INDOOR 3D POINT CLOUDS SEMANTIC SEGMENTATION BASES ON MODIFIED POINTNET NETWORK

J. Zhao[1,2,3],X. Zhang[1,3], Y. Wang[1,3,*]

[1] School of Geomatics and Urban Information, Beijing University of Civil Engineering and Architecture, 102616 Beijing ,China - zhaojiangh@bucea.edu.cn
[2] Key laboratory of Modern Urban Surveying and Mapping. National Administration of Surveying , Mapping and Geoinformation, 102616 Beijing, China
[3] Beijing Key Laboratory For Architectural Heritage Fine Reconstruction & Health Monitoring, 102616 Beijing, China-(zhangxiaoguang,2108521519005)@stu.bucea.edu.cn

KEY WORDS: 3D LiDAR Point Cloud; Point Cloud Segmentation; Semantic Segmentation; Deep Learning; Indoor Structural Elements; PointNet

ABSTRACT:

Indoor 3D point clouds semantics segmentation is one of the key technologies of constructing 3D indoor models,which play an important role on domains like indoor navigation and positioning,intelligent city, intelligent robot etc. The deep-learning-based methods for point cloud segmentation take on higher degree of automation and intelligence. PointNet,the first deep neural network which manipulate point cloud directly, mainly extracts the global features but lacks of learning and extracting local features,which causes the poor ability of segmenting the local details of architecture and affects the precision of structural elements segmentation . Focusing on the problems above,this paper put forward an automatic end-to-end segmentation method base on the modified PointNet. According to the characteristic that the intensity of different indoor structural elements differ a lot, we input the point cloud information of 3D coordinate, color and intensity into the feature space of points. Also,a MaxPooling is added into the original PointNet network to improve the ability of attracting and learning local features. In addition, replace the 1×1 convolution kernel of original PointNet with 3×3 convolution kernel in the process of attracting features to improve the segmentation precision of indoor point cloud. The result shows that this method improves the automation and precision of indoor point cloud segmentation for the precision achieves over 80% to segment the structural elements like wall,door and so on ,and the average segmentation precision of every structural elements achieves 66%.

1. INTRODUCTION

1.1 General Instructions

3D LiDAR technology has gradually became a important method of understanding the 3D indoor scene because of this advantages that it can acquire massive point cloud data with high speed,low cost and high precision. However, the semantic segmentation has became a hot spot of research fields like: 3D indoor modeling, indoor navigation, robot pattern etc. The traditional segmentation methods for point cloud has developed for a period of time. So there are amounts of classic segmentation algorithms, such as: segmentation methods base on edge(Himmelsbach et al.,2009), segmentation methods base on surface (Li et al., 2011; Zhang et al., 2015; Hu et al., 2012) segmentation methods base on clustering (Chen et al., 2012; Sun et al., 2006; Lin et al., 2016), segmentation methods base on machine learning (Rusu et al., 2008; Rusu et al., 2009; Aldoma et al., 2011). With the improvement form lots of researchers, the traditional segmentation methods for point cloud has been enhanced constantly. However the traditional segmentation methods for point cloud require manually designed feature descriptors, which demand the designers possess empiric knowledge. There are lots of thresholds needed to be selected for traditional segmentation methods during the process of point cloud segmentation which is complicated, only suitable for specified tasks and poor in generalization. To enhance the automation and intelligence of point cloud segmentation, the segmentation methods for point cloud base on deep learning has became a latest research hot spot. Deep learning is a kind of novel technology that automatically extract the high level features of the input data by the structure of deep neural network. Currently,the segmentation methods base on deep learning are mainly divided into 3 types: ①Convert point cloud into multi-view images then input the images into 2DCNN to realize the segmentation of 3D elements. (Su et al., 2015) put forward MVCNN network, which utilized 2DCNN network structure. However,converting point cloud to image will lose the spatial information of point cloud, which will affect the precision of segmentation; ②The neural networks that use voxel as input. Daniel and Sebastian (2015) put forward VoxNet network model base on point cloud voxelization and supervised 3DCNN. This network preprocess the point cloud into voxel, then use 3D convolutional kernel to carry out convolution operation, which is the original 3DCNN network but along with the disadvantages, such as: additional computation,"dimension explosion" etc. ③The deep neural network that directly use points as input. Qi et al. (2016), from Stanford, put forward PointNet network, which utilize multi-layer perception (MLP) to extract the global feature of point cloud and use maximum symmetric function to solve the problem of irregular format to achieve a good segmentation precision. However this network only pay attention to the global

* Corresponding author

feature and ignore the local feature. So this network has poor capacity of details segmentation.

