Structure Design of Convolutional Neural Network Based on Residual Theory for Face Recognition

Zhang Chenchen¹, Jin Hong¹
¹North University of China, Taiyuan, 030051, China
*Corresponding author’s e-mail: 871068198@qq.com

Abstract. Compared with the traditional face recognition methods, the deep convolution neural network model does not need to manually design complex and time-consuming feature extraction algorithms, but only needs to design an effective neural network model, and then carry out end-to-end, simple and efficient training on a large number of training samples, so as to obtain better classification accuracy. In this paper, based on the original VGG network model, combined with Residual theory, a deep-level residual convolutional neural network structure is designed and implemented, which not only reduces the computing power requirements of hardware computers, but also achieves better recognition results.

1. Introduction
In recent years, with the breakthrough of computer vision technology and the continuous improvement of computing capability of computer hardware, image recognition technology has developed rapidly in the application field of actual scenes, and face recognition technology has also made great progress as an important sub-direction in the field of computer vision.[1]. With the development of science and technology, face recognition and cognition is no longer the unique ability of human beings, and machines can also recognize faces. [2-3]. The steps required for data collection of face recognition are simple and the recognition accuracy is high. Compared with fingerprint recognition, speech recognition and iris recognition, face recognition is a better choice for general identity recognition.[4-5].

In this paper, based on the original VGG network model structure, combined with Residual theory, a deep-level residual convolutional neural network structure is designed and implemented, which can not only reduce the requirement of hardware computing ability, but also achieve a better recognition effect.

2. Deep-level Convolutional Neural Network
Convolutional neural network is a special deep neural network model, and its weight sharing network structure is similar to biological neural network. This feature can not only reduce the complexity of the network model, but also reduce the number of weights, so it has become one of the research hotspots in the field of image theory. When the input of the network is multi-dimensional image, the advantages of convolutional neural network are more obvious, and the image can be directly used as the input of the network, thus avoiding the complex feature extraction and data reconstruction process in the traditional recognition algorithm. Convolutional neural network mainly includes convolution layer, pooling layer, activation function, full connection layer and so on. According to the three characteristics of local receptive field, weight sharing and spatial down-sampling, the convolution neural network obtains a certain degree of high invariance of translation, scaling and tilt. The training
The process of convolutional neural network is mainly to learn network parameters in convolution layer such as convolution kernel parameters and interlayer connection weights, and the prediction process is mainly to calculate class labels based on input images and network parameters. The key points of convolution neural network are network structure (including convolution layer, down-sampling layer, full connection layer, etc.), back propagation algorithm and so on.

3. Design of Deep-level Convolutional Neural Network Based on Residual Theory

3.1. Design of Deep-level Convolutional Neural Network

The deepening of the convolution network model is beneficial to improve the recognition rate of the network, but it also makes optimization more difficult. Using residual learning theory, neural network model structure can improve the learning ability of extracting abstract features by optimizing the residual between input data and output data after mapping, thus improving the tuning ability of deep-level convolutional neural network.

A deep-level convolutional neural network structure model constructed by residual theory is shown in Figure 1. The model is mainly composed of A-type Residual Block and B-type Residual Block, which is composed of alternating unit structures of Residual Block, and ignores Batch Norm layer and ReLU layer, which can fit complex nonlinear transformation. Among them, each Block unit uses a branch layer. Whatever the branch directly transfers the input vector or transfers the result of mapping that transforming the input vector through 1*1 convolution kernel conforms to the residual theory. Among them, the branch of the solid arrow represents the A-type Residual Block, and the branch of the dotted arrow represents the B-type Residual Block. There are four columns of networks in the model, and each column has the same number of convolution cores, which are connected from left to right to form a complete Residual convolution neural network. The last Residual Block in the fourth column in the figure is connected with an average down-sampling layer, and finally the network is connected with a full connection layer with the number of nodes equal to the total number of classification S.

