Research on Face Recognition Based on Gabor-LeNet Convolutional Neural Network Model

Qinxuan Dai 1, a, Xiaoshu Luo 1, b, *, Zhiming Meng 2, c

1 College of Electronic Engineering Guangxi Normal University Guiling, China
2 College of Innovation and Entrepreneurship Guangxi Normal University Guiling, China

a daixq1997@163.com, *, b Corresponding author e-mail: 2237596628@qq.com,
c 54615437@qq.com

Abstract. Aiming at the problems of slowing convergence speed, and low recognition accuracy in the face recognition training of the existing LeNet-5 convolutional neural network, an improved LeNet-5 convolutional neural network model is proposed and applied for face recognition. The main improvement is the use of Gabor filter to initialize the first convolutional layer and the activation function of Parametric Rectified Linear Unit (PReLU). The recognition accuracy of the improved model on ORL and GT face datasets has reached more than 98%. At the same time, the recognition accuracy rate on the face data set AR with occlusion has reached more than 90%, indicating that the improved model has strong robustness.

Keywords: Face recognition; convolutional neural network; structural improvement; activation function optimization; convolutional layer initialization.

1. Introduction

Face recognition technology refers to the use of computer technology for analysis and comparison to recognize faces. Face recognition is a popular field of computer technology research. The traditional methods of face recognition rely on a combination of artificially designed features (such as edges and texture descriptions) and machine learning techniques, such as principal component analysis, linear discriminant analysis, or support vector machines[1]. It is difficult to artificially design robust features for different changes in an unconstrained environment, which makes past researchers focus on special methods for each type of change, such as methods that can cope with different ages[2][3], energy Methods to cope with different postures [4], methods to cope with different lighting conditions, etc.[5][6]. In recent years, with the development of deep learning algorithm[7] and hardware acceleration technology proposed by Hinton et al., Convolutional Neural Network (CNN) has the ability to "learn" to extract features and carry out The characteristics of classification have been widely researched and applied in the field of face recognition[8][9]. Facebook released Deepface[10] based on a convolutional neural network, which achieved a recognition accuracy rate of 97.35% in the LFW (Labeled Faces in the Wild Home, LWF) data set. The recognition rate of DeepID proposed by Yi Sun[11] in the LFW dataset has reached 97.45%.
CNN generally has strong face recognition performance, but because of the parameter update based on gradient descent algorithm, CNN will have the problems of slow convergence and low recognition accuracy during training [12]. In order to solve these problems, this paper first studies the face recognition performance of LeNet-5 convolutional neural network; Then the convolutional layer initialization improvement, activation function optimization and structural improvement are carried out for the problem. The proposed improved LeNet-5 convolutional neural network model is more robust.

2. Convolutional network model

The LeNet-5 convolutional neural network model is shown in Fig.1 [13]. It can be seen that the LeNet-5 network structure is relatively small, consisting of a downsampling layer (Convolutions), a pooling layer (Subsamplings), and a fully connected layer (Full connection). LeNet-5 has a total of 8 layers including the input layer, and each layer contains multiple parameters. The model in Figure 1 includes the input layer, C1 convolution layer, S1 pooling layer, C3 convolution layer, S2 pooling layer, F5 fully connected layer, F6 fully connected layer and Output layer. The input layer is a 32×32 pixel image. C1 convolution layer contains 6 feature maps (feature map), 5×5 convolution kernels, the neurons of each feature map are connected to the 5×5 convolution kernel regions, and the convolution operation is performed to obtain six 28×28 feature maps. The S2 pooling layer contains six 14×14 feature maps, and the neurons in each feature map are connected to the 2×2 area in the feature map corresponding to the C1 convolution layer. The purpose of the pooling layer is to reduce the degree of overfitting of network training parameters and models. The pooling method of the pooling layer is usually divided into average pooling and maximum pooling. The C3 convolution layer uses a 5×5 convolution kernel to process the S2 layer. The C3 convolution layer has 16 feature maps, and each feature map is composed of different combinations of the feature maps of the previous layer. The S4 pooling layer contains 16 feature maps of 5×5 size, and each neuron in the feature map is connected to the 2×2 size area of the corresponding feature map in C3. The C5 convolution layer also uses a 5×5 convolution kernel, and the neurons of each feature map are convolved with all the feature maps of the S4 layer to obtain a 1×1 feature map. The F6 fully connected layer has 84 feature maps, all connected to the C5 layer. The output layer is also a fully connected layer, with a total of 10 nodes, using radial basis function (RBF) network connection.