This paper used the modified PointNet to segment the point cloud of indoor structural elements. The main works are as follow:

(1)Currently, there are few indoor point cloud dataset which contain intensity information. So this paper constructed a indoor point cloud data set contains 8 types of indoor structural elements with intensity information for the experiment;

(2)Focusing on the indoor structural elements (door, wall, window etc.) with different intensity information, this paper input intensity information,coordinate information and color information into the neural network as tensor to improve the segmentation precision of PointNet to segment the indoor structural elements.

(3)Focusing the problem that the original PointNet only pay attention to extract the global features of point cloud but ignore local features, this paper modified the structure of PointNet network to let it has better ability to extract local features and improve the segmentation precision of structural elements.

2. CONSTRUCTING THE INDOOR POINT CLOUD DATA SET

Currently,there are only 2 public large-scale indoor 3D point cloud data set which are showed in table 1. But these 2 public data set contain no reflection intensity information. Hence,we created a data set contains 4 areas, 8 semantic elements and 70,000,000 labeled points for better segmentation result of indoor structural elements. The point cloud of this data set contains not only spatial coordinate information (X,Y,Z), color information (R,G,B), but also reflection intensity information. We mainly utilized Faro Focus3D X130 scanner to acquire point cloud and the technical index of it are showed in Table 2. After acquiring the original point cloud, preprocessed the original point cloud like: registering, denoising, getting sparse. After prepocessing, labeled the point cloud manually and divided the data set into training set and testing set which is showed in Figure 1 .

Data Set	Year	Format	Feature of Point Cloud	Producer
S3DIS (Armeni et al., 2017)	2017	Point Cloud	(X,Y,Z,R,G,B)	Stanford RGB-D
Scannet (Dai et al., 2017)	2017	Point Cloud	(X,Y,Z,R,G,B)	Stanford RGB-D

Table 1. Current open large indoor point cloud datasets

Configuration	Faro Focus3D X130
Ranging/m	0.5～130
Distance Accuracy Index	0.6mm/10m
Scanning View	360°×120°
Scanning precision	0.1mm/50m
Speed	1,200,000points/s

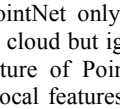

Table 2. Faro Focus 3D X130 scanner and main technical indicators

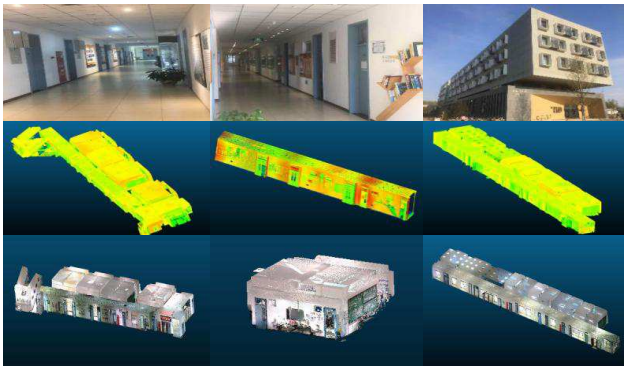

Figure 1. Dataset main scenarios
（Mainly divided into two parts :data set with reflection intensity information and data set without reflection intensity information ）

3. METHOD

3.1 Structure of PointNet segmentation network

The structure of PointNet network is showed in Figure 2. It utilized transformation network T-net for rigid transformation and use maximum symmetric function to handle the problem of disorder. The algorithm can be divided into 3 flowing steps:

(1)Utilize T-Net to learn a transformation matrix and then multiply the matrix with point cloud to make sure that the rigid transformation steadiness of point cloud.

(2)Utilize MLP to extract high dimensional feature from the point cloud which has been transformed by T-Net and use maximum symmetric to manipulate the disordered point cloud to extract global feature.

(3)Fuse the global and local features and then utilize MLP to downsample the features for the probability score to measure what type the points might belong to.

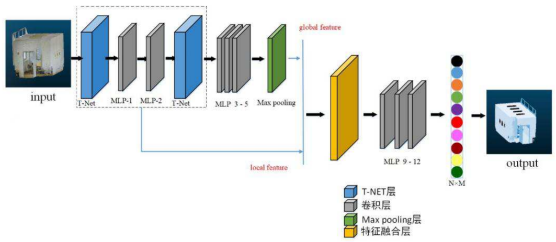

Figure 2. Structure of PointNet

3.2 Modify PointNet

We modified the original PointNet network. First, add reflection intensity information of the point cloud into the feature space. Then we add a layer to extract local feature and then fuse the global feature and local feature. The modified network structure is showed in Figure 3.