The parameters of Residual convolutional neural network mainly come from the convolutional layer, and the learning parameters can be greatly reduced without the fully connected layer. Although the number of network layers has reached 34 layers, compared with the original VGG model, the number of network parameters has been greatly reduced, which can greatly reduce the requirements for computing hardware.
3.2 network training

Residual convolutional neural network is a deep-seated network, and its learning ideas can make the network model easier to tune, but the setting of its learning parameters is particularly important for the training and convergence of the network in this process. Because the network training with ReLU as the activation function is easier to train, the parameter initialization method of Residual Convolutional Neural Network adopts random initialization method. Suppose a convolution layer is calculated as:

\[ y_i = Wx_i + b_i \]  

(1)

In this formula, \( x_i \) is a vector with the size of \( k^2c \), \( k \) is the size of convolution kernel, and \( y_i \) represents a value in the output feature map of the convolution layer, and its calculation diagram is shown in Figure 2.

![Figure 2 Schematic diagram of calculation](image)

If \( n = k^2c \), then \( W \) is a matrix of \((d,n)\), \( d \) represents the number of convolution kernels, and accordingly, the number of convolution kernels in the previous layer is \( d_{i-1} \), and \( b \) is offset. Gaussian sampling is used in the initialization of this layer, but it is not a general Gaussian sampling method, which the mean value is still 0, and the standard variance calculation formula is as follows:
In the formula, $l$ represents the current layer, $k_i$ is the size of convolution sum of the current layer, and $d_i$ is the number of convolution kernel of the current layer. Compared with general initialization methods, this random initialization method can converge faster, especially in deeper networks. The residual convolutional neural network established in this paper has not been pre-trained layer by layer, but has been retrained on a new database, so it is very important to initialize the parameters of the network reasonably.

An epoch in Figure 3 represents a cycle in which all training data are trained once. It can be seen from the trend chart of loss value that the initial loss value of the network is very small, around 9.0, which is due to the random initialization of network parameters, and it means that its parameters are not optimized. In several epoch at the beginning of training, the loss value of the network drops rapidly, which shows that the convergence speed of the network is fast, and the designed network model based on residual theory is effective. In the training process, the problems of gradient dissipation and gradient explosion are not appeared, and the convergence speed is fast, which also shows that the random initialization method of parameters in the training process is effective. In the subsequent epoch, the convergence speed of the network is still very fast, and it will slow down as the network optimization and the decline of learning rate.

\[
std = \sqrt{\frac{2}{k_i^2 d_i}}
\] (2)

**Figure 3** trend chart of loss value of residual network

**4. network accuracy analysis**

**Figure 4** Trend chart of residual network accuracy
After the network is trained on Casia-WebFace database, the accuracy of the multi-classification test of the verification set reaches about 78%. The reason is that the network model is difficult to fit well on the training set due to the interference of the wrong training set. While benefited by a large number of sample types, the network model still has a good ability to extract face features.

5. Conclusion
The deep-level convolutional neural network is beneficial to the recognition rate of the network, but it will make its optimization more difficult. Aiming at this deficiency of the deep-level convolutional neural network, this paper designs and implements the deep-level Residual convolutional neural network based on the original VGG network and the residual theory. The network model adopts the way of unit structure stacking, which can fit complex nonlinear transformation; In the process of network training, the parameter initialization method adopts random initialization method, which achieves fast convergence speed. The effectiveness of the designed network is verified by calculating the loss value of the network model, and the recognition effect of the model is verified by the accuracy curve of the network.

References
[1] Sharif M, Mohsin S, Javed M Y. A Survey: Face Recognition Techniques[J]. Research Journal of Applied ences Engineering & Technology, 2012, 4(23): 1-10.
[2] Turk M, Pentland A. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1):71-86.
[3] Ahonen T, Hadid A, Pietikinen M. Face Description with Local Binary Patterns: Application to Face Recognition[C] Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part I. IEEE, 2006.
[4] ROBERTO, B. Face Recognition : Feature versus Templates[J]. IEEE Trans. pattern Anal. & Mach. intell, 1993, 15(10):1042-1052.
[5] Bartlett, Marian, Stewart, Face Recognition by Independent Component Analysis.[J]. IEEE Transactions on Neural Networks, 2002..