Enhancing semantic segmentation of construction scenes using IFC-derived synthetic point clouds for geometric Digital Twins

Inshu Chauhan[1], Druhin Abrol[1,2], Olli Seppänen[3]

[1] School of Engineering, Aalto University, Espoo, Finland – inshu.chauhan@aalto.fi
[2] Nava Technologies Pvt. Ltd., Bangalore, India – druhin.abrol@gmail.com
[3] School of Engineering, Aalto University, Espoo, Finland – olli.sepannen@aalto.fi

Keywords: Construction, Digital Twin, Virtual Laser Scanning, Semantic Segmentation, BIM, point clouds.

Abstract

Accurate semantic segmentation of as-built point clouds is foundational for Digital Twin Construction, yet training deep learning models is constrained by data scarcity in the AEC sector. This study presents a method to create semantically labeled synthetic point clouds from IFC models, and employs geometric augmentation to bridge the domain shift to real-world LiDAR data. We developed an automated pipeline to generate semantically labeled synthetic point clouds directly from IFC models, and conducted a systematic hyperparameter search, testing the impact of point-level Gaussian noise (Jitter Standard Deviation) and maximum displacement (Clipping Range) on the Oneformer3D model's ability to segment complex Mechanical, Electrical, and Plumbing (MEP) components. Our results demonstrate that this augmentation approach significantly enhances segmentation performance compared to baseline model, particularly in real-world environments. The optimal configuration achieved a synthetic mIoU of 67.31% (up from 62.52%) and a real-world mIoU of 32.75% (improving significantly upon the 14.66% baseline). These results confirm that balancing high noise variance (for domain generalization) with strict displacement control (to retain feature integrity) is key to successful synthetic-to-real domain transfer. This approach offers a scalable alternative for generating the necessary training data for automated geometric Digital Twin realization.

1. Introduction

Digital Twin Construction has emerged as an important concept for construction management that combines digital models with physical representations of the real world. A digital twin construction has been defined as data-driven planning and control workflow for the design and construction of buildings and civil infrastructure that is founded on digital twin information systems (Sacks et al., 2020). The digital twin can include product information or process information. The physical product, in the case of construction, is a building under construction. The design intent is captured in the building model but all the elements of the design are not present at any given time. The real progress of construction thus can be sensed and represented digitally, for example, in a 3D model(Boje et al., 2020). An essential part of the twinning process is re-creating the geometry of the as-built status. Often, the as-built elements differ from the as-designed elements both in terms of progress and geometry, and these changes must be reflected in geometric Digital Twins (gDT).

Light Detection and Ranging (LiDAR) scanners have become an effective tool to obtain precise geometric representations of as-built infrastructure (Ramiya et al., 2014), which are the basis for gDTs. Semantically identifying the information from the point clouds generated by scanners is vital for downstream applications, such as progress monitoring, clash detection, and quality assurance (Sgrenzaroli et al., 2022). Significant research has been published on the processing of point clouds to monitor progress at construction sites (Turkan et al., 2012; Kavaliauskas et al., 2022). Most of these methods have relied on comparing as-built with as-designed elements to provide a situational picture of in-progress construction (Braun et al., 2020; Kim et al., 2020). But because this approach does not require semantic understanding of points, such methods fall short of creating an accurate gDT. Semantic understanding is also required for progress detections if any deviations occur between as-built and as-designed elements.

Deep learning models trained on construction site data offer an effective way to automatically extract meaningful information. Researchers have attempted to automatically process point cloud data for semantic segmentation using deep learning (Hu et al., 2023). Many studies have concentrated on Scan to BIM methods to generate Building Information Models (BIM) for existing buildings that lack them (Barazzetti, 2016; Wang et al., 2015). Most of these studies have labelled the point clouds manually, which is easier for simpler objects such as walls, columns, beams, and slabs. Some studies have experimented with more classes such as radiators, pipes, and valves, but their focus has been on point cloud quality estimation rather than addressing the automatic processing of data (Rebolj et al., 2017). However, complete automation is yet to be accomplished, which is a necessary step to fully realise gDT(Tao et al., 2024). To achieve greater automation, point clouds must be accurately annotated to establish a reliable ground truth. Manual labelling of point clouds is not only time-consuming, error-prone and labour-intensive but also requires expert knowledge for specialised scenarios such as construction sites (Xu et al., 2022). Such 3D labelled datasets are not only lacking in the architecture, engineering, and construction (AEC) sector but have become a bottleneck for stakeholders to adopt state-of-the-art (SOTA) remote sensing technologies like LiDAR (Wu et al., 2022).

Creating labelled synthetic point clouds from BIM model offer a plausible solution to reduce this scarcity of data in AEC. Few attempts have been made to convert CAD or BIM models to synthetic point cloud (Ma et al., 2021). Noichl et al. (2024) used industrial CAD models and HELIOS pipeline to get synthetic point clouds showing promising results for industrial semantic segmentation. Most interior building point cloud labeling methods rely on the S3DIS dataset, which includes CAD models and labelled point clouds (Czerniawski et al., 2019; De Geyter et al., 2022). These synthetic point clouds have fixed classes that are available in the S3DIS dataset for buildings. Most approaches also depend on manually transferring labels from BIM to a generated point cloud, a process that can be both error-prone and time-consuming(Zhou et al., 2025). This work aims to mitigate this research gap by developing a method to get labelled point clouds from IFC in Unity without the need to transfer labels manually. We also investigate the best augmentation-based parameters to help generalise the synthetic dataset to real-world scenarios. We place particular emphasis on Mechanical, Electrical, and Plumbing (MEP) systems within open ceiling structures, as these are challenging to segment due to their intricate installation processes and similar geometric configurations.

