

# Face Recognition Based on Convolutional Neural Network

Ganzhou Wu

School of science, Guangdong University of Petrochemical Technology, Maoming 525000, China.

---

**Abstract:** In this paper, we collect the picture data to be processed by gray scale, and use the python environment configured by computer to train and learn the processed training data. Finally, the accuracy rate is above 95% after the model training, which shows that the method of convolutional neural network is very efficient.

**Keywords:** Face recognition; Convolutional neural network; Gray scale.

---

## 1. Introduction

The study of coherent technology of face detection technology has very important academic value. Face image is a natural structure data model with quite rich details. The challenge of detection of these system models is: human face is composed of appearance, forehead, nose bridge, mouth, eyes and other facial parts, each part is variable; The problem of using eyes too long on campus has caused an increase in the number of people wearing glasses, which may exist in various eye accessories such as eyeglass frames and contact lenses; As the image formed by the face in three-dimensional space, under the action of a certain amount of light, the image can have a certain degree of influence. The above constraints of face recognition, such as finding some feasible methods to deal with these problems together, and relatively feasible to build a system model of face detection system and face tracking, can be extended to solve other similar and complex applications provide a good model reference, and provide a very important processing method for related detection and recognition problems. With the continuous advancement and development of science and technology, the current understanding of human visual recognition technology is sufficient to meet the needs of artificial intelligence in the application of face recognition. Constructing a complete face recognition system through skillful use of face recognition technology, and being able to use civil computers to display automatic recognition of face pictures still has a huge and expandable application space in various fields such as finance, medical care, and manufacturing.

With the continuous movement of time and the exponential increase of scientific and technological achievements, the development momentum of face recognition is particularly rapid. Before the 21st century, the United States entered this field very early and continued to develop deeply, and obtained a lot of research results., a variety of algorithms to solve various problems of face recognition. A German company has also set up a face image database and developed a recognition system [1]. Japan is also in a leading position in this regard. The research in this field in our country started relatively late, starting in the 1980s, but after decades of research and development and rapid development, many breakthrough results have been obtained. Huang Xiaoying proposed to improve pose correction based on sinusoidal transformation to optimize the face recognition model [2]. Tang Yongqiang discovered that in face recognition, the problem of data

processing is solved by extracting feasible features, so as to study the linear projection analysis in feature extraction more deeply [3]. Chang Xueping conducted a series of simulation experiments on some face databases, and through a large number of experimental results, proved that the algorithm proposed by the Institute can improve the single sample for face recognition, and the algorithm can significantly improve the recognition efficiency in the experimental process [4]. Yan Xiaona proposed the SVM classification algorithm for face recognition. Experiments have verified that the SVM algorithm can improve the recognition rate and shorten the recognition time [5]. Fitten proposed to unify 2DPCA recognition and image data enhancement processing, which greatly reduces the adverse effect of lighting on face recognition [6]. Zhang Cheng proposed a face recognition method based on Tengine edge intelligence, which reduces the reaction time of machines and equipment, and has a certain guarantee for data transmission, so as to achieve the performance of a face recognition system [7]. Ji Shengbo proposed a face recognition method based on adaptive enhancement of mask areas and identity feature constraints to generate adversarial network illumination changes to improve the accuracy of face recognition [8]. Lin Miaozen proposed to classify postures based on deep learning and within the neighborhood of postures. This method has obvious effects on posture classification [9].

## 2. Principle of Convolutional Neural Networks

### 2.1. Input layer

When using convolutional neural networks to process image data, the input layer of the network structure can be regarded as the pixel matrix of the image, and a three-dimensional matrix is a picture. In a three-dimensional matrix, its length and width represent the size of the image, and its depth represents the color channel of the image, for example when the depth of the black and white image is 1, but it is 3 in the depth of the color mode. Convolutional neural network has a wide range of applications in image recognition. Many researchers usually presuppose three-dimensional input data when discussing its network structure, that is, two-dimensional pixels and their color channels on the image plane. Convolutional neural network and other conventional neural network algorithms use gradient descent method for learning and training, but in the input layer of convolutional

neural network, when inputting feature data, the image data needs to be standardized to facilitate the subsequent model training. The specific operation of standardization is to normalize the pixel value and color channel of the input image data before the image data is input. For example, when the input data is an image pixel, the original pixel value in the [0-255] interval needs to be converted into the value in the [0-1] interval. Standardizing the input feature data can improve the convergence and accuracy of the convolutional neural network model.

