Face Recognition System Based On CNN

Di Wang
Northwest Minzu University
Lanzhou, China
wanganddi0918@gmail. com

Ding Wang
Northwest Minzu University
Lanzhou, China

Hongzhi Yu
Northwest Minzu University
Lanzhou, China

Guanyu Li
Northwest Minzu University
Lanzhou, China

***Abstract*—With the development of computer vision and artificial intelligence, face recognition is widely used in daily life. As one of the most concerned methods of biometric recognition, face recognition has become one of the research hotspots in the field of computer vision and artificial intelligence. However, face recognition is easily affected by internal and external differences, and it is often difficult for traditional face recognition methods to achieve ideal results. In order to further improve the recognition accuracy of current face recognition algorithms, this paper proposes a face recognition algorithm based on improved convolutional neural network. Experiments show that the improved algorithm can be effectively applied to the data set.**

Keywords-convolutional neural network CNN; face recognition; optimization algorithm

I. INTRODUCTION

Convolutional neural networks (CNN) was first proposed in the 1960s, when Hubel and Wiesel discovered its unique network structure, which can effectively reduce the complexity of feedback neural network, while studying the neurons used for local sensitivity and direction selection in the cat cerebral cortex.

The idea of CNN is developed from artificial neural network. The biological neural network is characterized by nonlinearity, concurrency, robustness and high fault tolerance. Because of these characteristics, artificial neural network has been widely used in pattern recognition, image processing and other aspects. The main idea of artificial neural network is to learn and analyze the laws between the corresponding relationships between inputs and outputs according to the corresponding relationships between inputs and outputs, and then use this law to predict the output results of new inputs. The learning and analysis process in the network is generally called training. At the same time, artificial neural network has high self-learning ability and self-adaptability.

CNN can directly input the original image, so that the image preprocessing becomes simple. CNN applied the techniques of local sensing field, weight sharing and pooling to greatly reduce the training parameters. At the same time, CNN also has image translation, rotation and distortion invariance. In the research of deep learning, CNN is specially applied to computer vision involving image classification and object recognition. Before CNN was widely adopted, most pattern recognition tasks were completed through the initial stage of manual feature extraction and classifier. However, the emergence of CNN completely changed pattern recognition. Instead of using a standard manual feature extraction method, it USES the raw pixel strength of the input image as a flat vector. In general, convolutional neural network has excellent processing ability for two-dimensional data, such as voice and image.

II. DATA SORTING

All data sets are collected in the network environment. When we collect data sets on the Internet, we need to manually screen out some unqualified images, such as: the picture is too fuzzy, the collected pictures are not consistent with the specific people and so on. Finally, we need to manually delete the image that cannot be repaired. After the images are collected, a specific label is added to each training set image through the algorithm to facilitate the recognition during training. Since the size of the images collected may be different each time, we need to preprocess all available images after the data set collection to unify the size of the images. The approximate process of data consolidation is shown in Fig.1.

In the whole process of data collection, we selected three Chinese actresses as our research objects, collected 200 pictures for each character, and finally collected 600 pictures, all of which were of excellent quality.

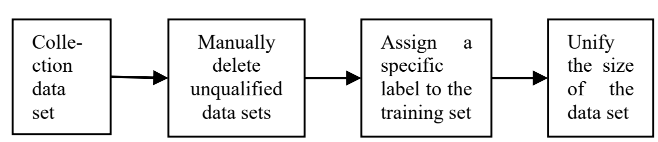

Figure 1. Data collection and collation process

978-1-7281-9837-8/20/$31.00 ©2020 IEEE
DOI 10.1109/CIBDA50819.2020.00111

III. TRAINING DATA

A. Introduction to the basic model

Our basic research model is the CNN model used in traditional cat species identification, with 8 layers. There are four convolutional layers, two pooling layers, one fully connected layer and one output layer.

The processed data is used as the input of the convolutional neural network for iterative training. The image size of the input convolution layer is uniformly adjusted to 160x120. When the training process converges or reaches a certain number of iterations, the training of convolutional neural network is completed.

