Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

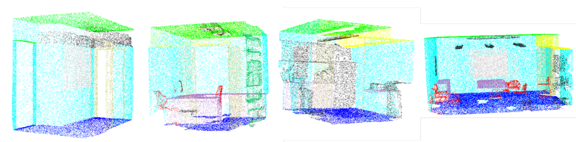
Random Walk Network for 3D Point Cloud Classification and Segmentation

Lei Zhu, Weinan Chen, Xubin Lin, Li He, Yisheng Guan*, Hong Zhang

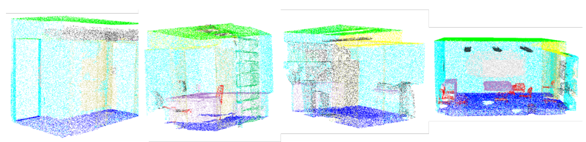
Biomimetic and Intelligent Robotics Lab (BIRL)
School of Electromechanical Engineering
Guangdong University of Technology
Guangzhou, China

***Abstract*— Object classification and segmentation via point cloud are essential for mobile robot navigation and operation. A lot of researches ranging from 3D voxels, mesh gird and multi-view were proposed based on point cloud. However, an accurate point cloud classification is still a challenging problem. Inspired by multi-label classification in images and convolutional neural networks (CNN), in this paper we present a novel network, named Random Walk Network (RWNet), which directly processes raw 3D point cloud data to classify points and, as a result, segment one 3D scene. State-of-the-art methods mainly focus on the features of one point while spatial relationships are also essential in point classification. To deal with this issue, we combine both appearance features and spatial information of feature points to restrain the point cloud processing. We employ PointNet first to generate initial point labels and adopt point labels with high confidence as seeds. We then construct the similarity matrix between seeds and low-confidence-labeled points according to their structural and spatial similarity and use Random Walk to obtain the final classification. We demonstrate our method in 3D classification task in various scenes and compare with some benchmark methods. Experimental results show that RWNet has a better performance than others.**

***Index Terms*— Point Cloud, Random Walk, Classification, Semantic Segmentation**



(a)



(b)

Fig. 1: Performance of RWNet. RWNet takes raw point cloud as input and outputs semantic segmentation. Subfigures show 3D scenes including hallway, office, pantry and conference room. (a) the ground truth of semantic segmentation. (b) the predicted labels of RWNet. Different colors represent different classes.

I. INTRODUCTION

Point cloud is widely used in mobile robot or auto-driving car and achieves more and more attention in last decade [1]. However, classification via point cloud is still a challenging task partially due to its high computation cost and lack of commonly used expression [2]. 2D semantic segmentation with CNN has shown good experimental results in [3, 4]. Unlike the well-studied 2D image deep learning, raw data from point cloud are more complex in nature: (a) point cloud is unstructured and hence not friendly for CNN filter, and (b) one point cloud may be saved in an unordered fashion.

Several challenging problems, such as missing data and sensor noise, disturb feature point expression of the local contexts. Many researchers adopt voxels [5-7], mesh grid [8, 9] and multi-view [7, 10] to form 3D structures. Works in [11] put forward a milestone contribution that pushes a point cloud as the input of a CNN. This network extracts point features directly from the input point cloud and shows its advantages in many scenes. Despite its superiority in 3D classification, the accuracy of works in [11] is not high enough for mobile robot navigation or recognition. To improve the accuracy of 3D point cloud classification, in this paper, we propose a novel network using both position and appearance features to process 3D point data, as depicted in Fig. 1. Our main idea is to employ PointNet first to obtain point labels with high confidence, and use Random Walk to generate labels on points with low confidence. We call this Random Walk Network (RWNet).

The input of our method is a raw point cloud and the output is the semantic label classified for each point. We first cut the 3D scene into several sub-clouds and adopt PointNet [11] to extract point features in each sub-cloud and extract point labels from PointNet. When the PointNet-

* Corresponding author, ysguan@gdut.edu.cn. W. Chen is also with the Department of Electrical and Electronics, Chinese University of Hong Kong. H. Zhang is also with the Department of Computing Science, University of Alberta, Canada.