This study is guided by the following hypotheses: Synthetic data derived from IFC models can provide a sufficient training foundation for segmenting real-world point clouds, provided that the training data is augmented with domain-specific geometric augmentations.

2. Related work

This section reviews studies that have used BIM/CAD models to label point clouds or to create synthetic labelled point clouds. We also explore segmentation methods from earlier research applied to building point clouds.

2.1 BIM-to-Scan

Numerous studies have employed BIM in conjunction with point clouds acquired on site to generate labels for these point clouds (Ma et al., 2020). The transfer of labels from BIM to point cloud relies on precise registration between the two datasets(Czerniawski et al., 2019) These analyses have focused on structural elements such as floors, ceilings, walls, beams, and columns, and have excluded more complex interior or MEP components. The method has limitations if the registration is inaccurate, which may result from differences between as-designed and as-built geometries. This situation is often seen with MEP elements. The registration of BIM and point cloud also depends upon setting certain parameters such as cloud-to-cloud (c2c) distance, which limits the method's generalisability for different buildings (De Geyter et al., 2022)

Another approach is to generate synthetic point clouds based on CAD/BIM models. Some studies have explored placing virtual scanners inside a BIM model, which converts BIM geometry into a point cloud (Ma et al., 2021). The labels are then transferred to the generated point clouds by aligning the CAD/BIM model to the synthetic point cloud. Although the point clouds are generated from the BIM/CAD model itself, synthetic point clouds can include occlusions not present in the CAD model due to overlapping geometries while scanning (Humblot-Renaux et al., 2023). As a result, these occlusions during label transfer can lead to mislabelling and alignment errors. The approach is sensitive to misalignment, where a slight error can lead to numerous mislabels in these point clouds.

Other works have produced labelled point clouds for different applications, for example infrastructure or industrial applications. Korus et al. (2023) developed a dynamo-based workflow for a bridge from its BIM model. The scanner in this case was placed facing the infrastructure, an approach that is quite different from working within the interior of a building where the scanner needs to be inside the scene to generate the synthetic point cloud. Some studies have developed synthetic point clouds for industrial facilities from CAD models (Noichl et al., 2024; Birkeland and Udnæs, 2024). Noichl et al. (2024) applied HELIOS for industrial facilities but the point clouds were formed separately for each class. Having a separate point cloud for each object in an industrial facility limits the method's generalisability for different applications. For example, this labelled data may not contain occlusions which are typical on real industrial or construction sites. Birkeland and Udnæs (2024) relied on CAD for point cloud alignment through visual inspection, with only 45 % correctly labelled points. These factors affect downstream deep learning model training, as the model' s accuracy depends on the accuracy of the labelled points and the structure of the point cloud.

Recently, Zhou et al. (2025) conducted experiments with both coloured and unicoloured point clouds, offering detailed insights into how parameters such as colour affect the semantic segmentation of interior building point clouds. The authors used BlenSor to generate synthetic point clouds from CAD models available in the S3DIS dataset. As they noted, their approach does not yet offer a scalable, automated method for generating labelled point clouds for training. Their pipeline relies on manual alignment of point clouds to transfer labels, which limits scalability and can lead to failures when the BIM model significantly differs from the as-built structure.

2.2 Semantic segmentation of building-related point clouds

Deep learning architecture has significantly enhanced the capability to process large-scale point clouds, enabling the automated semantic segmentation of complex built environments with increasing precision. To address the geometric intricacies of indoor scenes, Hu et al., (2023) utilised ResPointNet++, an architecture that integrates residual learning with PointNet++, to extract multi-scale local features. The approach allowed for the effective classification of structural elements such as walls and floors, achieving high segmentation accuracy even in cluttered environments. Similarly, the versatility of lightweight architectures was demonstrated by (Kim et al., 2022), who employed RandLA-Net to segment temporary scaffold structures. Liu et al. (2022) developed a Context-Aware Network (CAN) designed to capture long-range dependencies in urban environments through a global context aggregation module. While this architecture improved the understanding of large-scale structures, the authors observed that the model still struggled to correctly classify minority classes, such as poles or hardscapes, due to the inherent class imbalance in the training datasets. Perez-Perez et al. (2021) noted that their joint learning framework, which coupled Support Vector Machines (SVM) for semantic labelling with AdaBoost for geometric classification, frequently misclassified cylindrical columns as pipes. Therefore, significant challenges remain regarding class imbalance and feature ambiguity in geometrically similar classes.

