A Face Recognition System Based on Improved Convolutional Neural Network

Lan Jiangyu
Beijing Information Science and Technology University, Beijing, 100101, China
ljy960716@163.com

Xie Yinggang†
Beijing Information Science and Technology University, Beijing, 100101, China
xieyinggang@bistu.edu.cn

Wang Hui
Beijing Information Science and Technology University, Beijing, 100101, China
1374207430@qq.com

Liu Guangjun
Department of Aerospace Engineering, Ryerson University, Toronto, M5B 2K3, Canada
gjliu@ryerson.ca

ABSTRACT

Be aimed at the problems of slow training speed and poor precision of traditional face recognition methods, a face recognition method based on a convolutional neural network is researched and implemented based on classical network LeNet-5. The algorithm consists of three layers of a convolutional layer, three layers of the pooling layer, one layer of fully connected layer and output layer, and a new activation function is added to improve the accuracy and convergence speed of the model. The algorithm can automatically perform data dimensionality reduction, feature extraction, and classification results. The experimental results show that the verification and recognition rate of the algorithm is 99.87% in the ORL face database, 98.34% in the LFW face database, and 97.86% in the AR face database. The algorithm has better performance than the commonly used ANNs and CNNS recognition methods. Finally, based on this algorithm, the face recognition system is realized. The system can input facial information and perform face recognition, which can effectively enhance safety management.

CCS CONCEPTS

Computing methodologies → Artificial intelligence →Computer vision → Computer vision problems

KEYWORDS

face recognition, convolutional neural network, LeNet-5, Pattern recognition.

ACM Reference format:

Lan Jiangyu, Xie Yinggang, Wang Hui, and Liu Guangjun. 2019. A Face Recognition System Based on Improved Convolutional Neural Network. In *Proceedings of 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI ' 19). Sanya, China, 6 pages.* https://doi.org/10.1145/3377713.3377765

†Corresponding author: xieyinggang@bistu.edu.cn

ACAI '19, December 20–22, 2019, Sanya, China

ACM ISBN 978-1-4503-7261-9/19/12…$15.00
https://doi.org/10.1145/3377713.3377765

1 Introduction

Face recognition technology is a major application in machine vision and has received extensive attention from researchers in recent years [1, 2]. Face recognition is currently widely used in multiple scenes [3, 4], such as security monitoring, human-computer interaction, etc. [5].

To improve the recognition rate, recall rate, recognition speed, stability and other indicators of face recognition [6], new technologies and new algorithms have sprung up. There are several corresponding algorithms in the steps of face preprocessing, feature extraction and feature matching. Among them, Principal Component Analysis (PCA) [7], Support Vector Machine (SVM) [8], Random Forest (RF) [9] and other algorithms are commonly used.

However, the above conventional face recognition method is susceptible to occlusion, illumination and other factors [10, 11]. In recent years, the Convolutional Neural Network (CNN) has been increasingly used in the research of face recognition. CNN is inspired by the natural visual perception mechanism of biology [12]. It is an end-to-end algorithm that combines feature extraction, feature matching, and result classification [13]. CNN improves the recognition rate of face recognition because it has a simple learning process and high robustness to face illumination, occlusion, expression changes, and other factors.

In the 1990s, LeCun proposed the concept of deep convolutional neural networks and constructed the LeNet-5 model for handwritten Arabic digits recognition [14]. In 2012, Krizhevsky et al [15] used the AlexNet network for image classification on the ImageNet image dataset. The recognition rate is much better than the traditional method, and the convolutional neural network has once again attracted people's attention. Afterward, several new convolutional neural network models were proposed. For example, Facebook proposed DeepFace [16] on CVPR2014, and the recognition rate of the LFW face database reached 97.35%; Google's FaceNet [17] was further improved in LFW face database. The recognition rate reached 99.63% recognition rate.

In this paper, we have proposed a CNN model with eight layers of network structure based on the LeNet-5 model. And a face recognition system based on the algorithm is implemented. Experiments have proved that the algorithm proposed in this paper has better performance in face recognition than the LeNet-5 model. And the face recognition system based on this algorithm can run stably.

2 Related Work

2.1 Convolutional Neural Network

Neural network algorithms began to emerge in the 1980s [18]. The neural network algorithm is composed of a plurality of nodes (neurons) connected, and each node represents a specific output function called an activation function. The connection between every two nodes has a weight that is trained. The network structure, weight values, and stimulus functions determine the output of the network.