The work in this paper is partially supported by the Natural Science Foundation of China (Grant No. 61673125, 61703115), the Frontier and Key Technology Innovation Special Funds of Guangdong Province (Grant No. 2016B090910003), the Program of Foshan Innovation Team of Science and Technology (Grant No. 2015IT100072), and the Team of Modern Agricultural Industrial Technology System in Guangdong.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

predicted labels are confused or with a low predication probability, we adopt spatial relationship between points for further classification. Enlightened by [12] in 2D image labeling task, we feed labels of feature points with high confidence as seeds of random walk and run Random Walk (RW) to predict the labels of the low confidence points. For those feature points with low confidence, we employ both feature expression and Euclidean distance to its nearest high confident points to build a similarity matrix for operation. We compare our method with PointNet as well as previous works based on PointNet. RWNet shows better performance than other methods on S3DIS [13] , ModelNet40 [14] and ShapeNet [15] datasets.

The contributions of our work are that:

- We propose an efficient way to select high confidence points.
- We use both appearance and spatial features to construct the similarity matrix for RW.
- We obtain higher accuracy than the state-of-the-art.

The rest of this paper is organized as follows. Section II gives the reviews of related works in 3D point cloud processing. Our method is described in Section III. Section IV shows the experimental results on several data sets. We conclude our works in Section V.

II. RELATED WORK

Promoted by deep learning, there sprung up many networks dealing with classification and segmentation tasks. CNN possesses the ability of feature extraction, which avoid the boundedness of handcrafted features and promote the research in 2D image segmentation. However, for 3D point clouds, data increases rapidly and applicable sensors rises, many researches and applications extend to 3D area [16-18].

3D Voxels Considering pixels in 2D image, it is common to think of transiting ideals and methods in image processing to 3D point cloud analysis. Volume pixel is the smallest unit in 3D space base, which describes the resolution ratio and could be exploited to establish 3D rendering and extract the profile of 3D object. In early days, reachers make attempts to analysis end-to-end algorithms for 3D data of an object or environment in [5-7]. [6, 14, 19] give us examples that they divide point cloud with 3D volume pixels with handcrafted features and put them into end-to-end deep learning network. Compared with 2D image pixels, voxels possess one more dimensionality, which makes the network being complex in both time and space. Input in this condition will also be in low resolution ratio and lack of local information. All this factors influence volume pixels unsuitable for classification and segmentation tasks in big scenes.

Mesh Grid To simplify 3D models and make up cavity, mesh is a good choice whose unstructured grid was formed by a set of convex polygons and vertexes. This kind of methods is based on functional view and focus on local information connection of point clouds [8, 20-23]. These methods represent graph structure via neighborhood between point sets. Transition from raw point cloud data to polygonal mesh is helpful for 3D rendering and visualization. [21, 22] optimized non-Euclidean manifolds and minimized cost of specific task as well as the dimension of descriptors to learn the unchanged shape feature for description and verified in body and object models. Mesh grid is hard to catch up with the speed of 3D point cloud collection equipments to manage so much data and simplify the classification algorithm. In addition, variant of base model and loss of functional part will make mesh grid unreliable in classification and segmentation tasks.

Multi-View Acquired from an object model by virtual cameras, multi-view image processing collects 2D images to build a complete 3D model. In [10, 24], the authors classify point cloud in multi-view systems on the basis of graph cut. [10] estimates object and background according to the distance between the projection and image and the pre-weighted graph is used for judging the object from background. [25] transforms 3D shapes into panoramic picture and learns the depth information with CNN. Methods based on multi-view apply traditional CNN to multiple images in 2D viewpoints, features are gathered to form a 3D object. This expression of 3D data needs many images to structure a whole 3D model and it will lose efficiency because of self occlusion.

All these methods above have their pros and cons. As the 3D sensors develops, 3D point cloud models are getting more and more securable and its classification tasks is of great importance. In this work we focus on raw point cloud data and its classification and semantic segmentation.

III. RANDOM WALK NETWORK

We introduce our random walk network method for 3D scene classification and semantic segmentation. The input of our method requires 3D point clouds with x, y, z coordinates value and the output are classified labels of each feature point and semantic segmentation of the input 3D scene.

Diagram of our proposed network is shown in Fig. 2. We firstly deal with the feature extraction and evaluate their confidence of different labels. According to the evaluation, we add spatial connection to those feature points with low confidence and ensure the label class for RW prediction. At last, we merge the labels to form the semantic segmentation of 3D scene. Detailed description is presented as follows.

