Semantic Segmentation of building point clouds based on Point Transformer and IFC

Wei S.[1], Gao G.[1,2,*], Ke Z.[1], Fan G.[1], Liu Y. [1], Gu M. [1,2]
[1]School of Software, Tsinghua University, Beijing, China, [2]Beijing National Research Center for Information Science and Technology(BNRist), Tsinghua University, Beijing, China
gaoge@tsinghua.edu.cn

Abstract. In the process of building construction, the semantic understanding of building point clouds provides potential solutions for efficient building quality supervision, progress monitoring, and sub-system deviation analysis. However, the lack of suitable public labeled datasets, the low degree of color discrimination and the particularity of multi-system coexistence in construction scenes have brought certain challenges to semantic segmentation of point clouds. But rich semantic information in related Industry Foundation Classes(IFC) can be used to synthesize effective labeled data. Our main contributions are as follows:1) We propose a synthetic conversion method from BIM model to point cloud, and construct a synthetic dataset for construction scenes based on IFC. 2) We segment the colorless point cloud of the construction scene into five types (IfcSlab, IfcBeam, IfcWall, IfcColumn, IfcDistributionFlowElement) with Point Transformer and use focal loss to improve the segmentation accuracy of small-area components with synthetic data for data enhancement.

1. Introduction

In the building life cycle, the construction phase is a key part to control the quality of the entire building, but the traditional way of building supervision relies more on manual inspection. The development of 3D laser scanning technology has enabled us to collect and understand buildings more conveniently and quickly. Recently, there have been many related studies to achieve great progress on point cloud semantic recognition, but they either focus on the indoor scenes which contain too many components that are not applicable in the construction scene (such as furniture) or only focus on the segmentation of building structural systems (walls, beams, slabs, columns) without considering the piping system. In the existing building datasets, there are few public datasets only for construction scenes with rich semantic information annotations, and also meet the standard for the classification of component types in IFC. At the same time, the low degree of color discrimination and the special types of components included in the construction scene bring challenges to the point cloud segmentation in the construction scene. We found that many IFC models corresponding to construction scenes have not been used. If the rich semantic information in IFC can be used to synthesize effective labeled data, the problem of data shortage in real construction scenes can be solved.

Not only for the analysis of deviations between components within a single system, but also for subsystem component division, especially when MEP systems and structural systems coexist, to meet the regulatory requirements of different systems, we propose an IFC-based point cloud semantic segmentation system for building electromechanical coexistence construction scenarios.

The main contributions are as follows: 1) We propose a synthetic conversion method from BIM model to point cloud, and construct a synthetic dataset for construction scenes based on the

https://doi.org/10.7146/aul.455.c197

classification of component types in IFC. 2) We segment the colorless point cloud of the construction scene into five types (IfcSlab, IfcBeam, IfcWall, IfcColumn, IfcDistributionFlowElement) with Point Transformer and use focal loss to improve the segmentation accuracy of small-area components with synthetic data for data enhancement.

2. Related Work

Quality supervision, progress monitoring, and deviation analysis of buildings in the past relied more on manual inspection and has not yet formed a perfect automated process. With the development of 3D scanning technology, point clouds have gradually become a format for fast scanning and fully expressing building. Point cloud data of buildings under construction can be obtained directly or indirectly through laser scanners and RGB-D cameras. Without the limitations of image resolution and multi-perspective image processing, the point cloud data obtained by laser scanning is more accurate and suitable for the acquisition of construction data of large construction scenes, which is also the way we use in our method.

In order to propose a more automated and efficient solution in quality supervision, progress monitoring, and deviation analysis of buildings, many methods have been proposed to compare as-built and as-planned buildings. There are two main categories: one is Scan-vs-BIM (Bosché et al., 2014; Rebolj et al., 2017; Tran et al., 2019) and the other is Scan-to-BIM (Murali et al., 2017; Avetisyan et al., 2020; Bosché et al., 2015). The former does not require complete modeling of the building, focusing on the comparison of the corresponding structures; the latter requires a complete modeling of the building and further analysis based on the modeling results. But in both approaches, understanding the semantic information of architectural point clouds is crucial. Traditional point cloud segmentation methods are widely used such as methods based on area growth which are over-reliance on the selection of seed points and require trade-offs between accuracy and efficiency, and methods based on model fitting such as k-means clustering algorithms which still need to manually determine K values.

