

SEMANTIC SEGMENTATION OF INDOOR 3D POINT CLOUDS BY JOINT OPTIMIZATION OF GEOMETRIC FEATURES AND NEURAL NETWORKS

M. M. Yao^{1,2}, X. M. Li^{1,2}, W. X. Wang^{1,2}, L. F. Xie^{1,2}, S. J. Tang^{1,2*}

¹School of Architecture and Urban Planning, Research Institute for Smart Cities, Shenzhen University – 2060325013@email.szu.edu.cn; (lixming, wangwx, linfuxie, shengjuntang)@szu.edu.cn

²Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen, P.R. China

Commission IV, WG IV/9

KEY WORDS: Neural Network, Point Clouds, Plane Extraction, Region-growing Segmentation.

ABSTRACT:

Indoor navigation, indoor robotics, and other deep applications of interior space can be realized through semantic segmentation of 3D point clouds. We propose a semantic segmentation method for point clouds that uses geometric features of point clouds and neural networks to address the problem of incomplete and inconsistent segmentation objectives in existing semantic segmentation methods. Using neural networks, semantic labels are extracted from indoor structural information as the first step. The paper proposes a probabilistic model to cross-validate the initial segmentation results with the segmentation results of geometric features to achieve joint optimization of the results for semantic segmentation. Three sets of indoor point clouds data from simple to complex indoor scenes are used to test the accuracy and validity of the segmentation method proposed in this paper. The experimental results demonstrate that the method proposed in this paper can effectively improve the semantic segmentation accuracy of indoor 3D point clouds.

1. INTRODUCTION

As laser scanner equipment and depth sensors rapidly developed, the acquisition of indoor high-precision 3D point clouds has become more and more convenient, and it has gradually become the key to supporting various indoor space-oriented applications (Kang et al., 2020), including indoor navigation (Choi et al., 2014), indoor robotics (Taira et al., 2018), and augmented reality (Pomerleau et al., 2015). The raw point clouds lack semantic information and cannot be directly applied in deep indoor spatial applications (Bello et al., 2020). A major challenge in this research area is how to achieve high-precision semantic segmentation of indoor 3D point clouds.

In traditional 3D point cloud segmentation, geometric features of the point clouds are used first, followed by architectural features. To obtain accurate edge information about an object, Bao-Shun et al., (2013) developed a rolling circle-based edge detection model. Xu et al., (2009) proposed a method for segmenting laser point clouds using graph theory and region growth. Zhan et al., (2011) proposed a point segmentation method based on normal vector estimation and color clustering, which combines the advantages of geometric and color feature segmentation, but is prone to over segmentation. Biosca and Lerma, (2008) proposes an unsupervised file clustering method based on a fuzzy method to achieve plane extraction from laser scanned point clouds. There are two steps in this method: surface growth and clustering, and it can be used effectively on uneven terrain. There is, however, a risk of false planes being extracted using this method. Traditional point clouds segmentation methods can segment objects in specific scenes. However, they lack semantic information and require a large number of parameters.

Increasingly more scholars are using convolutional neural networks on 3D point clouds as convolutional neural networks have made significant progress on 2D images. Researchers used indirect point clouds processing methods to segment point clouds semantically in the past. The Squeeseg (Wu et al., 2017) method extracted features and semantic segments from point clouds projected onto a two-dimensional plane, but it compressed three-dimensional spatial information into two-dimensional space, resulting in a loss of three-dimensional spatial information. In O-CNN (Wang et al., 2020), the original point cloud data was voxelized, and then the octree structure was used to store the features on the leaf nodes. As a final step, it performs a convolution and pooling operation to learn feature information and segment the point clouds. The method does not involve a local geometric structure and does not learn the features of local point clouds. MVCNN (Feng et al., 2018) maps 3D objects into different view angles of 2D maps and uses CNN to extract features. By using CNN on 2D images and performing 3D recognition in real time, this method takes advantage of the maturity of CNN on 2D images. PointNet (Qi et al., 2017) uses spatial transformation, recurrent neural networks, and symmetric functions to solve point cloud disorder, pioneering semantic segmentation directly on point clouds. The local features between points cannot be obtained since each point's features are learned separately. To solve this problem, PointNet++ (Qi et al., 2017) uses a hierarchical structure to learn the features of point clouds based on PointNet to improve the segmentation accuracy, but at the same time increase the computational complexity. PointNet and PointNet++ use farthest distance downsampling to reduce the data volume, which preserves boundary information nicely, but the computational efficiency increases as the number of point clouds increases. The semantic segmentation methods based on deep learning can effectively handle multiple segmentations, but they lose point clouds feature information during data

* Corresponding author

preprocessing and are dependent on sample data accuracy, which is inefficient.