A. Feature Evaluation

RWNet first feeds the raw point cloud to a feature extraction network. This network outputs a feature vector F_i of each point with length c, where i represents the i_{th} point among all n points and c is the number of class labels. Each element in F_i is actually the class probability predicted by PointNet and we hence use this value to determine the confidence of the predicted label. Seeing statistics of feature vectors, especially the incorrectly predicted ones, we find out that when it comes to some of scores in little difference, the network is actually confused but still chooses the largest one corresponded class as the label of class. As Fig. 3 shows, in correctly predicted point classes, the peak value of feature vector F_i is much larger than the second one but in incorrect prediction, the second value is close to

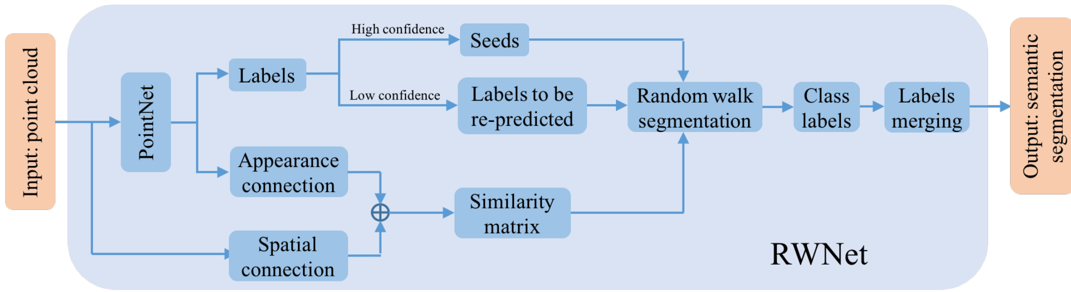


Fig. 2: Diagram of our 3D point cloud classification and semantic segmentation

the maximum. Incorrectly predicted labels come from this condition makes the network not so that efficiency, so we propose this step to judge the feature vector before real classification begins.

To get a proper threshold for ratio between the maximum class probability and the second in F_i , we analyze 40960 points in a 3D scene. We divide the correct / incorrect predicted points and compare the ratio range of them, as shown in Fig. 4. In an over view of the histogram, correct predicted points have large ratio and we show several example points with low ratio Fig. 3. We exhaustively search in the thresholds from 5 to 50 in a stride of 5 and calculate the final classification accuracy on a hallway scene in S3DIS dataset to select the optimal threshold with the maximum accuracy. Fig. 5 shows the classification accuracy with its corresponding threshold. In our experiments, we freeze the threshold learning from the this scene and use it in all tests.

We keep points with high confidence, i.e., of ratio larger than the threshold, as labeled points and the rest points as unlabeled points. After that, we employ Random Walk, in a semi-supervised fashion, to complete the final classification.

B. Similarity Matrix

Given the predicted labels via PointNet, we further categorize the points into two classes: high and low confident. We then use high confident points to predict labels of all points in a semi-supervised fashion. We employ Random Walk as our classifier. The main challenge of RW is the construction of a proper similarity matrix. In this section, we propose our way in building the similarity matrix.

Random Walk semi-supervised classification in [12] is proved efficiency in image segmentation based on the idea of graph-cut. Random walk in image segmentation tasks takes the handcrafted foreground and background as input, or known as $SEED$, and outputs the labels of under-predicted pixels. On the part of feature classification problem in 3D data, we extend 2D images to 3D point clouds.

To shape the relationship among all feature points, we establish a similarity matrix, which focuses on the local appearance and local structure of one feature point. Except for feature matrix M_{fm} formed by feature vectors, we import local distance to enhance the connection. We use the normalized Euclidean distance matrix to represent the position similarity among points which is marked as M_{dm}. Taking both feature and spatial space into consideration, we define our similarity matrix M as:

$$M = M_{fv} + |M_{dm}|$$

$M_{fv}(i,j) = Cov(F_i, F_j)$, where F_i is the point feature. $M_{dm}(i,j) = (P_i - P_j)^2$, where P_i is the point position.

We use M as the transfer probability matrix and take labels of high-confidence points as $SEED$, and employ Random Walk to generate labels of the points with low confidence.