Compared with traditional segmentation methods, methods based on deep learning can extract high-dimensional features from training data better and cluster better. TangentConv (Tatarchenko et al., 2018) projects the 3D point cloud onto the 2D plane, which will lose information and not make full use of the spatial features such as the sparseness of the point cloud. At the same time, the selection of the projection plane may also seriously affect the recognition accuracy. OCTNET (Riegler et al., 2017) converts irregular point cloud data into the expression of point cloud voxels and uses coefficient-based convolution to reduce computation and memory consumption. Despite this, the loss of geometric information due to point-to-voxel conversion is still unavoidable. The Pointnet series of deep learning networks that can directly input 3D point cloud output segmentation results is a point-level segmentation method which completely retains the geometric information of point clouds. Pointnet (Qi et al., 2017a)extracts global features for all point cloud data, Pointnet++ (Qi et al., 2017b) can extract local features at different scales, through multi-scale combination and multi-resolution combination which ensure better feature extraction. However, it’s still not sufficient to learn local features of points, so other methods such as graph convolution are proposed. DGCNN (Phan et al., 2018)introduced a new module Edgeconv to extract the local features of point clouds, and learn the semantic information of point sets by dynamically updating the graph structure. But the direction information between points is ignored.

These methods have been applied to building elements(such as walls, floors). Based on Stanford Large-Scale 3D Indoor Spaces(S3DIS) dataset, Collins (2020) uses DGCNN network to divide 4 classes (pipe system, wall, slab, stair), without taking into account the distinction between walls and columns, which are common and important structures in construction scenarios. Kim et al(2020) automatic pipe and elbow recognition from three-dimensional point cloud model of industrial plant piping system using the convolutional neural network without considering the segmentation of large-area structural components except for the specific subdivision of the pipe categories in the MEP system. Correspondingly, Kim et al. (2021) uses DGCNN network to divide five types(wall, beam, ceiling, floor and column), without considering the pipe system at all. However, in the construction scene, the structure and the pipe system are often integrated, and because the distribution of pipe is generally close to the wall or near the beam, it will have a certain impact on the division of the two systems. Therefore, how to accurately distinguish between two systems and multiple categories within the system also has certain challenges. At the same time, the classification criteria in these articles are not unified with the IFC standards, which are important in the field of architecture.

Furthermore, the above deep learning-based methods require a large number of point cloud datasets similar to the application scenario, but the existing datasets including S3DIS, Scannet,etc contain a large number of interior furniture and the scene color is diverse, which are quite different from the construction scene. Due to the high cost of the manual labeling point cloud, there are very few point cloud datasets only for construction scenarios that are public and meet IFC standards. Considering many IFC files including rich semantic information of the building itself in the design stage have not been used, IFC-based method of synthesizing data is a good choice. Ma, Czerniawski and Leite (2020) proposes a method of synthesizing data, from Revit to point cloud. But it still doesn't shed its reliance on the S3DIS dataset. Therefore, we propose a synthesis method for construction scenarios closely integrated with IFC, and use the synthesis dataset for data enhancement.

With the significant improvement achieved by the transformer and self-attention models in the image domain, the introduction of self-attention mechanisms on point clouds has been motivated. Point transformer (Engel et al., 2021) introduces a local-global attention mechanism to combine global and local features which are used to obtain the relationship and shape information between points. Point transformer has achieved state of the art in classification and segmentation tasks.

Therefore, to provide a better solution for multi-system and multi-component semantic understanding of construction scenes, we propose a method of synthesizing point cloud datasets to solve the problem of lack of labeled construction buildings dataset; On the basis of data enhancement of synthetic datasets, we use Point transformer to divide construction point clouds into five categories (IfcSlab, IfcBeam, IfcWall, IfcColumn, IfcDistributionFlowElement) to solve the problem of semantic understanding when multiple systems coexist in the construction scene; and focal loss is introduced to solve the problem of unbalanced component samples in the construction scene.

3. Methodology

The method is mainly divided into two parts: Synthetic dataset based on IFC and Semantic segmentation of building point clouds based on Point transformer.

