Lidar Image Classification based on Convolutional Neural Networks

Yang Wenhui*,
School of computer science and engineering Xi'an Techonolgical University,Xi'an ,Shananxi 710021
e-mail:1353678463@qq.com

Yu Fan
School of computer science and engineering Xi'an Techonolgical University,Xi'an ,Shananxi 710021
e-mail:yffshun@163.com
*The correspoding author

Abstract—This paper presents a new method of recognition of lidar cloud point images based on convolutional neural network. This experiment uses 3D CAD ModelNet, and generates 3D point cloud data by simulating the scanning process of lidar. The data is divided into cells, and the distance is represented by gray values. Finally, the data is stored as grayscale images. Changing the number of cells dividing point cloud results in different experimental results. Experiments show that the proposed method has higher accuracy when dividing the cloud with 27×35 cells. Comparison of point cloud cell image method with VoxNet method, *experimental results show that the classification method based on gray image and convolutional neural network has more advantages than the most advanced point cloud recognition network Voxnet.*

Keywords-Point Cloud; CNN; Gray Image; Lidar;

I. INTRODUCTION

Lidar has been widely used in the acquisition of point cloud data because of its high precision and wide range of visibility. The classification and recognition of point cloud images formed by lidar has been the focus of many domestic and foreign famous experts and scholars. The key technology and the final aim of this method are feature extraction and classification of point cloud images.

Domestic and foreign famous experts and scholars have done a lot of research work on it. Some scholars use manual extraction of features, and then select a classifier for classification and recognition methods. R. B. Rusu and others use the relation between the normal vectors of a region as feature [1], and classify objects by classification. Yasir, Salih and others use VFH as a feature, and SVM is used as classifier to classify and identify point clouds. [2]. Liu Zhiqing and others improved classifier, and used information vector machine to classify and identify point cloud [3]. The laser scanning point cloud data, sparse and incomplete, effective and accurate description of artificial feature selection is often difficult. n Researchers need professional knowledge and heuristic methods, which rely on personal experience to a great extent. While deep learning can automatically extract features and classify them, they are invariant to displacement, scaling, and other forms of rigid body change. Therefore, in recent years, some experts and scholars have begun to use deep learning to classify point cloud images. Daniel, Maturana and Sebastian Scherer processed the data into a physical form, the feature was extracted by using the three-dimensional convolution kernel, and the neural network was used to classify and recognize [5]. Good results were obtained.

Lidar data is difficult to obtain relative to visible image data. Some experts use Sydeny Urban Objects data sets to train the network, but this dataset is too small. It can not achieve the purpose of data-driven. Therefore, this experiment uses the ModelNet data sets with rich data types as the basis, and preprocess the data, then use the convolution neural network to classify and recognize. The VoxNet recognition method proposed by Daniel, Maturana and others achieves higher accuracy in the 3D data set, while the cloud point image is not fully used in the recognition of the laser point cloud image, thefore the accuracy is not very high.

In order to compare, the ModelNet data set is classified and identified by using the VoxNet method and the point

cloud cell image proposed in this experiment. By experiment , the cloud cell image method achieves higher accuracy.

II. MODEL CONSTRUCTION BASED ON CONVOLUTIONAL NEURAL NETWORK

In this paper, the formation and principle of convolutional neural networks are briefly introduced. The network structure of VoxNet and point cloud cell images is compared. The experimental scheme is designed, and the accuracy of the two networks is compared. Some useful conclusions are obtained.

A. Convolutional neural network

Convolutional neural network is inspired by the neural mechanism of visual system, it is a kind of deep learning ability of artificial neural network system. Compared with the traditional method, convolution neural network has the advantages of strong applicability, feature extraction and classification at the same time, strong generalization ability, can be used to recognize the change of rigid body displacement, zoom and other forms about two-dimensional or three-dimensional image. It has become the focus of the field of machine learning at present [6].The standard convolutional neural network is a special multilayer feedforward neural network. It has a deep network structure, which is generally composed of input layer, convolutional layer, down sample layer, full connection layer and output layer. Among them, the Convolutional layer, the down sample layer and the full connection layer are hidden layers. In the convolutional neural network, input layer is usually a matrix for receiving original image; convolution layer for image feature extraction, reduce the noise interference; convolution layer sharing local weights, the special structure is more close to the real biological networks, the CNN has a unique advantage in the field of image processing. Compared with the full connection layer, the shared weight reduces the network parameters and accelerates the training speed. On the other hand, the complexity of the network is reduced, and the multidimensional input signals (voice and image) can be input directly, thus avoiding the process of data rearrangement during feature extraction and classification [6]. The down sample layer reduces the amount of data to be processed according to the principle of local correlation of the image, and the output layer maps the extracted features to the predicted tags.