Using the above analysis, we can conclude: segmentation methods based on geometric features can quickly and effectively obtain the planar structure, while clustering algorithms can accurately segment different types of elements. These methods rely on geometric information only for simple semantic segmentation, such as walls, floors, ceilings, etc., and it is difficult to determine the complex structure of tables and chairs. With deep learning, multiple types of elements can be segmented based on a large number of samples. Due to the limitations of training samples and feature descriptions, generalizing a model trained in one region to another is difficult. Based on the limitations and advantages of the two methods, a jointly optimized indoor segmentation method is proposed by integrating geometric features and neural networks. In this method, point clouds are first semantically labeled using neural networks and then segmented using point cloud structural features. To optimize the semantic segmentation results, the "semantic labels of point clouds" and "precise segmentation results of point clouds" are used based on the probabilistic model.

2. INDOOR POINT CLOUDS SEMANTIC SEGMENTATION OPTIMIZATION METHOD

As shown in Figure 1, the method proposed in this paper can be divided into three main parts : semantic segmentation module of deep learning, point clouds segmentation module based on geometric and color features, and joint semantic label optimization module. In the semantic segmentation module, we use a large sample deep learning method to extract semantic information from the indoor structure initially; meanwhile, we extract the interior structures using geometric and color features extracted from the original point clouds. Finally, we examine the advantages and disadvantages of the two types of segmentation methods in terms of segmentation accuracy. To optimize both outcomes simultaneously, we propose a probabilistic model-based semantic labeling optimization method. In order to complete the semantic association, we use the K-nearest neighbor algorithm to identify geometric and color features, and then find the deep learning label point cloud corresponding to each point in the color cluster. Then the labels of all point clouds in the cluster are reassigned to the label category with the highest number of labels in each color cluster.

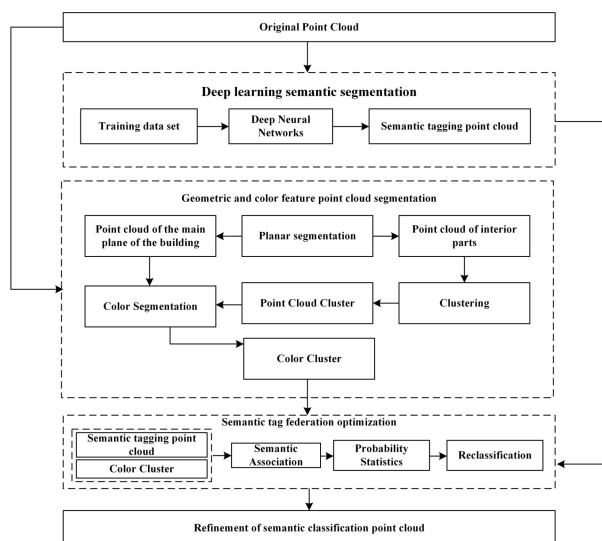


Figure 1. Workflow of our method.

2.1 Deep learning-based point clouds semantic segmentation

In the portion of this paper that focuses on deep learning, we adopt the RandLA-Net (Hu et al., 2020) framework as the basis for initial segmentation of point clouds. RandLA-Net is a lightweight neural network that is trained directly from point clouds and incorporates downsampling to achieve efficient processing of large-scale point clouds data in order to address the issue of the high computational cost of downsampling traditional neural networks, which does not apply to large-scale point clouds. Meanwhile, to solve the problem of losing geometric features of point clouds caused by random downsampling, a local feature aggregation module with an attention mechanism has been introduced in order to achieve fast downsampling of point clouds while retaining all the geometric features.