Based on traditional neural network models, convolutional neural networks introduce new network layers and combine them. Deeper network models tend to perform better than shallow networks [19-21]. There are three main types of network layers of convolutional neural networks: convolutional layer, pooled layer, and fully connected layer. Each connection layer represents a linear mapping of different types of data.

2.2 LeNet-5

The key to establishing a convolutional neural network model is the design of the network structure. If the structure is not good, it is difficult to match the application data. And this will result in the model being trained not converging.

The algorithm is based on LeNet-5 model, which is shown in Figure 1. As shown, it consists of three parts: input layer, hidden layer and output layer. The input layer is a 32 × 32 single-channel image, and the output layer outputs an integer representing the category. The function of the hidden layer is to extract and classify image features. Generally, it consists of a convolutional layer, a sampling layer, and a fully connected layer.

The specific structure and parameters of the LeNet-5 model are shown in Table 1, including the type of layers, convolution kernel size, number of cores, the number of parameters, and the output size.

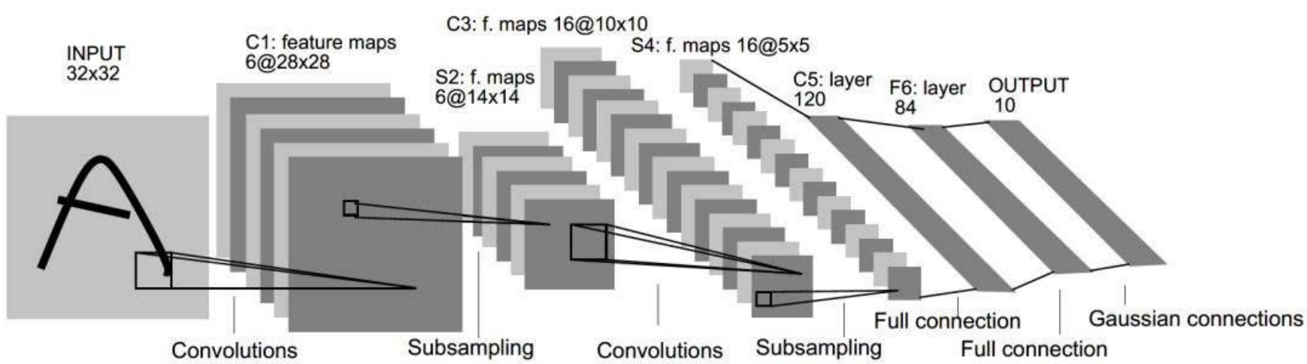

Fig. 1. Model structure of LeNet-5 using for handwritten Arabic digits recognition.

Table 1. LetNet-5 architecture specific configuration table.

Number of layers	Type	Convolution kernel size/number of cores	Number of parameters	Output size
C1	Convolutional layer	5×5/ 6	156	28×28
S2	pooling layer	————	————	14×14
C3	Convolutional layer	5×5/ 16	1516	10×10
S4	pooling layer	————	————	5×5
C5	Convolutional layer	5×5/120	48120	1×1
F6	full connection layer	84	10164	1×1
Output	————	————	10	10×1

3 Improved Algorithm

3.1 Network Structure Construction

The improvement of the network structure of the convolutional neural network can be adjusted from two aspects: changing the number of convolution layers and changing the size and number of convolution kernels. Therefore, the improved model starts with these two aspects.

The LeNet-5 model has seven layers. There are three convolutional layers, two pooled layers, one fully connected layer, and one output layer. The improved model is an eight-layer structure, as shown in Figure 2.

The LeNet-5 model is trained using a 5 x 5 convolution kernel. Previous studies have shown that the convolutional neural network performs best when the convolution kernel size is 3×3 [22], so the convolution kernel size is used for training.

The pooled layer in the improved convolutional neural network model uses a maximum pool of 2×2. Studies [22] have shown that the maximum pooling can achieve better training results in actual training.