Convolution of convolutional maps in convolution neural networks is discrete and can be expressed as Eq.（1）:

$$x_{\beta}^{\gamma} = f(\sum_{\alpha \epsilon M_{\beta}} x_{\alpha}^{\gamma-1} k_{\alpha\beta}^{\gamma} + b_{\beta}^{\gamma}) \qquad (1)$$

The weight and bias of CNN can be learned by back propagation algorithm, so it is not necessary to extract features manually. The convolution neural network uses the classical BP (back propagation) algorithm to adjust the parameters, and finally completes the learning task. BP network update weights for Eq.（2）:

$$\omega(n+1) = \omega(n) - \eta \frac{\partial C}{\partial \omega} \qquad (2)$$

$\omega(n)$ represents the nth map , and $\omega(n+1)$ represents the (n+1)th map. η represents Learning rate.$\frac{\partial C}{\partial \omega}$ represents the loss function, can be obtained by back propagation.

B. Model comparison

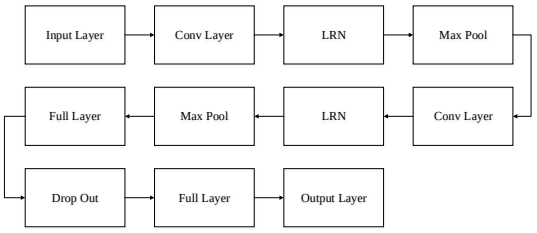

Figure 1. Point cloud cell network structure diagram

Figure. 1 is a convolutional neural network structure for testing in this paper. A total of two volumes are formed. The number of feature maps is 10 and 15, and the size of the convolution kernel is 8 and 5.Each convolutional layer has a 2×2 maximum pool layer for preventing over fitting and a LRN layer for local normalization. The discard rate of dropout layer is 0.5. The number of neurons in two fully connected layers was 256 and 10, respectively.

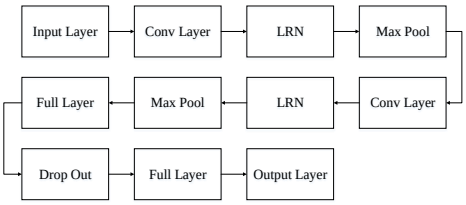

Figure 2. VoxNet network structure diagram

Figure. 2 is a network structure of the VoxNet, and the input layer accepts data in the form of $32 \times 32 \times 32$. A total of two volumes are laid, and the number of feature maps is 32, using $5 \times 5 \times 5$ and $3 \times 3 \times 3$ convolution kernels, respectively. The discard rate of the Dropout layer is 0.2 and 0.3, which can prevent overfitting and reduce the amount of computation. The largest pool layer uses $2 \times 2 \times 2$ filter. Finally, there is a full layer of neurons with a number of 128 and a dropout layer with a discard rate of 0.4. The seventh layer is the output layer, and the number of neurons is 10.

VoxNet divides the processed data into $32 \times 32 \times 32$ cells, which will lose some of information. In this paper, the data is projected into a point cloud cell image, which can save all the data and improve the utilization of data, thus improving the classification accuracy.

III. EXPERIMENTS

A. *Experiment environement and Datasets*

We use TensorFlow-gpu 0.12.1 open source software library, Windows 7 operating system, NVIDIA GTX 950 graphics card in our experiments. The experiment uses data for Princeton University's ModelNet which is a large 3D CAD model database similar to the ImageNet. ModleNet10 is a subclass of ModelNet that contains 3991 CAD models and 10 categories, all of which are distinct and located in the middle. This paper uses ModelNet10 as the experimental data set.

B. *Reference frame and resolution*

ModelNet stores data as OFF format, and OFF format represents the geometric structure of a model by describing polygons on the object surface. We save OFF format data as STL format by 3ds Max, STL using triangle to represent an object model is more conducive to the data processing in this experiment. We simulate the scanning process of laser radar, and use the gravity center method to determine whether the laser is scanned on the triangle [11], thus get a three-dimensional image with only one side. VoxNet divides the data into $32 \times 32 \times 32$ cells. The proposed method in this study need different treatment to the data, firstly using the plane grid of $216 \times 280, 108 \times 140, 54 \times 70, 27 \times 35$ to partition the point cloud data, the formation of the matrix corresponding to the size, because each grid unit may correspond to multiple point cloud data, each matrix element values for the corresponding grid unit is all the average distance of point cloud. In order to compare the recognition effect under different partition methods, the matrix of different partition method is transformed into a gray image with resolution of$216 \times 280, 108 \times 140, 54 \times 70, 27 \times 35$ and the point cloud cell image is obtained, as shown in Figure 3.

