

Face Recognition Based on Deep Learning Under the Background of Big Data

Hongbiao Ni

Department of Information Engineering, Jilin Police College, Changchun 130117, Jilin, China

E-mail: nhbhongb@yeah.net

Keywords: big data, deep learning, face recognition, CNN, loss function

Received: December 8, 2020

Face recognition has important value in real life. In this study, the application of the deep learning method in the field of face recognition was studied. The structure of LeNet-5 in convolutional neural network (CNN) was selected and improved; based on it, a face recognition method was designed. The performance of the method was analyzed taking CelebA as training set and LEW as testing set. The results showed that the improved LeNet-5 model which took A-softmax Loss as loss function not only had shorter training time, but also had higher recognition accuracy, its accuracy increased with the increase of sample size, and the highest accuracy rate reached 97.9%. The experimental results showed that the face recognition method designed in this study had good performance in large data background as it could effectively reduce the running time of the algorithm and improve the recognition accuracy. This study proves the reliability of deep learning methods such as CNN in face recognition, which is conducive to the further development of face recognition technology.

Povzetek: Opisano je prepoznavanje obrazov z metodami globokih nevronskih mrež in z velikimi podatki.

1 Introduction

With the development of computer technology and in the context of big data, people pay more attention to issues such as data security and personal privacy, and the social requirements for human identification are also increasing. Traditional identification methods based on identity cards and passwords have low reliability because they are easy to be counterfeited and lost. Therefore, biometric identification technologies such as fingerprints and voices have been widely recognized [1]. Face recognition is a kind of biometric recognition, which has attracted more and more research and attention. However, due to the difference of face pose and illumination, face recognition is difficult [2]. The deep learning method has excellent performance in face recognition, especially in big data processing [3], and relevant research is also deepening. Ding et al. [4] studied the recognition of face images with severe noise. Based on the deep neural network, an anti-noise network was designed, and the reliability of the network in face recognition with noise was proved by experiments. Lu et al. [5] proposed a deeply coupled ResNet model, which was composed of a relay network and two branch networks. It could extract various possible resolutions of images, and the reality of the model was proved by experiments in LFW and SCface databases. Jiang et al. [6] designed an unsupervised deep learning network by combining 2-D Gabor filter with PCA to improve the computing speed through short binary hashing and then proved the excellent performance of this method by testing in face database. Singh et al. [7] applied convolutional neural network (CNN) to neonatal recognition and found that CNN had a good accuracy in

neonatal recognition compared with conventional technology and CNN with two convolution layers and one hidden layer had the highest accuracy. In this study, deep learning was analyzed. Based on CNN in deep learning, a face recognition method was designed. The reliability of the method was proved by LFW data set, which provides some theoretical support for the further application of deep learning in face recognition.

2 Face recognition

Face recognition refers to extracting feature information from static or dynamic images collected by computer and then analyzing and matching to realize identity recognition. Compared with other biometric methods, face recognition image acquisition is more convenient, with rich personal characteristics, high recognition degree and good interaction. It has been widely used in surveillance video, intelligent consumption [8], criminal investigation [9] and so on.

Traditional face recognition methods include geometric features, template matching and so on, but there are also some shortcomings. Face feature extraction is a very important step in recognition, which has a great impact on the final results. In traditional recognition methods, feature extraction is mostly based on manual method. Under the background of massive data, the traditional recognition methods not only take a lot of time and energy, but also are difficult to recognize images because they are easily affected by illumination, occlusion and other factors. Deep learning can automatically extract

features, which is less affected by external factors, and it has been proved to have good recognition effect.

3 CNN

3.1 Overview of CNN algorithm

CNN is a common model of deep learning. Its basic structure is shown in Figure 1.

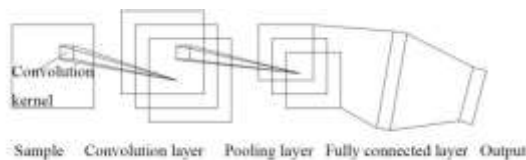


Figure 1: The structure of CNN.

