Face Recognition Based on Deep Features

Li Yi*, Sun Zhifang
Shandong University of Engineering and Vocational Technology, Jinan China, 250200
amgable@avc.edu

Abstract. A face recognition method based on deep feature decision fusion is proposed. First, a convolutional neural network (CNN) is designed for deep feature learning on face images. Then, all the feature are mapped to construct a single depth feature vector. In the classification stage, the sparse representation-based classification is used to characterize the constructed depth feature vector, and the face category of the object to be recognized is determined according to the overall reconstruction error. Experiments are carried out on ORL and Yale-B datasets and compared with several existing face recognition methods. The results verify the effectiveness and robustness of the proposed method.

1. Introduction
Face recognition is a widely studied problem in pattern recognition. Generally, algorithms are designed from two perspectives: feature extraction and classifier design. In terms of feature extraction, a large number of projection features based on manifold learning [1-3] and texture features based on local structure [4] [5] have been successfully used in face recognition. Classical classifiers, such as support vector machine (SVM) [6], sparse representation classification (SRC) [7] [8], have become powerful tools for face recognition. In recent years, face recognition methods based on deep models have become mainstream. Convolutional neural network (CNN) has become the most commonly used deep learning model in face recognition due to its own structural characteristics of processing two-dimensional data [9-13]. This paper proposes a face recognition method based on feature fusion of convolutional neural network. The network designs multiple convolutional layers to learn the multi-level deep features of face images. The feature maps learned by different convolution kernels reflect the different features of the original image and have good complementarity. Therefore, by combining multi-level deep features, attention should be paid to improve the final recognition performance. In this paper, SRC is used to characterize and classify the multi-level deep features obtained from face image learning. In the experiment, tests were carried out based on ORL and Yale-B face image datasets respectively, and the results showed their effectiveness.

2. Convolutional neural network and deep feature construction
2.1. Basic Theory
The core of CNN is to learn the multi-level features of the input image through convolution operation, and finally use these features to classify and determine its attribute category [9-13]. This paper designs the CNN structure shown in Figure 1 for the face recognition problem. There are 3 convolutional layers, 3 pooling layers, and two fully connected layers. Among them, the convolutional layer learns...
the multi-level features of the input image through different convolution kernels. The pooling layers use the maximum pooling to reduce the overall operation law of the network. The fully connected layer iteratively generates the final category label. The classic Rectified linear unit (ReLU) is used as the activation function to improve the nonlinear processing capability of the network. Compared with the CNN-based face recognition methods in the existing literature, the network structure designed in this paper is relatively simple. However, this article aims to make full use of the learned deep features and not just use CNN as an end-to-end classifier, so the complexity of the network itself is relatively low.

Based on the CNN shown in Fig. 1, this paper constructs depth feature vectors in each convolutional layer to provide support for subsequent classification decisions. On each convolutional layer, all feature maps are first converted into a feature vector by means of vector concatenation. Then, the dimension is reduced to a feature vector of length 200 through downsampling (the downsampling factor can be determined accordingly). Finally, based on the designed network structure, a unified deep feature vector can be obtained, which takes into account all the feature maps obtained by network learning. Therefore, it can better reflect the characteristics and attributes of the input face image, and help improve the subsequent recognition performance.

![Fig. 1 Architecture of designed CNN.](image)

### 2.2. Basics of SRC

SRC is used as the classifier for the fused deep features. For a test sample \(y\), it is represented on a global dictionary as follow:

\[
\hat{a} = \arg\min_a \|a\|_0 \\
\text{s.t. } \|y - A\hat{a}\|_2 \leq \epsilon
\]

(1)
where $A = [A^1, A^2, \ldots, A^C]$ is the global dictionary; $a$ denotes the sparse coefficient vector. With the estimation of $\hat{a}$, the reconstruction error of each training class is calculated as equation (2).

$$r(i) = \|y - y^i\|_2 (i = 1, 2, \ldots, C)$$

For the corresponding class to the test sample, it mostly probably produces the minimum reconstruction error. Therefore, to determine the target label of the test sample, the reconstruction errors of all the training classes are compared and assign it as the minimum one.

3. Experiments

3.1. Data for Experiments

The ORL and Yale-B face image datasets are used to test the performance of the proposed method. ORL data set collected 10 face images of 92 pixels×112 pixels each of 40 people, which came from different times and different lighting conditions. The Yale-B data set includes 45 face images of 10 people each under different lighting conditions, with an image size of 32 pixels×32 pixels. The partial face samples of the two data sets are shown in Figure 3. For the ORL data set, 5 images of each person are randomly selected for training, and the remaining 5 images are used as test samples. For the Yale-B data set, 20 images of each person are randomly selected for training, and 25 samples are left as test samples.