B. Wang et al. (2022) semantically segmented RGB images into MEP classes and transferred labels to point clouds to build a parametric BIM, yet their reliance on depth cameras made the technique's performance sensitive to indoor lighting conditions. A PointNet-based model was trained on manually extracted, labelled pipe points, achieving high pipe classification (Xu et al., 2022). The method lacked building context and left performance on mixed cylindrical MEP elements (e.g., sewers, water pipes, and ducts) and site conditions untested. Jing et al. (2024) similarly manually retained and labelled points related to MEP components instead of using full building point clouds. They expanded MEP categories and developed a well-performing model but again lost contextual information critical for real-world applications. Most recently, Lee et al. (2025) evaluated the efficacy of RandLA-Net in segmenting intricate MEP components in open ceiling designs. While the model achieved favourable overall metrics, the authors observed a distinct performance disparity: Intersection over Union (IoU) scores for complex categories such as ducts and pipes were notably lower than those for planar structural components. This finding suggests that limitations still exist, with the overlapping geometries and visual similarities inherent in MEP systems.

3. Method

The research is done in two steps; First, we generate labelled point clouds from an IFC-based BIM and then uses deep learning to evaluate their performance on a real construction site. The respective trades generally design the IFC files of a building; for example, the architectural file is designed by architects, and the MEP file is designed by the MEP designer. The project that was chosen for this work was an under-construction mall consisting of 7 floors. This project had 3 main IFC files ARK for the structural and other building elements, LVI for pipes, ducts and other MEP parts and SPR which is the sprinkler file, part of fire prevention system. All these systems are important for the functionality of a building and need to be monitored closely during the construction phase. For our synthetic point cloud generation these files were combined together with element properties preserved, floor-wise. To get correct labels from IFC, it' s necessary to understand the schema of an IFC file. The IFC file consists of child-parent relationships under various IFC schemas, for instance, IFCflowsegment contains all the pipes, ducts, sewerage pipes, etc. These names were either saved in name or in description or part-type. So while importing IFC file in Unity, this information needs to be preserved to enable labelling of such classes in point clouds.

For a unity-based simulation, the property extraction process utilises the BlenderBIM add-on, an open-source utility that enhances Blender' s native IFC capabilities by providing comprehensive access to IFC schema properties and relationships. BlenderBIM facilitates direct manipulation of IFC entities while preserving the semantic integrity of the original building model data structure. A custom script was developed to work in conjunction with BlenderBIM, systematically traversing the imported IFC model hierarchy to extract the semantic properties. The script captures component classifications, global unique identifiers (GUID), IFC specific property sets and other metadata and saved them into JSON files.

The structured JSON files that were generated in the previous step maintained the hierarchical organization of the original IFC model while providing a format compatible for subsequent Unity integration. Each JSON entry corresponds to an individual IFC entity and includes complete property inheritance from parent classes within the IFC schema. The geometric conversion and Unity integration process employs IFCConvert, a specialized utility developed for translating IFC geometry into game engine-compatible formats. IFCConvert processes the original IFC file independently of the Blender-based property extraction workflow, ensuring that geometric accuracy is maintained while generating Unity-compatible mesh assets. This process imports the geometry with GUIDs while discarding the IFC properties related to these geometries.

A dedicated Unity utility was developed to bridge the gap between the extracted IFC properties (JSON format) and the converted geometric assets. This utility implements an automated property assignment system that parses the JSON file that correlates IFC entity GUIDs with their corresponding Unity GameObjects, ensuring that semantic information is properly linked to the geometric representations. There are 4 such JSON files (electrical, plumbing, walls, ducts). Utilizing Unity's Physics systems, a LiDAR scanner simulation was developed using Raycasts. Table 1 describes the exact scan parameters considered by the system.

To replicate the behavior of a real-world LiDAR scanner, a simulation environment was developed in Unity using its physics-based raycasting system. The objective was to model the scanning process such that the simulated point cloud data closely resembles actual site conditions while maintaining control over the scan parameters. The virtual LiDAR scanner was configured based on the specifications of the instrument deployed at the construction site, with a maximum (when all rays hit an object) capacity of capturing approximately 4 million points per scan. Each emitted ray was generated as a Physics.Raycast along a defined angular step in both horizontal and vertical directions. The intersection of the ray with scene geometry was recorded to obtain the spatial coordinates of the corresponding point.

The scanning parameters were determined by setting the **horizontal resolution**, **vertical resolution**, and the vertical **field of view (FOV)** of the scanner. The angular step size was computed by dividing the scanning range, 360 degrees by the horizontal resolution (2000). The vertical step was computed by dividing the FOV with the vertical resolution as shown in Figure 1. Even though this will not give a square grid of points on a radial plane, it still gives similarly spaced grid points on the ceiling (where most of the MEP is located) that work for the scanner's range and position. The coordinates of each point scanned were directly calculated and returned by the Unity's raycast system, and the properties and GUID were stored. To increase realism, the virtual scanner advanced incrementally along a predefined set of control or scan points. At each control point along the trajectory, a full scan was performed before proceeding to the next position, thereby producing sequential frames of point cloud data. These were combined later to form full point clouds for each floor as training data was divided on the basis of each floor. The floor distribution is shown in Figure 2.

