Face Recognition Based on Deep Learning of Small Data Set

Lichun Yu¹ and Jinqing Liu²,*
¹School of Technology, Fuzhou University of International Studies and Trade, Fuzhou, China
²College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou, China

*Corresponding author email: jqliu8208@fjnu.edu.cn

Abstract. Face recognition based on deep learning relies on large-scale label data, which limits the application of face recognition technology. To solve this problem, this paper proposes an improved VGGNet algorithm for face recognition. The improved algorithm requires less training data size, faster training speed, and the face recognition rate reaches more than 95%, which can meet the practical application requirements of most scenes.

Keywords: Deep learning; LFW database; Label data; Recognition rate; Face recognition.

1. Introduction
Face recognition technology is currently widely used, such as identity authentication, network payment, public security monitoring, image tracking and so on. With the breakthrough of deep learning technology, face recognition technology is gradually changing from traditional machine learning to deep learning. In 2006, Hinton put forward the concept of deep learning [1] and proposed the deep confidence network, which brought a new method to solve the optimization problem of deep network [2]. Then the convolutional neural network proposed by LeCun et al. is the first real multi-layer structure learning algorithm. Researchers have made great achievements in deep learning of face recognition, such as Facebook DeepFace[3], Chinese University of Hong Kong DeepID[4], Google FaceNet. These deep learning algorithms rely on large-scale training data and high-speed computing capabilities, and have achieved excellent results. However, in most practical applications, it is difficult for small enterprises or ordinary individual researchers to obtain a large number of training data, they only have limited computational power. Therefore, a face recognition algorithm technology that can meet the practical application in the environment with limited computational power under the condition of small-scale training is needed.

In order to reduce the dependence of deep neural network on large-scale training data and computing power, we improved the VGGNet network structure, simplified the network structure, improved the loss function, and proposed a face recognition algorithm based on LiteVGGNet. The experimental results show that LiteVGGNet deep learning algorithm has a high recognition rate in the case of small training data and limited computing capacity, which can meet the application requirements of actual face recognition.

2. VGGNet
VGGNet is a famous image classification model of convolutional neural network, developed by the visual geometry group of the university of Oxford. It won the second place in the ImageNet large-scale visual recognition challenge in 2014. It performs better in multiple migration learning tasks than GoogleLenet, which ranked first in 2014. Therefore, the VGG model is a better image feature
The advantage of VGG model is that it has more than 138M parameters, needs larger storage space, and has a relatively long training time. The model structure of VGGNet[5] is shown in Figure 1.

**Figure 1. VGGNet model.**

The structure of the VGGNet model is described as follows:

- The input of the network is 224 x 224 RGB pictures, all of which are processed by means.
- There are 5 maximum pooling layers and 13 convolution layers in the network. Three full connection layers and one SoftMax classifier layer.
- The size of the convolution core in the convolution layer is 3 x 3, the step size is 1 (stride = 1), and the circle size is 0 (pad = 1).
- Max Pooling is used in the pooling layer, but not all convolution layers have pooling layer. The pooling window is aa, and the step is 2. That is to say, non-overlapping pooling is used.
- All hidden layers are equipped with ReLU layers.
- Dropout technology is also used behind the first and second full connection layers to prevent network over-fitting.

VGGNet is an extension of AlexNet, a well-known convolutional neural network model. Compared with AlexNet, VGGNet makes the following valuable improvements:

- Three consecutive 3×3 convolution cores were used to replace one 7×7 convolution core. For convolution cores of C channels, the convolution layer of 7×7 contains parameter 7²C, while the number of parameters of three 3×3 convolution layers is 3×3²C, so the number of parameters used can be greatly reduced.
- The reason for using 3×3 convolution kernel is that 3×3 is the smallest unit which can represent the modes of "left", "top and bottom", "center". At the same time, smaller convolution kernel is helpful to extract image details better. More convolution layers enhance the non-linear expression ability of the network.
- VGGNet does not use Local Response Normalization (LRN). Research experiments show that this standardization does not improve performance, but leads to more memory consumption and computing time[5].

3. Face Recognition Based on LiteVGGNet

3.1. The Defects of VGGNet

The deep neural network based on small data set will bring over fitting problem. Due to the lack of training data, it is easy to lead to over fitting of data training, that is, the effect of model on training samples may be good, but the generalization effect on test data set is not good. At this time, we can reduce the complexity of the model and fit the regular data to achieve better results.

Because VGGNet has good portability and promotional characteristics, this paper chooses VGGNet as the neural network model of face recognition. VGGNet has excellent effect in feature extraction of images, but it has the following shortcomings.

- There are many layers in the network. A large amount of training sample data is needed in training, which requires a large amount of calculation and slow convergence.
There are more network parameters, more memory for storage network and longer training time.

