Face Recognition Using Deep Convolutional Neural Network in Cross-Database Study

Mei Guo[1], Min Xiao[1(✉)], and Deliang Gong[2]

[1] College of Software and Communication Engineering, Xiangnan University, Chenzhou 423000, Hunan, China
xnxyxm@163.com

[2] Comprehensive Experimental Training Center, Xiangnan University, Chenzhou 423000, Hunan, China

Abstract. In this paper we study face recognition using convolutional neural network. First, we introduced the basic CNN neural network architecture. Second, we modify the traditional neural network and adapt it to another database by fine tuning its parameters. Third, the network architecture is extended to the cross database problem. The CNN is first trained on a large dataset and then tested on another. Experimental results show that the proposed algorithm is suitable for building various real world applications.

Keywords: Face recognition · Deep neural network · Image processing

1 Introduction

Face recognition has been a popular topic in computer science and artificial intelligence in the past decades [1–6]. There are many popular deep neural network models for image classification. Such as AlexNet [7], VggNet [8], GoogleNet [9], etc. These models are trained on the large image databases, and the fine-tuning on these models can be adapted to specific applications.

In many image classification problems, it is challenging to transfer one model to the other. The training process requires a large amount of data and the resulting model sometimes is lack of generalization ability. Therefore it is important to fine-tuning the existing deep models and adapt it to wider applications on cross-database situation.

We study the face recognition based on convolutional neural network. We modified the traditional network and adapt it to another database by fine tuning its parameters. The proposed algorithm is suitable for building various real world applications.

In this paper, we propose to use a modified convolutional neural network architecture for face recognition. A modified convolutional neural network architecture is demonstrated in Fig. 1. The rest of the paper is organized as follows: Sect. 2 gives the details on the basic network. Section 3 gives the modified face recognition algorithm. Experimental results are provided in Sect. 4. Finally, conclusions are drawn in Sec. 5.

F. Xhafa et al. (eds.), *Advances in Intelligent Systems and Interactive Applications*, Advances in Intelligent Systems and Computing 686,
https://doi.org/10.1007/978-3-319-69096-4_54

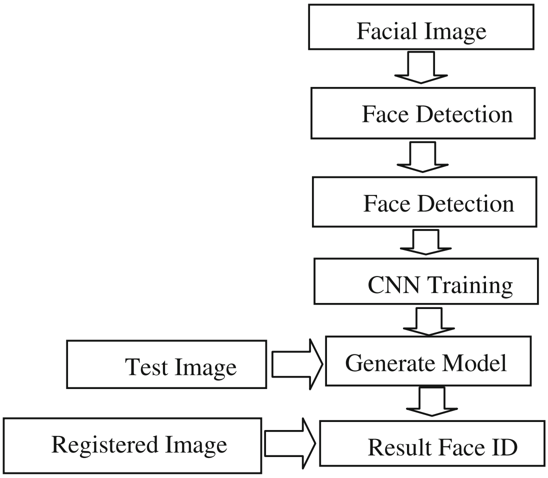

Fig. 1 A depiction of the face recognition system layout.

2 Deep Neural Network

The facial image features are constructed by the deep neural network. DNN has the ability to generate highly abstracted features that are advantageous compared to the traditional hand crafted features. The traditional face model is presented. Alex Krizhevsky in Toronto University proposed the deep convolutional neural network, also known as AlexNet [10]. It has a popular convolutional structure and it is widely used in the face recognition model. Many current models are inspired by AlexNet, such as Vgg-Face model. It extend the original structure to deeper levels. In the contest of Imagenet LSVRC-2012 contest, 1,200,000 images are classified into 1000 classes [10]. The convolutional neural network achieves satisfactory results. The structures are demonstrated in Fig. 2.

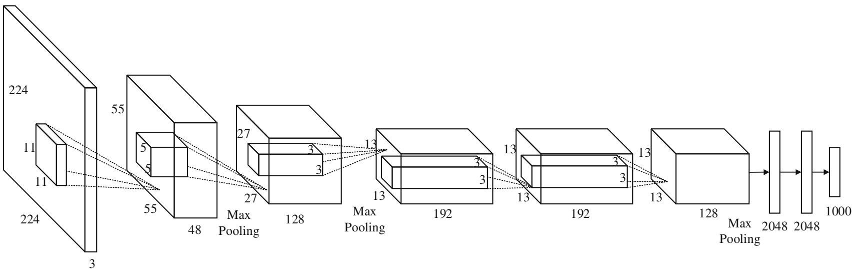

Fig. 2 Traditional convolutional neural network architecture [10].

The input image is first processed by five convolutional layers for feature abstraction. It is then send into the fully connected layers for modelling. The final

Softmax classifier is used for face classification. The convolutional layer involves convolution, activation, pooling and local response normalization.

3 Improved Neural network Architecture

The existing neural network model can be adapted to specific recognition tasks with modified network architecture and fine turning. The traditional network training method requires a very large number of training samples which is not practical in real application. In real world the training sample is limited to hundreds with reasonable human annotation. Therefore we propose to modify the network architecture to better suit this cross database problem. The CNN is first trained on a large dataset and then tested on another in the experimental section.

The basic neural network nodes are shown in Fig. 3.

