A Neural Network for Point Clouds Classification

Jian Liu[a], Ziyi Meng[b], and Di Bai[c]

Institution of Information and Control Engineering Shenyang Jianzhu University Shenyang, China

[a]644630900@qq.com, [b]keme@vip.qq.com, [c]1582729192@qq.com

Keywords: Deep learning; Point clouds; Classification network

Abstract: Semantic learning on 3D point clouds model using a deep network has received great interests to address the 3D object classification problem. The complete information of the 3D object is described by the disordered data structure. A new classification method based on deep learning is proposed, which combines 3D vision with deep network. The point clouds and 3D standard geometry are regarded as input in the process of training the network. The proposed network unifies feature extraction and classification into a stage, which removed the need of manual feature engineering for 3D point clouds in traditional method. Therefore, the 3D point clouds data of the identified object is directly input to the trained network and the classification result can be obtained. The specific classification experiments on target objects have been done. The training has been completed on the RW-4028GR-TRT2 experimental platform with six Nvidia GTX 1080Ti. The experimental results show the accuracy of the proposed classification network, which is significant to improve the efficiency of object classification. Furthermore, the network has been applied to the recognition of indoor scene in the intelligent building, and it will attribute to the wide application in many fields in the future.

1. Introduction

Point cloud classification has always been a challenging task in computer vision. There are also a wide range of applications in various fields, such as robotics and scene understanding. Despite the wide application of 3D sensors, 2D images still play an important role in various fields. Since a large number of data sets have been published and the labeling work has been carried out well, the application of neural networks in image classification has reached a very high level through the efforts of various teams, which has far surpassed that of humans. However, since the acquisition quality of the image sensor is greatly affected by the environment, the image acquired under the complicated environment of the light source cannot be used for the recognition task. The point cloud generated by the laser radar based on the time of fight(TOF) principle is not affected by environmental changes, so the task of identifying on the point cloud is also increasingly important.

Classic convolutional network structures require a high degree of regularization of input data, such as two-dimensional images or three-dimensional voxel formats for weight sharing and other kernel calculations. However, since the point cloud is not such a regular data type, the point cloud have a certain arrangement rule through an effective conversion method. A new classification method based on deep learning is proposed herein. Spherical projection is applied to the input layer of the neural network to map the point cloud to a unit sphere, only leaving the farthest point in a certain neighborhood. The effectiveness of the proposed algorithm is experimentally verified, when the original data is simply rotated or increased by random perturbations, the proposed method can still achieve the classification results accurately.

The remainder of this paper is organized as follows. In Section II, the related work is introduced. In Section III, the specifications of the network architectures are discussed. In Section IV, the experiments and analysis performed are presented. Finally, the conclusions are presented in Section V.

  DOI: 10.25236/icemc.2019.014

2. Related Work

The image is a regular arrangement of color features in a two-dimensional image plane, the video is a regular arrangement on the image time dimension, and the voice is a regular arrangement of audio features on the time dimension. These characteristics of the type of data are arranged in a certain order, and the reverse order of the data will bring different interpretation results. Convolutional neural network (CNN) brings about the rapid development of feature extraction of such rules [1]. Point cloud is distinguished from such data, although it can be generated by a visual sensor like an image, it does not have any permutation rules. Randomly disturbing the data sequence changes the content of the stored data, but does not change the actual information represented by the point cloud. There have been many efforts to handle the data without alignment rules, for example, converting a point cloud to a voxel and then calculating it with a 3D convolution layer [2,3,4], however, the representation of voxels can cause the data to be too sparse. Therefore, octree and other structures[5,6,7] have been proposed to try to solve the problem of data sparseness. Some research works[8,9,10] use spectral convolutional nerual network on polygonal meshes, this method converts convolution to multiplication and it's confined to manifold meshes. Other researchers have rendered point clouds from different perspectives into two-dimensional images[11,12,13], looking for relationships between multiple views. But it will ignore a lot of geometric details, which limits the performance for complex tasks. The method of transforming point cloud into a feature vector by extracting traditional geometric features[14,15] is proposed, and then classifying it with deep neural network (DNN). Since using PointNet [16] to solve the problem of point cloud disorder with symmetric functions, some network models have been used that directly use point clouds as input. For example, PointNet++ [17] uses PointNet to extract local features, PointCNN [18] uses X-Conv structure, ShapeContextNets [19] selects context regions based on self-attention, and PointSIFI [20] designs feature extraction modules based on SIFT features.

