Enriching Point Features to Improve Semantic Segmentation of Point Clouds

Yuandong Pan[1,2], M. Saeed Mafipour[1] and Mansour Mehranfar[1]
[1]Chair of Computational Modeling and Simulation, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
[2]Institute for Advanced Study, Technical University of Munich, Lichtenbergstrasse 2a, 85748 Munich, Germany
E-mail(s): yuandong.pan@tum.de, m.saeed.mafipour@tum.de, mansour.mehranfar@tum.de

Abstract: Recent advances in capturing technologies have resulted in the fast capturing process of existing assets with high measurement accuracy. The resulting point cloud data (PCD) can be processed and used for the digital twinning of existing facilities. Point cloud segmentation is a critical step, and it has been continuously improved by 3D deep learning approaches, which require point features as input and predict labels as output. In this paper, apart from those features from raw data such as spatial information (x-, y-, z-coordinates) and colour information (RGB values), we propose a set of new point features which considers points' local neighbouring information and the global information in the entire point cloud. We compute new point features in two different facilities (building and bridge) and evaluate the performance of two different networks with and without new features as input. The results show that the proposed feature enrichment can improve point cloud segmentation as well as digital twinning of assets.

Keywords: Semantic Segmentation, Digital Twins, Deep Learning, Point Cloud Data

1 Introduction

In the architecture, engineering, and construction (AEC) industry, digital methods have revealed tremendous applications in facilitating the operation and maintenance process of structures. In this domain, building information modeling (BIM) has been able to provide descriptive models to support structures through their life cycle. A building information model contains the geometry and corresponding data to represent the design and construction activities. As one of the recent advances in AEC, BIM has been used not only in the as-designed phase of the structure but also in as-built and as-is phases. In recent decades, there have been concerted research efforts to create and optimize BIM in the as-designed phase of structures. However, the lack of descriptive and digital methods in the as-built and as-is phases is still tangible.

As-built and as-is models correspond to the descriptive representation of existing assets from their post-construction phases. These models are utilized for monitoring and evaluation of the structure as well as its maintenance and operation. Recently, the concepts of as-built and as-is have been extended to digital twin (DT). In the domain of BIM, a DT can be defined as a geometric-semantic replica that represents the current status of the structure. DT of a structure inherits all the features of an as-is model and also demonstrates the interaction of humans with the structure. A DT can be updated in regular intervals depending on the type of the product. These intervals can be longer in the case of structures, as the physical features of the asset change gradually. The modeling process of an as-is or DT is generally started from the 3D model reconstruction of the existing structure. This model is then enriched with the information collected from the construction site. These models are generally created from the point cloud data (PCD) captured by laser scanning or photogrammetry. Semantic segmentation is an essential step in processing PCD and creating as-built, as-is, or DT models. Through semantic segmentation, the entire PCD is separated into the point clusters demonstrating the desired elements (classes). The point cloud of elements can then be used to create the 3D model of the structure. Semantic segmentation has two significant advantages: 1) the modeling process is simplified from the entire PCD to the cloud of elements, and 2) the element type of the segmented clouds is recognized. In the current practice, the semantic segmentation process of assets is performed manually, which is labor-intensive and error-prone. Recently, there have been efforts to automate the semantic segmentation process of PCD captured from buildings and bridges. Especially with the help of deep learning, the results of point cloud segmentation have been improved a lot. Apart from designing the new architecture of neural networks, point features also have a significant impact on the result.

This paper is organised as follows. Section 2 discussed some related research in point cloud segmentation by deep learning. The proposed feature design is introduced in Section 3. Results and discussions are shown in Section 4. In Section 5, we summarise the output and discuss potential future work.

2 Related research

Deep learning approaches have been widely applied to predict unknown information based on input features in the dataset. In general, predicting unknown labels from input features is the classification problem in deep learning [1]. When processing point cloud, we usually use the term segmentation. But it can also be seen as a classification problem for each point in the point cloud. Many networks are designed to adapt to both classification problems and segmentation problems in one architecture framework.