(1) Convolutional layer

Convolution layer is the core component of CNN. It extracts image features by convolution operation, generates different feature maps by different convolution kernels and superimposes them to obtain various features of input image. Its output calculation method is:

$$y_j^l = f \left(\sum_{i=1}^{N_j^{l-1}} w_{i,j} \otimes x_i^{l-1} + b_j^l \right), j = 1, 2, \dots, m$$

where l represents the current number of layer, w represents the convolution kernel weight matrix, x_i^{l-1} represents the output characteristic pattern matrix, f represents an activation function, \otimes represents convolution operations, and b_j^l represents the offset of the j -th characteristic pattern of the l -th layer.

(2) Pooling layer

The role of the pooling layer is to compress data and reduce the amount of computation. There are two common methods, average pooling and maximum pooling. Figure 2 shows an example of maximum pooling. The size of image is 4×4 , the size of pooling window is 2×2 , and the step length of maximum pooling operation is 2. In the first pooling window, the values are 5, 7, 9 and 2 and the maximum value is 9; thus the maximum pooling result can be obtained

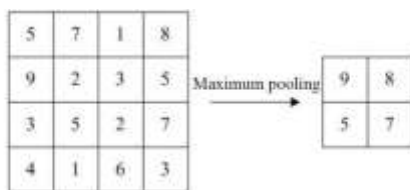


Figure 2: Maximum pooling.

by traversing the whole image.

(3) Fully connected layer

Fully connected layer plays the role of classification, and its calculation formula is:

$$\delta_j^l = f \left(\sum_{i=1}^n x_i^{l-1} w_{ij}^l + b_j^l \right)$$

where l represents the current level, n represents the number of neurons, w represents weights, b_j^l represents offset, and f represents an activation function.

3.2 Training process of CNN

The training process of CNN can be divided into two stages:

(1) Forward propagation

A sample (X, Y_p) is selected from the sample set, and X is input into CNN.

Actual output O_p of CNN is calculated.

(2) Reverse propagation

(1) The error between actual output O_p and expected output Y_p is calculated.

(2) The error is reversely propagated, weight matrix is adjusted, and parameters are optimized.

4 Face recognition based on deep learning

4.1 Experimental environment

The experiment was carried out on Ubuntu 16.04 operating system. The program was written in C++ language and Python language. The training and testing of CNN model was realized by Caffe framework, which supports GPU acceleration, runs faster and operates more simply.

4.2 Experimental data set

At present, data sets commonly used in face recognition include CAS-PEAL, CASIA-WebFace, LFW, MSCeleb, CelebA and so on. In this study, CelebA was selected as the experimental training set, and LEW was used as the testing set. CelebA can train the model well as it includes 200,000 face images of 10,177 people and there are changes in expression, posture, occlusion and illumination. LFW which has been widely used in the performance analysis of face recognition algorithms includes 13,233 images, a total of 6000 face combinations.

4.3 Data preprocessing

The main task of data preprocessing is face alignment. As the face image is partly inclined (Figure 3), the difficulty of recognition increases. Therefore, in order to obtain better recognition effect, image alignment is needed. The face images obtained after alignment are shown in Figure 4.



Figure 3: Face images.



Figure 4: Face images after preprocessing.

4.4 Improved LeNet-5

LeNet-5 is one of the most representative structures in CNN [10]. In order to improve the recognition performance of the network, the structure of LeNet-5 was improved in this study. Five convolution layers, four pooling layers and one fully connected layer were used. The specific parameters of each layer are shown in Table 1.

Type	Convolution kernel	Number of characteristic patterns	Number of neurons
Convolution layer 1	5×5	16	16128
Pooling layer 1	2×2	16	4032
Convolution layer 2	4×4	32	1536
Pooling layer 2	2×2	32	384
Convolution layer 3	3×3	64	128
Convolution layer 4	6×6	16	4032
Pooling layer 3	2×2	16	1008
Convolution layer 5	5×5	32	480
Pooling layer 4	2×2	32	256
Fully connected layer	-	-	192

Table 1: Improved LeNet-5.

In order to improve the training speed of the algorithm, an improved ReLU function, LReLU, was used as the activation function of the model:

$$LReLU(y) = \begin{cases} y, & \text{if } (y > 0) \\ ay, & \text{if } (y \leq 0) \end{cases}$$

where a represents a small constant, so that the function is not zero when the input is negative, preventing neuron necrosis.