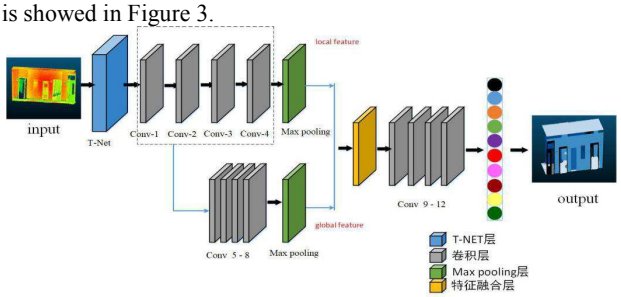

Figure 3. The Structure of Modified PointNet

3.2.1 The input Feature space

Because of the difference of distance between points and the sparse structure of point cloud, the point cloud need to be converted to the format that can be understand by neural network. The 3D coordinate （X,Y,Z） represents the spatial information and the color (R,G,B) represents the texture information. Also, the coordinates of original point cloud are normalized before input. The point cloud are divided according to the room they belongs to and then normalized base on the coordinate of the room. After acquiring the coordinate (X^0,Y^0,Z^0), every room are divided into the area of 1m×1m. In addition, the reflection intensity information is added, because different structural elements contain different information of reflection intensity. Finally,every points in the point cloud are input as $[X, Y, Z, R, G, B, X^0, Y^0, Z^0]$, the feature space of 10 dimension, into the neural network.

3.2.2 Local feature extraction

During the process of optimizing the original PointNet network, we found that adding MLP with 4 layers on the original MLP helps to improve the precision of segmentation. Hence,we add 2 more layers of MLP behind the MLP-2 layer of the original network as the extracting layer for local features. Also,Adding a MLP layer behind the original MLP-5 and parallel with the extracting layer for local features as the extracting layer for global features.

Compared with the original PointNet, our structure use more MLP layers to extracting the local feature, which have better ability to extract the correlation features between points. When optimizing the index of MLP layer network used to extract the high dimensional features, we found that the convolutional kernel in bigger size will get better segmentation precision during the semantic segmentation of indoor scene. So,we replace some of the 1×1 convolutional kernel of original PointNet network with 3×3 convolutional kernel to improve the segmentation precision of the network to segment the indoor structural elements. In addition, we add a MaxPooling layer behind the MLP-4 layer to extract the local feature for better ability to input the local feature of point cloud. Besides, let every point contain not only local feature but also global feature by fusing the local feature extracted by the first MaxPooling layer and the global feature extracted by the second MaxPooling layer. Utilizing the Concat operation of tensorflow framework to achieve the fusion of features. The algorithm of the Concat operation is showed in (1).

$$\begin{aligned} pool1 &= tf.\max\_pool2d(conv4,[4096,1], padding='VALID') \\ pool2 &= tf.\max\_pool2d(conv8,[4096,1], padding='VALID') \\ concat1 &= tf.concat(axis=3, values=[pool1, pool2]) \end{aligned} \quad (1)$$

4. RESULT ANALYSIS

After modifying the PointNet network, we utilized this network to segment the S3DIS data set and the indoor data set we created. We Selected the Area1, Area2, Area3, Area4, Area6 as training set, and chose Area5 as testing set. Besides, we chose Area1, Area2, Area3 as training set and Area4 as testing set. To validate the effectiveness of the modified PointNet, we compared the segmentation result of using the modified network and original PointNet on S3DIS data set. We also compare the result of using the data set we constructed with reflection intensity and without reflection intensity. In addition, compare the result of using different kernel size of 3×3 and 1×1. We used Tensorflow deep learning framework, adam optimizer to train the network. The learning rate is 0.001,batch siez is 16, epoch is 100 and the training consumed 21hours. The hardware and software we used to train the network are showed in Table3 and Table 4.