Random downsampling: Using the random sampling method, it is obtained the set of downsampled points by sampling uniformly k points from a total of N points to find the set of downsampled points. There is no dependence on the number of input points, but only a relation between the set of selected input points K and the computational efficiency, as shown in Equation 1.

$$P = \{N_1, N_2, N_3 \dots N_K\} \quad K \in N \quad (1)$$

Local feature aggregation: It consists of three parts: local space encoding, attention mechanism pooling, and dilated residual block. Firstly, to improve efficiency, the nearest neighbor algorithm based on euclidean distance is used to aggregate the proximity points. For each nearest neighbor K points of the centroid encode their relative positions A_i^k , as shown in Equation 2, where B_i and B_i^k is the spatial position of the point, \oplus is the cascade operation, $\|\cdot\|$ is the euclidean distance between the centroid and the nearest neighbor.

$$A_i^k = mlp(B_i \oplus B_i^k \oplus (B_i - B_i^k) \oplus \|B_i - B_i^k\|) \quad (2)$$

Secondly, The second step is to connect the encoded relative positions A_i^k to the features of each B_i^k , so that each point covers the features of the surrounding K nearest neighbors, resulting in data redundancy. In addition to that, an attention mechanism is introduced, which is designed to automatically learn the important local features so that the important feature information is retained as much as possible throughout the pooling process. Finally, a residual module is introduced to increase the acceptance domain of each point to retain more details of the point clouds during pooling.

We present an approach to train the RandLA-Net neural network using a training dataset, as well as to semantically segment the original data using the trained model to obtain semantic labels for point clouds by using the trained model. There are 13 categories in the segmentation result: ceiling, floor, wall, beam, column, window, door, chair, sofa, bookcase, board, and miscellaneous, as shown in Figure 2.

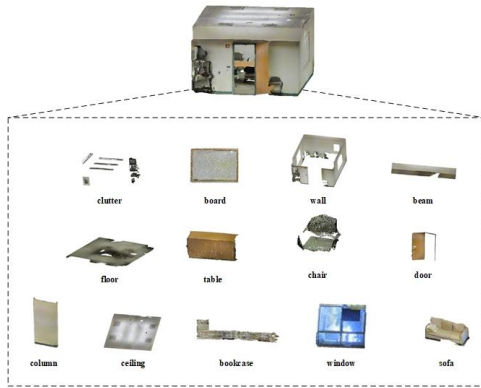


Figure 2. Indoor component segmentation results based on deep learning.

2.2 Point clouds segmentation based on geometric and color features

2.2.1 Multi-level plane extraction: Due to the large differences in shape and geometric characteristics of the interior components of the building, two-component categories of building main planes and the interior components of the building are defined according to the similarity of geometric characteristics of these components in this paper. The main planes of the building includes the floor, ceiling, walls, and wall attachments such as windows and blackboards. Interior components are other objects that remove the main plane of the building. For example, tables, chairs, miscellaneous objects, etc. We first use the plane extraction method to extract the main plane of the building. The traditional RANSAC plane segmentation (Tarsha-Kurdi, 2007) can accurately extract the planes in the point clouds data, but it is influenced by the number of iterations, and it is easy to produce insufficient iterations and poor model fitting for complex scenes. Therefore, We design a multi-stage plane extraction method to extract the main planes of indoor point clouds for the shortage of traditional RANSAC plane extraction, which contains two parts: Planar coarse extraction and Planar fine extraction.

Planar coarse extraction: Finding the maximum value P_{max} and minimum value P_{min} of the original indoor point clouds coordinates. Based on the original point clouds of the square interior building, the building's main planes are close to the outer edge of the rectangular enclosing box formed by P_{min} and P_{max} . The six planes of the enclosing box are therefore calculated, and the plane model is shown in Equation 3.

$$ax + by + cz + D = 0 \quad (3)$$

Following that, the point clouds empty set $\{a, b, c, d, e, f\}$ of six enclosing boxes are created. We traverse the original point clouds data P , and for each point clouds $P_i \in P$ in turn to the six planes of the enclosing box to determine the distance. It is assumed that the distance between P_i and the enclosing box plane is less than the threshold value T , then the point clouds that correspond to the enclosing box plane will be added to the set of point clouds that correspond to the enclosing box plane, thus completing the coarse extraction of the indoor point clouds building the body planes. The distance judgment is shown in Equation 4.

$$d = \frac{|ax + by + cz + D|}{\sqrt{a^2 + b^2 + c^2}} \quad (4)$$

Planar fine extraction: Three points $\{p_1^t, p_2^t, p_3^t\}$ are randomly selected from each S_i ($S_i \in \{a, b, c, d, e, f\}$) to fit the planar model L^t . As a result of this procedure, the distance between the remaining points and the plane model in S_i is calculated, and the points that are smaller than the distance threshold σ are used as the score. By iterating this step Z times, we will be able to select the plane that has the highest score and complete the extraction of the main building plane $\{S1, S2, S3, S4, S5, S6\}$ from the main building site. The multi-level plane extraction process is shown in Figure 3.