3.1 Constructive Conversion Method from BIM to Point Cloud

Nowadays, more and more related products in the construction industry provide data exchange interfaces based on the IFC standard. Especially in construction scenarios. Synthetic point cloud data based on IFC files can effectively make up for the lack of data in construction scenarios.

The flow chart of the entire BIM to point cloud conversion is as follows:

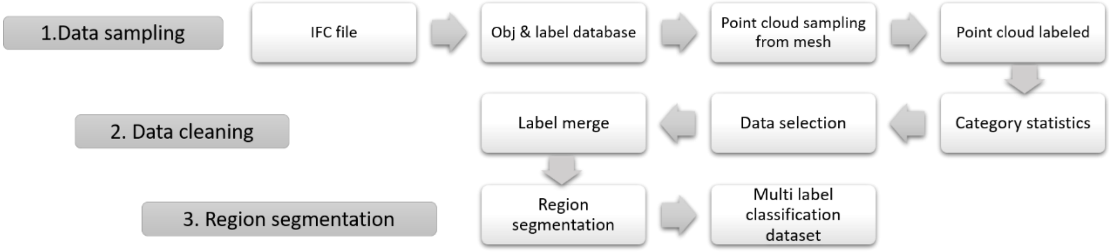

Figure 1: The flow chart of the entire BIM to point cloud conversion.

Synthesize Point Clouds from IFC. The first step is to construct building foundation components datasets automatically. We start from the IFC file of the actual building and generate a synthetic data set for training. In the first step, we convert the files from the industry foundation class format (.ifc) to obj format files (.obj) for each object based on CBIMS software platform. Second, we performed point cloud sampling and labeling from the mesh of the obj files for each foundation components class (such as wall, column, pipe, door, window, etc) through open3D. The sampling method is not random sampling but uniform sampling according to the size of the area. At the same time, the unit of length and area in IFC is unified into meters.

Label Classification and Data Cleaning Based on IFC Standards. There are more than 20 common types of IFC in construction scenarios according to statistics, but the main data types are IfcSlab, IfcBeam, IfcWall, IfcWallStandardCase, IfcColumn, IfcDoor, IfcWindow, IfcFlowFitting...These are also the main component in the construction process. According to the business needs of component identification in the construction scene, the IFC type is merged and finally focused on the following five types: IfcSlab, IfcBeam, IfcWall, IfcColumn, IfcDistributionFlowElement. IfcWall is a mixed representation of IfcWallStandardCase, and IfcDistributionFlowElement is a mixed representation of all categories related to IfcFlow. IfcBuildingElementProxy is deleted because of not a key category during construction, IfcGrid is deleted due to the existence only in IFC and not in actual buildings, and the remaining types are included in the other category.

Extract Different Small Blocks from Point Clouds. Since the point clouds extracted from the multi-story building are too large, which is not conducive to segmentation, we divide the point cloud data obtained in the previous step into blocks. Different from the point cloud voxelization operation, we only convert a large-scale point cloud into a small-scale point cloud according to certain rules, and there is no change for base objects.IfcSpace and floor elements are two kinds of available information. IfcSpace-based separation methods have certain limitations due to the lack of IFC files containing complete IfcSpace information, as well as inaccuracies at the edges of the room. Therefore, we first cut multi-story buildings into single-story buildings, then randomly cut on the XY plane of single-story buildings. Finally, the optimal division results are obtained after testing different block sizes through experiments.

3.2 Semantic Segmentation Based on Point transformer and Focal Loss

Semantic Segmentation Network. Point transformer is the classification and segmentation network for point clouds proposed by Oxford University et al, which has been verified to perform well on many point clouds datasets like S3DIS. And the most important design layer is Point transformer layer. This layer is invariant to arrangement and cardinality, so it is essentially suitable for point cloud processing. Besides, U-net structure is used in semantic segmentation, including 5 encoders and 5 decoders. The encoder uses Transition Down and Point transformer Block to down-sampling and extracts features, and the decoder uses Transition Up and Point transformer Block to up-sampling and mapping feature. Figure 3 shows the structure of Point transformer.

