Face Recognition Research Based on Fully Convolution Neural Network

YangWang[1,a]„Jiachun Zheng[1,b,*]

[1]Information Engineering College,Jimei University,Xiamen 361021,China

a1510109901@qq.com,

[b]jchzheng@jmu.edu.cn.

Keywords: Face recognition; Caffe; CNN; feature extraction

Abstract: Face recognition technology has always been a hot topic, and has a widely application in our daily life. In recent years, with the rapid development of deep learning and the birth of various frameworks, face recognition is provided a new platform and ushered in a new opportunity. In this paper, a fully Convolution Neural Network(CNN) based on the Caffe framework and GPU is put forward for face recognition. The accuracy rate of it reaches 99% in the test set. Compared with the traditional extraction feature combined with the classifier method, CNN has a unique advantage, and doesn't need to manually extract features.

1. Introduction

In 2006, Geoffrey Hinton, a professor of machine learning, published a paper on the world's top academic journal Science, which led to the development of deep learning in the field of research and application. In 2011, Microsoft proposed a voice recognition system based on neural network, the existing results completely changed by the results of voice recognition system. Google and Baidu also launched their own network model. The NEC Research Institute first proposed neural networks into natural language processing. In 2012, Alex used the deeper convolution neural network model to achieve the best results in the Image Net game, reducing the original error rate by 9%, making a big step forward in the study of image recognition. Alex builds deep neural network model training by using the original natural image without using any artificial feature extraction methods. This result makes the deep neural network by more and more people's attention.

Face recognition is a hot topic in deep learning and machine vision. The development of deep neural network provides a new platform for face recognition. Face recognition has a wide range of applications: public places demographic and video surveillance, electronic passports and identity cards, safe driving reminders and other related areas. The traditional method of face recognition is to extract human facial features combined with the classifier to identify the face area. Common features include: Haar[1], HOG[2], LBP[3], and Adaboost [4], SVM[5] and other classifiers. The convolution neural network (CNN) has unique advantages: local perception, weight sharing and multi-convolution kernel, and has achieved great success in image processing. With the advance of deep learning, there are many deep learning framework. Each frame has its unique characteristics. In the paper, I choose the Caffe, a clear, flexible deep learning framework, Caffe was developed by Yangqing Jia during his Ph.D. Caffe itself provides a complete toolkit for training, testing, and deploy models. Tt's easy to get started.

2. Experimental principle

2.1 Fully Connected Convolution Neural Network

The traditional image recognition methods based CNN typically use an image around block as CNN's input for training and prediction. This method has several shortcomings: First, storage space is very large. Second, the calculation efficiency is poor. The adjacent pixel blocks are substantially

repeated calculation. Third, the size of the pixel block limits the size of the perceived area. Usually the size of the pixel block is much smaller than the size of the whole image, and only some local features can be extracted, resulting in the performance of the classification being limited.

In response to this question, we use Fully Convolutional Networks (FCN). FCN has two obvious advantages: first, it can accept any size of the input image, without having to require all the training images and test images have the same size. Second, FCN is more efficient because it avoids the problem of repeated storage and computational convolution due to the use of pixel blocks.

2.2 The calculation of Receptive Filed

Receptive field is some nodes of output feature map corresponding the input image area. Output layer can be calculated by the following equations:

$$start_{out} = start_{in} + \left(\frac{k-1}{2} - p\right) \qquad (1)$$

$$start_{in} = s_i * p_{i+1} \qquad (2)$$

There are the kernel size k, the padding size p, and the stride size s, i-th layer input a point p_i.It calculates the center position of the receptive field of the output feature,
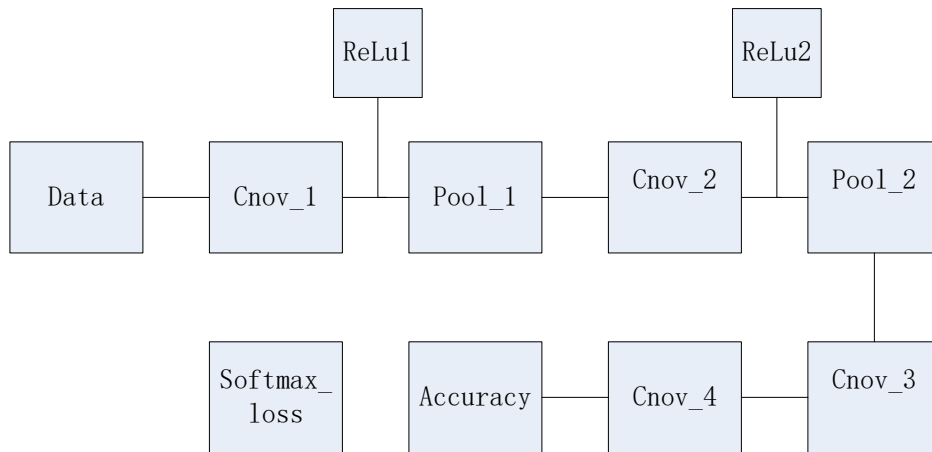
2.3 FCN Model

Fig.1 Fully Convolutional Networks

In the literature [6] proposed FVPN neural network, which has achieved good results in the vehicle detection. FVPN neural network has three branches of the last layer, respectively, on behalf of different tasks output. In this paper, the FVPN neural network is improved, leaving only one layer of Conv_fc_class for a task of 2 categories.The conv_1 contains 32 kernels of size 5*5. The conv_2 takes as input the feature maps obtained from the previous layer and filters them with 64 kernels of size 5_5. Max pooling and ReLU layers are configured after the first two convolutional layers.Pool_2 and pool_1 has the same parameters, step is2, kernel_sizeis 1.Theconv_3 includes 64 kernels of size 3*3,step is 1. Conv_4 contains 2kernels of size 10*10，stride is 1.Finally, it's Accuracy and Softmax-loss layer, output loss value.