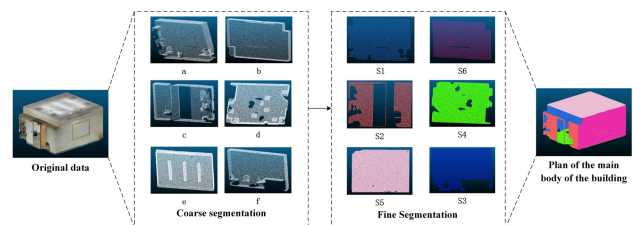


Figure 3. Multi-level planar extraction process.

2.2.2 Indoor component clustering segmentation: In order to compute the exact point clouds of the main planes of the building, multi-level plane extraction is performed. In order for the indoor component clustering to be free from interference from the building's main planes, the point clouds set of the building's main body planes $\{S1, S2, S3, S4, S5, S6\}$ are first eliminated from the original point clouds in order to make them free of interference. First of all, using the coordinates of each $p_p \in \{S1, S2, S3, S4, S5, S6\}$ as the search center and the euclidean distance as the unit of measure, the KNN algorithm is used in order to search for the closest point in the original point clouds to the search center. Following this, the nearest point searched is deleted in order to obtain the point clouds of indoor component. The distance calculation formula is shown in Equation 5, $v \in V$.

$$D(u, v) = \sqrt{(u_x - v_x)^2 + (u_y - v_y)^2 + (u_z - v_z)^2} \quad (5)$$

When the main planes of the building are rejected, the interior component point clouds are retained. An algorithm for coarse segmentation of indoor component point clouds based on euclidean clustering is used to extract the components of the indoor component point clouds into separate clusters according to their properties, as shown in Figure 4.



Figure 4. The process of indoor component clustering and segmentation.

Euclidean distance-based cluster segmentation proceeds as follows:

- (1) Find any point W in the space, use the point as the search center, and set the euclidean distance threshold.
- (2) Find the nearest K points to the point W .

(3) Iterate through each neighboring point and save those whose distance from point W is less than the euclidean distance threshold d_{th} in cluster A .

(4) Select another point in the cluster except for point A . The process of clustering will continue until the number of points in cluster A no longer increases. Then the process of clustering will be completed.

(5) Until all points are clustered, select the unclustered points for the next clustering.

2.2.3 Color-based region growing segmentation: It is important to note that multilevel planar segmentation and Euclidean distance clustering are mainly based on spatial geometric features for segmenting point clouds, and do not take color information about the point clouds into account. When it comes to complex indoor scenes, color plays an important role in the segmentation of point clouds as a feature that is important to it (Zhan et al., 2011b). Consequently, in this paper, using the segmented point clouds of the building body plane and the clustered point clouds of the interior parts of the building, the point clouds are segmented again using the color-based region growth method to obtain the final segmentation point set based on geometric as well as color features in order to obtain the final segmentation of the building, as shown in Figure 5.

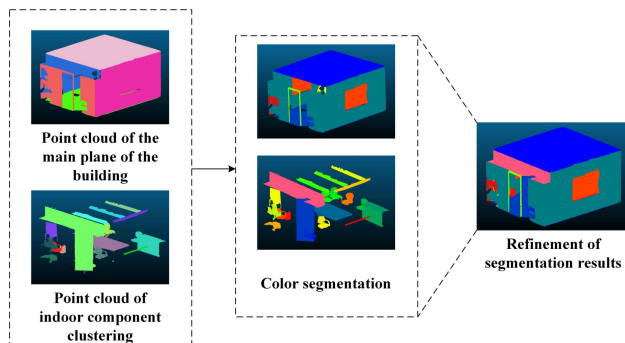


Figure 5. Color region based growth segmentation process.

The color-based region growing algorithm flow is as follows.

(1) Pick a random point from each point cloud and set it as the seed point n in the next step.

(2) For each seed point, query K nearest neighbors.

(3) Cluster A is formed by merging nearby points of similar color. Re-select any point except A in the cluster and repeat the above operation until all of the points are included in the point cloud cluster.