According to the current research results of the relevant literature [14][15], LeNet-5 convolutional neural network has demonstrated excellent recognition performance for character recognition, generally the recognition rate of characters has reached more than 98%. However, the research of this neural network for face recognition is still rare. This paper first studies the original LeNet-5 convolutional neural network for face image recognition. The training process uses Sigmoid activation function and stochastic gradient descent (SGD) (stochastic gradient descent)[16] to update the weight parameters for network optimization. Every 100 times the network is trained, the recognition rate is counted. After 9900 trainings, the recognition rates were only 76.61%. From the results, it can be seen that the original LeNet-5 convolutional network is used for face image recognition, and its recognition rate is too low and has no application value. The following describes how to improve the LeNet-5 convolutional network and study its application in facial image recognition.

3. Improvement of lenet-5 convolutional network model

Aiming at the problems of LeNet-5 convolutional neural network model applied to face recognition, this paper makes the following improvements based on the LeNet-5 convolutional network topology results:

(a) Change the first convolutional layer to initialize based on Gabor filter;
(b) Use PReLU activation function for network optimization;
(c) Reduce the number of C3 convolutional layer feature maps in the original network to 10, and connect each feature map in S2 to each C3 feature map;

(d) Change the C5 convolutional layer in the original network to a fully connected layer F5, F5 and F6 have 1024 and 120 neurons respectively. The F6 layer calculates the dot product and offset between the input vector and the weight vector, and then passes it to the PReLU function to complete the feature classification.

3.1. Convolutional layer initialization based on Gabor filter

This paper first designed an improved filter library based on Gabor, and then used the Gabor filter in the library to initialize the first layer representing the basic properties of the improved neural network before training the network model. The Gabor function is a linear filter for edge extraction, used in various computer vision applications, such as edge detection and texture analysis. Similar to the human visual system, a filter bank is created from the Gabor filter, and if it changes, it responds to frequency and direction. The Gabor filter in the spatial domain is generated by a complex two-component Gabor function. These two components are Gaussian and sine plane wave functions. The formula of this Gabor function is given in (1):

\[ g(x, y) = w_r(x, y)s(x, y) \]  

(1)

\( w_r(x, y) \) and \( s(x, y) \) are Gaussian and sine functions, respectively. To convert the Gabor function in (1) into a two-dimensional filter, the Gabor function can be reconstructed in the manner of (2):

\[ g(x, y, \sigma, \theta, \lambda, \gamma, \psi) = \exp\left(-\frac{x^2 + y^2 \gamma^2}{2\sigma^2}\right) \exp\left(i(2\pi \frac{x}{\lambda} + \psi)\right) \]  

(2)

Where \( \sigma \) is the standard deviation of the Gaussian function, \( \theta \) is the direction of the Gabor filter, \( \lambda \) is the wavelength parameter of the cosine function, \( \gamma \) is the spatial view ratio factor, and \( \psi \) is the phase parameter of the cosine function.

In the method proposed in this paper, a Gabor filter library is created for the first layer of convolutional layers, which represents the basic level attributes of the improved neural network. The total number of Gabor filters created in the library is equal to the first layer of the improved neural network. The number of channels in the layer (i.e. the number of feature maps). Since the first convolutional layer of the improved neural network used in this paper has 6 feature maps, a total of 6 Gabor filters are generated in the library. The core parameters required for the created Gabor filter are determined by the distribution of the value ranges given in Table 1. Fig.2 is a partial example of the Gabor filter.

Gabor filters are similar to the basic level attributes learned in the first layer of the pre-trained improved neural network. Therefore, in the method proposed in this paper, instead of using a pre-trained improved neural network, a pre-generated Gabor filter and target are used to initialize the training data used by the first layer convolution layer of the relevant model for training.