4. Experiments and results

4.1 Experimental Setup

To evaluate the effectiveness of neural networks for segmenting small-scale Mechanical, Electrical, and Plumbing (MEP) elements, we selected the Oneformer3D architecture (Kolodiazhnyi et al., 2024) within the MMDetection3D framework. Our experimental design employs a two-pronged evaluation protocol: (1) Synthetic validation, where the model is trained on a synthetic dataset derived from floors 1, 2, 3, and 7, and subsequently validated on Floor 5 to test floor-wise generalization; and (2) Real-world performance assessment, where the model, trained on the synthetic dataset, is deployed on a separate, independent real-world LiDAR dataset captured from Floor 2 to quantify the domain transfer efficacy. To facilitate integration with the framework, the generated synthetic dataset was formatted to match the standard S3DIS dataset structure.

https://doi.org/10.5194/isprs-annals-XII-4-W1-2026-73-2026

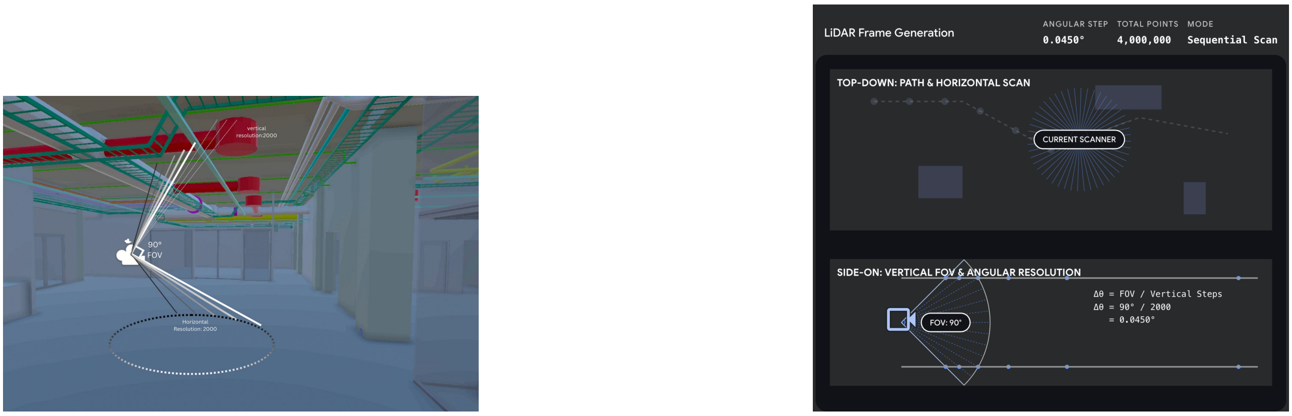

Figure 1. Visualisation of the scanner and scanning process.

Property name	Description	Value
Horizontal resolution	Number of data points (or rays) sampled across the horizontal axis per scan	2000
Vertical resolution	Number of data points (or rays) sampled across the vertical axis per scan	2000
FOV (field of view)	Angular field of view (vertically) of the LiDAR sensor	90°

Table 1. Virtual scanner parameters.

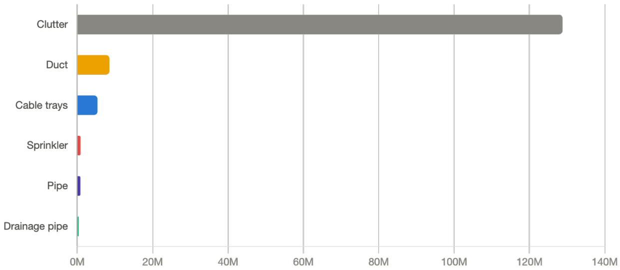

Figure 2. Class distribution of the synthetic dataset. Clutter dominates all scenes.

We evaluate model performance using mean Intersection over Union (mIoU) across six specific classes: cable trays, drainage pipes, ducts, sprinklers, pipes, and clutter. The “clutter” class encompasses non-MEP structural elements, such as walls, floors, and ceilings, which provide context to the MEP installations. Throughout the results, we present the top 5 performing configurations for each domain to illustrate the impact of hyperparameter optimization.

The ’6-class mIoU’ refers to the mean Intersection over Union calculated across the six target MEP and clutter classes. The ’ Clutter’ class specifically includes structural context such as walls, floors, and ceilings.

4.2 Data Augmentation

Two primary data augmentations were applied to bridge the gap between synthetic data and noisy real-world scans:

Gaussian Noise (Jitter): Applied to individual points to simulate sensor noise, controlled by the standard deviation of displacement (jitter_std) and a maximum displacement threshold (clip_range). Jitter represents sensor noise present in real data but absent from clean synthetic data

Random Rotation: Applied to the entire point cloud to account for the fact that real LiDAR scans are rarely perfectly axis-aligned, unlike the raw synthetic data exported from BIM.

These augmentations are also needed to account for class imbalance favoring classes smaller in numbers. Apart from these augmentations, some preprocessing was also done on the raw point clouds.

Color Removal: All color and texture information was stripped from the point clouds, restricting the input features strictly to spatial coordinates (X, Y, Z) as synthetic data derived from BIM does not have true color information.