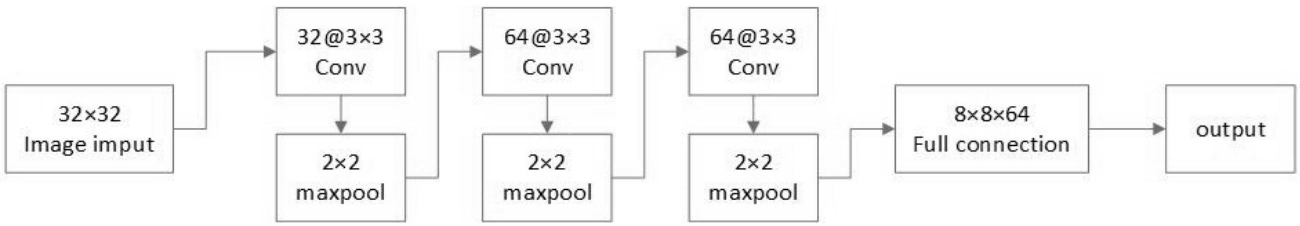

Fig. 2. Improved Convolutional Neural Network Model.

The output layer uses the SoftMax function for multiple classifications. Related studies have shown that SoftMax functions can achieve better results in multiple classifications than the ones used in LeNet-5 [23].

The LeNet-5 model is trained using a 5 x 5 convolution kernel. Studies have shown that the convolutional neural network performs best when the convolution kernel size is 3×3 [22], so this convolution kernel size is used for training.

The pooled layer in the improved convolutional neural network model uses a maximum pool of 2×2. Previous studies [22] have shown that the maximum pooling can achieve better training results in actual training.

The output layer uses the SoftMax function for multiple classifications. Related studies have shown that SoftMax functions can achieve better results in multiple classifications than the ones used in LeNet-5 [23].

3.2 Activation Function Adjustment

The LeNet-5 model uses the Sigmoid function as the activation function. However, the Sigmoid function has a convergent function characteristic, which may make the neural network easily fall into the gradient disappearance phenomenon.

If the value of x is too large or too small, the function value approaches a constant, and the first derivative value approaches zero. As a result, the neural network will disappear from the gradient.

The activation function of this paper uses the ReLU function. The function value of the ReLU function linearly increases as the value of X increases, preventing the occurrence of gradient disappearance. And because of its functional nature, the neural network can learn faster and accelerate the training process.

3.3 Add Dropout to Prevent Overfitting

Overfitting is a common problem in many machine learning algorithms. If an algorithm model is over-fitting, the resulting model will useless. Dropout can effectively alleviate the occurrence of over-fitting and achieve regularization [24, 25].

In each training, Dropout ignores half of the feature detectors (half the hidden layer node value is set to 0), which reduces the interaction between feature detectors (hidden layer nodes). Detector interaction means that some detectors rely on other detectors to function. Simply put, Dropout reduces the number of model nodes and makes the model more general because it doesn't rely too much on some local features.

In this paper, dropout is set to 0.5.

4 Experiment

4.1 Training Data

This paper uses the ORL face database to test the algorithm. The ORL face library was created by the AT&T Labs at the University of Cambridge, England. It contains a total of 400 facial images for 40 people. Some volunteer images include gestures, expressions, and changes in facial accessories.

Each of the collection objects of an acquisition object in the ORL face database contains 10 normalized grayscale images. The image size is 92×112 and the background of the images is black. The facial expressions and details of the collected objects vary, such as laughing and not laughing, eyes squatting or closing, and wearing or not wearing glasses. The poses of different face samples also change, with depth rotation and plane rotation up to 20 degrees. Some faces in the ORL face database are shown in Figure 3.

Fig. 3. ORL face library part of the face.

To facilitate the training of the model and the comparison of different algorithms, this paper first preprocesses the face in the face database. In this paper, the image processing function in the dlib library is used to grayscale the face, and the face is uniformly preprocessed to 32×32, which is convenient for importing the model for training and performance comparison with the LeNet-5 model.

4.2 Evaluation on Face Recognition

The LeNet-5 model and the improved convolutional neural network proposed in this paper are trained in the ORL face database, and cross-validation is performed to obtain the training results. The data set is randomly divided, and the entire data set is divided during the training.

After several rounds of iterations, the accuracy of both models tends to be stable. The recognition rate of the LeNet-5 model is about 96.08%, and the recognition rate of the improved model is about 99.87%. The experimental iteration process of the algorithm training is shown in Figure 4.