As point clouds are usually unstructured data, some deep learning approaches requires a pre-processing step. By voxelisation, the unstructured point cloud data are converted to voxel structure. In [2], 3D Convolutional Neural networks (CNNs) are applied to solve a binary classification problem. A more general network with 3D CNN architecture, VoxNet [3] is proposed to detect classes of objects for 3D point cloud data. A volumetric grid that can represent the spatial occupancy is calculated first and then applied to 3D CNNs. However, some original information from the raw point cloud could be lost in the voxelisation process.

Many networks try to work with points directly. One architecture named PointNet, proposed in [4], is the pioneer in this field. It takes point clouds as input directly and outputs labels for the entire input (classification task) or labels for each point (segmentation task). The novel design of PointNet considers each point (x-, y-, z-coordinate) independently at the first stage. Apart from the three spatial coordinates values (x,y,z), it can also use additional features like colours, normals, etc. An improved network based on PointNet considering spatial information of point sets proposed by the same author is called PointNet++ [5]. In PointNet++, points are grouped into overlapping local regions by their distance. Then local features are extracted by extracting fine geometric structures from small neighbour regions. These local features are then grouped together into larger units and processed to make relative higher-level features. Many more advanced networks are proposed to improve the prediction performance. In [6], a module that learns spatial contextual features from the point cloud is implemented and embedded in an encoder-decoder architecture. In [7], the local information of points is extracted, and the distinctness of the points from multiple resolutions is considered to achieve semantic segmentation. In recent years, the transformer architecture [8], originated from natural language processing (NLP), is applied in point cloud segmentation, such as [9], [10], and [11]. These networks prove the transformer architecture is very powerful in 3D deep learning.

With regard to processing point cloud in built environment, [12] employed a heuristic top-down approach to detect the bridge elements in the point cloud of concrete bridges. [13] elaborated a hybrid image-based-geometric point cloud segmentation method to extract features from images and train a multilayer perceptron (MLP). [14] applied a density-based heuristic algorithm to detect elements in the point cloud of bridges. As a geometric approach, [15] added contextual features of points to improve the accuracy of PointNet, and deep graph-convolutional neural network (DGCNN) [16]. [17] introduced a heuristic algorithm based on the existing connection rules in the point cloud of steel bridges for semantic segmentation. [18] defined a local descriptor to calculate the local features of points for semantic segmentation of bridges through a geometric method. In [19], the authors design a network for the Scan-to-BIM process to segment the structural, architectural, and mechanical components.

3 Proposed Method

Depending on device that used to collect point clouds, point features available in the point clouds can be differentiated. The basic features that usually exist in most of the point clouds are 3D coordinates (x,y,z) and colour values (r,g,b). In this section, seven more features are calculated from the point clouds and the input feature vector is enriched from the 6 original features to 13 features.

3.1 Features representing local information

The components of normal vectors in 3D are the first three values that are added to the feature vector. Basically, the normal of a point can to be estimated from the surrounding neighbouring points by estimating the normal of a plane tangent to the surface. normal vector of points can be a valuable property that differentiates various points easily. For example, the normal of a point on the vertical wall is horizontal while this normal vector is vertical on the horizontal floor points. Since estimated normals can only represent the neighbouring geometric information of points, more features that include other information are considered.