3. Network Architectures

3.1 Overview.

A network architecture is designed, the input 3D point cloud is automatically mapped to a sphere, features are abstractly extracted, and then the sphere is expanded into the form similar to a two-dimensional image, and the classification problem is solved by the mature image classification framework. Due to the unpredictability of feature mapping, the network is trained by maximizing the classification accuracy. The mapping method is different from multiple viewing angles, and the experimental results prove that this method is feasible.

For the input point cloud P, the point O is first learned as the center of the sphere to be mapped (the point can also be specified manually), and then a spherical map is generated, the features of the corresponding points are calculated, and the spherical surface is expanded into a form similar to a two-dimensional image to perform point cloud classification.

The network architecture is divided into three modules, the first module uses the point cloud P as the input to predict the spherical center O for mapping (Section 3.2), the second module calculates the spherical mapping characteristics based on the input point cloud P and the center O. (Section 3.3), and the last module generates a predictive label representing the class based on the spherical features (Section 3.4). The network is trained in a complete architecture with the final prediction probability and the softmax cross entropy of the tag as the loss.

3.2 Sphere Center Prediction.

The point cloud P is took as input and generate the spherical center O of the mapping sphere, in particular, this step can also be specified by the user without prediction through the network. The mapping of the point cloud on the sphere is sufficiently scattered rather than a certain area of concentration is expected. When a large amount of 3D point data is used as the original input, calculating a suitable spherical center O requires a large amount of calculation. In order to improve

the speed of this link, 512 points are randomly selected from the input for real participation in the calculation of this module. The position of the sphere center O is predicted by standard Multilayer Perceptron (MLP). The point cloud is mapped to the coordinate system with the spherical center O as the origin. After the transformation, the point cloud becomes a representation of the content on the spherical surface. The sphere center prediction process is shown in Fig. 1.

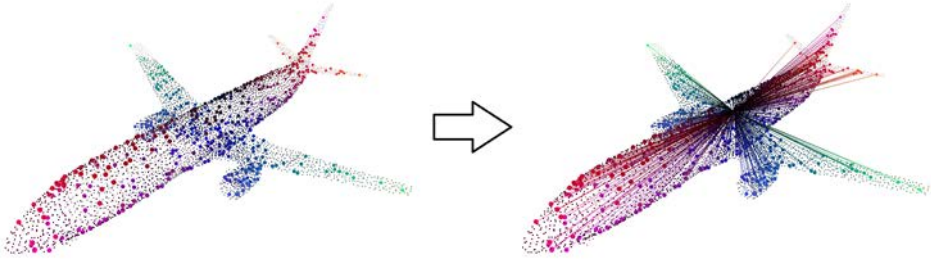

Fig. 1 The sphere center prediction process chart

3.3 Spherical Mapping.

Perform mapping operations according to (1) and (2).

$$x' = \arctan\left(\frac{z}{x}\right) - \frac{y}{|y|} * \left(\frac{x}{|x|} - 1\right) * \frac{\pi}{2} \quad (1)$$

$$y' = arccos\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right) \quad (2)$$

In the calculation process, the spherical surface is divided into n parts along the zenith angle, divided into 2*n parts along the azimuth angle, and then the point cloud is distributed into 2n*n discrete areas. There is no way to guarantee that subsequent calculations involve all the points of the input, because this allocation will always occur when there are multiple points in one area. The strategy adopted is to retain only the point farthest from the center of the sphere O for subsequent feature calculations. The original data is not processed in the mapping process, the original 3D coordinates of the point cloud are still stored in the 2n*n area, and the area not mapped is filled with 0. And then feature extraction is performed using a multi-layer convolutional network. The spherical mapping process is shown in Fig. 2.

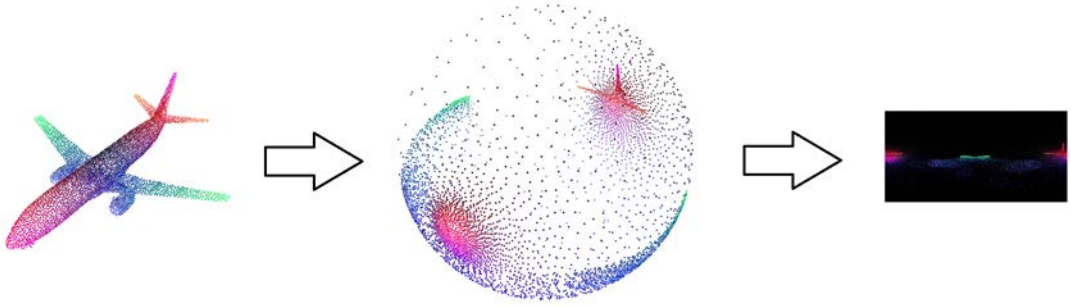

Fig. 2 The spherical mapping chart

3.4 Feature Classification.

The generated spherical map is input to achieve classification function, which combines SENet and ResNet v2. The mapping operation maps the point cloud to the two-dimensional spherical surface, the point cloud is represented by a method similar to a planar image, and the purpose of processing a three-dimensional point cloud by two-dimensional convolution is achieved. A network