The convolution layer is composed of multiple convolution kernels. The purpose is to realize the convolution operation on the input data and construct different feature images. Different convolution kernel functions can express all the characteristics of the data in the previous layer differently in the later layer. Kernel of 3x3 is selected for the first two convolution layers, of which strides is 1 and the number of filters is 32. The difference between the last two convolution layers and the previous one is that the number of filters used is 64.

At the same time, in order to reduce the image resolution and thus reduce the computation, the convolution layer is pooled according to the image local correlation principle to increase the recognition speed. The pooling layer can sample under the feature graph, compress the image to make the feature smaller, extract the main feature, discard the relatively minor feature, reduce the computational complexity of the network, and reduce the scale of the convolutional neural network after pooling. The features after pooling still contain the information of the original image. The pooling operation retains the original location information and reduces the number of parameters. The pooling operation is relatively simple with low computational complexity and easy to implement. The features of an area can be extracted by pooling, but sometimes the extraction features are not complete, affecting the accuracy of the network. The maximum function is used for the two pooling layers, with the pooling window of 2x2 selected, of which strides is 1. Kernel is the size of the selected convolution kernel and strides can be understood as the strides of the convolution kernel moving on the matrix.

After convolution, it is connected to a layer of full connection layer, which is used to train and classify the features extracted from the convolution layer. Finally, output results through the output layer. We used Python 3.5 to build the CNN subsystem and the "Keras" deep learning framework to build the convolutional neural network. The required python libraries are NumPy, OS, PIL, and Keras.

B. Model improvement

1) Selection of convolution layers

In order to only compare the effect of the number of layers on the recognition rate, other settings were kept unchanged in the experiment, such as the number of samples, optimization algorithm, et al. The optimization algorithm adopted the random gradient descent algorithm of the basic model. The CNN model used by the base model has 8 layers. On the original basis, we add three convolution layers, wherein kernel is 3x3, strides is 1, and the number of filters is 128. At the same time, another pooling layer is added, using the maximum function, a pooling window of 2x2 is selected, where strides is 1, and it is named model 1. Then, on the basis of model 1, three convolutional layers and one pooling layer are added, which are required to be consistent with the above and named model 2. The accuracy of the three models after training is shown in Table I.

TABLE I. MODEL ACCURACY COMPARISON

Model Type	Accuracy After Training
Basic model	78.64%
Model 1	81.84%
Model 2	68.27%

2) Activation function selection

In image recognition, the activation functions are relu function, tanh function and elu function. If the excitation function is not used, in this case the input of each layer node is a linear function of the output of the upper layer. This makes it easy to verify that no matter how many layers the neural network has, the output is a linear combination of inputs, which is comparable to no hidden layer. This is the case with the primitive perceptron, which makes the network's approximation quite limited. In order to only compare the effect of activation function on recognition rate, other settings were kept unchanged in the experiment, such as network structure, optimization algorithm, etc. The optimization algorithm choose the random gradient descent algorithm, and the network structure use the structure of the improved model 1. The effect of the above three activation functions is as follows:

a) The tanh function

Tanh is a function that maps the values from (-∞ , +∞) to (-1，1) .

The function is described as:

$$g(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \tag{1}$$

To some extent, the tanh function makes up for the shortcomings of the sigmoid function, but in essence, the gradient disappears when the derivative of the function approaches 0. The tanh function is the optimization of the sigmoid function. It can be seen from the figure that the accuracy rate of training with tanh function is less than 80%.

b) The relu function

Relu function is a piecewise linear function, also known as the modified linear unit, is now the mainstream of the image recognition activation function. Relu function has the following advantages: solve the gradient vanishing problem (in the positive interval),the computation is very fast, you just need to determine whether the input is greater than 0,the convergence rate is much faster than sigmoid and tanh.

The function is described as:

$$g(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases} \tag{2}$$

To a large extent, relu does not exist gradient disappearance problem. From the training results, the effect of relu function is slightly better than that of tanh function, and the accuracy is more than 80%.

c) The elu function

Elu function is an improvement of relu function. The most important thing is to improve the gradient when the input value is negative and avoid the gradient disappearing when the input value is negative.