## 2.2. Convolution layer

The convolution of the convolution layer is the effect that the data received by the input layer is processed by the convolution kernel-matrix product and then transmitted to the next level of neurons. In a convolution layer of the convolution neural network, the image data will be processed. After processing, several feature graphs will be generated, and the feature graphs are all composed of rectangular arrangements of neurons. Neurons with the same feature graph will have a shared weight, which is the convolution kernel. The weight parameters of the convolution kernel are the same as those of the conventional neural network. The convolution kernel is multiplied with the matched rectangular pixel values and then summed to obtain the numerical matrix above the convolution layer. Usually, the convolution layer uses the convolution kernel to extract and map the feature data, and the specific calculation process of the convolution layer is shown in Figure 1.

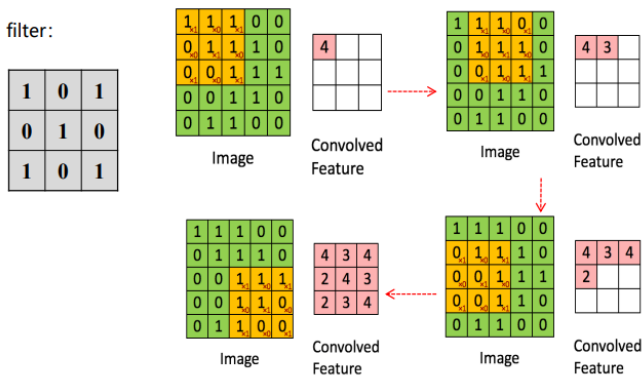


Fig. 1 Data prediction

The initialized convolution kernel usually appears as a simple matrix value. If the convolution kernel learns the weight value to a reasonable state during the training process of the model, it will help to reduce the complex neuron connections between the network layers, thereby reducing overfitting.

The nonlinear mapping of the convolutional neural network is the same as that of the feedforward neural network, both after linear combination and migration transformation, in order to enhance the fitting ability of the model. The activation function in the convolution layer can simplify and express complex features, and the mathematical formula is as follows:

$$A_{i,j,k}^l = f(Z_{i,j,k}^l) \quad (1)$$

Convolutional neural networks are similar to other deep learning algorithms. The commonly used activation function is the linear rectification function (ReLU). Before ReLU, the commonly used activation functions are Sigmoid function and tanh function. Under normal circumstances, activation

function operations are performed after the convolution kernel calculation is completed, but some early algorithms using pre-activation technology put the activation function before the convolution kernel calculation. For example, the LeNet-5 network, which studied the structure of convolutional neural networks in the early stage.

## 2.3. Pool layer

Subsampling is usually called pooling, and there are two common ways of subsampling: mean and maximum, and it can also be regarded as a special convolution operation. After the image data is processed by the convolution kernel of the convolution layer, the obtained feature maps will enter the pooling layer. Theoretically, these data can skip the pooling layer for feature training directly, but direct training will cause huge data calculations. The amount makes it difficult for the model to proceed, and it is easy for the model to overfit. In order to reduce the training parameters of the model and avoid overfitting, it is necessary to perform sub-sampling operations on the data passing through the convolution layer. There are two types of subsampling commonly used: (1) Maximum pooling layer. The maximum value of the selected partition in the feature extraction layer is selected as the sampling value (Max-Pooling). (2) Average pooling layer. The selected partition values in the feature extraction layer are added and averaged (Mean-Pooling), and the sampling value is expressed by the mean.