	CPU	GPU	RAM
Hardware	i7-6700	Nvidia GTX1060(6G)	16G

Table3.Hardware of the experimental platform

	Operating system	Deep learning framework	GPU accelerator	Programme language
Software	Ubuntu 16.04	Tensorflow 1.01	CUDA 8.0 CuDNN 5.1	Python 3.5

Table 4. Software of the experimental platform

4.1 Result

The segmentation results of using modified PointNet network to segment S3DIS data set are showed in Fig 4, and the results of segmenting the data set we constructed are showed in Figure5. The upper half part of the result figure shows input data, and the lower half part shows the segmentation result. The data set we constructed consist with the part contains reflection intensity and the part without that and Figure 6.shows the comparing result. The upper half part shows the result of using the intensity information, and the lower half part shows the result without using intensity information. The objects in the red pane are important objects needed to be compared Figure 6. shows that the result of upper half part is better than the lower one.

Figure 4. Segmentation Results of Area5 in S3DIS Dataset

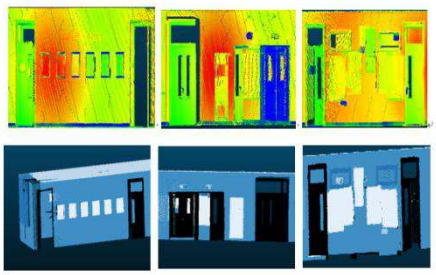

Figure 5. Segmentation results of indoor dataset

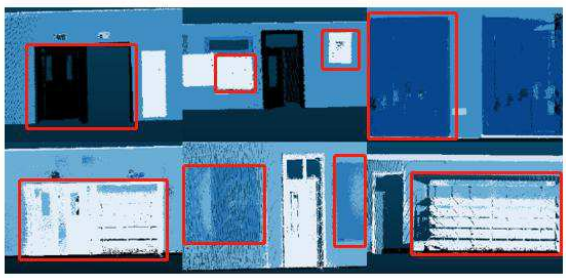

Figure 6. Comparison of Reflective including intensity and without intensity

4.2 Precision norms

The Intersection over (IoU), mean Intersection over Union (mIoU), overall accuracy (OA) are frequently-used norms in

Network	mIoU	ceiling	floor	window	door	board	lamp	Publicity cabinet	others
PointNet (with I)	65.46	92.06	95.89	79.29	82.53	40.06	25.36	62.42	50.09
PointNet (without I)	64.97	91.27	94.76	77.21	81.39	38.52	24.79	61.63	50.16
Ours（with I）	**68.24**	**92.93**	**96.23**	**80.68**	**85.37**	**42.63**	**26.43**	**63.56**	**50.58**
Ours (without I)	66.78	92.13	95.53	78.36	83.12	41.73	25.63	62.26	50.45

Table 7. The Segmentation Accuracy of our Dataset

both domestic and foreign countries to evaluate segmentation precision of every semantic elements. Therefore, this paper use IoU, mIoU and OA as norms to evaluate the precision. IoU is the ratio of the area of segmentation result to the area of ground truth, mIoU is the average of IoU, OA is the ratio of the number of the points accurately segmented to the number of total points. The algorithm of IoU is showed in (2),TP means true positive, FP means false positive,FN means false negative. The algorithm of mIoU is showed in (3), the algorithm of OA is showed in (4), 0 to K means there are K+1 types, P represent every point and P_{ii} represents the number of the points are correctly segmented, P_{ij} represents the points belong to type i but segmented as type j, P_{ji} represents the points belong to type j but segmented as type i.

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k} IoU_i \tag{3}$$

$$OA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \tag{4}$$

4.3 Result analysis

Figure 5. shows the mIoU of every categories of using modified PointNet algorithm and original PoingNet algorithm to segment the Area5 in S3DIS. We found that the modified network got higher precision than PointNet except chair, which showed in Table 5. Consequently, the modification for PointNet of this paper is effective.

Network	**PointNet**（IoU）	**Ours**（IoU）
mIoU	53.87	**56.84**
ceiling	90.80	**91.09**
floor	95.33	**96.13**
wall	70.56	**73.35**
beam	10.25	**0.15**
columu	13.92	**26.51**
window	66.26	**67.04**
door	78.62	80.36
table	55.32	62.86
chair	54.75	54.53
sofa	15.32	29.26
bookcase	50.76	51.14
board	45.52	52.35
clutter	52.86	54.17

Table 5. IoU of Area5 in S3DIS Dataset

Size of convolutiona l kernel in PointNet	MIoU(%)	OA(%)
1×1（original PointNet）	54.12%	80.59
3×3	**54.96**%	**81.03**

Table 6. Segmentation Results of PointNet with Different Convolution Kernels on S3DIS