The performance of the face recognition model trained by SoftMax is poor. VGGNet uses SoftMax classifier to classify images. SoftMax is an extension of logistic regression to solve multi-classification problems [6]. SoftMax is simple and effective for ordinary image classification, but it can be applied to face image recognition. Because the characteristics of face images are large differences within classes and small differences between classes, the features trained by SoftMax have large inter-class distance, but the intra-class distance is not compact enough, and the intra-class distance may be larger than the inter-class distance, so it is easy to cause facial recognition error.

In view of the existing problems of VGGNet, an improved face recognition algorithm of VGGNet is proposed, which is collectively referred to as LiteVGGNet.

3.2. Optimization of the Network Architecture

In order to reduce the computation of neural network and the dependence on a large number of training samples, we must reduce the network parameters. VGGNet has effectively extracted network features. In order to reduce the useful features being discarded too early, the underlying network has not changed, and the upper network should be modified as much as possible. In the network model, the full connection layer has more connections and more parameters, so network parameters can be reduced by reducing the full connection layer. In order to reduce parameters, three full connection layers were reduced to one full connection layer according to model DeepID and GoogLenet, and the last maximum pooling layer in front of the full connection layer was modified to represent pooling layer. The core size of the average sampling layer is 7×7. By using average pooling layer instead of maximum pooling layer, image extraction features can be maintained to the maximum extent, and network parameters can be effectively reduced while network feature extraction capability can be maintained as much as possible.

By reducing the total connection layer and replacing the maximum pooling layer with the mean pooling layer, the number of parameters is reduced from 138M to 16M, which will effectively reduce the calculation time and storage space of model parameters. The comparison of structural parameters before and after improvement is shown in Table 1.

| Layer no. | VGGNet       | LiteVGGNet  |
|-----------|--------------|-------------|
| 0–17      | No Change    | No Change   |
| 18        | Max Pooling  | Ave pooling |
| 19        | Full Connect | Full Connect|
| 20        | Full Connect | -           |
| 21        | Full Connect | -           |
| 22        | SoftMax      | SoftMax     |
| Total parameters | 138M       | 16M         |

3.3. Improvement of Loss Function

VGGNet uses SoftMax classifier to classify images. The features trained by SoftMax have large distance between classes, but the distance between classes is not compact enough, and the distance between classes may be larger than the distance between classes. Therefore, for ordinary image classification, SoftMax is simple and effective, but because of the similarity of face images, it is easy to cause recognition errors. So we can enhance the classification effect by improving the loss function. In order to get better classification results. Different categories should be separated as far as possible to minimize intra-class distance. SoftMax Loss can separate different classes, and Center Loss can increase the distance between classes and reduce the distance within classes. Center Loss is based on
SoftMax Loss classification and maintains a class center for each classification. During training, the distance between class members and class center is gradually reduced, and the distance between other class members and class center is increased. SoftMax Loss loss function[6] is shown in formula (1).

\[
L_s = - \sum_{i=1}^{m} \log \frac{e^{w^T_i x_i + b y_i}}{\sum_{j=1}^{n} e^{w^T_j x_i + b y_i}}
\]  

(1)

Center Loss Function[7] is shown in formula(2).

\[
L_c = \frac{1}{2} \sum_{i=1}^{m} \left\| x_i - c_{y_i} \right\|_2^2
\]  

(2)

The mixed loss function:

\[
L = L_s + \lambda L_c
\]  

(3)

the\(\lambda\) is the weight of CenterLoss.

Referring to the model Caffe Face [7], the final SoftMax classification layer of VGGNet is adjusted and modified to the joint monitoring training of SoftMax Loss and CenterLoss in order to achieve better training effect. The improved LiteVGGNet network structure is shown in Figure 2 below.

4. Training and Testing

4.1. Training

CASIA-WebFace is a large-scale face data set published and established by Institute of Automation, Chinese Academy of Sciences in 2014. It contains nearly 500,000 face pictures, totaling 10575 people. CASIA-WebFace is used as the training data set, and more than 60 face images of the same person in the database are selected as the training data set. There are 2007 qualified face categories, of which 2000 people are selected. Each person selects 60 face images as training data, and the final output category number is 2000. We use DeepLearning4J as the training tools. DeepLearning4J is a set of neural network toolkit based on Java language, which can build, train and deploy neural networks.