There were two choices of loss function for the model: Softmax and A-softmax Loss:

(1) Softmax: For input x , it is divided into k classes, then the probability of sample belonging to class i can be expressed as:

$$g_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}; \theta) \\ p(y^{(i)} = 2|x^{(i)}; \theta) \\ \dots \\ p(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \dots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}$$

where $g_{\theta}(x)$ is a hypothetical functions and θ_i is a model parameter.

(2) A-softmax Loss: A-softmax Loss is an improvement of Softmax, which introduces angular distance and angular margin, and its expression is:

$$L_{ang} = \frac{1}{N} \sum_i -\log \left(\frac{e^{\|x_i\| \cos m \theta_{y_i, i}}}{e^{\|x_i\| \cos m \theta_{y_i, i}} + \sum_{j \neq y_i} e^{\|x_j\| \cos m \theta_{j, i}}} \right)$$

where m represents an integer, which is used for controlling the angular distance.

5 Experimental results

Images of 100 people were selected from CelebA to train the model, ten images each people. The training time of different models is shown in Table 2.

CNN model	Loss function	Training time
LeNet-5	Softmax	59.27 s
LeNet-5	A-softmax Loss	57.46 s
Improved LeNet-5	Softmax	45.39 s
Improved LeNet-5	A-softmax Loss	42.18 s

Table 2: Comparison of training time of models.

It was found from Figure 2 that the training time of LeNet-5 model was longer than that of the improved LeNet-5 model when using the same samples. In the same CNN model, the training time of the model which used A-softmax Loss as the loss function was shorter than that of the model which used Softmax function, and the training time of the improved LeNet-5 model with A-softmax Loss as the loss function was the least.

Taking A-softmax Loss as the loss function, two CNN models were tested using LFW data sets. 100 pairs, 500 pairs, 1000 pairs and 2000 pairs of matched face images were taken as positive samples; as shown in Figure 5, the two images matched each other, which was called a pair of positive samples. Mismatched face images were taken as negative samples; as shown in Figure 6, the two images did not match, which was called a pair of negative samples. The recognition results of the model can be divided into four cases, as shown in Table 3.

The recognition accuracy of the model = (TP+TN)/the total number of samples.

Under different number of samples, the recognition accuracy of the two models is shown in Figure 7.

It was found from Figure 7 that the recognition accuracy of the model increased with the increase of the



Figure 5: An example of positive sample.



Figure 6: An example of negative sample.

	Identified as positive samples	Identified as negative samples
Actual positive sample	True Positive (TP)	False Positive (FP)
Actual negative sample	False Negative (FN)	True Negative (TN)

Table 3: Classification of recognition results.

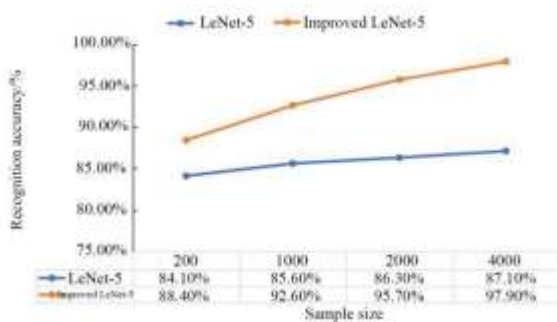


Figure 7: Comparison of recognition accuracy between models.

sample size, which showed that CNN model had excellent performance in recognizing massive face data and could accurately recognize large-scale data. From the comparison of the two models, it was found that the accuracy of the improved LeNet-5 was higher than that of LeNet-5. When the sample size was 4000, the recognition accuracy of LeNet-5 was 87.1%, while that of the improved LeNet-5 was 97.9%. The results showed that the improved CNN model could extract face features more comprehensively and obtain better recognition effect.

6 Discussion and conclusion

Deep learning is an important part of machine learning. It is based on big data and can automatically extract feature information from massive data by certain algorithms instead of traditional manual feature acquisition. It has higher accuracy than shallow learning and better

performance in dealing with non-linear problems. It has shown great advantages in fields such as computer vision and semantic analysis. CNN is one of the deep learning methods, which has been widely used in object recognition and detection. With the support of massive data, face recognition based on CNN has excellent performance [11].