Table 7. shows the comparison of the results of using both modified network and original PointNet to segment the data set with reflection intensity and without intensity. Table 6. shows that adding reflection intensity information into the point cloud helps to improve the segmentation precision of segmenting door, window, board and so on no matter we use the modified network or the original PointNet. However,the improvement is not obvious for ceiling, floor and others. The segmentation results of using modified network and the data set we constructed are better than using original PointNet on segmenting ceiling, floor, window, door and so on. However the segmentation results of segmenting board and lamp are not good enough, which the precision only achieve 45%. According to the analysis, we think there are two main reasons: ①The ability of the modified network base on the PointNet is weak to learn the local details of wall and board, so the segmentation results are poor. ②The number of different semantic elements are not well-distributed, so the segmentation precision of the elements which contain small quantity, like lamp, will be lower. Hence, the following research could follow these two following aspects: (1)Keep on improving the structure of network for a better ability to learn the local features of point cloud and improve the segmentation precision. (2)Enlarge the size of data set and make sure the uniformity of the number of every categories.

5. CONCLUSION

Focusing on the problem that PointNet network has poor ability to extract local feature of point cloud,this paper put forward a modified end-to-end deep neural network base on PointNet for semantic segmentation of indoor 3D LiDAR point cloud. Base on the structure of PointNet, we replac the original 1×1 convolutional kernel to 3×3 convolutional kernel. And add several layers of MLP and MaxPooling layer to improve the ability of extracting local feature. Also add a layer to fuse features for the segmentation task of indoor structural elements. At last,we compare the results of using S3DIS and the data set we constructed and the original PointNet.
The modified network base on PointNet acquires higher precision compared with the original PointNet according to the result, which proves that the modification are effective.

REFERENCES

HIMMELSBACH M, Luettel T, Wuensche H., 2009. Real-time object classification in 3D point clouds using point feature histograms. *International Conference on IEEE*: 994-1000.

LI Na, MA Yi-wei, YANG Yang et al., 2011. Segmentation of Building Facade Point Clouds using RANSAC. *Science of Surveying and Mapping*, 36(5):144-145

ZHANG Fan, GAO Yunlong, HUANG Xianfeng et al.,2015. Spherical Projection Based Straight Line Segment Extraction for Single Station Terrestrial Laster Point Cloud. *Acta Geodaetica et Cartographica Sinica*, 44 (6):655-662.

HU Wei, LU Xiaoping, LI Cheng et al.,2012. Extended RANSAC Algorithm for Building Roof Segmentation from Li DAR Data. *Bulletin of Surveying and Mapping*,11:31-34

CHEN Xiangyang, YANG Yang, XIANG Yunfei, 2012. Measurement of Point Cloud Data Segmentation Based on Euclidean Clustering Algorithm. *Bulletin of Surveying and Mapping*, (11): 31-34+39.

SUN Hongyan, SUN Xiaopeng, LI Hua, 2006. 3D Point Cloud Model Segmentation Based on K-means Cluser Analysis. *Computer Engineering and Applications*, (10): 42-45

LIN Xiangguo, NING Xiaogang, XIA Shaobo, 2016. A method For powerline LiDAR point cloud segmentation using K-meaning clustering of a feature space. *Science of Surveying and Mapping,* 41(5):60-63

RUSU R B, Blodow N, Marton Z C et al., 2008. Aligning point clouds views using persistent feature histograms. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3384-3391.

RUSU R B, Blodow N, Beetz M., 2009. Fast point feature histograms (FPFH) for 3D registration. *IEEE International Conference on Robotics & Automation*.

ALDOMA A, Vincze M, Blodow N et al., 2011. CAD model recognition and 6 DOF pose estimation using 3D cues. IEEE International Conference on Computer Vision Workshops, ICCV 2011 Workshops, Barcelona, Spain, November:6-13.

Su H, Maji S, Kalogerakis E et al., 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. *IEEE International Conference on Computer Vision*, 945-953.

Daniel M and Sebastian S., 2015. VoxNet:A 3D convolutional neural network for real-time object recognition. Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, Congress Center Hamburg, Sept 28-Oct 2. Hamburg, Germany, 922-928.

C.R.Qi, H.Su, K.Mo et al., 2016. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Computer Vision and pattern Recognition*, 217-223.

Armeni I, Sax S, Zamir A R et al.,2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *Computer Vision and pattern Recognition*, 113-118.

Dai A, Chang A X, Savva M et al., 2017. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes. *Computer Vision and pattern Recognition*, 261-268

APPENDIX

Foundation Support: the National Natural Science Foundation of China,No.41601409; National Natural Science Foundation of China,No.41501495