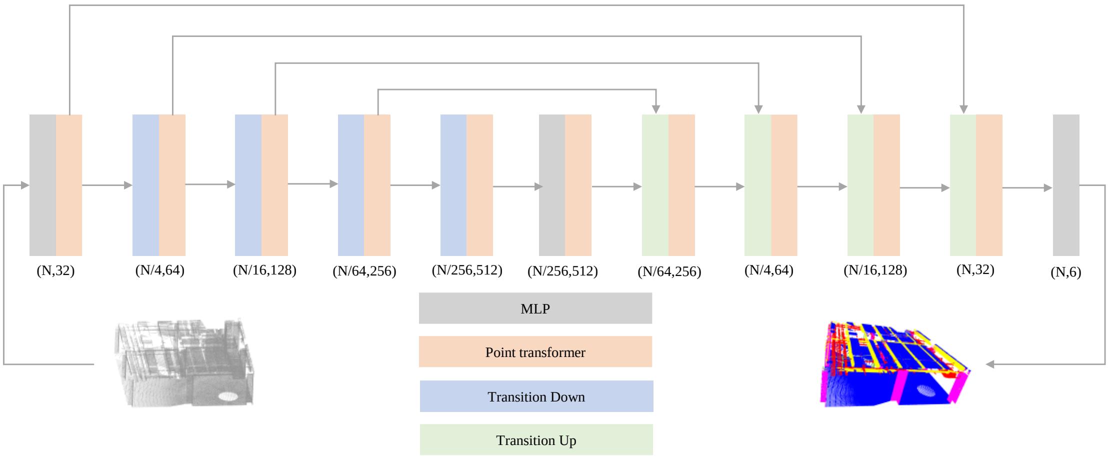

Figure 4 : The structure of Point transformer.

Based on Point transformer, we segment the construction scene point cloud into the above five types of components with the synthetic point cloud dataset obtained in the previous step as data enhancement and compared the effects of different ways of division on semantic segmentation through experiments.

Focal Loss Replaces Cross-entropy Loss. In the construction scene, the number and area of slabs and walls are large, and the relative sample number of columns, beams, and pipes is relatively small, which leads to the problem of multi-class sample imbalance. Therefore, we replace the cross-entropy loss function with the focal loss function, and give different weights to different categories according to the number of sample points and the difficulty of segmentation, so as to improve the segmentation effect.

$$Loss_{cross\ entropy} = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{1}$$

$$Loss_{focal} = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic} * \alpha_c(1-p_{ic})^{\gamma}\, log(p_{ic}) \tag{2}$$

Equations (1) and (2) represent the loss calculation formulas of cross-entropy loss and focal loss in multi-classification where M is the number of label, y_{ic} (symbolic function)is 1 if ground truth is c, else 0, p_{ic} is predicted probability that the observed sample i belongs to the class c, a_c is the weight of class c and γ is the hyperarameter adjusted according to experiments. In the experiment, we set $\gamma = 0.2$, and α is adjusted according to the number of samples.

4. Experiments

4.1 Point Cloud Collection on Real Construction Scenes

The Trimble x7 laser scanner was used in the point cloud collection process with a scanning accuracy of 2mm. There are two types of scanned construction scene buildings. The first collection scene is on the tenth floor of a residential building under construction, covering an area of about 490 square meters. Most of the scenes are wall, beam, and slab structures, without electromechanical pipelines and column components; the second scene is a square in a square. On the first floor of the basement, the actual scanned model area is about 8,000 square meters. In addition to the basic wall, beam, and slab structures, it includes a large number of column structures and electromechanical pipe structures.

Due to the large amount of data collected on-site results in the high memory usage, the point cloud data was down-sampled under the premise of meeting the accuracy requirements, and the minimum point distance of the final point cloud was 0.03m. The region is also divided to eight sub-scenarios.

4.2 Synthetic Dataset for Construction Scenes

According to the synthetic conversion method from BIM to point cloud proposed in Section 3, the design model(IFC) of the construction building is converted to a synthetic point cloud dataset on the construction site. Each scene is a single-story building, not containing decorations such as furniture, and the basic components such as beams, columns, and walls occupy the majority. Figure 4 shows one of these synthetic data scenarios.