In this study, CNN was analyzed firstly. Traditional recognition methods, such as SVM [12], can only extract shallow features when extracting image features, which is easily affected by other factors, and the recognition rate is not high. Deep learning methods such as CNN can extract abstract and conceptual features in depth [13], which is less disturbed by illumination, gesture and expression. CNN can extract multiple image features by convolution operation, then reduce the dimension by pooling layer to reduce the amount of calculation, and finally classify them. Based on LeNet-5 in CNN, the network structure was improved to make it more suitable for face image processing. Then, the improved ReLU function, LReLU, was used as activation function, and the influence of loss function on the performance of the model was analyzed. In the experiment, CelebA was used as training set to train the model, and then LEW was used as testing set to test the performance. The results showed that the improved LeNet-5 model using A-softmax Loss had shorter training time among LeNet-5 models using softmax and A-softmax Loss as the loss function and the improved LeNet-5 models, which showed that it had higher convergence speed. Then in the processing of LFW testing set, A-softmax Loss was used as the loss function, and the recognition accuracy of the improved LeNet-5 was significantly higher than that of LeNet-5. The recognition rate of the two models increased with the increase of sample size, and the gap between the two models increased as well. When the sample size was 4000, the recognition accuracy of LeNet-5 was 87.1%, while that of the improved LeNet-5 was 97.9%.

In summary, the face recognition method designed in this study has short training time and high recognition accuracy. It has excellent performance when facing a large number of face images. The reliability of deep learning methods such as CNN is proved, which makes some contributions to their further application.

7 References

- [1] Galbally J, Marcel S, Fierrez J (2014). Biometric Antispoofing Methods: A Survey in Face Recognition. *IEEE Access*, 2, pp. 1530-1552. <https://doi.org/10.1109/ACCESS.2014.2381273>.
- [2] Zhang K, Zhang Z, Li Z, et al (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 23, pp. 1499-1503. <https://doi.org/10.1109/LSP.2016.2603342>.
- [3] Pang SC, Yu Z (2015). Face recognition: a novel deep learning approach. *Journal of Optical Technology C/c of Opticheskii Zhurnal*, 82, pp. 237.
- [4] Ding Y, Cheng Y, Cheng X, et al (2017). Noise-resistant network: a deep-learning method for face

- recognition under noise. *Eurasip Journal on Image & Video Processing*, 2017, pp. 43.
- [5] Lu Z, Jiang X, Kot C (2018). Deep Coupled ResNet for Low-Resolution Face Recognition. *IEEE Signal Processing Letters*, pp. 1-1.
<https://doi.org/10.1109/LSP.2018.2810121>.
- [6] Jiang M, Lu R, Kong J, et al (2017). GB (2D) 2 PCA-based convolutional network for face recognition. *Neuroreha*, 06, pp. 131-135.
- [7] Singh R, Om H (2017). Newborn face recognition using deep convolutional neural network. *Multimedia Tools & Applications*, 76, pp. 1-11.
<https://doi.org/10.1007/s11042-016-4342-x>.
- [8] Smith DF, Wiliem A, Lovell BC (2015). Face Recognition on Consumer Devices: Reflections on Replay Attacks. *IEEE Transactions on Information Forensics and Security*, 10, pp. 736-745.
<https://doi.org/10.1109/TIFS.2015.2398819>.
- [9] Ghiass RS, Arandjelovic O, Bendada H, et al (2014). Infrared face recognition: a comprehensive review of methodologies and databases. *Pattern Recognition*, 47, pp. 2807-2824.
<https://doi.org/10.1016/j.patcog.2014.03.015>.
- [10] Zhao ZH, Yang SP, Ma ZQ (2010). License Plate Character Recognition Based on Convolutional Neural Network LeNet-5. *Journal of System Simulation*, 22, pp. 638-641.
<https://doi.org/10.3724/SP.J.1187.2010.00953>.
- [11] Wu W, Yin Y, Wang X, et al (2018). Face Detection With Different Scales Based on Faster R-CNN. *IEEE Transactions on Cybernetics*, PP, pp. 1-12.
<https://doi.org/10.1109/TCYB.2018.2859482>.
- [12] Zhang L, Zhou WD, Li FZ (2015). Kernel sparse representation-based classifier ensemble for face recognition. *Multimedia Tools & Applications*, 74, pp. 123-137.
<https://doi.org/10.1007/s11042-013-1457-1>.
- [13] Li Y, Lu Z, Jing L, et al (2018). Improving Deep Learning Feature with Facial Texture Feature for Face Recognition. *Wireless Personal Communications*, pp. 1-12.