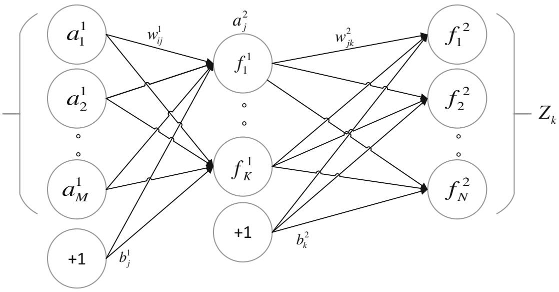

Fig. 3 Neural network structure

The input layer is denoted as a_i^1, the hidden layer is denoted as a_i^2. b_i^1 and b_i^2 is the bias parameter and f_i^1 and f_i^2 are the non-linear activation function of each nodes. The output results is denoted as Z_i.

The calculation of Z_i is:

$$\begin{cases} a^2 = f^1\left(w^1 * a^1 + b^1\right) \\ a^3 = f^2\left(w^2 * a^2 + b^2\right) \\ \qquad\qquad\vdots \\ Z = f^i\left(w^i * a^i + b^i\right) \end{cases} \tag{1}$$

The activation function is sigmoid:

$$y = f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

tanh:

$$y = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} \tag{3}$$

The neural network approximates the relation between target and feature through the activation function [11, 12].

In order to improve the cross database performance and adapt to different real world data, the modified structure is presented in Table 1.

Table 1. The modified neural network structure

Input (224 × 224 RGB)
Conv3-64
Conv3-64
MaxPooling
Conv3-128
Conv3-128
MaxPooling
Conv3-256
Conv3-256
Conv3-256
Conv3-256
MaxPooling
Conv3-512
Conv3-512
Conv3-512
Conv3-512
MaxPooling
Conv3-512
Conv3-512
Conv3-512
Conv3-512
MaxPooling
FullContact-4096
FullContact-4096
FullContact-4096
FullContact-1000
SoftMax

4 Experimental Results

The fine tuning is based on the Vgg_face model and retraining each layer in the model is not practical. Therefore we make use of the convolutional parts in the model ant using the new training samples to adjust the parameters. The Vgg-Face model is trained on 2.6M images and the retraining is carried out on the CASIA database [13], examples of the database is shown in Fig. 4. It includes different subjects with different expression and poses. In each class there are around 400 images and the least coherent has 20 images. We preprocess the coherents for balanced training sets. The classes that have more then 50 samples are chosen and form the training set. There are 10,575 people in total in the original database.

Fig. 4 CASIA-WebFace Image Examples

The established model is tested on LFW database [14]. In this database the faces are captured in the wild. It is not some posed images used in the lab, but real data from real world. Therefore it is more accurate for the face recognition study compared to other simple face images taken in lab environment which is not suitable for real applications.

There are illumination changes in the image sample in LFW database. We first prepocess the images by equalization techniques. Then we use the proposed neural network to extract the face features. The feature vectors are then compared in distance and the final ID is achieved in view of similarity.

In our test, the hardware platform configuration is given in Table 2. The LFW database provides test dataset that contains 10 groups. In each group, there are 600 pairs of images from the same people and 300 pairs from different people. We calculate the image feature vectors from CNN and the distance measure. After normalize the feature vectors to [0, 1], we test the true positive and false negative rates, as well as the false positive and true negative rates. Results are shown in Table 3. We can see that the proposed algorithm is effective for the adaptation on LFW dataset.

Table 2. Hardware configuration of the face recognition experiment.

Item	Parameter
CPU	Intel i5 3450 3.1GHz 6M
Memory	Kinston DDR3 16G
GPU	NVIDIA GeForce GTX Titan X 12G

Table 3. Face recognition results

Normalized threshold	True positive/False negative	False positive/True negative
0.3700	2732/280	53/2946
0.4263	2833/141	154/2846
0.5200	2931/94	240/2760
0.6320	2932/13	1270/1730

5 Conclusion

In this paper, we propose to use a modified convolutional neural network architecture for face recognition. Gives the details on the basic network, and gives the modified face recognition algorithm. Experimental results show that the proposed algorithm is effective for the adaptation on LFW dataset.

References

1. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
2. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. BMVC **1**(3), 6 (2015)
3. Haig, N.D.: The effect of feature displacement on face recognition. Perception **42**(11), 1158–1165 (2013)
4. Sagar, G.V., Barker, S.Y., Raja, K.B., et al.: Convolution based face recognition using DWT and feature vector compression. In: Third International Conference on Image Information Processing (ICIIP), pp. 444–449. IEEE, New York (2015)
5. Lu, J., Tan, Y.P., Wang, G.: Discriminative multimanifold analysis for face recognition from a single training sample per person. IEEE Trans. Pattern Anal. Mach. intell. **35**(1), 39–51 (2013)
6. Sun, Y., Liang, D., Wang, X., et al.: Deepid3: Face recognition with very deep neural networks. arXiv preprint arXiv:1502.00873 (2015)
7. Ballester, P., Araujo, R.M.: On the performance of GoogLeNet and AlexNet applied to sketches. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
8. Wu, X., He, R., Sun, Z.: A lightened cnn for deep face representation. In: IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR) (2015)
9. Zhong, Z., Jin, L., Xie, Z.: High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. In: 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 846–850. IEEE, New York (2015)

10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural information Processing Systems, pp. 1097–1105 (2012)
11. Deng, L., Hinton, G., Kingsbury, B.: New types of deep neural network learning for speech recognition and related applications: An overview. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8599–8603. IEEE, New York (2013)
12. Kingsbury, B., Sainath, T.N., Soltau, H.: Scalable minimum Bayes risk training of deep neural network acoustic models using distributed Hessian-free optimization, pp. 10–13. In: Interspeech (2012)
13. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014). Available: http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html
14. Huang, G.B., Ramesh, M., Berg, T., et al.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, University of Massachusetts, Amherst (2007)