Indoor Object and Scene Recognition Based on Deep Learning

ZHU Lu-lin[1], WANG Ya-qi[1], CUI Han[1], GUO Li-ru[1]

(1.North China Electric Power University,BaoDing,071003,China)

Abstract：Deep learning is an important breakthrough in artificial intelligence this decade, convolution neural network (CNN) of deep learning theory is increasingly used to solve the problem of object recognition. In this paper, we propose a method of deep learning to achieve accurate image of indoor object and scene recognition from the complex background. Firstly, the convolutional neural network model of Pre-ResNet for feature extraction and recognition is introduced; Secondly, we build the indoor object and scene image data sets. Experimental results show that in indoor object and scene recognition accuracy reaches 87.84% and 91.80% respectively, that proves the method has high recognition accuracy in image of complicated background, and has certain feasibility and practical value.

Key words：deep learning; convolutional neural network; indoor object recognition; indoor scene recognition; Pre-ResNets

1 Introduction

The development of artificial intelligence in recent years has brought great changes to production and life, and the rapid development of deep learning has brought about a major breakthrough in artificial intelligence. It has been a great success in speech recognition, natural language processing, computer vision, image and video analysis, multimedia and many other fields. The recognition of indoor objects and scenes is of great significance to indoor positioning, such as indoor robot navigation, smart home and so on, strengthening the robot's ability to identify indoor objects for indoor robot navigation, supporting the further development of smart home, making people's lives more comfortable, and realizing indoor scene recognition for indoor safety monitoring and navigation. In the past 5 years, deep convolutional neural networks have revolutionized the field of computer vision, with encouraging results in computer vision tasks, in particular the convolutional neural network (CNN) AlexNet proposed in 2012 to achieve the best results in the classification of objects in the ImageNet dataset, and Zhou proposed a feature extraction method for training CNN in image dataset Places on NIPS in 2014. Places-CNN, with a structure similar to AlexNet's, gets the best performance across multiple scenario classification datasets. Convolutional neural networks have been developing, the number of convolutional layers is increasing, from the 8th floor of AlexNet, to the 19th floor of VGG, 3, and 22 layers of GoogleNet, to the thousands of layers of ResNets, intelligent network learning ability is constantly improving, image recognition performance is also improving. Therefore, this paper applies the convolutional neural network in deep learning to the identification of indoor objects and scenes, establishes the image database of indoor objects and scenes, and uses the pre-ResNets network with superior recognition performance to extract and identify the indoor images.

2 Pre-ReNets NetWork

Deep convolutional neural networks have problems such as gradient disappearance, He7 and others thus proposed residual network (ResNets), ResNets outside the original convolution layer to add cross-layer connection branches to form the basic residual block, so that the original mapping H(X) is represented as H (X) , F (X) x, where F (X) is called residual mapping, x is the input signal. ResNets transforms network-to-H(X) learning into F(X)

learning through residual block structures. Based on the more learnable nature of residual blocks, ResNets alleviates degradation and gradient disappearance in deep CNN training by stacking the residual mappings sequentially.

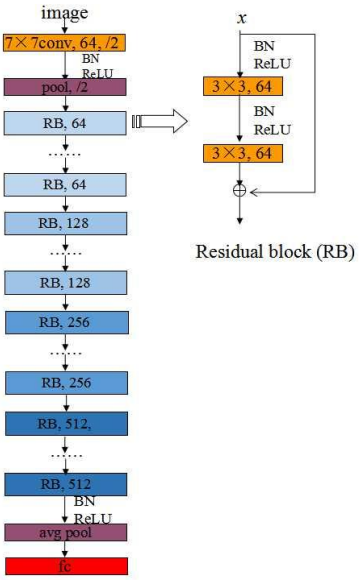


Figure 1 34-layer Pre-ReNets network

This paper uses the improved residual network Pre-ResNets. Unlike ResNets, Pre-ResNets places ReLU before the convolution layer, passing data by constructing a channel in the network, making training easier, alleviating the problem of gradient disappearance, and training to optimize thousands of layers of network. In this paper, the 34-layer Pre-ResNets network is used, the structure of which is shown in Figure 1, entering an RGB picture of 224×224, and after a convolution core of 7×7 and a step of 2 convolution layers, the output feature graph is 112×112 The maximum pooling layer is followed by a downsampling, followed by four sets of residual blocks, each containing a number of residuals of 3, 4, 6, and 3, and the same output dimensions for each group, 64, 128, 256, and 512, respectively Finally, the classification results are output for the average pooling layer and the full-connection layer softmax.

3 Data Set

Because CNN has high order of magnitude requirements for data sets, it is difficult to obtain data sets manually, and this paper uses existing large image datasets to filter the desired indoor scene and indoor object images. The indoor scene dataset is obtained from the Places dataset, and the indoor road sign object dataset is obtained from the ImageNet dataset. Using these filtered images to build a dataset, and the data set of five-sixths of the training set, one-sixth of the test set, the image in terms of angle, background and light condition quality are very different, so recognition is very challenging.