The function is described as:

$$g(z) = \begin{cases} z & z \geq 0 \\ a(e^z - 1) & z < 0 \end{cases} \tag{3}$$

It can be said that the elu function is a successful improvement of the relu function, which greatly avoids the gradient disappearance. Although elu function is always better than relu function in theory, there is no good evidence that elu function is always better than relu function in practice.

d) Summary

From this experiment, it can be seen that compared with the above three activation functions, relu function is more efficient in terms of accuracy compared with the other two activation functions.

After 32 batches of training on the face recognition model, the relatively highest accuracy rate was more than 81%, so we chose relu function. Table II is the comparison table of the accuracy of the activation function we selected after training. From the accuracy after training, we can see that the relu function has significantly better effect than the other two activation functions. We will use the relu function as our activation function in the following experiments.

3) Dropout choice

Dropout is to prevent the over-fitting in the process of model training, which leads to the decline of model accuracy, and randomly disconnects a certain proportion of neurons to make them inactive. Disconnected due to random selection is more, we chose the 25%, 80% and 100%, the proportion of the three experiments. In order to only compare the influence of dropout on the recognition rate, other settings were kept unchanged in the experiment, such as network structure, optimization algorithm, et al. The optimization algorithm selected the random gradient descent algorithm, the network structure used the structure of the improved model 1, and the activation function used the relu function. The accuracy is shown in Table III. According to the above experiment, we can see that 100% Dropout ratio can make the training set correct the most.

TABLE II. ACTIVATION FUNCTION ACCURACY COMPARISON

The Name Of The Function	**Accuracy After Training**
tanh	78.33%
elu	80.24%
relu	81.84%

4) Optimizer selection

At present, optimization algorithms can be divided into two categories, one is first-order optimization algorithm, the other is second-order optimization algorithm. The first-order optimization algorithm refers to the use of the first-order derivative to optimize the network. In other words, the first-order derivative to solve the gradient to calculate through the gradient descent method to calculate the update parameters. First-order optimization algorithm only need to compute a derivative, gradient, but the second order optimization calculation is very big, need to calculate the second derivative, which is calculated Hessian matrix and its inverse matrix, the algorithm complexity is much higher than that of first-order optimization algorithm, so the convolution neural network, the quadratic optimization algorithm has not been widely used.

So this paper mainly studies the first-order optimization algorithm. There are many commonly used first-order optimization algorithms in the application, this article selects the typical stochastic gradient descent(SGD) method, RMSprop algorithm and adaptive moment estimation(Adam) three experiments, in order to only compare the optimization algorithm of recognition rate and the effect of experiment the other Settings unchanged, such as network structure and activation function, such as the activation function selection relu function, choose 100% Dropout, network structure using the improved structure of model 1.

The advantage of SGD is that when large data sets are used, the training speed is very fast. For example, take hundreds of data points from millions of data samples, calculate an SGD gradient, and update the model parameters. Compared with the standard gradient descent method, it is much faster to update the parameters with each input sample. However, at the same time, its disadvantage is that SGD will introduce noise when randomly selecting the gradient, so that the direction of weight update may not be correct. In addition, SGD cannot overcome the problem of local optimal solution alone. Since the neural networks are all non-convex, RMSProp has better results under non-convex conditions. Empirically, it has proved to be an effective and practical deep learning network optimization algorithm. It adds a attenuation factor to control how much historical information is acquired.

TABLE III. DROPOUT ACCURACY COMPARISON

Dropout Rate	**Accuracy After Training**
25%	81.84%
80%	72.41%
100%	85.39%

TABLE IV. OPTIMIZATION ALGORITHM ACCURACY COMPARISON

Optimization Algorithm	Accuracy After Training
SGD	85.39%
RMSprop	87.19%
Adam	88.41%

Adam is an update of RMSProp optimizer, which dynamically adjusts the learning rate of each parameter by using the first-order moment estimation and second-order moment estimation of the gradient. The advantage is that the learning rate of each iteration has a clear range, so that the parameter changes very smoothly. Adam is to add bias-correction and momentum on the basis of RMSprop. As the gradient becomes sparse, Adam is better than RMSprop.

