Fast 3D Point Cloud Segmentation Using Deep Neural Network

Yunbo Rao，Menghan Zhang，Zhanglin Cheng，Junmin Xue，Jiansu Pu，Zairong Wang
电子科技大学

三等奖

摘要

Accurate segmentation of entity categories is the key step for 3D scene understanding. In this paper, we present a fast deep neural network model with Dense Conditional Random Field(DCRF) as a post-processing method, which can perform accurate semantic segmentation for 3Dpoint cloud scene. On this basis, a compact but flexible framework is introduced for performing segmentation to the semantics of point clouds concurrently, contribute to more precise segmentation. Moreover, based on the labels of semantics, a novel DCRF model is elaborated to refine the result of segmentation. Besides, without any sacrifice to accuracy, we apply optimization to the original data of point cloud, allowing the network to handle fewer data. In the experiment, our proposed method is conducted comprehensively through four evaluation indicators, proving the superiority of our method.

提出的方法

1. 框架

Given a 3D point cloud, we first map the point cloud into critical points through voxelization, and then extract more interdependent effective feature preparing for importing it into network for further feature extraction, after which the mapped point cloud scene is imported to network to perform semantic segmentation. To finish this task, and predict the semantic label of each point cloud of the scene, we design our network structure based on PointNet. In the end, those labels will be merged into the CRF model to optimize semantic segmentation.

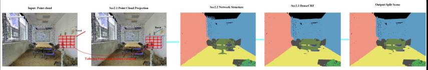

图1. The overview of our overall framework

2. 实现过程

A. Point cloud mapping and feature extension

In the incipient network, it utilizes only the information of coordinate, ignoring the remaining details. Other from that basic information we extend a brand-new series of attribute for describing 3D point cloud based on it geometric characteristic, including scattering(Sc), linearity(Li), planarity(Pl) and verticality(Ve).

$$\text{Linearity} = \frac{\lambda_1 - \lambda_2}{\lambda_1}$$

$$\text{Planarity} = \frac{\lambda_2 - \lambda_3}{\lambda_1}$$

$$\text{Scattering} = \frac{\lambda_3}{\lambda_1}$$

Linearity reflects the extent of neighborhood growth while planarity represents if the neighborhood can be fitted by a plane, and best for loose categories like indoor plants, scattering corresponds to the disorder degree of the point clouds within its spherical neighborhood. A new attribute called verticality is introduced here, which is critical for distinguishing between planes and elevation. And is calculated as in equation below.

$$\text{Ve} = [\hat{u}]_i \propto \sum_{j=1}^{3} \lambda_i \left|[u_j]_i\right| \quad i = 1,2,3 \quad \|\hat{u}\| = 1$$

B. New neural network for semantics segmentation

The network design of 3D point cloud is determined by its two diverse features: 1) It is not sensitive to the order of points which does not affect the collection, since 3D point cloud itself is collection of disordered points; 2) The rotation of 3D point cloud should not change the result of classification, which makes it necessary to transform the 3D point cloud data.

As shown in figure below, based on PointNet structure, we construct a network for segmentation of 3D point cloud.

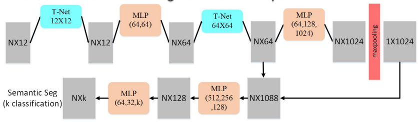

图2. Neural network for semantics segmentation

In our work, the loss function is calculated by summing the lost value of two parts.

$$L = L_{pred} + L_{emb}$$

The L_{pred} function is a regular cross entropy function that is used directly from the PointNet network. The L_{emb} function is a conventional cross entropy function, which is inspired by the ASIS network. The loss function L_{emb} is formulated as follows:

$$L_{emb} = L_{var} + L_{dist} + \alpha L_{reg}$$

And α is set to 0.001 in our experiment. In detail, the specific formulation of each item is as follows:

$$L_{var} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N_k}\sum_{j=1}^{N_k}[\|\mu_k - e_j\| - \delta_v]_+^2$$

$$L_{dist} = \frac{1}{K(K-1)}\sum_{k=1}^{K}\sum_{m=1,m\neq k}^{K}[2\delta_d - \|\mu_k - \mu_m\|]_+^2$$