architecture for flat images can be thought of as a network that can handle input with two-dimensional structural features, and so the mapped point cloud is processed by using a network architecture with planar image classification.

4. Experimental Results

4.1 Dataset Description.

The ModelNet10 and ModelNet40 data sets were used to evaluate our network. The ModelNet10 dataset consists of ten kinds of CAD models with category labels, a total of 4,900 are divided into 3,991 for training and 909 for testing. The ModelNet40 dataset consists of forty kinds of CAD models with category labels, a total of 12,311 are divided into 9843 for training and 2,468 for testing. During the training process, the point cloud samples are dynamically enriched by randomly rotating a certain angle along the X and Y axes and adding a Gaussian noise with a mean of 0 standard deviation of 0.01 for each point. The point cloud classification experiment was performed on the RW-4028GR-TRT2 platform with six Nvidia GTX 1080Ti.

4.2 Point Cloud Classification.

For the ModelNet10 and ModelNet40, uniform sampling and random sampling point clouds of different points were prepared, which were 1024, 2048, 4096, and 8192 sampling points respectively, to test the influence of sampling points on our network performance. The accuracy and the loss are shown in Fig. 3.

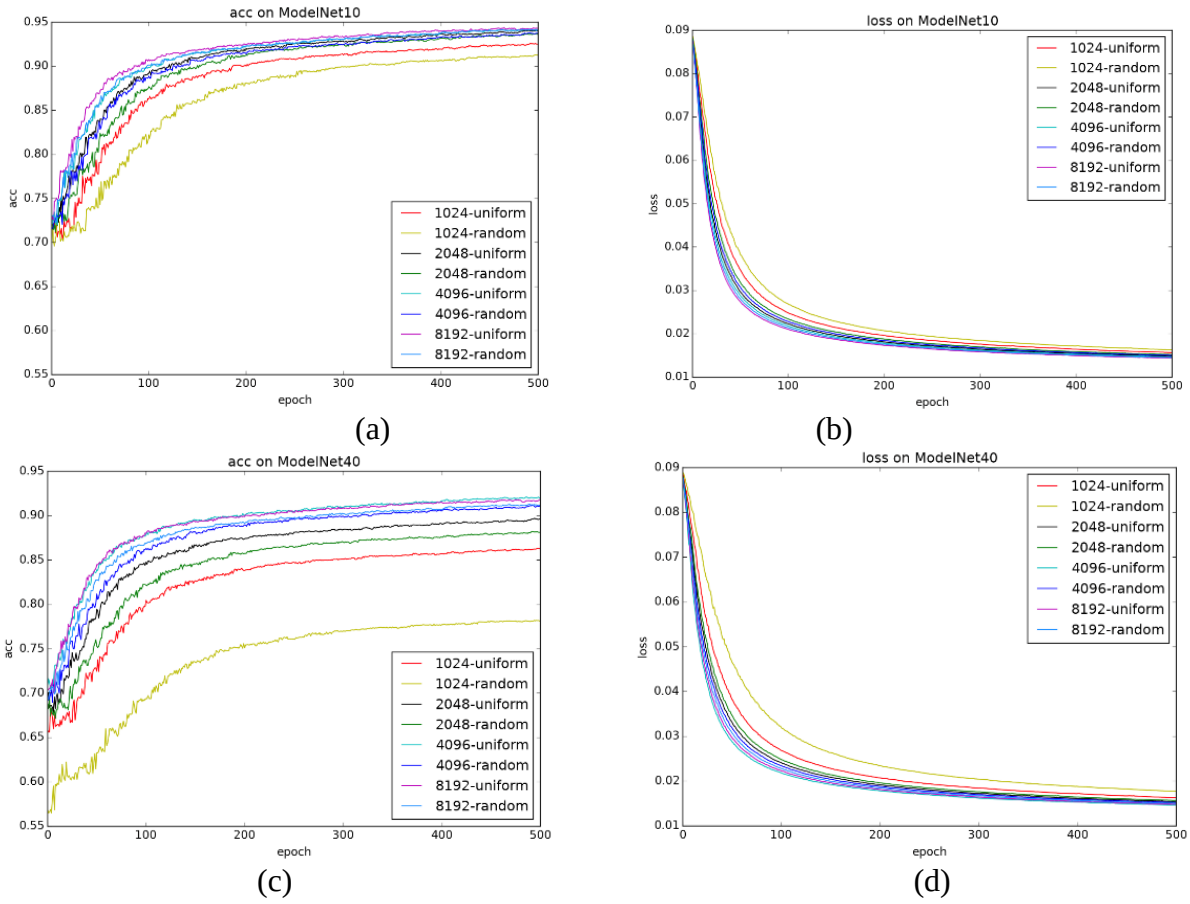

Fig. 3 (a) Accuracy on ModelNet10; (b) Loss on ModelNet10;(c) Accuracy on ModelNet40; (d) Loss on ModelNet40

The training results are shown in table 1.

Table 1 Table Type Styles

Data Set	uniform				random			
	1024	**2048**	**4096**	**8182**	**1024**	**2048**	**4096**	**8192**
ModelNet10	92.6%	94.1%	94.2%	94.4%	91.3%	93.7%	93.8%	94.2%
ModelNet40	86..3%	89.7%	92.1%	91.8%	78.2%	88.2%	91.2%	91.3%

ModelNet10:The result of uniform sampling is better than the result of random sampling. The lowest accuracy is 91.3% of random sampling of 1024 points, the best is to evenly sample 8192 points to achieve an accuracy of 94.4%. The closest set of random sampling to uniform sampling is 8192 points, which is only 0.2% worse. Overall, the accuracy rate increases as the number of sampling points increases.