Spatial Cropping: Because the majority of MEP elements are suspended from or affixed to the ceiling, the lower 80% of the volumetric scene was cropped. This drastically reduced the voxel volume, allowing for a finer voxel size during training without exceeding memory limits. It also affected the inherent class-imbalance in such scenes.

4.3 Hyperparameter tuning

To optimize the model’ s ability to generalize to the inherent noise of real-world LiDAR scans, a systematic grid search was conducted focusing on data augmentation parameters. While the learning rate (0.0003) and voxel size (0.03) were kept constant, we systematically varied the parameters controlling point-level Gaussian noise (jitter).The unit for voxel size, jitter and clipping range is in m.

The two primary hyperparameters investigated were:

Jitter Standard Deviation (jitter_std): Tested at values of 0.001, 0.003, 0.005, and 0.008. This controls the spread of the random noise applied to the spatial coordinates of each point.

Clipping Range (clip_range): Tested at +/-0.001, +/-0.0025, and +/-0.005. This enforces a maximum spatial displacement limit to prevent extreme outliers from excessively warping the structural geometry.

4.4 Results

4.4.1 Synthetic Validation. The results for the top 5 performing configurations from the grid search on synthetic data validation are presented in Table 2 and Table 3. Results show that while clutter remains stable across configurations, complex classes like sprinklers and pipes benefit significantly from the jitter-rotation augmentation combination, indicating the model’

Synthetic rank	Augmentation	Jitter std.	Clip range	Overall acc.	Mean class acc.	6-class mIoU
Baseline	None	None	None	96.02	72.72	62.52
1	Jitter + rotation	0.003	+/-0.0025	95.93	76.97	67.31
2	Jitter + rotation	0.003	+/-0.005	95.63	77.22	67.16
3	Jitter + rotation	0.005	+/-0.005	95.93	78.44	67.14
4	Jitter + rotation	0.005	+/-0.0025	95.54	75.96	66.18

Table 2. Synthetic-validation summary metrics.

Class	Baseline	0.003/+/-0.0025	0.003/+/-0.005	0.005/+/-0.005	0.005/+/-0.0025	0.008/+/-0.001
Cable trays	81.50	75.46	76.92	74.66	77.42	74.05
Drainage pipe	17.48	21.03	22.57	24.11	33.90	24.15
Duct	83.48	77.40	75.70	79.30	77.86	77.37
Sprinkler	64.35	83.36	83.07	76.43	71.29	76.07
Pipe	31.12	49.28	48.09	50.95	40.24	40.62
Clutter	97.16	97.35	96.58	97.38	96.39	96.46
Overall mIoU (%)	62.52	67.31	67.16	67.14	66.18	64.79

Table 3. Synthetic-validation class-wise IoU (%).

s increased sensitivity to geometry-dependent features. The best configuration (jitter 0.003, clipping range +/-0.0025) improved overall mIoU from 62.52% to 67.31%.

4.4.2 Real Data Validation. The model selected independently on synthetic validation achieved 28.40% six-class mIoU and 24.51% MEP mIoU. Notably, the highest real-world performance was produced by a different configuration (jitter 0.005, clip +/-0.0025), reaching 32.75% six-class mIoU. The differing rankings between synthetic and real domains confirm that while geometric augmentation improves transfer, a domain gap persists that jitter and rotation alone cannot fully resolve. Table 4 summarizes the top 5 performing configurations on the real LiDAR dataset. Figures 3, 4, and 5 show the comparison of model performance when trained on synthetic data without augmentation and with augmentations.

4.4.3 Ablation Study. The results in Table 6 demonstrate that point-level jitter is the primary driver of transferability. While rotation helps with global alignment, it provides negligible benefits to MEP class identification compared to the noise-robustness afforded by jitter.

As can be seen from the figures 3 and 5, a model augmented with Jitter (noise) and rotation is able to perform better on real data as compared to a model with no augmentation at all. Some random wires on the construction site are labelled as Sprinklers as they are slender and close to the ceiling. Besides this confusion, the model is able to differentiate between all MEP elements such as Ducts, Cable trays and pipes very well (Table 3).

5. Discussion

This study presented an automated pipeline for generating labeled point clouds from IFC data, demonstrating its potential for segmenting complex MEP systems to support geometric Digital Twin construction. By training on synthetic data and testing on real-world LiDAR scans, we observed a performance gap, 67.31% synthetic versus 32.75% real-world mIoU, which indicates that while synthetic training is a scalable foundation, it does not fully replicate the geometric complexity of real-world construction sites.

The ablation study identified that jitter-based noise augmentation is the primary driver of successful domain transfer. This mechanism forces the model to learn features capable of handling sensor-related inaccuracies, whereas rotation provides only marginal benefits for global alignment. These results should be interpreted within the context of the dataset structure: although the target data contained 11 classes, this study focused on 6 specific MEP components. This selection influences the performance metrics, as a broader class set could introduce different patterns of geometric ambiguity.

Ultimately, by automating the segmentation of MEP components, this workflow addresses a critical bottleneck in the creation of geometric Digital Twins. By converting raw point cloud data into semantic information, this pipeline bridges the gap between LiDAR acquisition and as-built modeling. This semantic segmentation is a necessary prerequisite for downstream tasks, such as clash detection and progress monitoring, where accurate identification of individual MEP elements is required to maintain reliable as-built geometric representations.