Table 1. Gabor parameter range

| parameter | range |
|-----------|-------|
| \( \sigma \) | [2,21] |
| \( \theta \) | [0,360] |
| \( \lambda \) | [8,100] |
| \( \gamma \) | [0,300] |
| \( \psi \) | [0,360] |
3.2. Activation function optimization

The reason why the convolutional neural network based on the LeNet-5 model has a slow convergence rate in face recognition training is because the Sigmoid activation function used in the convolutional neural network has a gradient disappearance phenomenon when the gradient descent method is used to train the network. Equation (3) is the expression of Sigmoid function:

\[
\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}
\]  

(3)

In the back-propagation algorithm, to derive the activation function, equation (4) is the first derivative Sigmoid'(x) expression of Sigmoid(x):

\[
\text{Sigmoid}'(x) = \frac{1}{1 + e^{-x}} - \left( \frac{1}{1 + e^{-x}} \right)^2
\]  

(4)

Fig.3 is a graph of Sigmoid(x) and its first derivative Sigmoid'(x). From the figure, we can see that once the input falls into the saturation zone, Sigmoid'(x) will become close to 0, resulting in a very small gradient transmitted to the bottom layer. At this time, the network parameters are difficult to be effectively trained, and the gradients are easy to disappear, resulting in the network unable to converge.
In order to solve the problem that the gradient disappears and the network is not easy to converge, this experiment uses a PReLU\[17\] that adaptively learns parameters from the data. PReLU is shown in the following (5), where \( i \) represents a different channel.

\[
PReLU(x) = \begin{cases} 
  x_i & x_i > 0 \\
  a_i x_i & x_i \leq 0 
\end{cases}
\]  

(5)

PReLU only adds a very small number of parameters, which means that the amount of network calculation and the risk of overfitting have only increased a little. When back propagation updates \( a_i \), the update method with momentum is adopted, as shown in (6), where \( \mu \) is momentum and \( \varepsilon \) is the learning rate. In this article, \( a_i \) is initialized to 0.1.

\[
\Delta a_i = \mu \Delta a_i + \varepsilon \frac{\partial \varepsilon}{\partial a_i}
\]  

(6)

The Sigmoid and PReLU activation functions are compared and analyzed. The images of these two functions are shown in Fig.4. According to Fig.4, the PReLU function has one-sided suppression of \( x<0 \) input and a relatively wide excitation boundary, so the nonlinear mapping is sparse. The PReLU activation function is not saturated at both ends, and the gradient will not disappear in the CNN training of gradient descent, so the convergence rate of the network will be faster than the Sigmoid activation function. PReLU is an improved version of ReLU and LReLU. It has non-saturation. The negative half-axis slope \( a_i \) can be learned rather than fixed. It can adaptively learn parameters from the data. PReLU has the characteristics of fast convergence speed and low error rate. It can also be used for back propagation training and can be optimized simultaneously with other layers.

3.3. Improvement of LeNet-5 model network structure

The improved convolutional neural network topology proposed in this paper is shown in Fig. 5. It can be seen from Fig. 5 that the first convolution layer of the model used has 6 5×5 size convolutions, with 6 channel outputs, and each channel is activated by PReLU. The generated channel is down-sampled at the second pooling layer using an average pooling of 2×2 size and step size of 2. Among them, the convolution at the time of downsampling first performs zero padding with a value of 1, fills the input image with a layer of a boundary with a value of 0, and enhances the extraction of the edge and contour features of the face. The third convolutional layer has 10 convolutions of 5×5 size. These convolutions will produce 10 channels, and these channels are also activated by PReLU. The generated channel is similarly down-sampled at the fourth pooling layer using an average pooling of 2×2 size and step size of 2. In the fifth fully connected layer, regularization is added and the Dropout function is used to avoid over-fitting problems. 120 filters of 1x1 size are used and 120 channels of output are generated. In the sixth fully connected layer, 40 channels are generated for the ORL dataset. The resulting output will be transmitted to the fully connected layer for classification and training using Softmax to give a probability estimate for each category label. In the convolutional layer and the fully connected layer, except for the first convolutional layer, the filters of other layers are initialized according to the values obtained from the truncated normal distribution. The filters in the first convolutional layer are initialized by the method described in Section A, and the deviation values of all layers are initialized to zero.