ModelNet40:The results of uniform sampling are generally good. The minimum uniform sampling of 1024 points can finally reach 86.3% accuracy, and the uniform sampling of 4096 points achieves the highest accuracy of 92.1%. The results of random sampling are slightly worse than the corresponding uniform sampling results. The random sampling of 1024 points has the worst effect is only 78.2%. The other three groups are close to each other. The closest result is that the random sampling of 8192 points is 91.3%. It is only 0.5% lower than uniform sampling, and the result of random sampling of 4096 points is 91.2%, which is only 0.1% lower than the random sampling of 8192 points.

According to the above experimental results, it is reasonable to select 4096 points when sampling in the range of 1024-8192. The effect of uniform sampling is best, but the time for uniform sampling is several times that of random sampling and it is not suitable for applications that are time sensitive.

4.3 Comparisons to Others.

The test results of ModelNet10 and ModelNet40 in Table 2 were compared with the results of uniform sampling and random sampling of 4096 points. The proposed method has an accuracy of 92.1% on ModelNet40 and 94.4% on ModelNet10. G3DNet[21] adopt a graph based methods to 3D point clouds to introduce a generic vector representation of 3D graphs, whose accuracy is 1.3% and 0.97% lower than the proposed method testing on the ModelNet10 and ModelNet40 data sets. binVoxNetPlus[22] transforms the inputs and weights in the network to binary values through binary transformation, whose performance is 0.78% and 5.66% lower compared to G3DNet testing on the ModelNet10 and ModelNet40 data sets, respectively. LightNet[23] using the beam search, whose performance on the ModelNet10 and ModelNet40 data sets is 0.46% lower than the proposed method and 3.17%. MVCNN-MultiRes[24] introduce multi-resolution filtering in 3D, and the improvement tested on the ModelNet40 dataset was 0.27% higher than G3DNet. PointNet[16] use a single symmetric function to deal with unordered input set, and the accuracy rate reached 89.2% tested on the ModelNet40 dataset. KCNet[25] construct k nearest neighbor graphs to utilize the neighborhood information for kernel correlation and to recursively conduct the max-pooling operations in each nodes neighborhood, and the test results on the ModelNet40 data set are 1.1% worse than the proposed method.

