Feature Relation based Graph Convolution for 3D Point Cloud Analysis

Yang Wang*, Shunping Xiao

1College of Electronic Science, National University of Defense Technology, Changsha, Hunan, 410073, China
*Corresponding author’s e-mail: wangyangs4@nudt.edu.cn

Abstract. 3D point cloud recognition is still a challenge task since the shape implied in irregular points is difficult to capture. Standard convolution is inherently limited for these tasks due to its isotropy about features. In this paper, a novel graph convolutional network is introduced for point cloud classification and segmentation task. The proposed convolution adds an additional layer on the basis of relation-shape convolutional neural network, which can obtain more information and make the representation of point cloud more robust. At the same time, a feature relation method is proposed instead of the coordinate relation used in relation-shape convolutional neural network. The experimental results on challenging classification and segmentation datasets shows that the proposed method can learn discriminating features for recognition and semantic segmentation.

1. Introduction
In recent years, it has been caught too much attention for applications which rely on 3D point cloud data, e.g., autonomous driving, robotic manipulation and virtual reality [1]. Thus, there are critical needs for approaches to effectively and efficiently 3D point cloud processing. However, this task is very challenging, since inferring the underlying shape formed by these irregular points is difficult.

To solve this problem, many works are focused on replicating the remarkable success of convolutional neural network (CNN) on regular grid data analysis to irregular point cloud processing [3]. They can be divided into two categories. Some works transform point cloud to regular voxels [4] or multi-view images [5] for easy application of classic grid CNN. However, this kind of methods induce a loss of information and excessive consumption of memory and high computation cost.

The other is to directly process point cloud. PointNet [6] independently learns on each point and gathers the final features for a global representation and it ignores local structures. Local structures have been proven to be important for abstracting high-level visual concepts in image. Thus, A lot of improvement work has been proposed. Nevertheless, this extremely relies on effective inductive learning of local sub-sets, which is quite intractable to achieve.

Liu et. al [7] introduces a Relation-Shape Convolutional Neural Network (RS-CNN) for 3D point cloud analysis. The key of RS-CNN is learning from relation, i.e., the geometric topology constraint among points, which can encode meaningful shape information in 3D point cloud. However, because of down sampling is used from the first layer, RS-CNN lost some important information prematurely. To address this problem, an additional layer is used to extract the feature information in advance, and then the feature information is added to RS-CNN, so more feature information can be extracted. The experimental results show that the proposed extra layer can be used to improve the performance of RS-CNN.
2. Method

In this section, the RS-CNN [7] and our improvement are both introduced.

2.1. RS-CNN

RS-CNN learns contextual shape-aware representation by a novel relation-shape convolution (RS-Conv) which extending regular grid CNN to irregular configuration. Given \( n \) points in a point cloud \( X = \{ x_i, x_j, ..., x_n \} \) and \( N(x_i) \) is the neighbourhood of \( x_i \). RS-Conv can be formulated as:

\[
f_p = \sigma \left( A \left( \left\{ M \left( h_y \right) \cdot f_j, \forall x_j \in N(x_i) \right\} \right) \right)
\]

where \( x_i \) is a 3D point and \( f \) is a feature vector. Function \( M, A \) and \( \sigma \) are shared multi-layer perceptron (MLP), aggregation function and nonlinear activator respectively. The core part \( h_y = \{3D - Ed, x_i - x_j, x_j, x_k\} \) and \( 3D - Ed \) represents the three-dimensional Euclidean distance between points. The flow of RS-CNN is shown in Fig. 1 [7].

![Figure 1. The flow of RS-CNN [7]](image)

2.2. Our improvement of RS-CNN

Although RS-CNN has achieved excellent results, it only uses coordinates and spatial information to learn shape information and ignores feature information. Thus, formula (1) is modified as follows:

\[
f_p = \sigma \left( A \left( \left\{ M \left( f_i \right) \cdot f_j, \forall x_j \in N(x_i) \right\} \right) \right)
\]

Relative to formula (1), coordinates are replaced with more powerful features which can learn better for the relationship of objects. Compared with only using coordinates and spatial information to learn information, features contain more abundant information, and the learning results are more robust.

Moreover, in order to capture more feature information, we add an extra layer of convolution. As shown in Fig. 2, our classification network has four layers, while RS-CNN has only three. The \( L_0 \) layer is used to learn feature information from coordinates and serve as the input of next layer. The purpose of this is to be able to get stronger expression because the expression ability of features is much stronger than coordinates. In the segmentation task, the learned features are interpolated back to the finest scale layer by layer to obtain the feature map that has the same number of points as the original input.

3. Results

We evaluate our method in the task of object classification and shape part segmentation. All the experiments are run on a GTX 1070 GPU with CUDA 11.3 and CuDNN 7.6.5. Experimental environment is python 3.6.5 and pytorch 0.4.1.
3.1 Classification on ModelNet40 classification benchmark
ModelNet40 classification benchmark [8] is composed of 9843 train models and 2468 test models in 40 classes and 1024 points are sampled uniformly [6]. As the same in [9], during training in the classification task, the input data is augmented with random anisotropic scaling in the range [-0.66, 1.5] and translation in the range [-0.2, 0.2]. Meanwhile, dropout technique with 50% ratio is applied in FC layers and ten voting tests with random scaling are performed to obtain the average of predictions [6]. As shown in Table 1, the proposed method achieves an accuracy of 92.6%, improved by 0.4%. The final result of using ten voting tests is 92.8%, which is a competitive result.