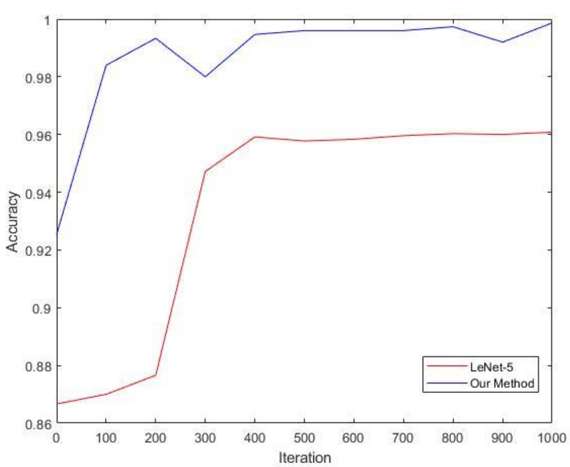

Fig. 4. LeNet-5 model training results.

Comparing the recognition rate of the two algorithms with the convergence speed, the following results show that the improved convolutional neural network model has a better recognition rate, as shown in Table 2.

Table 2. ORL face database algorithm comparison.

Model	Recognition rate
LeNet-5	96.08%
Our Method	99.87%

To further test the improved convolutional neural network in this paper, this paper selects the LFW face database and AR face database to test the algorithm.

Among them, the AR face database has more than 4,000 face images of 126 people, about 26 pairs per person, and the image is a 24-bit color map of 576*768 pixels. The image of the database is a positive face image with changes in expression, illumination, and occlusion. The LFW dataset mainly tests the recognition rate of face recognition. The database randomly selects 6000 pairs of faces to form a face recognition image pair. 3000 pairs of 2 faces of the same person, 3000 pairs of faces belonging to different people each.

In this paper, two face databases are processed, and the size of the face is also processed to 32×32. Then use the algorithm proposed in this paper to train.

It can be seen from Table 3 that the algorithm can achieve a higher recognition rate for different face databases. Since the three face libraries selected above are all obtained in the laboratory environment, they contain different face poses and lighting conditions. Therefore, the model has better robustness, higher recognition rate, and relatively satisfactory results.

Table 3. Comparison of different face database algorithms.

Face Database	Number of Face Categories	Total Number of Faces	Recognition Rate
ORL	40	400	99.87%
LFW	3000	6000	98.34%
AR	126	4000	97.86%

5 Face Recognition System

5.1 System Design

The face recognition system proposed in this paper can be used for security monitoring in schools, factories, etc. The system can acquisition face information and record the characteristics of the face. Face recognition can be performed when identity verification is required to prevent potential danger.

The system is technically divided into two parts, face acquisition, and facial recognition. The face acquisition part uses the camera to collect the face, extracts the feature information, and saves it to the feature library of the system. The face recognition part uses the camera to detect the face, preprocess, extract information, and feature comparison.

The specific process flow of the face recognition part is shown in figure 5.

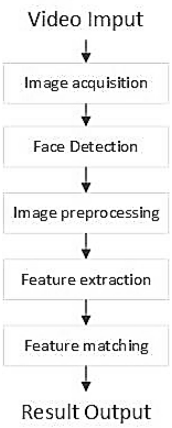

Fig. 5. Face recognition process flow.

5.2 System Implementation

First, the system uses the camera to obtain the face to be recognized. To achieve a better recognition effect, the system acquires multiple faces of different angles at a time and performs feature extraction. To verify the effect of the system, the program is run to obtain the face. Some faces after processing are shown in Figure 6.

Fig. 6. The collected author faces.

After completing the model training, we can use the system to recognize our face. The system will call the camera function of the OpenCV library, detect the face and frame the face. Finally, the system will return the user tag (filename) to show the result of the recognition. We can view the results in a text box called "Recognition Results".

After testing, the system can run normally, and the face can be detected and recognized normally. The interface and test results of the face recognition system are shown in Figure 7.

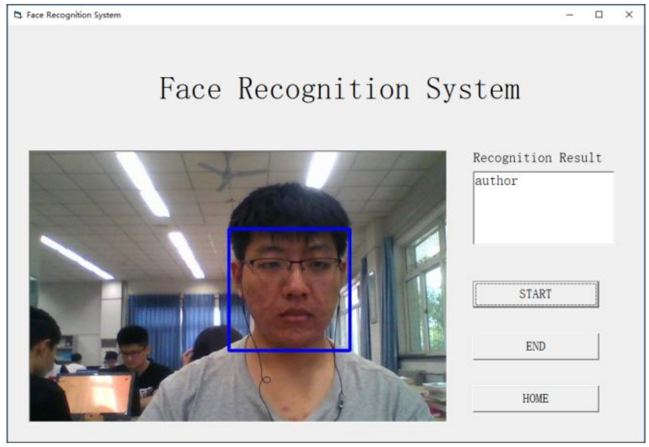

Fig. 7. Face recognition system.