$$L_{reg} = \frac{1}{K}\textstyle\sum_{k=1}^{K}\|\mu_k\|_2$$

C. Improving segmentation with DenseCRF

In this paper, we directly change the neighboring pixel to a point cloud, and the segmentation result obtained from the deep network is used as the input of the DCRF model. At the same time, a model can be established based on the relationship between point clouds in the space to fully grasp the context of the entire space, and its energy function is defined as:

$$E(l_i^S|V) = \sum_j \phi(l_j^S) + \sum_{(j,k),j<k} \phi(l_j^S, l_k^S)$$

$$\phi(l_j^S) = -\frac{e^{\left[-\frac{1}{2}(e_j-\mu_i)^T \Sigma_i^{-1}(e_j-\mu_i)\right]}}{\sqrt{(2\pi)^d|\Sigma i|}} - \log[\sum_k (l_j^S = i)]$$

$$\phi(l_j^l, l_k^l) = w_{i,j}\exp\left(-\frac{\|l_j - l_k\|^2}{\lambda_1^2} - \frac{\|c_j - c_k\|^2}{\lambda_2^2} - \frac{\|d_j - d_k\|^2}{\lambda_3^2}\right)$$

实验

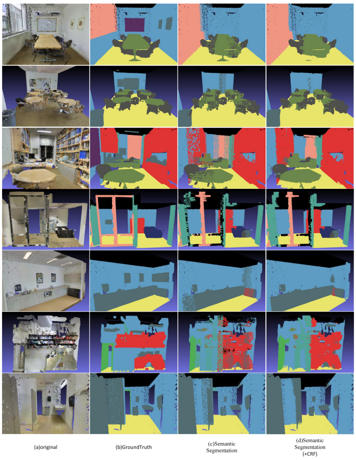

图3. This figure shows the different scenarios in the Area5 area of the S3DIS data set

The results of our method are shown in figure right. The first column represents the original 3D point cloud scene; the second column represents the ground truth of the scene appeared in the first column; the third column represents the results processing the original scene without optimizing through CRF; the fourth column represents the segmentation result with CRF optimization.

Method	mIoU	mAcc	mRecall
PointNet	47.6	52.1	-
PointNet++	50.8	58.3	-
ASIS	55.7	59.3	-
Ours	53.7	65.6	0.338
Ours(+CRF)	56.2	67.5	0.372

表1. Semantic segmentation mIoU, oAcc and mRecall on S3DIS.

结论

This paper proposes a novel network structure that improves semantic segmentation results. By applying semantic segmentation to produce a segmentation model, our network achieves high accuracy on semantic segmentation taking CRF as a post-processing step. At the same time, performing mapping and feature extension to point cloud data in the stage of preprocessing, the number of network arguments are sharply cut off without decreasing the overall performance. Compared to other methods, we evaluate our proposed methods on S3DIS indoor dataset, which shows that the proposed method has good segmentation performance and can be easily generalize into other segmentation tasks within various point cloud scenes.

主要参考文献

[1] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: deep learning on point sets for 3D classification and segmentation. Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 1(2):4, 2017. 1, 2,

[2] QI C. R., YI L., SU H., GUIBAS L.J. Pointnet++:deep hierarchical feature learning on point sets in a metric space. In Advances in Neural Information Processing Systems, Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., (Eds.). Curran Associates, Inc., 2017, pp. 5099–5108

[3]WANG X., LIU S., SHEN X., SHEN C., JIA J.: Associatively segmenting instances and semantics in point clouds. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019), pp. 383 4091–4100.

[4]ARMENI I., SENER O., ZAMIR A. R., JIANG H., BRILAKIS I., FISCHER M., SAVARESE S.: 3D semantic parsing of large-scale indoor spaces. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016), pp.1534–1543

[5]HUA B., TRAN M., YEUNG S.: Pointwise convolutional neural networks. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018), pp.984–993.

[6]TCHAPMI L., CHOY C., ARMENI I., GWAK J., SAVARESE S.: Segcloud:semantic segmentation of 3D point clouds. In 2017 International Conference on 3D Vision (3DV)(2017), pp.537–547.2,8,9

联系方式

联系人：饶云波
手机：15908177003
邮箱：raoyb@uestc.edu.cn