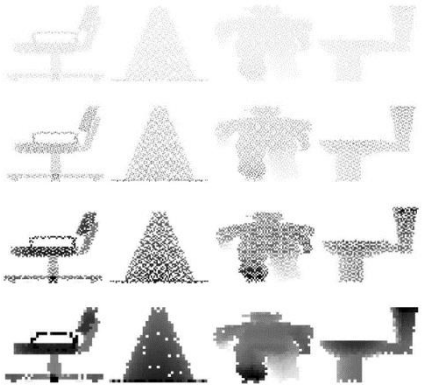

Figure 3. Gray image of point cloud divided by different cells

Recognition result

When the number of cell division, and the point cloud image sparse number of point cloud, it is easy to produce some grid corresponding to zero, which represents the corresponding grid distance value is zero, but the distance of adjacent grid corresponding to the actual situation value should be smooth distribution. If used directly in network training, there will be greater deviations. In order to get better results, the experimental data will be divided into different cells, were divided into 216×280, $108 \times 140, 54 \times 70, 27 \times 35$ cells in different partition methods

have different results. The accuracy of the experiment, as shown in Table 1, can be seen:

TABLE I. DIVIDES THE ACCURACY RATE INTO DIFFERENT CELLS

Number of cells	accuracy rate
$\mathbf{216 \times 280}$	0.735
$\mathbf{108 \times 140}$	0.808
$\mathbf{54 \times 70}$	0.776
$\mathbf{27 \times 35}$	0.858

1) *The cells are divided differently, and the results are different. When the image is mapped to* 216×280 *cells, the accuracy is lowest, and the accuracy is the highest when it is mapped to* 27×35 *cells. As the cells are smaller, the number of pixels in each cell increases and the number of points in the space is less, so a higher accuracy can be achieved.*

2) *The* 108×140 *cells have higher accuracy than* 54×70 *cells, because the distribution of the points in the* 108×140 *cells is more uniform, and better results can be obtained.*

When a point cloud is divided into grayscale images of different cells, its accuracy rate with the number of training times as shown in Figure 4. At the beginning of the network training, the network model is not fully trained, so the accuracy is low. With the increase of the number of iterations, the parameters of the network are also learning constantly, and the accuracy of classification recognition will gradually increase. Finally, the classification accuracy fluctuates in a small range, which means that the convolutional network is convergent and the classification accuracy is stable.

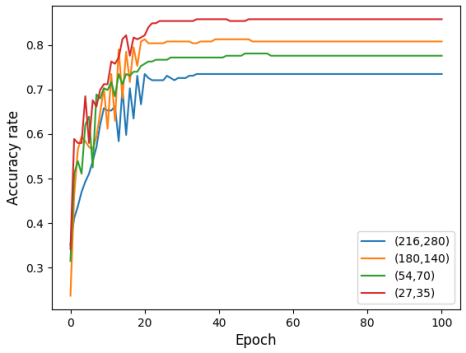

Figure 4. Relation between training times and recognition rate

C. Comparison of point cloud cell image method with VoxNet method

In order to test the accuracy of the experimental method, compare with VoxNet network which performs well in 3D recognition. VoxNet uses 3D point cloud data to classify and convert them into stereo cells, and the proposed network uses 3D point cloud data to divide gray images into input data. The accuracy of VoxNet classification recognition is 78.5%, and the recognition rate of the best segmentation results of cloud cell images is about 6% higher than the VoxNet. The experimental results are shown in table 2.

TABLE II. COMPARISON OF ACCURACY BETWEEN POINT CLOUD CELL IMAGE AND VOXNET RECOGNITION

Network type	VoxNet	Point cloud cell image
Recognition accuracy	0.785	0.858

IV. CONCLUSION

A classification and recognition method of point cloud based on convolutional neural network is proposed in this paper. Firstly, the point cloud data is processed as point cloud cell image, and then the network is used to classify and recognize the image. Experiments show that the classification method based on gray image and convolutional neural network has more advantages than the most advanced point cloud recognition network Voxnet.

REFERENCES

[1] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In Robotics and Automation, 2009. ICRA'09. IEEE International Conferenceon, pages 3212–3217. IEEE, 2009.

[2] Yasir Salih, A.S. Malik, D. Sidibé, M.T. Simsim, N. Saad and F. Meriaudeau .ompressed VFH descriptor for 3D object classification. 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2014.

[3] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015, pp. 1912–1920.

[4] Daniel Maturana,Sebastian Scherer.VoxNet: A 3D Convolutional Neural Network for real-time object recognition. Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on.

[5] J. Bergstra, F. Bastien, O. Breuleux, P. Lamblin, R. Pascanu, O. Delalleau, G. Desjardins, D. Warde-Farley, I. Goodfellow, A. Bergeronet al., "Theano: Deep learning on gpus with python," in NIPS 2011,BigLearning Workshop, Granada, Spain, 2011.

[6] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao 3D ShapeNets: A Deep Representation for Volumetric Shape Modeling Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015).

[7] The HDF Group. Why Use HDF?. Retrieved January 4, 2012, from https://www.hdfgroup.org/why-hdf/.

[8] C. R. Qi, H. Su, M. Nießner, A. Dai, M. Yan, and L. Guibas.Volumetric and multi-view cnns for object classification on 3d data. In Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 2016.

[9] http://www.cnblogs.com/graphics/archive/2010/08/05/1793393.html The Princeton ModelNet. http://modelnet.cs.

[10] The Princeton ModelNet. http://modelnet.cs.

[11] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proc. CVPR, 2014.