6 Conclusion

This paper proposes an improved convolutional neural network algorithm based on LeNet-5 and uses Python language and TensorFlow library to complete the algorithm. The algorithm is tested in the ORL, AR, and LFW. The experiment proves that the improved algorithm proposed in this paper has a high recognition rate in the face database. Besides, this paper implements a face recognition system based on the algorithm. The system calls the camera to detect and recognize the face to get the corresponding matching result.

The algorithm and face recognition system of this paper is mainly based on the face database under ideal conditions. In the real world, face recognition is more complicated. Light conditions, face poses, and occlusions all affect the recognition rate of face recognition. In future research, the robustness of the system can be improved by further improving the model and performing face recognition tests under various conditions.

ACKNOWLEDGMENTS

This work is supported by Beijing Natural Science Foundation (Grant No.4192023); The National Natural Science Foundation of China (61603047); The Qin Xin Talents Cultivation Program of BISTU. (Grant No. QXTCPC201704)

REFERENCES

[1] Lahasan B, Lutfi S L, San-Segundo R. A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression[J]. ARTIFICIAL INTELLIGENCE REVIEW. 2019, 52(2): 949-979.

[2] Hassaballah M, Aly S. Face recognition: challenges, achievements and future directions[J]. IET COMPUTER VISION. 2015, 9(4): 614-626.

[3] Guo Y, Zhang J, Cai J, et al. CNN-Based Real-Time Dense Face Reconstruction with Inverse-Rendered Photo-Realistic Face Images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019, 41(6): 1294-1307.

[4] Chen J, Chen Z, Chi Z, et al. Facial Expression Recognition in Video with Multiple Feature Fusion[J]. IEEE Transactions on Affective Computing. 2018, 9(1): 38-50.

[5] Ding C, Tao D. Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018, 40(4): 1002-1014.

[6] He K, Sun J, IEEE. Convolutional Neural Networks at Constrained Time Cost[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[7] Low C, Teoh A B, Ng C. Multi-Fold Gabor, PCA, and ICA Filter Convolution Descriptor for Face Recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology. 2019, 29(1): 115-129.
[8] Nguyen D T, Pham T D, Baek N R, et al. Combining Deep and Handcrafted Image Features for Presentation Attack Detection in Face Recognition Systems Using Visible-Light Camera Sensors[J]. 2018, 18(3): 29.
[9] Fanelli G, Gall J, Van Gool L. Real Time Head Pose Estimation with Random Regression Forests[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[10] Zhang W, Zhao X, Morvan J, et al. Improving Shadow Suppression for Illumination Robust Face Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019, 41(3): 611-624.
[11] Best-Rowden L, Jain A K. Longitudinal Study of Automatic Face Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018, 40(1): 148-162.
[12] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition. 2018, 77: 354-377.
[13] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascad-ed Convolutional Networks[J]. IEEE Signal Processing Letters. 2016, 23(10): 1499-1503.
[14] El-Sawy A, El-Bakry H, Loey M. CNN for Handwritten Arabic Digits Recognition Based on LeNet-5[C]. 2017.
[15] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM. 2017, 60(6): 84-90.
[16] Taigman Y, Yang M, Ranzato M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[17] Schroff F, Kalenichenko D, Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[18] Schmidhuber J. Deep learning in neural networks: An overview[J]. NEURAL NETWORKS. 2015, 61: 85-117.
[19] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[20] Peng C, Wang N, Li J, et al. DLFace: Deep local descriptor for cross-modality face recognition[J]. PATTERN RECOGNITION. 2019, 90: 161-171.
[21] Gorgel P, Simsek A. Face recognition via Deep Stacked Denoising Sparse Autoencoders (DSDSA) [J]. APPLIED MATHEMATICS AND COMPUTATION. 2019, 355: 325-342.
[22] Du Tran, Bourdev L, Fergus R, et al. Learning Spatiotemporal Features with 3D Convolutional Networks[C]. 2015.
[23] Wang F, Cheng J, Liu W, et al. Additive Margin Softmax for Face Verification[J]. IEEE Signal Processing Letters. 2018, 25(7): 926-930.
[24] Shen X, Tian X, Liu T, et al. Continuous Dropout[J]. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. 2018, 29(9): 3926-3937.
[25] Ranjan R, Patel V M, Chellappa R. HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019, 41(1): 121-135.