3. Experiment results and analysis

The computer configuration in this experiment is:Intel（R）Core i7,3.6GHz，RAM 8GB,GPU：GeForce GTX 760. Windows 7, 64-bit operating system.

In this paper, we select CelebA dataset as positive sample. Crop the picture, part of the pictures is shown in the Fig.2. The pictures are divided into two groups according to 2: 1, respectively, as a training set of positive samples and test set positive samples. At the same time，we select about 300,000 non-human face images as a training set negative samples and 80,000 non-human face pictures as a test set negative sample. The picture is normalized to a size of 60 * 60.
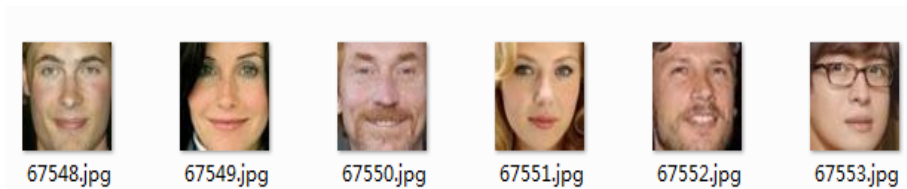
Fig. 2 Partial crop images

In order to reduce the training time and improve the accuracy, we use the GPU model. Training time is about 132 minutes. we get train loss curve and accuracy curve.
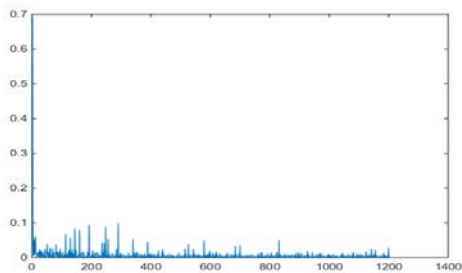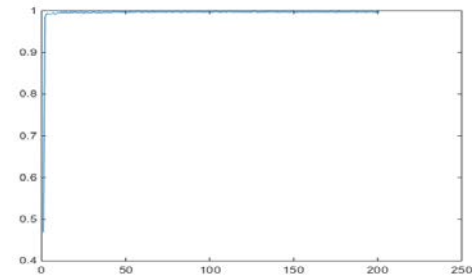

Fig.3 Train loss curve

Fig.4 Accuracy curve

According to the train loss curve and accuracy curve, face recognition has achieved good results and correct rate reaches 99%.In order to reduce the impact of the background and improve the detection rate, I use non-maximum suppression algorithm to eliminate cross-repeated window，detect and mark out face. Of course, for the complex background of image，test result is not very good, There will be a situation of false detection, the selected samples need to be further optimized. For deep learning, sample selection is more important than the algorithm.

Fig. 5 The results show

4. Conclusions

We design a FCN model and achieve good results. Face recognition correct rate reach 99%. Compared with the traditional method of face recognition, it's doesn't need to manually extract features. Face detection has also made a good effect, face detection will appear false detection in complex background. While, It's lay the foundation for the follow-up study. Face recognition is widely used, it is worth further exploration.

5. Acknowledgments

*Corresponding author: Jia-chun ZHENG; Tel: +86-13806073677; Fax: +865926180950; E-mail: jchzheng@jmu.edu.cn.; Address: No.185 Yinjiang Rd., Jimei District, Xiamen,361021, Fujian China.

References

[1] C.zheng,B.xiaojuan,W.Yu, Fatigue driving detection based on Haar feature and extreme learning machine, The Journal of China Universities of Posts and Telecommunications.23(2016),pp.91-100.

[2] R. Abduljalil, Z. Nasharuddin, A.Shahrel, Automated segmentation of iris images acquired in an unconstrained environment using HOG-SVM and GrowCutOriginal Research Article, Digital Signal Processing, 64(2017) , pp. 60-70.

[3] L. Yiming, Z. Ximei, W. Guodong, Y. Kexin u, et al, A novel approach for cirrhosis recognition via improved LBP algorithm and dictionary learning Original Research, Biomedical Signal Processing and Control, 38(2017), pp. 281-292.

[4] D. Cui-huan, Z. Hong , L. Li-ming, Face detection in video based on AdaBoost algorithm and skin model, The Journal of China Universities of Posts and Telecommunications, 20(2013), pp. 6-9.

[5] B. Saman, G.Eric, S.Robertetal, Dynamic ensembles of exemplar-SVMs for still-to-video face recognitionOriginal Research Article, Pattern Recognition, 69(2017), pp. 61-81.

[6] Z.Yi Zhouy, L.Li, S. Ling Shaoy, DAVE: A Unified Framework for Fast VehicleDetection and Annotation, 14th European Conference on Computer Vision,2016.