valves. These components were selected because they have the largest sample sizes (n = 11 and n = 12, respectively), providing statistical reliability, and they represent the two critical extremes in infrastructure: standardised versus irregular geometries. The remaining components, while successfully segmented, have limited sample sizes ($n \leq 3$) due to their lower count within the test site; thus, they are reported but not emphasised in comparative statistical tests.

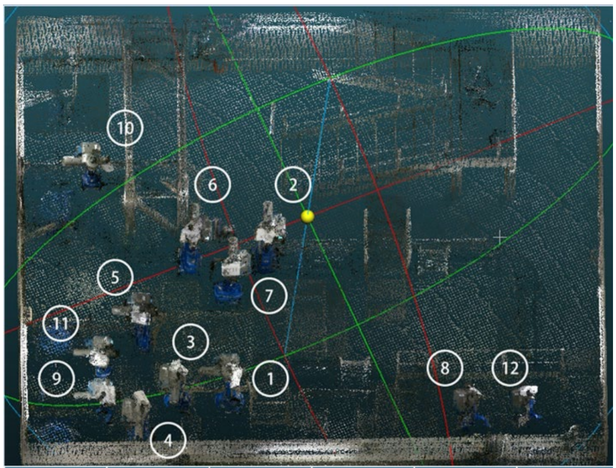
The results reveal a distinct performance profile for each component type. Light tubes achieved near-perfect geometric completeness (IMR = 99.9%) but a higher mean positional error (ED = 0.166 m). This is attributed to environmental noise, such as light diffusion on reflective surfaces, which can inflate the point cloud cluster and shift the calculated centroid of these elongated objects. Conversely, gate valves showed lower but robust completeness (IMR = 84.6%) with superior positional accuracy (ED = 0.116 m). Their compact, dense installations yield more stable point clouds, leading to more consistent centroid estimation despite their irregular shape.

Independent two-tailed t-tests were conducted to compare the performance between light tubes and gate valves. The difference in mean IMR was found to be statistically significant ($p = 0.0007$). This confirms that the framework achieves significantly higher geometric completeness for standardised components (light tubes) compared to irregular ones (gate valves). In contrast, the difference in mean ED was not statistically significant ($p = 0.1617$). This indicates that while light tubes exhibited a numerically larger average positional deviation (0.166 m vs. 0.116 m), this difference is not statistically conclusive given the sample sizes and variability. The result suggests that YOSO provides a consistent level of positional accuracy for both standardised and irregular components, with both mean ED values falling within a practically acceptable range for clash-free MiC/MiMEP installation.



Light Tube	IMR (%)	ED (m)	Light Tube	IMR (%)	ED (m)
No.1	100	0.165	No.7	99.3	0.101
No.2	100	0.216	No.8	100	0.284
No.3	100	0.248	No.9	100	0.044
No.4	100	0.297	No.10	100	0.203
No.5	100	0.155	No.11	100	0.081
No.6	100	0.031			

Figure 8. Light tube annotations and evaluation results.



Gate Valve	IMR (%)	ED (m)	Gate Valve	IMR (%)	ED (m)
No.1	90.2	0.144	No.7	88.8	0.067
No.2	89.9	0.073	No.8	65.9	0.120
No.3	95.0	0.160	No.9	90.2	0.014
No.4	87.0	0.100	No.10	97.2	0.021
No.5	84.9	0.169	No.11	83.8	0.077
No.6	90.2	0.259	No.12	52.6	0.187

Figure 9. Gate valve annotations and evaluation results.

These findings validate the complementary utility of the proposed dual-metric framework. A high IMR ensures that a component's geometry is fully captured for accurate off-site prefabrication, while a low ED guarantees that its position is correctly located for clash-free on-site assembly. The high IMR and low ED for light tubes (Figure 8) and gate valves (Figure 9) collectively demonstrate YOSO's ability to balance automation with precision, fulfilling the study's objective of generating reliable as-built BIM models for wastewater infrastructure.

5.1.2. Comparative Analysis with Manual Workflows

Benchmarking YOSO against traditional manual Scan-to-BIM processes reveals transformative improvements in efficiency and operational scalability, as summarised in Table 2.

Table 2. Efficiency comparison between manual and YOSO workflows for a 90 m² pumping station.

Metric	Manual Workflow	YOSO Workflow
On-site Scanning Time	1 day	10 minutes (~99% reduction)
Total Processing Time	1 week	1 day (80% reduction)
Primary Human Role	100% manual segmentation and modelling	Manages scanning and verifies low-confidence AI predictions

The most substantial gain is in regard to time efficiency. YOSO reduces the on-site scanning phase from one day to 10 minutes, enabled by the portability and real-time feedback of the handheld MindPalace Pocket. The integrated AI–human pipeline then condenses the total office-based processing from one week to a single day. While manual workflows can achieve high precision, they are vulnerable to inconsistencies from human fatigue. YOSO's AI-driven approach provides consistent, repeatable outputs. Operationally, YOSO replaces a suite of disjointed software tools with a unified pipeline, enabling a single operator to manage the entire process from scanning to validated BIM. This integration reduces error propagation and is critical for resource-constrained projects requiring rapid delivery.