Limitations and Future Work

While this research provides an automated pipeline for generating labeled training data, it is subject to several limitations:

Validation Scope: The study is validated on a single construction project, which may not capture the full variance in building types, MEP configurations, and architectural styles.

Occlusion: Certain building components or system elements cannot be adequately captured during real-world scanning due to inherent line-of-sight constraints. For example, vertical ducts located behind shaft walls remain difficult to identify, as they are inherently occluded or only partially visible in LiDAR point clouds. This limits the workflow's ability to achieve comprehensive geometric completeness independently. Within the proposed framework, such cases necessitate either manual intervention, specific capture strategies prior to the installation of concealing elements, or the integration of supplementary "as-designed" BIM information to ensure full geometric representation in the final Digital Twin.

Domain Generalization and Class Scope: Despite employing geometric augmentation techniques such as jitter and rotation, a substantial domain gap persists between synthetic and real-world performance metrics. This indicates that current augmentation methods are insufficient for fully bridging the shift inherent in synthetic-to-real transfer. Furthermore, as this evaluation focused on a subset of 6 MEP classes from the original 11, these results may vary when generalized to a broader class set. Future research should investigate advanced domain generalisation strategies to further mitigate these discrepancies and ensure robustness across a wider range of construction objects.

As-Designed vs. As-Built: The current pipeline relies exclusively on as-designed BIM data to generate synthetic training sets and facilitate segmentation. The scope of this research is limited to enhancing semantic segmentation to derive meaningful geometric information for construction Digital Twins, and thus the framework is not designed to explicitly differentiate between legitimate field design changes (e.g., pipe rerouting) and

Rank	Augmentation	Jitter std.	Clip range	MEP mIoU	6-class mIoU	Overall acc.	Mean class acc.
Baseline	None	None	None	10.24	14.66	40.66	31.68
1	Jitter + rotation	0.005	+/-0.0025	30.43	32.75	57.30	53.84
2	Jitter + rotation	0.008	+/-0.0025	25.49	28.75	56.41	48.45
3	Jitter + rotation	0.005	+/-0.005	24.85	28.46	57.97	47.93
4	Jitter + rotation	0.003	+/-0.0025	24.51	28.40	58.15	49.18
5	Jitter + rotation	0.001	+/-0.0025	25.16	28.34	54.90	48.61

Table 4. Real-data summary metrics (manually labelled Floor 2).

Class	Baseline	0.005/+/-0.0025	0.008/+/-0.0025	0.005/+/-0.005	0.003/+/-0.0025	0.001/+/-0.0025
Cable trays	9.99	36.82	25.69	19.52	31.69	29.90
Drainage pipe	1.82	23.07	13.17	2.62	9.74	13.66
Duct	25.17	41.05	44.83	52.06	45.32	40.77
Sprinkler	11.76	37.07	33.62	37.67	20.37	33.37
Pipe	2.46	14.14	10.13	12.40	15.44	8.10
Clutter	36.75	44.34	45.05	46.47	47.85	44.25

Table 5. Real-data class-wise IoU (%).

Domain	Augmentation	MEP mIoU	6-class mIoU	Overall acc.	Mean class acc.
Synthetic validation	None	55.59	62.52	96.02	72.72
Synthetic validation	Jitter only	55.83	62.76	96.06	71.70
Synthetic validation	Rotation only	56.43	63.09	94.84	75.86
Real Floor 2	None	10.24	14.66	40.66	31.68
Real Floor 2	Jitter only	17.39	20.73	47.72	35.02
Real Floor 2	Rotation only	11.48	15.43	42.67	29.05

Table 6. Augmentation ablation on synthetic and real data.

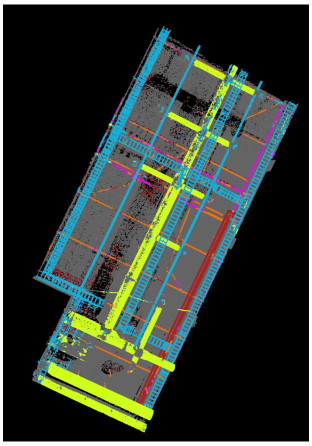

Figure 3. Qualitative prediction on real data with jitter and rotation. Duct = yellow, pipes = red, sprinkler pipe = orange, cable trays = blue, and drainage pipe = magenta.

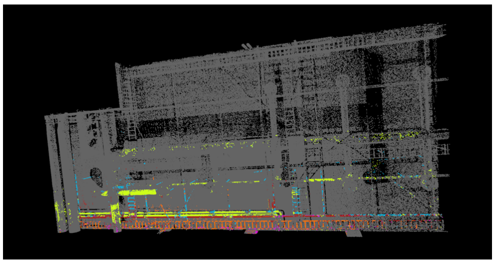

Figure 4. Same scene as in Figure 3, predicted by a model trained with no jitter and no rotation.