C. Class Aggregation

Our RWNet only deals with points that are confused about their labels in feature vectors, so there are two kinds of labels that required by the class aggregation for semantic segmentation. We utilize the predicted labels with high confidence from PointNet as seeds L_P and the rest are predicted by RWNet. RWNet returns the max probability of candidate labels of points that are classified by $SEED$ and M, which are recorded as L_{RWNet}. The result of the whole points classification are aggregated as L, the union of high confident labels and RWNet-predicted labels, where $L = L_P \bigcup L_{RWNet}$, and it will be used for semantic segmentation. The algorithm of RWNet is attached as Algorithm 1.

Algorithm 1 Random Walk Network for Classification and Semantic Segmentation

Input: Raw 3D point cloud data with x, y, z coordinate
Output: Class labels of feature points and semantic segmentation of 3D scene

```
feature point extraction with PointNet
for each feature points do
    if the feature is of high confidence then
        save this point as SEED
        save the label as L_P
        establish feature correlation as M_fv
    else
        establish spatial relationship of points as M_dm
    end if
end for
M = M_fm + |M_dm|
RWNet processes M and output L_RWNet
Return L = L_P ⋃ L_RWNet
```

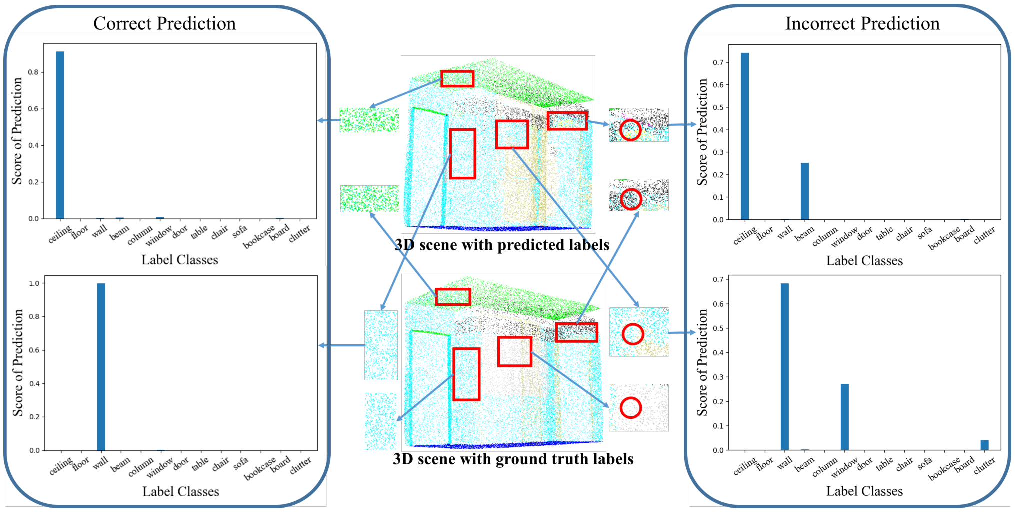


Fig. 3: A simple example of predicted labels with high or low confidence. Typically, a prediction is with high confidence if unimodal is occurred. And incorrect predictions always company with multimodal.

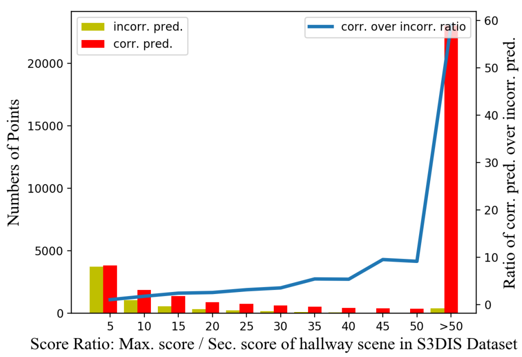


Fig. 4: Probability ratio of points in $Hallway$. The ratio is defined as the PointNet-predicted maximum probability over the second of a point. The yellow bars indicate the volume of incorrectly predicted points and the red of correctly predicted. Blue line represents the ratio of volumes of correct over incorrect. In this example, 6806 points were incorrectly predicted in a total of 40960 points.