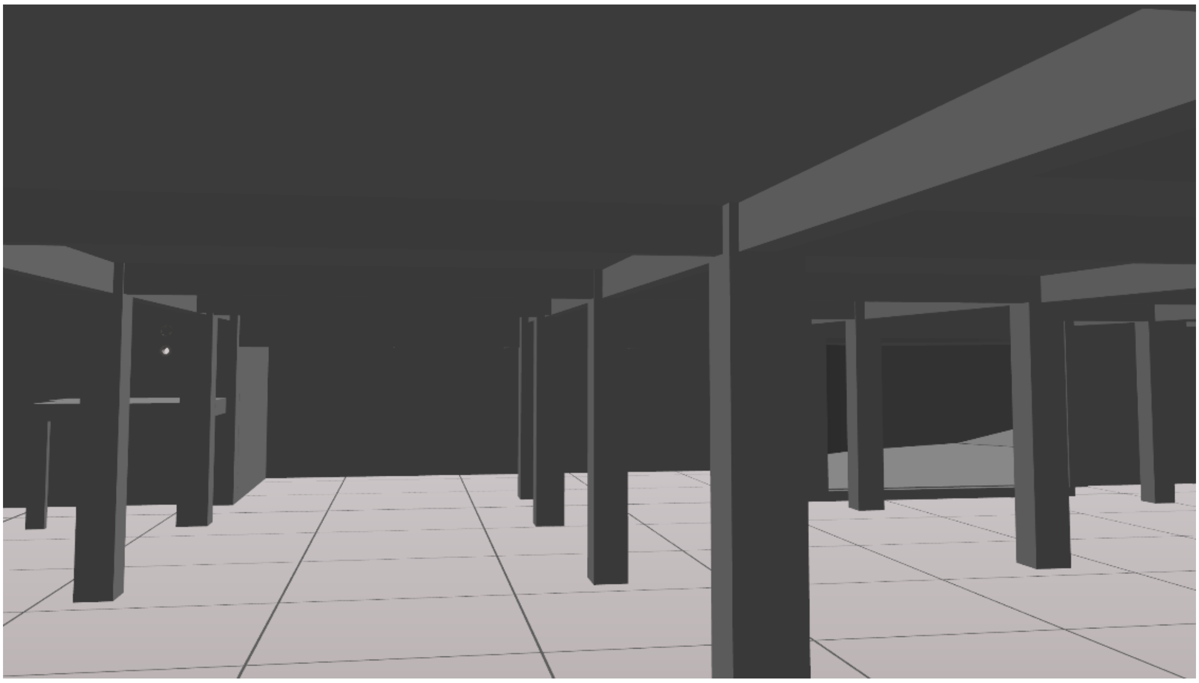

Figure 4: One Sysnthetic Scene.

We optimized the size of the block and slide into the Point transformer on synthetic datasets. Considering whether the column can be included in a block facade, we set the minimum block size to 1m, and test the segmentation accuracy on the synthetic dataset under different parameter configurations(see Table 1). In the experiment, block=2.5m and slide=1m achieved the highest accuracy.

Table 1: The effect of different blocks and slides on the segmentation accuracy of synthetic datasets.

Block size	Slide	Accuracy
5	2.5	0.74
2.5	1	0.88
1	0.5	0.62

4.3 Semantic Segmentation Results

Through comparison of different block divisions, we choose the optimal parameter (block = 2.5m, slide size = 1m) to generate the segmentation dataset, and then segment on the synthetic data and the real dataset scanned in Section 4.1.

Table 2: Compare results using S3DIS for enhancement and synthetic data for enhancement.

Class	Acc(S3DIS+real)	Acc(synthetic+real)
IfcSlab	0.990	0.990
IfcBeam	0.758	0.783
IfcWall	0.960	0.966
IfcColumn	**0.559**	**0.905**

Table 2 shows the results of comparative experiments using S3DIS for enhancement and synthetic data for enhancement. Since there is no pipe in S3DIS, we only chose four real sub-scenarios including only four types (IfcSlab, IfcBeam, IfcWall, IfcColumn), to compare the experimental results obtained by using S3DIS for enhancement and synthetic data for enhancement. It shows that indoor datasets like S3DIS that contain a large number of furniture, stairs, and other irrelevant components and have high color discrimination are not suitable for construction scenarios. However, when the synthetic dataset proposed in this paper is used for data enhancement, the segmentation accuracy on beams and columns is improved to a certain extent.