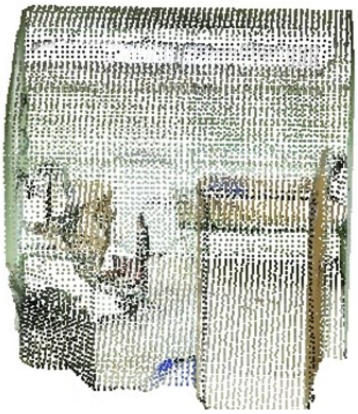

Figure 1: Original point cloud of room

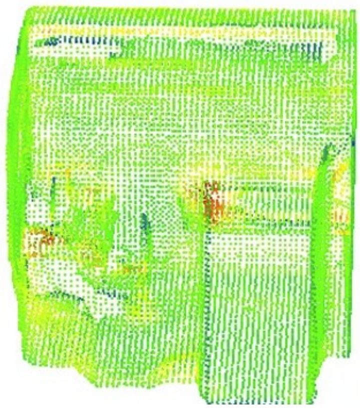

Figure 2: Colour field for number of neighbouring points

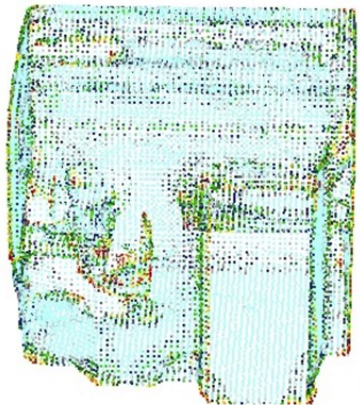

Figure 3: Colour field for mean shift

The number of neighbouring points within a given radius of a point can also differentiate some points. For those point located in dense region (for example room corner), their value would be higher than others. The colour-coded scalar field for the number of neighbouring points is shown in Figure 2, in which the neighbouring number for red points is larger. Another local feature that we use is the mean shift of a point. We compute the centroid of a point's neighbouring points within a radius first and then compute the distance between the centroid and the point. The colour-coded scalar field for the mean shift value is shown in Figure 3. We can see the value for those points on edges is usually larger.

3.2 Features representing density after projection

Apart from the features incorporating the local information of points, two more features representing the density of points after projection are added. The first one, called the 2D density, is the number of neighboring points after projection. This feature is calculated after projection of all points onto the XY plane and computing the number of neighbours within a circle with a predefined radius. The colour-coded scalar field for this value is shown in Figure 4. As we can see, this value is larger for those points on the vertical plane (yellow/red points) while smaller on the horizontal plane (blue points). This value shows the distribution of all points in the point cloud on the XY plane.

Similarly, for the second density-based feature, all the points are projected onto the z-axis and the number of neighbours within a range on the axis is computed. The colour-coded scalar field for this value is shown in Figure 5. As we can see, this value is larger for those points on the horizontal plane (yellow/red points) while smaller on the vertical plane (blue points).

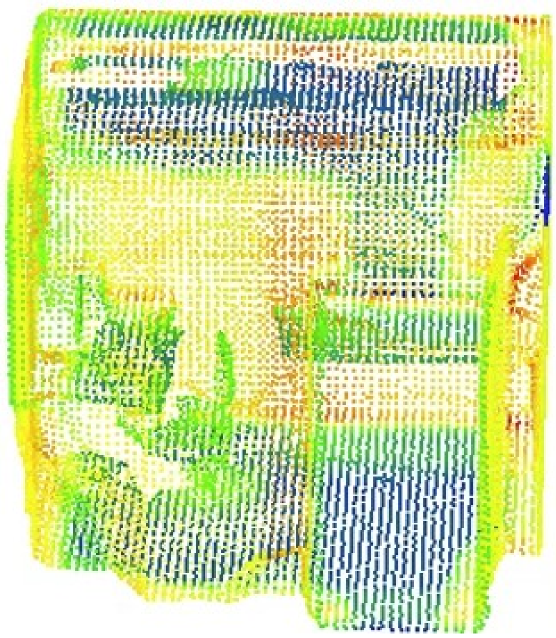

Figure 4: Colour field for number of neighbours in XY plane

Figure 5: Colour field for number of neighbours in vertical direction

4 Experiment and Result

The designed features are extracted for two datasets including the S3DIS dataset [20] which is the point cloud dataset of large scale indoor environment, and a birdge dataset that contains 10 bridges [12]. Also, the performance of enriching features are evaluated in two networks (Pointnet++ [5] and RandLA-Net [21]).