Despite these advancements, YOSO's current limitations, such as fragmented results for non-standardised objects (e.g. air ducts), mirror the challenges inherent to manual workflows. However, unlike static manual methods, YOSO's iterative learning framework and localised dataset allow continual refinement, ensuring adaptability to evolving infrastructure needs. By integrating human oversight for low-confidence predictions (e.g., the misclassified flow meter in Figure 5(b)), YOSO bridges the gap between automation and precision, setting a new standard for Scan-to-BIM that prioritises both speed and reliability.

5.2. Limitations and Future Directions

While the YOSO framework demonstrates high potential for automating Scan-to-BIM, several limitations must be acknowledged and addressed to realise its full scalability and robustness. A primary constraint lies in the time-intensive dataset creation, requiring four days of manual annotation to curate a limited-sized dataset. While this one-time investment enables future automation, it delays immediate deployment in new facilities lacking standardised component libraries. To mitigate this, future efforts will focus on developing a more comprehensive semantic library, automating the labelling of common components (e.g., valves, pipes) to accelerate dataset generation and enhance adaptability across diverse sites.

A second limitation stems from the fragmented workflow separating on-site scanning and off-site BIM editing. Although MindPalace Pocket reduces on-site scanning to 10 minutes, post-processing remains office bound, limiting real-time model validation, which is critical for dynamic environments. To bridge this gap, cloud-edge computing integration will unify scanning, segmentation and BIM generation into a cohesive workflow, enabling iterative adjustments during fieldwork and fostering collaborative updates between on-site and off-site teams.

A third limitation is that the framework struggles with non-standardised objects such as air ducts, whose irregular geometries and occlusions result in fragmented segmentation. For instance, light diffusion on reflective walls introduced noise into light tube clusters (Figure 6), slightly elevating positional errors (ED = 0.166 m). Similarly, dense spatial configurations of gate valves challenged segmentation completeness (IMR = 84.6%). To address this, UAV-LiDAR systems will be deployed to capture elevated or occluded areas, while advanced AI models will prioritise context-aware segmentation, leveraging spatial and semantic context to improve the detection of non-modular components.

Finally, the dependency on manual intervention for low-confidence predictions (e.g., the misclassified flow meter in Figure 5(b)) introduces variability in processing times. While the human-in-the-loop mechanism ensures reliability, it exposes the need for self-improving systems. Future research will integrate continual learning algorithms, enabling YOSO to iteratively refine its models using verified corrections, thereby reducing human input while maintaining accuracy.

By addressing these challenges, YOSO will evolve into a seamless, end-to-end solution for Scan-to-BIM automation. Its enhanced scalability and precision will directly support a transition to Modular MiC/MiMEP, ensuring rapid, clash-free upgrades essential for climate resilience.

6. Conclusion

This study introduces YOSO (You Only Scan Once), a novel Large-Small Model Co-Adapter framework that bridges edge-cloud AI with human expertise to revolutionise Scan-to-BIM automation for ageing wastewater infrastructure. By integrating real-time spatial capture via the lightweight MindPalace Pocket, adaptive AI segmentation and cloud-based clustering, YOSO addresses critical bottlenecks in traditional workflows, labour-intensive manual processing, fragmented toolsets and generic training data. The framework's confidence-driven interface reduces manual input by 80% while ensuring accuracy through human verification of low-certainty predictions. Validation at TKRSPS No. 8 demonstrates YOSO's efficacy: It achieves a 1-day semi-automated workflow, a stark improvement over traditional 1-week manual processes. It also achieves near-complete component capture (99.9% IMR for light tubes, 84.6% for gate valves) and precise spatial fidelity (mean ED of 0.166 m and 0.116 m, respectively), directly supporting MiC/MiMEP workflows for prefabrication and clash-free installation.

YOSO's success in balancing automation with accuracy demonstrates its potential to modernise ageing wastewater infrastructure globally. As climate risks escalate, YOSO provides a critical pathway for rapid, resilient upgrades, aligning with digital transformation goals and advancing sustainable asset management practices. By transforming Scan-to-BIM from a bottleneck into an enabler, YOSO sets a new benchmark for AI–human collaboration in infrastructure resilience.