Table 1 Indoor object dataset

Indoor object	Training set	The test set	total
bed	1063	212	1276
desk	1084	216	1300
bookshelf	1084	216	1300
table	1084	216	1300
Windows	1084	216	1300
TV	1084	216	1300
Sofa	1084	216	1300
trash can	1084	216	1300

Table 2 Indoor scene dataset

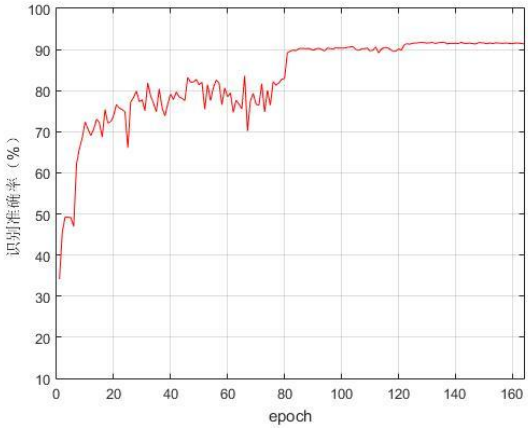
Indoor scene	Training set	The test set	total
bedchamber	12500	2500	15000
restaurant	12500	2500	15000
corridor	5144	1148	6892
playroom	4942	988	5931
office	11535	2307	13842

The built indoor object set contains 8 types of indoor common objects, the specific data information as shown in Table 1, the indoor scene set contains a total of 5 types of indoor common scenes, the specific data information as shown in Table 2.

4 Experiment results and analysis

In order to verify the validity of this method, this paper uses 34 layers of Pre-ResNets to train and test on indoor object set and indoor scene set. Trained with Titan X GPU, the network implementation environment is Torch 7. The image is batched (batch), the batch size is 32, the network parameters are updated by random gradient drop algorithm, the maximum Epoch for training is 164, the learning rate

starts at 0.1, the learning rate starts at 0.01, and then drops to 0.001 after 121 Epoch. The recognition accuracy curve obtained from each Epich training test set is shown in Figures 2 and 3, and the experimental results show that after the network converges, the recognition accuracy of indoor objects and scenes is good, and the recognition accuracy reaches 87.84 percent and 91.80 percent, respectively.



F

Figure 2 Indoor object recognition accuracy

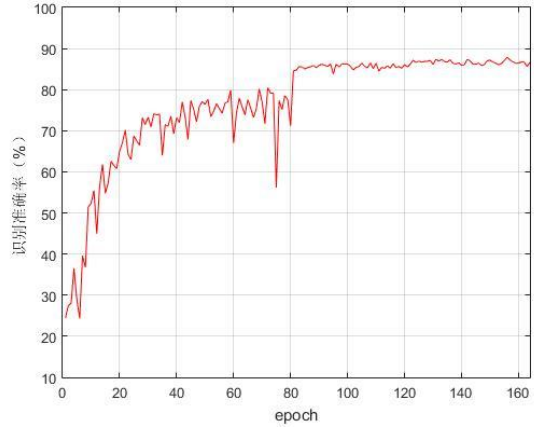


Figure 3 Indoor Scene Recognition Accuracy Rate

5 Conclusion

In this paper, the recognition of indoor objects and scene images is realized by using convolutional neural network through GPU platform. First, the convolutional neural network model Pre-ResNet is introduced for feature extraction and recognition, and secondly, the indoor object and scene image data set is constructed. The experimental results show that the accuracy of test sample recognition in indoor object and scene recognition reaches 87.84 percent and 91.80 percent respectively.

Bibliography:

[1] 郑胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. 中国图象图形学报, 2014, 19(2):175-184

[2] A. Krizhenvshky, I. Sutskever, and G. E. Hinton, Imagenet: classification with deep convolutional networks, in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097-1105.

[3] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2015 .

[4] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[5] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[6] Zhou B, Garcia A L, Xiao J, et al. Learning Deep Features for Scene Recognition using Places Database[J]. Advances in Neural Information Processing Systems, 2014, 1:487-495.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[8] He K, Zhang X, Ren S, et al. Identity mappings n deep res idual networks[C]//European Conference on Computer Vi sion. Springer International Publishing, 2016: 630-645.

Fund project: Central College Fund（2016MS99）

Received date: 2017-4-9

About the Author:

Zhu Lulin, female, born in Baoding City, Hebei Province, student, undergraduate, main research direction is deep learning;

Cui Han, female, from Baoding City, Hebei Province, student, undergraduate, main research direction is deep learning;

Wang Yaqi, female, from Shuozhou City, Shanxi Province, student, undergraduate, main research direction is deep learning;

Guo Liru, female, born in Baoding City, Hebei Province, student, master, main research direction is deep learning and computer vision.

Contact: Zhu Lulin

Address: North China Electric Power University, No. 689, Huadian Road, North District, Baoding City, Hebei Province

Postcode: 071003

Email: 744904583@qq.com

Tel: 13722977073