Table 3: Compare segmentation accuracy of five types under cross-entropy loss and focal loss.

Class	Acc(cross entropy loss)	Acc(focal loss)
IfcSlab	0.960	0.951
IfcBeam	**0.679**	**0.794**
IfcWall	0.915	0.837
IfcColumn	**0.122**	**0.422**
IfcDistributionFlowElement	**0.695**	**0.746**

In order to better distinguish pipe in the MEP system and important structural components, we added four construction sub-scenarios containing pipe and structural categories. At the same time, in order to improve the segmentation effect of beams and columns, we use focal loss function. We compared the segmentation accuracy of five types under cross entropy loss and focal loss(see Table 3). We compared IOU of five types under cross-entropy loss and focal loss(see Table 4).

Table 4: Compare IOU of five types under cross entropy loss and focal loss .

Class	IOU(no focal loss)	IOU(focal loss)
IfcSlab	0.905	0.913
IfcBeam	0.470	0.431
IfcWall	0.677	0.640
IfcColumn	**0.106**	**0.248**
IfcDistributionFlowElement	**0.612**	**0.655**

Table 4 shows that after replacing the loss function with our segmentation network, the segmentation accuracy of beams, columns, and pipes is improved, especially for columns, while the segmentation accuracy of beams and walls remains unchanged. Figure 5 shows three real sub-scene segmentation results, the left column is ground truth, and the right column is the segmentation result.

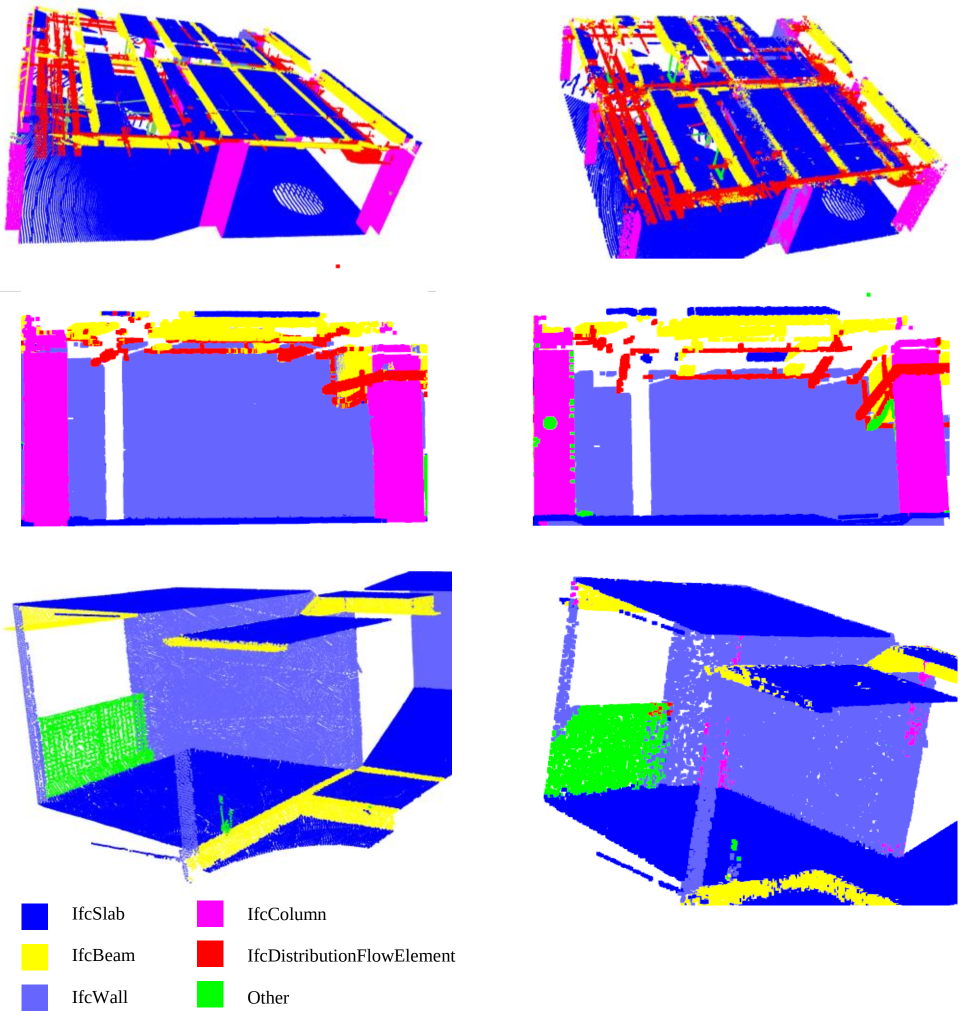

Figure 5: Three real sub-scene segmentation results, the left column is ground truth, and the right column is the segmentation result.