Test results of RandLA-Net [21] on the bridge dataset are shown in Figure 6 to Figure 9. It can be seen that the prediction with feature enrichment is closer to the ground truth. This can be also seen in the Table 1, where Intersection over Union (IoU), as a statistical indicator, is higher for the dataset with feature enrichment. It is obvious that the results of all classes are improved by enriching features of points.

Figure 6: Original point cloud of bridge

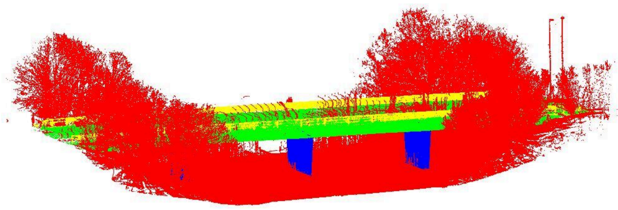

Figure 7: Semantic ground truth for bridge

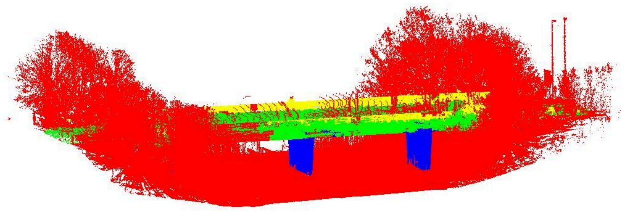

Figure 8: Prediction result without feature enrichment

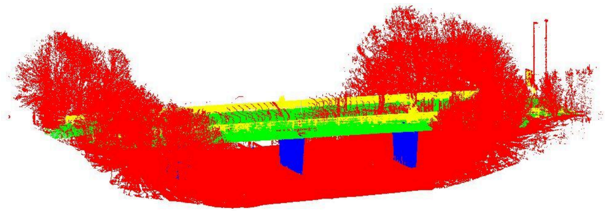

Figure 9: Prediction result with feature enrichment

Table 1: Prediction result with/without feature enrichment in bridge dataset (IoU:%)

	mIoU	Deck	Pier	Railing	Background
without	74.4	77.6	78.5	48.4	93.0
with	87.8	93.6	88.3	72.7	96.7

In the dataset of indoor environments, Pointnet++ [5] is trained and evaluated based on IoU as well. The quantitative evaluation result is shown in Table 2. It can be seen that the results of most classes of structural elements such as ceiling, floor, wall, and column are improved. However, the prediction accuracy for some classes such as sofa (from 42.2% to 26.6%) and bookcase (from 59.8% to 56.2%) decreases.

The potential reason is that most of the structural elements in buildings and bridges are either horizontal or vertical. Horizontal elements such as slabs or decks generally provide the required surfaces for sustaining and transferring the applying live loads to the structure. Vertical elements such as columns or piers on the other side convey the dead and live loads of the structure to the ground. The newly added features can represent the information of vertical and horizontal planes. However, they cannot illustrate geometric information of furniture points which are not horizontal or vertical. Hence, the newly added information makes the model more confused in distinguishing those points.

Table 2: Prediction result with/without feature enrichment in building dataset (IoU:%)

	mIoU	Ceiling	Floor	Wall	Beam	Column	Window	Door	Table	Chair	Sofa	Bookcase	Board	Clutter
without	52.7	86.0	96.2	70.3	0	4.2	37.2	26.8	66.2	71.4	42.2	59.8	47.1	40.2
with	54.7	88.8	96.4	74.4	1.1	15.4	38.2	26.6	67.2	78.8	26.6	56.2	45.6	38.8

5 Conclusion

The paper presents a new feature enrichment design for semantic segmentation of point cloud data. The results of testing two different network architectures on the dataset of two different facilities shows that feature enrichment can improve the performance of the deep learning models. In addition, the results show higher improvement in the prediction accuracy of the bridge dataset than that of the building dataset. It seems that there are two potential reasons behind that. Firstly, the class number in the bridge dataset is smaller, and this makes the segmentation relatively easier. Secondly, the occlusion in the bridge point cloud is not as much as that in the building where lots of furniture exists.