As a comparison, some face recognition methods in the existing literature are selected for comparison experiments, including the method in the literature [6] (Method 1); the method in the literature [8] (Method 2) and the method in the literature [9] (Method 3). Subsequent experiments are first based on the original training and test samples to investigate the basic recognition performance of the proposed method. Then, the method of noise addition is used to investigate the recognition performance of the proposed method under noise interference conditions to verify its robustness.

3.2. Results and Analysis

In the actual process, the face image to be recognized may come from a noisy environment. With the continuous improvement of the noise level, the facial features in the image are gradually submerged in the noise, which brings greater obstacles to correct recognition. The face images in the original ORL and Yale-B datasets mostly come from high signal-to-noise ratio (SNR) conditions. In order to test the robustness of the proposed method under noise interference conditions, this paper applies additive white Gaussian noise to the original test images in two data sets, and constructs test samples under different signal-to-noise ratios by controlling the added noise level. Based on the noise samples, the recognition performance of various methods is tested and the results shown in Fig. 2 are obtained. It can be seen that the method in this paper can maintain the strongest performance robustness on both data sets. Although the results of the Method 3 are better than Method 2, it no longer has a performance advantage under the low signal-to-noise ratio conditions in this experiment, which is mainly due to the robustness of the sparse representation to noise interference. This method finally improves the overall noise robustness of the recognition method by combining multi-level deep features and joint sparse representation.
4. Conclusion
This paper proposes a face recognition method based on feature fusion of convolutional neural network. First, we analyze the multi-level feature maps learned by the neural network, and construct deep feature vectors in each convolutional layer. Then, SRC is used for decision fusion of the constructed deep features, and the category of the face image to be recognized is determined according to the final reconstruction error. The proposed method can make full use of depth features, thereby promoting the improvement of overall recognition performance. Experiments are carried on ORL and Yale-B face data sets. Experimental results show that the proposed method has strong effectiveness and robustness for face recognition problems.

References
[1] Turk, M. A. and Pentland, A. P. 1991. Face recognition using eigen-faces. Proc. Of IEEE Conf. Computer Vision and Pattern Recognition, 1991, 586–591.
[2] He, X., Yan, S. and Hu, Y., et al. 2005. Face recognition using Laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell., 27, 3, 328–340.
[3] Wang, Y. and Wu, Y. 2010. Face recognition using Intrinsicfaces. Pattern Recognition, 43, 3580-3590.
[4] Wen, Y., He, L., and Shi, P. 2012. Face recognition using difference vector plus KPCA. Digital Signal Processing, 22, 1, 140-146.
[5] Choi, J. Y., Ro, Y. M., and Plataniotis, K. N. 2012. Color local texture features for color face recognition. IEEE Transactions on Image Processing, 21, 3, 1366-1380.
[6] Lu, J., Liong, V. E., and Zhuang, X., et al. 2015. Learning compact binary face descriptor for face recognition. IEEE Trans. Pattern Anal. Mach. Intell., 37, 10, 2041–2056.
[7] Zhang, Y., Liu, C. 2003. Face recognition based on support vector machine and nearest neighbor classifier. Journal of Systems Engineering and Electronics, 14, 3, 73–76.
[8] Wright, J., Yang, A., and Ganesh, A., et al. 2009. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell., 31, 5, 210–227.
[9] Lu, J., Wang, G., and Deng, W., et al. 2017. Simultaneous feature and dictionary learning for image set based face recognition. IEEE Trans Image Process, 26, 8, 4042-4054.
[10]Wagner, A., Wright, J., and Ganesh, et al. 2012. Toward a practical face recognition system: robust alignment and illumination by sparse representation. IEEE Trans. Pattern Anal. Mach. Intell., 34, 2, 372-386.
[11]Ran, H., Zheng, W. S., and Hu, B. G., et al. 2012. Two-stage nonnegative sparse representation for large-scale face recognition. IEEE Transactions on Neural Networks & Learning Systems, 24, 1, 35-46.
[12] Ding, C. and Tao, D. 2015. Robust face recognition via multimodal deep face representation [J]. *IEEE Trans. Multimedia*, 17, 11, 2049–2058.

[13] Gao, S., Zhang, Y., and Jia, K., et al. 2017. Single sample face recognition via learning deep supervised autoencoders. *IEEE Transactions on Information Forensics & Security*, 10, 10, 2108-2118.

[14] Wimalajeewa, T., Varshney, P. K. 2014. OMP based joint sparsity pattern recovery under communication constraints. *IEEE Trans. Signal Process.*, 62, 19, 5059-5072.