5. Conclusion & Outlook

In this paper, a more efficient solution is proposed for multi-system multi-component semantic understanding, from generating data, to proposing methods and solving special difficulties in construction. In view of the lack of synthetic data of construction buildings, we propose a method for generating synthetic point cloud datasets from IFC, which automatically generates annotated synthetic point cloud datasets. Aiming at the problem of non-uniform component classification standards, we use IFC as a reference to determine the category of segmentation objects to ensure the unification of upstream and downstream tasks. Aiming at the inaccurate segmentation of multi-category components in multiple systems, we use synthetic datasets for data enhancement, segment the construction point cloud into five categories(IfcSlab, IfcBeam, IfcWall, IfcColumn, IfcDistributionFlowElement) with a better segmentation network Point transformer, and introduce focal loss to solve the problem of unbalanced component samples in construction scenes. We hope that more good methods can be proposed in the future to solve the problem of semantic information understanding in more subdivided categories for construction scenarios.

Acknowledgement

This work was supported by the 2019 MIIT Industrial Internet Innovation and Development Project " BIM Software Industry Standardization and Public Service Platform".

References

Avetisyan, A., Khanova, T., Choy, C., Dash, D., Dai, A., and Nießner, M. (2020, August). Scenecad: Predicting object alignments and layouts in rgb-d scans. In European Conference on Computer Vision. Springer, Cham., pp. 596-612.

Bosché, F., Ahmed, M., Turkan, Y., Haas, C. T., and Haas, R. (2015). 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, pp.201-213.

Bosché, F., Guillemet, A., Turkan, Y., Haas, C. T., and Haas, R. (2014). Tracking the built status of MEP works: Assessing the value of a Scan-vs-BIM system. Journal of computing in civil engineering, 28(4), 05014004.

Collins, F. (2020). Encoding of geometric shapes from Building Information Modeling (BIM) using graph neural networks.

Engel, N., Belagiannis, V. and Dietmayer, K. (2021). Point transformer. IEEE Access, 9, pp.134826-134840.

Kim, H., and Kim, C. (2021). 3D as-built modeling from incomplete point clouds using connectivity relations. Automation in Construction, 130, p.103855.

Kim, Y., Nguyen, C. H. P., and Choi, Y. (2020). Automatic pipe and elbow recognition from three-dimensional point cloud model of industrial plant piping system using convolutional neural network-based primitive classification. Automation in Construction, 116, p.103236.

Ma, J., Czerniawski, T. and 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, p.103144.

Murali, S., Speciale, P., Oswald, M. R., and Pollefeys, M. (2017, September). Indoor Scan2BIM: Building information models of house interiors. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)., pp. 6126-6133.

Phan, A.V., Le Nguyen, M., Nguyen, Y.L.H. and Bui, L.T. (2018). Dgcnn: A convolutional neural network over large-scale labeled graphs. Neural Networks, 108, pp.533-543.

Qi, C.R., Su, H., Mo, K. and Guibas, L.J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition.,pp. 652-660.

Qi, C.R., Yi, L., Su, H. and Guibas, L.J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30.

Rebolj, D., Pučko, Z., Babič, N. Č., Bizjak, M., and Mongus, D. (2017). Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring. Automation in Construction, 84, pp.323-334.

Riegler, G., Osman Ulusoy, A. and Geiger, A.(2017). Octnet: Learning deep 3d representations at high resolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition.,pp. 3577-3586.

Tatarchenko, M., Park, J., Koltun, V. and Zhou, Q.Y. (2018). Tangent convolutions for dense prediction in 3d. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.,pp.3887-3896.

Tran, H., and Khoshelham, K. (2019). Building change detection through comparison of a lidar scan with a building information model. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, pp.889-893.