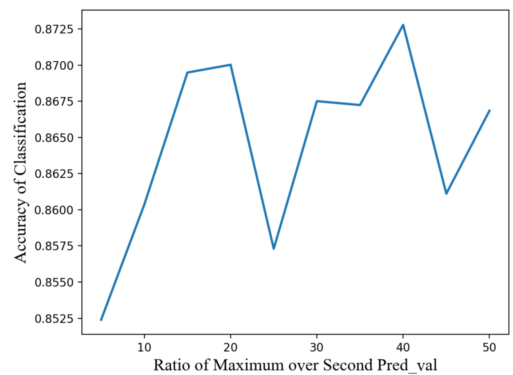


Fig. 5: The relationship of classification accuracy with prediction ratio.

IV. EXPERIMENTS

To demonstrate the performance of RWNet, we carry out comparison experiments with the state-of-the-art methods and evaluate the efficiency on benchmark datasets of Stanford 3D Indoor Semantics Dataset (S3DIS) [13], ModelNet40 [14] and ShapeNet dataset [15]. We use the strategy in [11] as our baseline to process datasets in our experiments.

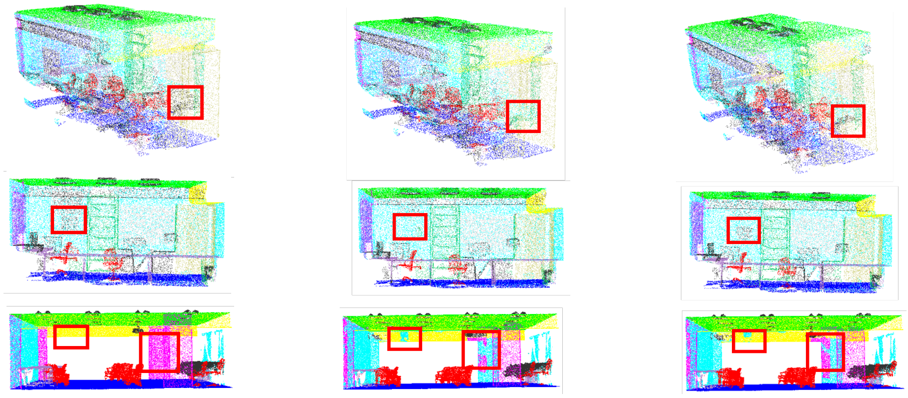
We firstly verify the ratio threshold value described in Sec. III-A to separate correct / incorrect labels. We operate our RWNet on one 3D scene of 40960 points with different ratios ranging from 5 to 50 with step 5 and output the classification accuracy, shown in Fig. 5. In our test, we set the ratio value of 40, as the highest classification accuracy, in our experiments.

A. RWNet on S3DIS Dataset

We only use x, y, z coordinates values of points as our input. When training, we sample 4096 points in each sub-room, which is cut as $1m$ x $1m$ from the raw point cloud scene and each point is labeled as one of the 13 classes. Fig. 6 shows the semantic segmentation results on S3DIS dataset with RWNet and the corresponding evaluation are shown in TABLE I. For detailed information about the classification, TABLE II presents the comparison performance with other state-of-the-art methods. Our RWNet shows good performance in many types of 3D point cloud classification and segmentation tasks in this dataset.

TABLE I: Evaluation of 3D scene semantic segmentation. Metrics are mean IoU (%) over 13 classes as presented in S3DIS dataset and classification accuracy.

	mIoU	accuracy
PointNet	47.71	78.62
RWNet	**78.81**	**92.49**



(a) (b) (c)

Fig. 6: Samples of comparison experiments. From left to right (a) the semantic segmentation of ground truth, (b) PointNet semantic segmentation result and (c) our RWNet. Red bounding boxes show improvement of ours over PointNet.

Different 3D point clouds from S3DIS dataset are sampled, and the visualized results with red bounding boxes are shown in Fig. 6. Our method rectified some of the point labels that were incorrectly predicted by PointNet and further improves the performance of semantic segmentation.

We adopt two widely used metrics, Intersection over Union (IoU) and mean accuracy (mAcc), to evaluate our RWNet in classification and semantic segmentation tasks. We compare RWNet with PointNet. As is shown in TABLE I, our method achieves an improvement of 31.3% on IoU and the accuracy is improved from 78.62% to 92.49%, indicating a better performance of our method than PointNet.