(4) Combining clusters with similar colors, and merging two neighboring clusters whose average color variability is lower, will form clusters with similar colors.

(5) The total number of point clouds for each cluster should be verified, and if the total number of point clouds is less than the threshold K , then merge the current cluster with its nearest neighbor cluster.

2.3 Semantic label optimization based on statistical information

In this paper, point clouds with semantic information can be obtained by using the deep learning semantic segmentation method, and geometric segmentation point clouds without semantic information can be obtained by using the geometric and color feature segmentation method. It consists of using statistical information to correlate two segmentation results, and then by using a probabilistic model to cross-fuse the correlation results, fine segmentation results can be obtained with semantic information. A specific process involved in this process is that,

first of all, each point P_{si} in the geometric segmentation result point clouds P_s is used as a query points, and the deep learning segmentation result point clouds P_d is used as a search point, which means that the semantic labels of the nearest points in P_d to P_{si} can be obtained by using the K -nearest neighbor query algorithm to obtain the corresponding semantic label set L . Secondly, the statistical histogram of semantic labels in each cluster in P_s is counted, and the percentage of each type of label is calculated as Equation 6 shows.

$$N_i = \frac{L_i}{L_1 + L_2 + \dots + L_k} \quad (6)$$

L_i represents the total number of labels per category in L . K represents the number of categories contained in L , and N_i is the calculated percentage of each label. The point clouds category with the largest label share L_{max} is calculated, as shown in Equation 7.

$$L_{max} = MAX\{N_1, N_2, \dots, N_k\} \quad (7)$$

It is determined based on the percentage value of semantic labels in each cluster. The tags with $N_i < K$ (K is the ratio threshold) are reclassified as the tags with a maximum ratio L_{max} of the tags with K , as shown in Figure 6. The final semantic coarse segmentation results are optimized and the high precision semantic segmentation results are obtained by optimizing the final semantic coarse segmentation results.

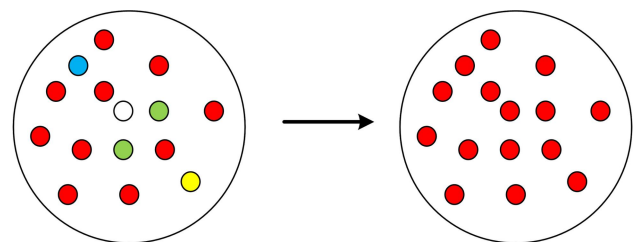


Figure 6. Schematic diagram of label resegmentation .

3. EXPERIMENT AND ANALYSIS

3.1 Experimental data and accuracy evaluation

S3DIS is an indoor 3D spatial dataset collected by Stanford University using a patternport scanner in 2016. It contains 6 large-scale 3D spatial indoor scenes covering a total of about 6000 m² that are labelled with 13 common semantic categories of indoor (walls, tables, ceilings, chairs, etc.) that can be directly used for semantic segmentation of point clouds. Three indoor point clouds from indoor scenes are selected for experimentation in order to verify the effectiveness of the proposed method, as shown in Figure 7, which include an office, a corridor, and a conference room. Scenes range from simple to complex indoor scenes, which can be used to validate the effectiveness and accuracy of the segmentation method. Room_1 is a simple scene, containing the basic building body plane; Room_2 is a fairly complex scene, containing cabinets, tables, chairs, as well as other indoor parts of the building in addition to the body plane of the building, all of which make up the conference room; Compared to Room_2, Room_3 is a more complex office scene, with a messy scene, and stacked items, as well as some serious adhesion problems.

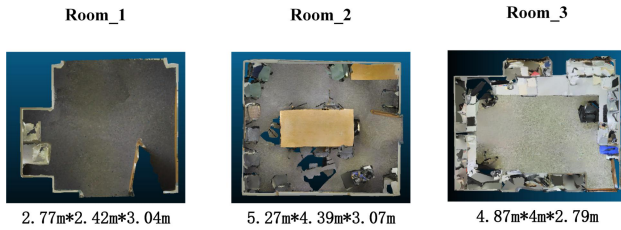


Figure 7. Experimental data set.