Despite improvements in the prediction accuracy of deep learning models, the impact of each feature on the performance of the model is still unclear.

In the future, more features are added and tested separately for improving the performance of 3D deep learning models. Also more hand-crafted features are explicitly designed for the models.

References

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[2] D. Prokhorov, "A convolutional learning system for object classification in 3-d lidar data", *Trans. Neur. Netw.*, vol. 21, no. 5, pp. 858–863, May 2010, ISSN: 1045-9227. DOI: 10.1109/TNN.2010.2044802. [Online]. Available: http://dx.doi.org/10.1109/TNN.2010.2044802.

[3] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition", in *Ieee/rsj International Conference on Intelligent Robots and Systems*, 2015, pp. 922–928.

[4] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation", *arXiv preprint arXiv:1612.00593*, 2016.

[5] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space", *arXiv preprint arXiv:1706.02413*, 2017.

[6] S. Fan, Q. Dong, F. Zhu, Y. Lv, P. Ye, and F.-Y. Wang, "Scf-net: Learning spatial contextual features for large-scale point cloud segmentation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 504–14 513.

[7] S. Qiu, S. Anwar, and N. Barnes, "Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 1757–1767.

[8] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need", *Advances in neural information processing systems*, vol. 30, 2017.

[9] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "Pct: Point cloud transformer", *Computational Visual Media*, vol. 7, no. 2, pp. 187–199, 2021.

[10] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 16 259–16 268.

[11] N. Engel, V. Belagiannis, and K. Dietmayer, "Point transformer", *IEEE Access*, vol. 9, pp. 134 826–134 840, 2021.

[12] R. Lu, I. Brilakis, and C. R. Middleton, "Detection of structural components in point clouds of existing rc bridges", *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 3, pp. 191–212, 2019, ISSN: 1093-9687. DOI: http://dx.doi.org/10.1111/mice.12407.

[13] F. Hu, J. Zhao, Y. Huang, and H. Li, “Structure-aware 3d reconstruction for cable-stayed bridges: A learning-based method”, *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 89–108, 2021, ISSN: 1093-9687.

[14] G. Qin, Y. Zhou, K. Hu, D. Han, and C. Ying, “Automated reconstruction of parametric bim for bridge based on terrestrial laser scanning data”, *Advances in Civil Engineering*, vol. 2021, p. 8 899 323, 2021, ISSN: 1687-8086. DOI: https://doi.org/10.1155/2021/8899323.

[15] J. S. Lee, J. Park, and Y.-M. Ryu, “Semantic segmentation of bridge components based on hierarchical point cloud model”, *Automation in Construction*, vol. 130, p. 103 847, 2021, ISSN: 0926-5805.

[16] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, “Dynamic graph cnn for learning on point clouds”, *ACM Transactions on Graphics (TOG)*, vol. 38, pp. 1–12, 2019.

[17] Y. Yan and J. F. Hajjar, “Automated extraction of structural elements in steel girder bridges from laser point clouds”, *Automation in Construction*, vol. 125, p. 103 582, 2021, ISSN: 0926-5805.

[18] T. Xia, J. Yang, and L. Chen, “Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning”, *Automation in Construction*, vol. 133, p. 103 992, 2022, ISSN: 0926-5805.

[19] Y. Perez-Perez, M. Golparvar-Fard, and K. El-Rayes, “Scan2bim-net: Deep learning method for segmentation of point clouds for scan-to-bim”, *Journal of Construction Engineering and Management*, vol. 147, no. 9, p. 04 021 107, 2021.

[20] I. Armeni, O. Sener, A. R. Zamir, *et al.*, “3d semantic parsing of large-scale indoor spaces”, in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2016.

[21] Q. Hu, B. Yang, L. Xie, *et al.*, “Randla-net: Efficient semantic segmentation of large-scale point clouds”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.