Experiments on 3D point cloud scenes are also taken into consideration and the results are presented in TABLE II. We randomly take samples from all dataset scenes to perform. In this table the accuracy of semantic segmentation raises around 2% compared with PointNet in all scenes and the accuracy from all rooms increased by 3.83%. In TABLE II, we employ another metric, Callback, which means the percentage of point clouds that are correctly recognized by RWNet who were incorrectly recognized by PointNet. It is around 25% in all 3D scenes except open-space, because of some points, such as single pillar, do not have strong connection with their neighbors both in distance and feature hence our RW may fail to construct reasonable similarity values for those points.

B. RWNet on ModelNet40 Dataset

We test our method on ModelNet40 dataset in this section. ModelNet is a large-scale 3D CAD dataset and we choose 40 common object categories with 100 unique CAD models per category. The performance of RWNet as well as the state-of-the-art method on ModelNet40 dataset are reported in TABLE III. For a fair-play, we use the same parser arguments, batch size 8, of PointNet. We calculate the average accuracy of classification and overall accuracy.

TABLE III: Results on ModelNet40 dataset, accuracy.

Methods	accuracy avg.class	overall accuracy
PointNet	72.6	77.4
RWNet	**76.7**	**80.3**

From TABLE III, we can see that with the help of random walk to rectify PointNet classification, both metrics are increased and our RWNet performances better. This experiment verified efficiency of our work.

C. RWNet on ShapeNet Dataset.

We also implement the classification and semantic segmentation tasks of RWNet on ShapeNet dataset, shown in TABLE IV. Presentation includes results from PointNet, KD-net, and the method proposed by Yi [26]. Notice that since several of the competing methods are not open available, we report their performance according to their original papers. IoU is adopted in this experiment for evaluation.

In some models, RWNet does not perform the best such as $guitar$ and $lamp$ compared with Yi [26] or $eraphone$ and $knife$ compared with KD-net [27], however, we achieves better performance in most models, 11 out of 16 in particular. Especially, compared with PointNet, the state-of-the-art work, RWNet shows better performance. In terms of IoU as shown in TABLE IV, RWNet outperforms other methods and compared with prior segmentation tasks on ShapeNet dataset, RWNet improves the mean IoU by $2.6\% \sim 4.9\%$.

We compare our work with PointNet by accuracy, which is shown in TABLE V. In all these models, RWNet performances better, and the mean accuracy is increased by 1.4%.

V. CONCLUSIONS

This paper proposed a novel method for 3D point cloud classification and semantic segmentation. We divide the PointNet-predicted points according to their prediction confidence and set the high-confident points as seeds for Random Walk. We then combine both appearance features and spatial relationship among points to construct a similarity matrix

TABLE II: Results on S3DIS 3D scene dataset, all room eval accuracy.

Methods	all room eval accuracy	conferenceRoom	copyRoom	hallway	office	openspace	pantry
PointNet	91.05	94.83	94.46	88.79	93.12	94.39	95.11
RWNet	94.88	96.43	95.66	91.44	95.38	96.62	96.72

TABLE IV: Results on ShapeNet, mIoU.

Methods	mIoU	aero	bag	cap	car	chair	era phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
Yi [26]	81.4	81.0	78.4	77.7	75.7	87.9	61.9	**92.0**	85.4	**82.5**	95.7	**70.6**	91.9	**85.9**	53.1	69.8	75.3
KD-net [27]	82.3	80.1	74.6	74.3	70.3	88.6	**73.5**	90.2	**87.2**	81.0	94.9	57.4	86.7	78.1	51.8	69.9	80.3
PointNet	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
RWNet	**86.3**	**84.8**	**79.3**	**83.3**	**78.1**	**96.5**	**73.5**	91.9	86.2	82.4	**96.0**	69.7	**93.7**	81.5	**64.6**	**75.3**	**81.4**

TABLE V: Results on ShapeNet, mAcc accuracy.

Methods	mAcc	aero	bag	cap	car	chair	era phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
PointNet	93.6	91.5	95.6	91.7	91.5	94.2	91.6	96.7	92.4	90.0	97.9	86.0	99.3	95.7	83.3	95.2	94.8
RWNet	**95.0**	**92.5**	**96.3**	**92.0**	**92.1**	**97.6**	**91.7**	**96.9**	**92.8**	**91.6**	**98.2**	**88.2**	**99.4**	**96.2**	**85.7**	**96.2**	**95.5**

and finally employ Random Walk to re-predict labels for low-confidence points. Comparison experiments on different datasets show that RWNet has a better performance than the state-of-the-art methods on widely used metrics.