4.2. Testing

After the training, the LFW face database was used to validate the model results. The collection contains 13233 face images, 5749 people, all from the Internet, not the laboratory environment. Every face is standardized by a person’s name. Among them, 1680 people have more than two faces. This data set is often used to evaluate the performance of face recognition algorithms. Firstly, the tested faces are aligned. The input of LFW is two face images, and the output is to judge whether the two face images are the same person. The face recognition verification process is shown in Figure 3.
The official LFW test data set is pairs.txt, of which 3000 pairs are the same person and 3000 pairs are different people. The experiment uses 6000 pairs of faces given by pairs.txt to test the accuracy of the network model. Face images are extracted by in-depth learning neural network. The distance between two face features is calculated. The cosine similarity is chosen as the feature distance. At the same time, the threshold is used to judge whether two face images belong to the same person. Before testing, we use Facetool to align the pictures. If alignment fails without finding a face, the aligned face image is replaced by the original image. By setting a threshold value T, the calculated result greater than T is regarded as a positive class, otherwise it is regarded as a negative class. If the threshold is reduced, more positive classes can be identified and more negative classes can be mistaken for positive classes. The true positive rate (TPR) and false positive rate (FPR) can be improved simultaneously. If the threshold is increased, the true positive rate and false positive rate will decrease at the same time. Therefore, by gradually increasing the threshold, a curve about TPR and FPR is obtained, which is called receiver operating characteristic curve. ROC curve is a functional curve describing the sensitivity of face recognition. Most face recognition models use ROC as the evaluation criterion. The ROC curve of the model is drawn by experiment as shown in Figure 4.

The model was tested with the test data set paris.txt. The test results are shown in the table 2 below.

**Table 2. LFW recognition rate.**

| Input                  | Threshold | Correct | Wrong | Recognition rate |
|------------------------|-----------|---------|-------|------------------|
| 3000 paris (same person) | 0.8       | 2883    | 117   | 96.10%           |
| 3000 paris (different person) | 0.8       | 2836    | 164   | 94.53%           |
| total                  | 0.8       | 5719    | 281   | 95.32%           |

We compare the recognition rate of this model with other famous face recognition models. The LiteVGGNet1 model removes two full connection layers from the VGGNet model and replaces the maximum pooling layer with the average pooling layer. LiteVGGNet2 model improves the loss function of vggnet. LiteVGGNet model not only modifies the network structure, but also improves the loss function. The comparison results are shown in Table 3.
| Model            | No. of training samples | No. of pictures(k) | LFW Accuracy |
|------------------|-------------------------|-------------------|--------------|
| DeepFace[2]      | 4030                    | 4400              | 97.35%       |
| CaffeFace[7]     | 17189                   | 700               | 99.28%       |
| DeepID2[3]      | 10177                   | 200               | 99.15%       |
| VGGNet           | 2000                    | 120               | 91.45%       |
| LiteVGGNet1      | 2000                    | 120               | 93.59%       |
| LiteVGGNet2      | 2000                    | 120               | 94.27%       |
| LiteVGGNet       | 2000                    | 120               | 95.32%       |

It can be seen from the table that after VGGNet improved the network structure, the recognition rate of LiteVGGNet1 model increased by 2.14%, and that of LiteVGGNet2 model improved the loss function, the recognition rate increased by 2.82%. At the same time, the network structure and loss function were optimized, and the recognition rate of LiteVGGNet model reached more than 95%. Although it is slightly lower than other famous deep learning network models, the number of training samples and pictures is much smaller than other models. Therefore, the LiteVGGNet model can be applied to the application scenarios with less training data sets and higher hardware computing power requirements.

5. Conclusion
At present, deep learning algorithms rely on large-scale training data and computing resources. To solve this problem, an improved face recognition method is proposed based on image classification deep learning algorithm VGGNet. The improved algorithm use the CASIA-WebFace as the training data, with LFW face library as the test data, the experimental results show that the improved model reduces the parameters relative to the original model, the improved loss function, not only reduces the dependence on large scale training database and computing resources, reduce differences in the class at the same time, increasing the difference between classes, face recognition performance is good, the improved algorithm can meet the requirements of practical application.

6. Acknowledgements
The project is supported by The National Natural Science Foundation of China(61179011). We are also mostly grateful to *Journal of Physics* for providing this chance.

References
[1] Zhang Junyang, Wang Huili, Guo Yang, et al. Research review on deep learning [J]. Computer application research, 2018,35 (07) : 1921-1928.
[2] Liu Fangyuan, Wang Shuihua, Zhang Yudong. Deep confidence network model and application research overview [J]. Computer engineering and application, 2018,54 (01) : 11-18.
[3] Zhou Feiyan, Jin Linpeng, Dong Jun. Review of studies on convolutional neural networks [J]. Journal of computer science, 2017,40 (06) : 1229-1251.
[4] Sun Y, Wang X, Tang X. Deep Learning Face Representation by Joint Identification-Verification[J]. Advances in neural information processing systems, 2014.
[5] Wang Ting, Li Hang, Hu Zhi. Design and implementation of a VGGNet image style transfer algorithm [J]. Computer applications and software, 2019,36 (11) : 224-228.
[6] Ge Ting, Mou Ning, Li Li. Brain tumor segmentation algorithm based on softmax regression and graph resection [J]. Acta electronics sinica, 2017,45 (03) : 644-649.
[7] Wang QC, Guo GD. Benchmarking deep learning techniques for face recognition[J]. Journal of Visual Communication and Image Representation,2019,65.