Table II Table Type Styles

Method	ModelNet10	ModelNet40
G3DNet[21]	93.1%	91.13%
binVoxNetPlus[22]	92.32%	85.47%
LightNet[23]	93.94%	88.93%
MVCNN-MultiRes[24]	-	91.4%
PointNet[16]	-	89.2%
KCNet[25]	94.4%	91.0%
The proposed Net	94.4%	92.1%

5. Summary

A new classification method based on deep learning is proposed, which combines 3D vision with deep network. First, map the point cloud to the origin of the sphere, and then the farthest point is retained by performing the spherical map, finally, the convolution network is used to extract the features according to the two-dimensional arrangement of the spherical expansion. This simplifies the complexity of feature extraction and improves the accuracy of point cloud classification. The proposed method unifies feature extraction and classification into a stage, which removed the step of manual feature engineering for 3D point clouds in traditional network. Therefore, the 3D point clouds data is input to the trained network and the classification result can be obtained directly. Specific experiments have been performed on the ModelNet10 and ModelNet40 datasets. The experimental results show the accuracy of the proposed classification network, which is significant to improve the efficiency of object classification. Furthermore, apply the network to point cloud semantic segmentation is expected.

Acknowledgment

This research is supported by the scientific research projects in National Natural Science Foundation of China (11704263), Liaoning Province Natural Science Foundation (201602616) and Liaoning Province Department of Education Scientific Research Project (2015443).

References

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521,7553 (2015), 436–444.

[2] Zhirong Wu, S. Song, A. Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and J. Xiao. 3d shapenets:A deep representation for volumetric shapes. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1912–1920, June 2015.

[3] D. Maturana and S. Scherer. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. In IROS, 2015

[4] Charles R Qi, Hao Su, Matthias Niessner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. arXiv preprint arXiv:1604.03265, 2016.

[5] M. Tatarchenko, A. Dosovitskiy, and T. Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In IEEE International Conference on Computer Vision (ICCV), 2017.

[6] Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. Octnet: Learning deep 3d representations at high resolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[7] Yangyan Li, Soeren Pirk, Hao Su, Charles R Qi, and Leonidas J Guibas. Fpnn: Field probing neural networks for 3d data. arXiv preprint arXiv:1605.06240, 2016.

[8] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. CoRR, abs/1312.6203, 2013.

[9] J. Masci, D. Boscaini, M. M. Bronstein, and P. Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In 2015 IEEE International Conference on Computer Vision Workshop (ICCVW) pages 832–840, Dec 2015.

[10] Li Yi, Hao Su, Xingwen Guo, and Leonidas Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. arXiv preprint arXiv:1612.00606, 2016.

[11] H. Su, F. Wang, E. Yi, and L. Guibas. 3d-assisted feature synthesis for novel views of an object. In 2015 IEEE International Conference on Computer Vision (ICCV), pages 2677–2685, Dec 2015.

[12] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape

[13] Charles R Qi, Hao Su, Matthias Niessner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. arXiv preprint arXiv:1604.03265, 2016.

[14] K. Guo, D. Zou, and X. Chen. 3d mesh labeling via deep convolutional neural networks. ACM Transactions on Graphics (TOG), 35(1):3, 2015.

[15] Y. Fang, J. Xie, G. Dai, M. Wang, F. Zhu, T. Xu, and E. Wong. 3d deep shape descriptor. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2319–2328, 2015.

[16] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. arXiv preprint arXiv:1612.00593, 2016.

[17] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413, 2017.

[18] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, Baoquan Chen.PointCNN: Convolution On X-Transformed Points. arXiv preprint arXiv:1801.07791, 2018.

[19] Liu, S. (2018). Attentional ShapeContextNet for Point Cloud Recognition. UC San Diego. ProQuest ID: Liu_ucsd_0033M_17627. Merritt ID: ark:/13030/m51020r9. Retrieved from https://escholarship.org/uc/item/624107gk

[20] Mingyang Jiang, Yiran Wu, Cewu Lu.PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation.arXiv preprint arXiv:1807.00652, 2018.

[21] Miguel Dominguez,Rohan Dhamdhere,Atir Petkar,Saloni Jain, Shagan Sah, Raymond Ptucha, General-Purpose Deep Point Cloud Feature Extractor. WACV 2018.

[22] Chao Ma, Wei An, Yinjie Lei, Yulan Guo. BV-CNNs: Binary volumetric convolutional neural networks for 3D object recognition. BMVC2017.

[23] Shuaifeng Zhi, Yongxiang Liu, Xiang Li, Yulan Guo Towards real-time 3D object recognition: A lightweight volumetric CNN framework using multitask learning Computers and Graphics (Elsevier)

[24] Charles R. Qi, Hao Su, Matthias Niessner, Angela Dai, Mengyuan Yan, and Leonidas J. Guibas. Volumetric and Multi-View CNNs for Object Classification on 3D Data. CVPR 2016.

[25] Yiru Shen, Chen Feng, Yaoqing Yang and Dong Tian Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling. CVPR 2018