The accuracy is shown in Table IV.

5) Algorithm to compare

In this paper, 600 face recognition benchmark data sets collected from the network are used to verify the effectiveness of the algorithm, and the model is improved by studying and comparing the number of model layers, activation function, dropout ratio and optimization algorithm. Since it is difficult to collect more data in practice, we guess that the data set used this time has some limitations in some aspects. However, it can still effectively reflect the efficiency of the improved algorithm.

After many experiments, we finally find that three convolution layers and one pooling layer were added to the original basic model. We choose to use the relu activation function, and set the dropout proportion to 100%, and optimize the algorithm to Adam. Its accuracy rate has been greatly improved. The comparison graph of the accuracy after training between the basic model and the improved model is shown in Fig. 2.

The comparison table of the accuracy after testing between the basic model and the improved model is shown in Table V. As can be seen from the figure, the accuracy of our improved model increased from 68.85% to 79.41% after the test set.

It is emphasized again that the use of convolutional neural network for face recognition has good practical application ability.

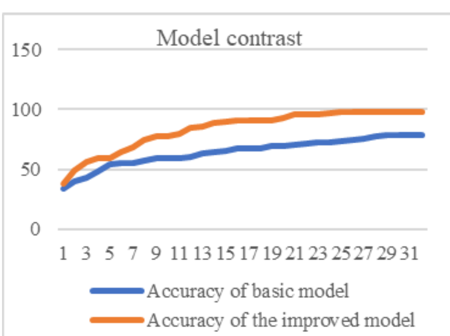

Figure 2. Accuracy after training

TABLE V. MODEL IMPROVEMENT ACCURACY COMPARISON

Model Type	Accuracy After Testing
Basic model	68.85%
Improved model	79.41%

IV. CONCLUSION

Convolutional neural network is an important kind of deep learning. In image processing, convolutional neural network has unique advantages by virtue of local connection weight sharing and other characteristics. The training in CNN is very important, which directly determines the success of network training and the final recognition rate.

In this paper, the number of model layers, activation function, dropout value and optimization algorithm are studied. Based on the original data set, the model is improved, and finally the recognition rate is greatly improved. Because it is difficult to collect a large number of data in practice, the data set used in this study has some limitations in some aspects, but it can still reflect the efficiency of the improved algorithm effectively.

The improved model still has room for improvement, and its recognition accuracy can be further improved in our future work.

ACKNOWLEDGMENT

Thanks to the Northwest Minzu University for providing us with a comfortable learning environment and rich learning resources. In addition, I would like to thank my teachers for their selfless help during this period. In the paper writing and experiments, they gave me a lot of suggestions and ideas for improvement. At the same time, I would like to thank my family for their love, so that I can study harder and study more academically. Every time I met with setbacks or difficulties, they stood firmly behind me and gave me confidence. In addition, I also want to thank the laboratory students and my seniors. Because of them, just let me harvest a lot of knowledge and fun.

References

[1] Huang G. , Liu Z. , Weinberger K Q. “Densely connected convolutional networks,” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, 1(2): 3.

[2] RAMALINGAM S, SHENOY A, VIET N T. “Fundamentals and advance in 3d face recognition,” Biometric-Based Physical and Cybersecurity Systems. Springer,2019:125-162.

[3] LU Z. , JIANG X. , KOT A. “Deep couple resnet for low-resolution face recognition,” IEEE signal Processing Letters,2018,25(4):526-530.

[4] Trnovszky T, Kamencay P, Orjesek R. “Animal recognition system based on convolutional neural network,” Advances in Electrical and Electronic Engineering,2017, 15(3): 517.

[5] Chenkai J. , Tao S. , Lei Z. “A review of face recognition based on deep convolutional neural network,” Computer applications and software, 2018, 35(1):223-231.

[6] Wang S H, Muhammad K, Hong J. “Alcoholism identification via convolutional neural network based on parametric Relu, dropout, and batch normalization,” 2018,20(6):43-48.

[7] Feiyan Z. , Linpeng J. Jun D. “A review of the research on convolutional neural networks. Acta computerica,”2017, 40(6):1229-1251.