We use the intersection and merging ratio (Iou) in conjunction with the accuracy (Acc) in this paper to assess the accuracy of point cloud segmentation based on intersection and merging ratios. In this case, iou represents the ratio of the intersection and merging of the true value FN of the segmentation of the data and the predicted result FP as a result of it, as shown in Equation 8.

$$Iou = \frac{FN \cap FP}{FN \cup FP} \quad (8)$$

The ratio of TP(FN ∩ FP) to FP is called Acc, as shown in Equation 9.

$$Acc = \frac{TP}{FP} \quad (9)$$

3.2 Experimental results and accuracy analysis

In the context of three groups of indoor scene point clouds with different levels of complexity, the preliminary deep learning segmentation point clouds were obtained by RandLA-Net, the refined geometric segmentation point clouds were obtained using geometric and color features, and finally both deep learning segmentation point clouds and geometric segmentation point clouds were cross-fused to achieve the overall optimization of semantic segmentation point clouds. The specific experimental results for the three groups of scenes are presented in Figure 8, 9 and 10. Room_1 is clearly optimized for misclassified walls, while room_2 is also optimized to varying degrees for misclassified walls, slabs, and beams. There are also different degrees of optimization for the interior

component point clouds in Room_3. The real labels represent the semantic information of the real parts in each room, the deep learning segmentation labels represent the preliminary segmentation results of the RandLA-Net neural network, and the optimization results represent the combined optimization results of geometric features and neural networks.

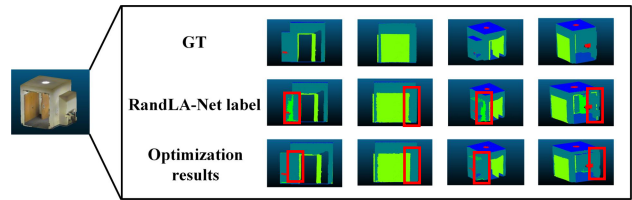


Figure 8. Results of point clouds segmentation in room_1.

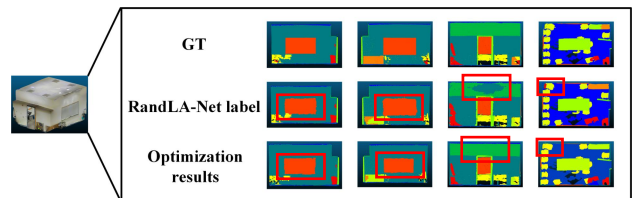


Figure 9. Results of point clouds segmentation in room_2.

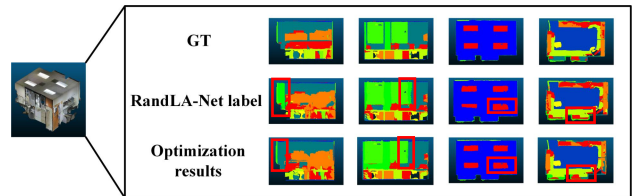


Figure 10. Results of point clouds segmentation in room_3.

Based on Table 1, this table represents the iou of the RandLA-Net method and our optimization method on 13 categories. Using the optimization method proposed in this paper, the semantic labels of the three sets of data can be corrected. There is a 2%-10% improvement in accuracy for walls, doors, and chairs. There is a 10% improvement in segmentation accuracy for the categories with poor segmentation, and a 5% loss of accuracy for the categories with higher segmentation accuracy, such as floor and ceiling.

		ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board	clut.
Room_1	RandLA-Net	0.980	0.967	0.920	0.769	/	/	0.935	0.690	0.826	/	0.301	0.931	0.771
	Our methods	0.980	0.971	0.936	0.941	/	/	0.954	0.793	0.839	/	0.598	0.872	0.707
Room_2	RandLA-Net	0.842	0.973	0.881	/	/	/	0.946	/	/	/	/	/	0.856
	Our methods	0.837	0.972	0.945	/	/	/	0.964	/	/	/	/	/	0.830
Room_3	RandLA-Net	0.962	0.984	0.592	/	0.590	0.597	0.613	0.616	0.891	/	0.670	/	0.593
	Our methods	0.967	0.969	0.642	/	0.692	0.595	0.668	0.613	0.926	/	0.664	/	0.589

Table 1. Evaluation of the accuracy of the segmentation results of deep learning and the segmentation results.