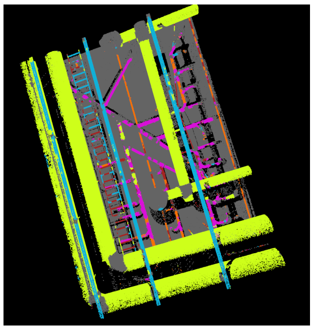

Figure 5. Another real point cloud scene segmented using the augmented model trained on synthetic data.

unintended construction deviations. Consequently, the framework lacks the capability to contextualize these discrepancies without access to an updated "as-built" BIM reference. Future research should investigate strategies for reconciling as-built point clouds with BIM models, enabling the system to contextualize these differences.

Comparative Benchmarking: While our results indicate improvement over a baseline, future work will involve comparative analysis against other established VLS strategies (e.g., Helios) to determine relative performance benchmarks across different industrial environments.

6. Conclusion

This work presented an automated framework for generating semantically labeled point clouds from IFC models and evaluated its effectiveness for segmenting complex Mechanical, Electrical, and Plumbing (MEP) systems in construction Digital Twins. Our method successfully mitigates the data scarcity bottleneck in the AEC sector. A critical aspect of this work was bridging the domain gap between synthetic data and noisy real-world LiDAR scans through systematic hyperparameter optimization of Gaussian Noise (Jitter) and Clipping Range. We demonstrated that the model' s performance is distinctly sensitive to these augmentation parameters, which simulate real-world sensor noise. The optimal segmentation performance, achieving a synthetic mIoU of 67.31% and a real-world mIoU of 32.75%, was achieved by applying a Jitter Standard Deviation of 0.005 coupled with a Clipping Range of +/-0.0025. This finding underscores a crucial principle: while the model requires exposure to significant point perturbation to generalize to sensor noise, restricting the maximum displacement is necessary to preserve the fine geometric features essential for accurate identification of small, intricate MEP components. Overall, the use of jitter-augmented synthetic data significantly enhanced the OneFormer3D model' s ability to differentiate between MEP elements such as ducts, cable trays, and pipes on real-world scans, confirming that this approach is a viable pathway for automated as-built geometric Digital Twin creation. Future work should focus on including more classes such as floor, ceiling, walls, columns to the generated dataset to understand their effect on segmentation accuracy and ambiguity in complex construction scenes.

Acknowledgements

The work was supported by the 'Digital Twins of processes for construction of buildings' project funded by Research Council of Finland, grant number 357307. We acknowledge the computational resources provided by the Aalto Science-IT project, which maintains and hosts the high-performance computing cluster with multiple GPUs. We acknowledge the help provided by research assistants Faraz Ahmed and Alexander Urbina for manually labelling the real construction-site point clouds

References

Barazzetti, L., 2016: Parametric as-built model generation of complex shapes from point clouds. *Advanced Engineering Informatics* 30, 298–311. doi.org/10.1016/j.aei.2016.03.005

Birkeland, A.A., Udnæs, M., 2024: Semi-automated dataset creation for semantic and instance segmentation of industrial point clouds. *Comput Ind* 155. doi.org/10.1016/j.compind.2023.104064

Boje, C., Guerriero, A., Kubicki, S., Rezgui, Y., 2020: Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction* 114, 103179. doi.org/10.1016/J.AUTCON.2020.103179

Braun, A., Tuttas, S., Borrmann, A., Stilla, U., 2020: Improving progress monitoring by fusing point clouds, semantic data and computer vision. *Automation in Construction* 116, 103210. doi.org/10.1016/j.autcon.2020.103210

Czerniawski, T. and Leite, F., 2019: Semantic segmentation of building point clouds using deep learning: a method for creating training data using BIM to point cloud label transfer. In *ASCE International Conference on Computing in Civil Engineering,* 410–416, Reston, VA: American Society of Civil Engineers.

De Geyter, S., Bassier, M., Vergauwen, M., 2022: Automated training data creation for semantic segmentation of 3d point clouds, in: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. International Society for Photogrammetry and Remote Sensing, pp. 59–67. doi.org/10.5194/isprs-archives-XLVI-5-W1-2022-59-2022

Hu, D., Gan, V.J.L., Yin, C., 2023: Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors. *Automation in Construction,152, 109949*.doi.org/10.1016/j.autcon.2023.104949

Humblot-Renaux, G., Jensen, S.B., Møgelmose, A., 2023: From CAD Models to Soft Point Cloud Labels: An Automatic Annotation Pipeline for Cheaply Supervised 3D Semantic Segmentation. *Remote Sensing (Basel)* 15. doi.org/10.3390/rs15143578

Jing, S., Cha, G., Maru, M.B., Yu, B., Park, S., 2024: Improved building MEP systems semantic segmentation in point clouds using a novel multi-class dataset and local–global vector transformer network. *Journal of Building Engineering* 96. doi.org/10.1016/j.jobe.2024.110311

Kavaliauskas, P., Fernandez, J.B., McGuinness, K., Jurelionis, A., 2022: Automation of Construction Progress Monitoring by Integrating 3D Point Cloud Data with an IFC-Based BIM Model. *Buildings* 12. doi.org/10.3390/buildings12101754

Kim, J., Chung, D., Kim, Y., Kim, H., 2022: Deep learning-based 3D reconstruction of scaffolds using a robot dog. *Automation in Construction, 134, 104092*. doi.org/10.1016/j.autcon.2021.104092