In convolutional neural networks, the pooling operation can sparse the feature graph and greatly reduce the complexity of the network structure in the model, reduce the number of neuron nodes in the subsequent fully connected layer, so as to reduce the calculation of training data and prevent overfitting, thereby improving the performance of the model. The fully connected layer of convolutional neural network has the classification function. After several feature processing of the convolution layer and pooling layer, the image data is trained and learned by the convolution neural network model, and finally the classification results are obtained by the fully connected layer. The previous convolution and pooling operations can be regarded as extracting useful information features from image data, and the full connection layer completes image classification through these information features.

## 3. Experimental results

First of all, The picture data in this article comes from the photos taken by myself and friends around me, and the training equipment of the model uses a commonly used notebook computer. All the data that can be collected is eighty people, 100 gray pictures per person.

Preprocess and partition the dataset The train \_ test \_ split function is called to achieve data partition, and the training set and test set are randomly divided into 8: 2. Further, we want to convert the picture into a gray-scale image. Grayscale mode is also known as single tone. In the gray-scale image, each pixel value uses a brightness value from 0 to 255, of which 0 is black and 255 is white. This mode is a common black-and-white photo. Compared with color pictures, it can highlight the content of the picture. The details and binarization also reduce the image by a certain amount of data to highlight the outline shape of important targets.

The convolutional neural network structure model consists of ten layers, including two input layers, pooling layer and full connection layer, and one flattening layer and three

discarding layers. In addition, after the convolution, pooling, flattening, and full connection layers, a discarding layer is added, and the discarding layer can prevent the model from overfitting. The complex definition sequence model is Sequential, and the following is the specific construction content of the model:

(1) Level 1 Convolution layer, the first convolution operation, the number of convolution kernels is 32, the size of convolution kernels is (3, 3), and the activation function is relu;

(2) Level 2 the pooling layer, the first pooling operation, using the maximum value pooling, the pooling layer convolution core size is (2, 2);

(3) Level 3 The dropout layer, the core function of dropout is to freeze some weights or thresholds so that they are not updated in this round of training to prevent model overfitting;

(4) Level 4 Convolution layer, the second convolution operation, the number of convolution kernels is 64, the size of convolution kernels is (3, 3), and the activation function is relu;

(5) Level 5 The pooling layer, the second pooling operation, uses the maximum value pooling, and the pooling core dimension is 2 \* 2;

(6) Level 6 The dropout layer freezes part of the data;

(7) Level 7 Flattening layer, flattening the output data of the upper layer into a one-dimensional vector structure;

(8) Level 8 Fully connected layer, the number of neurons is 512, and the activation function is relu; (9) Level 9 The dropout layer freezes part of the data;

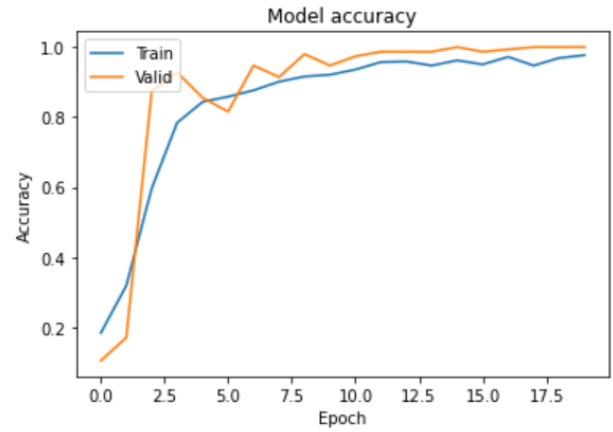
(10) Level 10 In the output layer, the number of neurons is 8, and the activation function is sigmoid;

The output of the computer's model compilation results is shown in Table 1.