References

[1] W. Chen, S. Gu, Y. Guan, H. Zhang, and H. Tang, "A multi-layered path planning algorithm for truss climbing with a biped robot," in *2016 IEEE International Conference on Information and Automation (ICIA)*, 2016.

[2] W. Chen, S. Gu, L. Zhu, H. Zhang, H. Zhu, and Y. Guan, "Representation of truss-style structures for autonomous climbing of biped pole-climbing robots," *Robotics and Autonomous Systems*, vol. 101, pp. 126–137, 2018.

[3] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012.

[4] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," 2015.

[5] E. H. Lim and D. Suter, "3d terrestrial lidar classifications with supervoxels and multi-scale conditional random fields," *Computer-Aided Design*, vol. 41, no. 10, pp. 701–710, 2009.

[6] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," 2017.

[7] C. R. Qi, H. Su, M. NieBner, A. Dai, M. Yan, and L. J. Guibas, "[ieee 2016 ieee conference on computer vision and pattern recognition (cvpr) - las vegas, nv, usa (2016.6.27-2016.6.30)] 2016 ieee conference on computer vision and pattern recognition (cvpr) - volumetric and multi-view cnns for object classification on 3d," 2016.

[8] S. Katz, G. Leifman, and A. Tal, "Mesh segmentation using feature point and core extraction," *Visual Computer*, vol. 21, no. 8-10, pp. 649–658, 2005.

[9] L. Yi, H. Su, X. Guo, and L. Guibas, "Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation," in *Computer Vision and Pattern Recognition*, 2017.

[10] R. Pan and G. Taubin, "Automatic segmentation of point clouds from multi-view reconstruction using graph-cut," *Visual Computer*, vol. 32, no. 5, pp. 601–609, 2016.

[11] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[12] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[13] I. Armeni and A. R. Zamir, "3d semantic parsing of large-scale indoor spaces," in *Computer Vision and Pattern Recognition*, 2016.

[14] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," 2014.

[15] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, and H. Su, "Shapenet: An information-rich 3d model repository," *Computer Science*, 2015.

[16] B. Li, "3d fully convolutional network for vehicle detection in point cloud," 2016.

[17] S. Song and J. Xiao, "Deep sliding shapes for amodal 3d object detection in rgb-d images," in *Computer Vision and Pattern Recognition*, 2016.

[18] X. Chen and Y. Zhu, "3d object proposals for accurate object class detection," in *International Conference on Neural Information Processing Systems*, 2015.

[19] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner, "Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks," 2017.

[20] Z. Ren and E. B. Sudderth, "Three-dimensional object detection and layout prediction using clouds of oriented gradients," in *Computer Vision and Pattern Recognition*, 2016.

[21] K. Guo, D. Zou, and X. Chen, "3d mesh labeling via deep convolutional neural networks," *Acm Transactions on Graphics*, vol. 35, no. 1, pp. 1–12, 2015.

[22] J. Masci, D. Boscaini, M. M. Bronstein, and P. Vandergheynst, "Geodesic convolutional neural networks on riemannian manifolds," in *IEEE International Conference on Computer Vision Workshop*, 2015.

[23] L. Kobbelt, P. Schrder, M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," *Proc.eurographics/acm Siggraph Symp.on Geometry Processing*, vol. 43, no. 2, pp. 156–164, 2003.

[24] M. Sormann, C. Zach, and K. F. Karner, "Graph cut based multiple view segmentation for 3d reconstruction," in *International Symposium on 3d Data Processing*, 2006.

[25] B. Shi, S. Bai, Z. Zhou, and X. Bai, "Deeppano: Deep panoramic representation for 3-d shape recognition," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2339–2343, 2015.

[26] L. Yi, L. Guibas, V. G. Kim, D. Ceylan, I.-C. Shen, M. Yan, H. Su, A. Lu, Q. Huang, and A. Sheffer, "A scalable active framework for region annotation in 3d shape collections," *Acm Transactions on Graphics*, vol. 35, no. 6, p. 210, 2016.

[27] R. Klokov and V. Lempitsky, "Escape from cells: Deep kd-networks for the recognition of 3d point cloud models," 2017.