---

**Fig.5** Face recognition model based on Gabor-LeNet convolutional neural network
4. Research on Face Recognition Based on Gabor-lenet

4.1. Data set selection and preprocessing
In this paper, the face database ORL dataset is used. The ORL dataset has a total of 40 objects of different ages, genders, and ethnicities. Each object has 10 grayscale images, a total of 400 grayscale images, and the image size is 92×112 pixels. There are changes in facial expressions, such as laughing or not, eyes open or not, glasses on or not, etc., which is currently the most widely used standard database. The preprocessing of the data set requires face detection, face cropping, and standardization of the face database. Since the ORL dataset is already a cropped face, it can be processed only by standardizing it. The face database is processed, and finally a 32×32 library is standardized, as shown in Fig.6:

![Sample ORL data set after processing](image)

After that, the data set is divided into two sets, the training set and the test set, and the corresponding ratio is set to 8:2. Then, the ORL face database collection is divided into 40 categories, that is, the label is set to 1~40.

4.2. Experimental conditions
Due to the need for convolution calculation in model training, the use of GPU will greatly shorten the training time and speed up the training compared to the CPU. The computer configuration used in this experiment is dual E5-2637 v4 CPU, operating system For Ubuntu 16.04, GTX1080Ti graphics card and 12GB memory are also used to speed up training. The platform used is Tensorflow1.9.0, a machine learning framework developed by Google.

4.3. Model parameter setting
Firstly, the input image is normalized, so that the gray value range of the input image pixel is linearly mapped to [0,1]. This article first pre-trains the initialization model. The regularization coefficient is $10^{-4}$, the learning rate is 0.01, the learning decay rate is 0.99, the batch training sample number is 50, the iteration number is 10000, and the result is output once every 100 iterations in the training process.

4.4. Experimental results
In the experiment conducted on the ORL data set, Fig.7 is the comparison result of the loss curve trained using the original network and the improved Gabor-LeNet network, where the loss value is equal to the sum of the cross-entropy function and the regularized loss. It can be seen from Fig. 7 that the loss of the original network has been oscillating, with a small decrease in the first 2000 iterations, a faster decline in 2000-6000 iterations, and a shock of 0.3-0.7 after 6000 iterations; However, the improved Gabor-LeNet network's loss decreased rapidly in the first 1000 iterations and remained at 0.02 after 1000 iterations. It can be seen that its loss declines faster than using the original network, there is basically no oscillation, the value of the loss function is smaller, and finally it tends to zero to reach a stable.

Fig.8 is a comparison between the recognition rate curve trained using the original network and the improved Gabor-LeNet network. It can be seen from Fig.8 that the recognition rate of the original network is very low, only 76.61%, and there is shock; After the improvement, the recognition rate is significantly higher than that of Gabor-LeNet network, and the recognition rate reaches more than 99.5%, and there is basically no shock. This indicates that the face recognition performance of the improved model is greatly improved compared to the original network LeNet-5 model.

Table II show of the proposed network model in this paper with other models in ORL face database.
In order to verify the versatility of the improved model for face recognition in this paper, the improved Gabor-LeNet network is applied to the GT face dataset and the AR face dataset with sunglasses and scarf occlusion at the same time to perform the recognition experiment. The experimental results of the two data sets show that the loss function of the original network on both data sets is far greater than 1, and the recognition rate is only about 70%; The improved Gabor-LeNet network has relatively small loss values on both data sets, both of which are less than 0.04, and the recognition rate has also been greatly improved. For the GT data set, its face recognition rate has reached 98.56%; for AR data The face recognition rate on the set is 91.25%.