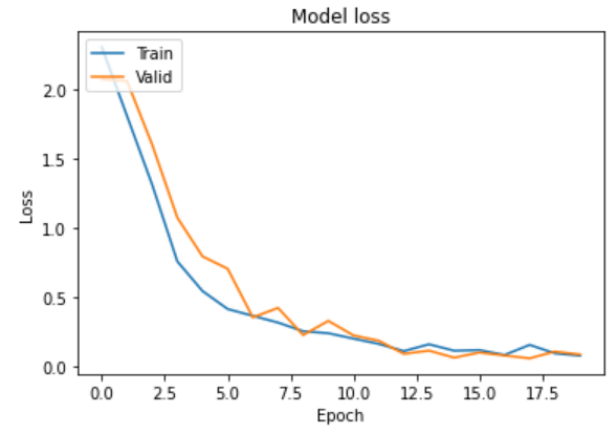
**Table 1.** Model structure data

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	320
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
dropout (Dropout)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 14, 64)	0
flatten (Flatten)	(None, 12544)	0
dense (Dense)	(None, 512)	6423040
dropout_2 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 8)	4104

The model compilation optimizer optimizer is set to adam, the loss function loss is set to SparseCategoricalCrossentropy, and the evaluation function metrics is set to accuracy. The number of model training iterations epochs is set to 20, and the training effect is shown in Figs. 2 and 3.



**Fig.2** The Accuracy of Face Image Data Training



**Fig.3** The Loss Value of Face Image Data Training

It can be seen from Fig.2 and Fig. 3 that as the number of training times increases, the accuracy rate of recognition is also constantly improving. When the number of training times reaches ten times, the accuracy rate tends to be stable, and the loss value of model training also increases with the number of iterations. The increase decreases, and finally tends to a relatively stable position. In the convolutional neural network model, the accuracy rate of the training set and the test set is more than 95%, and the loss value is infinitely close to zero.

## 4. Summary

This paper mainly studies the face recognition technology based on convolutional neural network. Through the continuous learning of the basic theoretical knowledge of neural network, we further study the convolutional neural network model, and practice the face recognition technology through common computers. The research on the convolutional neural network model starts with the early simple methods such as template matching, geometric features and sparse representation to fully understand the main technical difficulties of face recognition. Later, we understand that the neural network algorithm based on the neural model can truly realize face recognition. However, because the traditional neural network model has high hardware requirements and the model training time is long. After studying and exploring the convolutional neural network, the application of the technology based on the convolutional neural network in face recognition is proposed. After a period of parameter debugging, the training and testing results of face recognition can be obtained, and the training time of the convolutional neural network model is

relatively short, and the accuracy of picture recognition is high, and the testing results are more than 95%. It can be seen that the convolutional neural network model has excellent effects on face image recognition. The reason why face recognition based on convolutional neural network can get high accuracy is summarized. The most important thing is to build a good model for the neural network based on the valuable experience accumulated by predecessors. Secondly, it is necessary to effectively preprocess the data before training and testing, and then combine the theoretical knowledge learned by oneself to grayscale the picture and adjust the model parameters.

## References

- [1] Xiong Xin. Face Recognition Technology and Application [M]. Zhengzhou: Yellow River Water Conservancy Press, 2018.
- [2] Huang Xiaoying. Research on Multi-pose Face Recognition Method Based on Subspace [D]. Chongqing University, 2009.
- [3] Tang Yongqiang. Application of Linear Projection Analysis Algorithm in Face Recognition [D]. Northeast Petroleum University, 2010.
- [4] Chang Xueping. Research on Single Sample Face Recognition Based on Sparse Theory [D]. Zhejiang Normal University, 2011.
- [5] Yan Xiaona. Application Research of Support Vector Machine Kernel Method in Face Recognition [D]. Ocean University of China, 2012.
- [6] Feiteng. Research on PCA-based Face Recognition [D]. Jilin University, 2016.
- [7] Zhang Cheng. Face recognition method based on Tengine edge intelligence [D]. Nanjing University of Posts and Telecommunications, 2020.
- [8] Ji Shengbo. Research on Illumination Change Face Recognition Method Based on Generative Adversarial Network and Fuzzy Masking [D]. Xi'an University of Science and Technology, 2021. [9] Lin Miaozen. Research on Face Recognition Based on Deep Learning [D]. Xidian University, 2013.