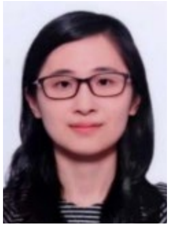
Notes on contributors

Mr Winson TC Leung is an Electrical and Mechanical Engineer at the Drainage Services Department, the Government of the Hong Kong Special Administrative Region (HKSAR) of the People's Republic of China. He is currently serving in the electrical and mechanical projects division of the department and pursuing the application of robotics and artificial intelligence in storm water and drainage services.

Mr CH Wong is an Electrical and Mechanical Engineer at the Drainage Services Department, the Government of the Hong Kong Special Administrative Region (HKSAR) of the People's Republic of China. He has over 15 years' experience in the electrical and mechanical industry and is currently serving in the BIM Support Team of the department.

Mr Senna CS Ng is a Senior Electrical and Mechanical Engineer at the Drainage Services Department, the Government of the Hong Kong Special Administrative Region (HKSAR) of the People's Republic of China. He is currently responsible for the cavern project for the relocation of the Shatin Sewage Treatment works.

Mr Simon YO Lai (MHKIE, CCBM, Eng BIM Pro) is a civil engineer who has over 30 years' experience in the construction industry including project planning, design & construction, operation & maintenance, information technology and BIM management. He is currently heading the BIM Support Team for the Drainage Services Department, the Government of the Hong Kong Special Administrative Region (HKSAR) of the People's Republic of China.



Ms Carry PC Cheung is an Electrical and Mechanical Engineer at the Drainage Services Department, the Government of the Hong Kong Special Administrative Region (HKSAR) of the People's Republic of China.

Mr CY Lam earned his bachelor's degree in Architecture from Tsinghua University in 2022. He worked in AEDAS for 2 years, concentrating on residential and commercial project coordination and design.

Dr YM Qin received a B.E. in Electrical Engineering, in 2018, from the Virginia Polytechnic Institute and State University in the United States. He received a Ph.D. degree in Mechanical Engineering from the University of Hong Kong, Mechanical Engineering. His research interests include mechatronics and robotics, VTOL and multirotor UAV system design, autonomous vehicles, and control.

Mr LF Ren received a M.S. in Software Engineering in 2025 from Carnegie Mellon University, and a B.S. in Machine Learning in 2023 from Virginia Polytechnic Institute and State University, both in the United States. His research interests include robotics and related machine learning algorithms.

References

[1] Jiang, F., Ma, L., Broyd, T. & Chen, K. (2021), 'Digital twin and its implementations in the civil engineering sector', *Automation in Construction* 130, 103838.

[2] Justo Dom´ınguez, A., Soil´an, M., S´anchez Rodr´ıguez, A. & Riveiro, B. (2021), 'Scan-to-bim for the infrastructure domain: Generation of ifc-compliant models of road infrastructure assets and semantics using 3d point cloud data', *Automation in Construction* 127, 103703.

[3] Kim, S., Yajima, Y., Park, J., Chen, J. & Cho, Y. K. (2022), '*A hybrid semantic-geometric approach for clutter-resistant floorplan generation from building point clouds*'. URL: https://doi.org/10.48550/arXiv.2305.15420

[4] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y. et al. (2023), Segment anything, in *'2021 IEEE/CVF International Conference on Computer Vision (ICCV)'*, IEEE.

[5] Lee, T., Chan, K., Chan, H. & Kok, M. (2012), 'Projections of extreme rainfall in hong kong in the 21st century', *Acta Meteorologica Sinica* 25.

[6] Maxime, Q., Stefan, B., Aymeric, H. & Laure, D. (2024), 'Scan-to-bim: Unlocking current limitations through artificial intelligence', in *'2024 Proceedings of the 41st ISARC, Lille, France, ISBN 978-0-6458322-1-1, ISSN 2413-5844'*, pp. 779-788. URL: https://www.iaarc.org/publications/fulltext/134_ISARC_2024_Paper_197.pdf

[7] Perez-Perez, Y., Golparvar-Fard, M. & El-Rayes, K. (2021), 'Scan2bim-net: Deep learning method for segmentation of point clouds for scan-to-bim', *Journal of Construction Engineering and Management* 147(9). Publisher Copy- right: © 2021 American Society of Civil Engineers.

[8] Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016), You only look once: Unified, real-time object detection, in *'2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)'*, pp. 779–788.

[9] Smith, J., Brown, K. et al. (2017), 'Impacts of extreme weather on infrastructure and society: A review of recent research', *Journal of the Air Waste Management Association* 67(4), 445–460. URL: https://www.tandfonline.com/doi/full/10.1080/10962247.2017.1401017

[10] Xiong, X., Adan, A., Akinci, B. & Huber, D. (2013), 'Automatic creation of semantically rich 3d building models from laser scanner data', *Automation in Construction 31*, 325–337. URL: https://www.sciencedirect.com/science/article/pii/S0926580512001732