![Fig.7 Comparison of loss curves between the original network and Gabor-LeNet network](image1)

![Fig.8 Comparison between the original network and Gabor-LeNet network recognition accuracy curve](image2)

| Method      | Accuracy |
|-------------|----------|
| ICA[18]     | 90.80    |
| 2DPCA[19]   | 97.14    |
| Gabor-LeNet | 99.50    |

Table 2. The comparison of accuracy for ORL database

5. Conclusion
In this paper, by improving the LeNet-5 convolutional neural network model, the first convolutional layer is initialized with a Gabor filter, and the activation function is a parameterized rectified linear unit (PReLU), an improved face recognition model of Gabor-LeNet convolutional neural network is proposed. The experimental results of the three face data sets show that the improved Gabor-LeNet convolutional neural network model can achieve a higher face recognition rate, and the loss function recognition rate of the network is relatively stable during training, and there is basically no shock. The recognition rate of AR face data set with sunglasses and scarf occlusion needs further study to improve its recognition rate.
Acknowledgments
This work is supported by Guangxi Science and Technology Major Project under grant NO. Guike AA18118004, Guangxi Humanities and Social Sciences Development Research Center "Scientific Research Engineering • Innovation and Entrepreneurship" Special Major Commissioned Project under grant NO.ZDCXCY01, Guangxi Postgraduate Education Innovation Program Project under grant NO. XYCSZ2020056.

References
[1] Noori R, Karbassi A R, Moghaddamnia A, et al. “Assessment of input variables determination on the SVM model performance using PCA,” Gamma test, and forward selection techniques for monthly stream flow prediction. Journal of Hydrology, 401(3-4), pp. 177-189, 2011.
[2] U. Park, Y. Tong and A. K. Jain, “Age-Invariant Face Recognition,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pp. 947-954, May 2010.
[3] Z. Li, U. Park and A. K. Jain, “A Discriminative Model for Age Invariant Face Recognition,” in IEEE Transactions on Information Forensics and Security, vol. 6, no. 3, pp. 1028-1037, Sept. 2011.
[4] Ding C, Tao D, “A comprehensive survey on pose-invariant face recognition,” ACM Transactions on intelligent systems and technology, 7(3), pp.1-42, 2016.
[5] A. Z. Kouzani, F. He, K. Sammut and A. Bouzerdoum, "Illumination invariant face recognition," SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218), San Diego, CA, USA, vol.5, pp. 4240-4245, 1998.
[6] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," in IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1635-1650, June 2010.
[7] Hinton G E, Salakhutdinov R, “Reducing the dimensionality of data with neural networks,”. Science,313 (5786), pp. 504-507,2006.
[8] Zhang Guoyun, Xiang Canqun, Luo Baitong, et al. “Research on an improved CNN structure for face recognition,” Computer Engineering and Applications,53(17), pp. 180-185, 2017.
[9] Wu Yao, Qiu Weigen, “Face recognition based on improved deep convolutional neural network,” Computer Engineering and Design, 38(08), pp. 2246-2250, 2017.
[10] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 1701-1708, 2014.
[11] Sun Y, Wang X, Tang X. “Deep learning face representation from predicting 10,000 classes,” Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1891-1898,2014
[12] Yang Wei, Ni Qian, Wu Jun, “Research on Initial Value and Convergence of BP Neural Network Weights,” Journal of Electric Power Systems and Automation, 014(001), pp. 20-22, 2002.
[13] LeCun, Yann, et al. “Gradient-based learning applied to document recognition,” Proceedings of the IEEE 86.11, pp. 2278-2324,1998.
[14] Zhao, Z. H., S. P. Yang, and Z. Q. Ma. “License plate character recognition based on convolutional neural network LeNet-5,” Journal of System Simulation 22.3, pp. 638-641, 2010.
[15] Zhao Z H, Yang S P, Ma Z Q, “License plate character recognition based on convolutional neural network LeNet-5,” Journal of System Simulation, 22(3), pp. 638-641, 2010.
[16] Wang Gongpeng, Duan Meng, Niu Changyong. “Stochastic gradient descent algorithm based on convolutional neural network,” Computer Engineering and Design,39(2), pp. 441-445,2018.
[17] K. He, X. Zhang, S. Ren and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,” 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, pp. 1026-1034, 2015.
[18] Yang, Ming-Hsuan, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," Fgr, Vol. 2, pp. 1457-1464, 2002.

[19] Jian Yang, D. Zhang, A. F. Frangi and Jing-yu Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 131-137, Jan. 2004.