Table 1. Shape classification results (%) on ModelNet40 benchmark (“vot.”: with ten voting tests, “no vot.”: without ten voting tests)

| Methods          | Input          | Accuracy |
|------------------|----------------|----------|
| DGCNN [10]       | 1k points      | 92.2     |
| SO-Net [11]      | 2k points      | 90.9     |
| PointNet++ [12]  | 5k points + normal | 91.9 |
| PCNN [13]        | 1k points      | 92.3     |
| FPConv [14]      | 1k points      | 92.5     |
| KPConv [15]      | 1k points      | 92.9     |
| RS-CNN [7]/vot.  | 1k points      | 93.6     |
| RS-CNN [7]/no vot.| 1k points     | 92.2     |
| InterpCNN [16]   | 1k points      | 93.0     |
| **Ours/no vot.** | 1k points      | **92.6** |
| **Ours/vot.**    | 1k points      | **92.8** |

3.2 Segmentation on ShapeNet part benchmark
ShapeNet part benchmark contains 16881 shapes with 16 categories, and is labelled in 50 parts in total. In the experiments, 2048 points are randomly picked as the input [12]. The results of part segmentation are shown in Table 2. The proposed method still achieves a reasonable performance since it can learn a certain distinguishable information.

Table 2. Shape part segmentation results (%) on ShapeNet part benchmark

| Method   | Class mIoU | Instance mIoU |
|----------|------------|---------------|
| KCNet [17]| 82.2       | 84.7          |
| PointNet [6]| 80.4     | 83.7          |
| PointConv [18]| 82.8    | 85.7          |
| PCNN [13] | 81.8       | 85.1          |
| DGCNN [10]| 82.3       | 85.1          |
| RS-CNN [7]| **84.0**   | **86.2**      |
| **Ours**  | **82.2**   | **85.8**      |
4. Conclusion
In this paper, an improved RS-CNN is proposed for 3D point cloud recognition and semantic segmentation. The proposed network can learn much more structure information in the recognition and semantic segmentation task. Compared with single scale RS-CNN, the performance of our method on the classification dataset is improved by 0.4%. Moreover, the proposed method also achieves a reasonable performance in part segmentation task. Although the proposed method have obtained an improvement, but the experiments are performed on a simulated dataset in which each object is complete. Thus, applying this method on an incomplete dataset is the future work.

References
[1] Qi, C. R., Liu, W., Wu, C., Su, H. and Guibas, L. J. (2018) Frustum PointNets for 3D Object Detection from RGB-D Data. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp. 918-927.
[2] Rusu, R. B., Marton, Z., Blodow, N., Dolha, M. and Beetz, M. (2008) Towards 3D Point Cloud based Object Maps for Household Environments. Robotics and Autonomous Systems. 56(11): 927-941.
[3] Jiang M., Wu, Y., Zhao, T., Zhao, Z. and Lu, C. (2018) PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation. arXiv preprint arXiv:1807.00652.
[4] Gadelha, M., Wang, R., and Maji, S. (2018) Multiresolution Tree Networks for 3D Point Cloud Processing. In: European Conference on Computer Vision. Munich. pp. 105-122.
[5] Feng, Y., Zhang, Z., Zhao, X., Ji, R., and Gao, Y. (2018) GVCNN: Group-view Convolutional Neural Networks for 3D Shape Recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp. 264-272.
[6] Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2016) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas. pp. 77-85.
[7] Liu, Y., Fan, B., Xiang, S. and Pan, C. (2019) Relation-Shape Convolutional Neural Network for Point Cloud Analysis. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach. pp. 8887-8896.
[8] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. (2015) 3D ShapeNets: A Deep Representation for Volumetric Shapes. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Boston. pp. 1912-1920.
[9] Klokov, R. and Lempitsky, V. (2017) Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models. In: IEEE International Conference on Computer Vision. Honolulu. pp. 863-872.
[10] Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M. and Solomon, J. M.. (2019) Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics. 38(5): 146: 1-12.
[11] Li, J., Chen, B. M. and Lee, G. H. (2018) SO-Net: Self-Organizing Network for Point Cloud Analysis. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp. 9397-9406.
[12] Qi, C. R., Yi, L., Su, H. and Guibas, L. J. (2017) PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: Thirty-first Conference on Neural Information Processing Systems. Long Beach. pp. 5099-5108.
[13] Atzmon, M., Baron, H. and Lipman, Y. Point Convolutional Neural Networks by Extension Operators. (2018) ACM Transactions on Graphics. 37(4): 71: 1-12.
[14] Lin, Y., Yan, Z., Huang, H., Du, D., Liu, L., Cui, S. and Han, X. (2020) FPCConv: Learning Local Flattening for Point Convolution. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle. pp. 4292-4301, 2020.
[15] Thomas, H., Qi, C. R., Deschaud, J., Marcheuy, B., Goulette, F. and Guibas, L. (2019) KPConv: Flexible and Deformable Convolution for Point Clouds. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach. pp. 6411–6420.
[16] Mao, J., Wang, X. and Li, H. (2019) Interpolated Convolutional Networks for 3D Point Cloud Understanding. In: IEEE/CVF International Conference on Computer Vision. Seoul. pp. 1578-1587.

[17] Shen, Y., Feng, C., Yang, Y. and Tian, D. (2018) Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp. 4548-4557.

[18] Wu, W., Qi, Z. and Li. F. (2019) PointConv: Deep Convolutional Networks on 3D Point Clouds. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach. pp. 9613-9622.