Using our optimization method, we will achieve a stable improvement in overall accuracy for three sets of indoor scene data of varying complexity, with MIOU improvement of around 1.9% and mAcc improvement of around 1.6%.

		mIou	mAcc
Room_1	RandLA-Net	0.853	0.946
	Our methods	0.891	0.953
Room_2	RandLA-Net	0.807	0.922
	Our methods	0.811	0.954

Room_3	RandLA-Net	0.700	0.830
	Our methods	0.717	0.839

Table 2. Overall accuracy evaluation of each component.

4. CONCLUSION

Based on the optimization of geometric features and neural networks, a semantic segmentation method is proposed for indoor 3D point clouds. First, the method obtains a preliminary semantic labeling of the indoor components using deep learning segmentation, and then geometric and color features are utilized to segment the entire scene. By combining the results of the

former and the latter, it is possible to improve segmentation accuracy by combining the two results.

An experiment is conducted on the S3DIS using three different sets of scenes with varying levels of complexity. We evaluate the effectiveness and accuracy of the proposed method and compare the experimental results with those of the original deep learning methods. Experimental results demonstrate that the proposed method can take full advantage of deep learning and feature segmentation to improve the accuracy of semantic segmentation. Segmentation accuracy is significantly improved over the original deep learning method.

However, the method proposed in this paper has some limitations, including inaccurate clustering in complex structures and inconsistent parameters across scenes. As of now, our method can only be applied to indoor scenes where the room configuration is square. We will further investigate the adaptive parameter setting during scene segmentation in our future work and further enhance its robustness.

ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China (Projects Nos. 2019YFB210310, 2019YFB2103104) and in part by a Research Program of Shenzhen S and T Innovation Committee grant (Projects Nos. JCYJ20210324093012033, JCYJ20210324093600002), the Natural Science Foundation of Guangdong Province grant (Projects No. 2121A1515012574), the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR(No. KF-2021-06-125), and the National Natural Science Foundation of China grant (Projects Nos. 71901147, 41901329, 41971354, 41971341).

REFERENCES

Bao-Shun, L.I., Cen, H.Y., Bao, Y.P. and Yi-Feng, L.I., 2013. Edge detection algorithm for point cloud based on rolling-circle. *Computer Engineering and Design*.

Bello, S.A., Yu, S., Wang, C., Adam, J.M. and Li, J., 2020. Review: Deep Learning on 3D Point Clouds. *Remote Sensing*, 12(11): 1729.

Biosca, J.M. and Lerma, J.L., 2008. Unsupervised robust planar segmentation of terrestrial laser scanner point clouds based on fuzzy clustering methods. *Isprs Journal of Photogrammetry & Remote Sensing*, 63(1): 84-98.

Choi, J., Choi, J. and Kim, I., 2014. Development of BIM-based evacuation regulation checking system for high-rise and complex buildings. *Automation in Construction*.

Feng, Y., Zhang, Z., Zhao, X., Ji, R. and Yue, G., 2018. GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Hu, Q., Yang, B., Xie, L., Rosa, S. and Markham, A., 2020. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Kang, Z., Yang, J., Yang, Z. and Cheng, S., 2020. A Review of Techniques for 3D Reconstruction of Indoor Environments.

International Journal of Geo-Information, 9(5): 1-31.

Pomerleau, F., Colas, F. and Siegwart, R., 2015. A Review of Point Cloud Registration Algorithms for Mobile Robotics. *Foundations & Trends in Robotics*, 4(1): 1-104.

Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *IEEE*.

Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.

Taira, H., Okutomi, M., Sattler, T., Cimpoi, M. and Torii, A., 2018. InLoc: Indoor Visual Localization with Dense Matching and View Synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Tarsha-Kurdi, F., 2007. Hough-Transform and Extended RANSAC Algorithms for Automatic Detection of 3D Building Roof Planes from Lidar Data, ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007.

Wang, F., Zhuang, Y., Gu, H. and Hu, H., 2020. OctreeNet: A Novel Sparse 3-D Convolutional Neural Network for Real-Time 3-D Outdoor Scene Analysis. *IEEE Transactions on Automation Science and Engineering*, 17(2): 735-747.

Wu, B., Wan, A., Yue, X. and Keutzer, K., 2017. SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud.

Xu, W., Kang, Z. and Jiang, T., 2009. A Segmentation Approach for Terrestrial Point Clouds based on the Integration of Graph Theory and Region Growing, *IEEE*.

Zhan, Q., Yu, L. and Liang, Y., 2011a. A point cloud segmentation method based on vector estimation and color clustering, International Conference on Information Science & Engineering.