Kim, S., Kim, S., Lee, D.E., 2020: 3D Point cloud and BIM-based reconstruction for evaluation of project by as-planned and as-built. *Remote Sensing (Basel)*. doi.org/10.3390/RS12091457

Korus, K., Czerniawski, T., Salamak, M., 2023: Visual programming simulator for producing realistic labeled point clouds from digital infrastructure models. *Automation in Construction,*156,105126.doi.org/10.1016/j.autcon.2023.105126

Kolodiazhnyi, O., and others, 2024: OneFormer3D: One Transformer for All Tasks in 3D Semantic Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11623–11633. doi.org/10.1109/CVPR52733.2024.01099

Lee, J., Xiao, J., Koo, D., Kang, K., 2015: Deep Learning-based 3D Point Cloud Semantic Segmentation for Mechanical, Electrical, and Plumbing Systems in Open Ceiling Building Structure. In *CIB Conferences*, Vol. 1, No. 1, p. 419.

Liu, C., Zeng, D., Akbar, A., Wu, H., Jia, S., Xu, Z., Yue, H., 2022: Context-Aware Network for Semantic Segmentation Toward Large-Scale Point Clouds in Urban Environments. *IEEE Transactions on Geoscience and Remote Sensing* 60, 5703915. doi.org/10.1109/TGRS.2022.3182776

Ma, J.W., Czerniawski, T., Leite, F., 2020: Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds. *Automation in Construction,* 113, 103144. doi.org/10.1016/j.autcon.2020.103144

Ma, J.W., Han, B., Leite, F., 2021: An Automated Framework for Generating Synthetic Point Clouds from as-Built BIM with Semantic Annotation for Scan-to-BIM, in: *Proceedings - Winter Simulation Conference*. Institute of Electrical and Electronics Engineers Inc. doi.org/10.1109/WSC52266.2021.9715301

Noichl, F., Collins, F.C., Braun, A., Borrmann, A., 2024: Enhancing point cloud semantic segmentation in the data-scarce domain of industrial plants through synthetic data. *Computer-Aided Civil and Infrastructure Engineering* 39, 1530–1549. doi.org/10.1111/mice.13153

Perez-Perez, Y., Golparvar-Fard, M., El-Rayes, K., 2021: Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment. *Automation in Construction,* 125, 103584. doi.org/10.1016/J.AUTCON.2021.103584

Ramiya, A.M., Nidamanuri, R.R., Krishnan, R., 2014: Semantic labelling of Urban point cloud data, in: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. International Bureau of Photogrammetry and Remote Sensing, pp. 907–911. doi.org/10.5194/isprsarchives-XL-8-907-2014

Rebolj, D., Pučko, Z., Babič, N.Č., Bizjak, M., Mongus, D., 2017: Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring. *Automation in Construction,* 84, 323-324 . doi.org/10.1016/j.autcon.2017.09.021

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S., Girolami, M., 2020: Construction with digital twin information systems. *Data-Centric Engineering* 1. doi.org/10.1017/dce.2020.16

Sgrenzaroli, M., Ortiz Barrientos, J., Vassena, G., Sanchez, A., Ciribini, A., Mastrolembo Ventura, S., Comai, S., 2022: Indoor mobile mapping systems and (BIM) digital models for construction progress monitoring, in: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. International Society for Photogrammetry and Remote Sensing, pp. 121–127. doi.org/10.5194/isprs-archives-XLIII-B1-2022-121-2022

Tao, B., Bosché, F., Li, J., 2024: Mixed Reality-based MEP construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison. *Automation in Construction,*168, 105852. doi.org/10.1016/j.autcon.2024.105852

Turkan, Y., Bosche, F., Haas, C.T., Haas, R., 2012: Automated progress tracking using 4D schedule and 3D sensing technologies. in: *Automation in Construction,* 22, 414-421. doi.org/10.1016/j.autcon.2011.10.003

Wang, B., Wang, Q., Cheng, J.C.P., Song, C., Yin, C., 2022: Vision-assisted BIM reconstruction from 3D LiDAR point clouds for MEP scenes. *Automation in Construction* 133, 103997. doi.org/10.1016/j.autcon.2021.103997

Wang, C., Cho, Y.K., Kim, C., 2015: Automatic BIM component extraction from point clouds of existing buildings for sustainability applications. *Automation in Construction* 56, 1–13. doi.org/10.1016/j.autcon.2015.04.001

Wu, L.T., Lin, J.R., Leng, S., Li, J.L., Hu, Z.Z., 2022: Rule-based information extraction for mechanical-electrical-plumbing-specific semantic web. *Automation in Construction*, 135, 104108. doi.org/10.1016/j.autcon.2021.104108

Xu, Z., Kang, R., Li, H., 2022: Feature-Based Deep Learning Classification for Pipeline Component Extraction from 3D Point Clouds. *Buildings* 12. doi.org/10.3390/buildings12070968

Zhou, S., Lin, J.R., Pan, P., Pan, Y., Brilakis, I., 2025: Impact of color and mixing proportion of synthetic point clouds on semantic segmentation. *Automation in Construction,* 171, 105963. doi.org/10.1016/j.autcon.2025.105963