Wireless Pers Commun (2018) 103:1195–1206
https://doi.org/10.1007/s11277-018-5377-2

Improving Deep Learning Feature with Facial Texture Feature for Face Recognition

Yunfei Li[1,2] · Zhaoyang Lu[1] · Jing Li[1] · Yanzi Deng[1]

Published online: 9 February 2018

Abstract Face recognition in the reality, is a challenging problem, due to varieties in illumination, background, pose etc. Recently, the deep learning based face recognition algorithm is able to learn effective face features to obtain a very impressive performance. However, this kind of face recognition algorithm completely relies on the machine learning based face features, while ignores the useful experience in hand-craft features which have been studied in a long period. Therefore, a face recognition based on facial texture feature aided deep learning feature (FTFA-DLF) is proposed in this paper. The proposed FTFA-DLF is able to combine the benefits of deep learning and hand-craft features. In the proposed FTFA-DLF method, the hand-craft features are texture features extracted from the eyes, nose, and mouth regions. Then, the hand-craft features are used to aid deep learning features by adding both deep learning and hand-craft features into the objective function layer, which adaptively adjusts the deep learning features so that it can better cooperate with the hand-craft features and obtain a better face recognition performance. Experimental results show that the proposed face recognition algorithm on the LFW face database to achieve the accuracy rate of 97.02%.

Keywords Face recognition · Convolution neural network · Facial texture feature · Feature fusion

✉ Yunfei Li
wnlff@126.com

Zhaoyang Lu
zhylu@xidian.edu.cn

Jing Li
jinglixd@mail.xidian.edu.cn

Yanzi Deng
dyzamour@163.com

1 School of Telecommunications Engineering, Xidian University, Xian 710071, China

2 School of Network Security and Information, Weinan Normal University, Weinan 714099, China

1 Introduction

Due to faces are intuitive, face recognition has wide applications in practice, such as human computer interaction [1], human identification [2], attendance system [3] and multimedia management [4], etc. The face recognition under unconstrained conditions is a challenging problem, since varieties in illumination, background, pose etc. At present, the research focus of face recognition is to design an efficient face feature representation, making the distance between the same subjects as small as possible, and the distance between different subjects as large as possible, which is beneficial for face recognition.

The face feature representation methods can be roughly divided into two categories: (1) hand-crafted features [5–11] based on manually designs; (2) deep feature learning methods [12–16]. From the perspective of the face feature development history, the hand-crafted feature design methods are more traditional methods. Because hand-crafted feature design methods based on manually experience are researched in a long period and the deep feature learning methods are really popular recently. There are many hand-crafted features used in face recognition, such as local binary pattern (LBP) [5] and its variants multi-block local binary pattern (MBLBP) [6], histogram of oriented gradients (HOG) [7], Gabor [10] and scale-invariant feature transform (SIFT) [9]. These face feature representations are local descriptors and hold illumination robustness or scale invariance, which are suitable to describe face images. In practice, in order to further improve face features, multiple features are fused together to represent a face image. For example, the affine invariant key-points descriptors and Gabor ternary pattern (GTP) are fused in [11].

The aforementioned hard-crafted face feature representation methods are used to extract the low-level features of face images. For example, the LBP and Gabor features are used to describe the texture information, HOG features are used to describe the gradient information and SIFT features are used to describe the key-point information, respectively. The discriminative ability of these low-level features cannot meet the demand of face recognition directly, thus it is need to build face classifiers on these low-level features based on supervised machine learning algorithms to obtain a better face recognition performance. The commonly used supervised learning algorithms for face recognition include AdaBoost [17] algorithm, support vector machine (SVM) [18], partial least square (PLS) [19], Joint Bayesian [20] and metric learning algorithm [21], etc.

From the perspective of face recognition performance, deep learning based face recognition methods [15, 16, 33–37] obtain better results than the traditional face methods that based on hand-craft features and supervised machine learning algorithms on some public face databases, such as LFW [22], YTF [23]. The key to the success of deep learning based face recognition methods is a group of effective face features learned by deep learning algorithms. There are two kinds of network structure using for face feature learning: (1) unsupervised network structure [12]; (2) supervised network structure [13, 15, 16]. At present, the supervised network structure achieves better face recognition performances, since it applies the supervision information during the training progress.

Face recognition can be divided into two sub-tasks: verification, it is used to judge whether a pair of face images belong to the same subject; (2) identification, it is used to identify a face image belongs to which subject. Obviously, the verification sub-task is a binary classification problem and the identification sub-task is a multi-class classification problem, respectively. For the face verification sub-task, Sun et. al [13] proposed a face verification method based on a convolutional neural network (CNN) and restricted Boltzmann machine(RBM). In this method, the CNN is lied on the bottom for feature extraction, while the RBM is lied on the top for classification (1-class represents a pair of images belong to the same subject, 0-class represents a pair of images belong to different subjects). For the face identification sub-task, the DeepID [15] and DeepID2 [16] are proposed. In the DeepID method, the CNN is used to extract face features and the Softmax function is used to predict the identification of a face image. The main difference between [13] and [15] is that the supervision signal used in [15] is much powerful than that used in [13], since the supervision signal in [15] is a multi-class label rather than a binary label used in [13]. It is worth mentioning that the DeepID is the first face recognition method which has an identification capability closing to human on the LFW [22] dataset. Based on DeepID, DeepID2 [16] jointly combines the verification and identification supervision signals into the objective function and obtain a better face recognition result.

Most deep learning based face recognition methods [15, 16, 24] learn face features completely dependent on machine learning and ignore the useful experience of hand-crafted face feature design. However, the role of the hand-crafted face features are still effective, for example the high-dim LBP [31] is able to be comparable with several deep learning based features. For this, in this paper a face recognition based on facial texture feature aided deep learning feature (FTFA-DLF) is proposed. The proposed FTFA-DLF is able to combine the benefits of CNN learning and hand-craft feature to improve the performance of face recognition. In the FTFA-DLF, the hand-craft features are texture features extracted from the eyes, nose, and mouth regions. Then, the hand-craft features are used to aid deep learning features by adding both deep learning and hand-craft features into the objective function layer, which adaptively adjusts the deep learning features so that it can better cooperate with the hand-craft features and obtain a better face recognition performance. The contributions of this paper are follows. First, an effective facial texture features aided deep neural network is designed for improving face recognition performance. Second, the face recognition performance of the proposed FTFA-DLF is superior to state-of-the-art methods. The rest of this paper is organized as follows. Section 2 introduces the details of the proposed face recognition method. Section 2.4 presents experimental results. Section 3 concludes this paper.

2 Proposed Face Recognition Method

2.1 Facial Features

The overall outline of the human face and the facial components are generally similar, consisting of the eyes, eyebrows, nose and mouth, and the shapes of these components are similar. These facial features are the decisive factors to distinguish between different faces, face recognition is mainly based on the overall facial contour and facial features of the shape and location to distinguish between different people [3].

All the features in the face are related to each other with a certain relationship, such as a big mouth not matching a pair of small eyes, and a big face not going with a small nose. In order to find out the relationship between face features, we labeled over 1700 individuals eye corners, nose wings and mouth corners by hand which were from the ORL database (40 persons *10 images), lab self-built database (20 persons *15 images) and Chinese ID cards (1000 persons*1 image) (the second version), as shown in Fig. 1a, and then the width of all eyes, nose wings and the mouth corners were investigated as shown in Fig. 1b. The results show that the ratio of one individuals eye, nosewing and mouth is relatively fixed, namely, the individuals eye, nose wing and mouth is proportional. So theoretically in the

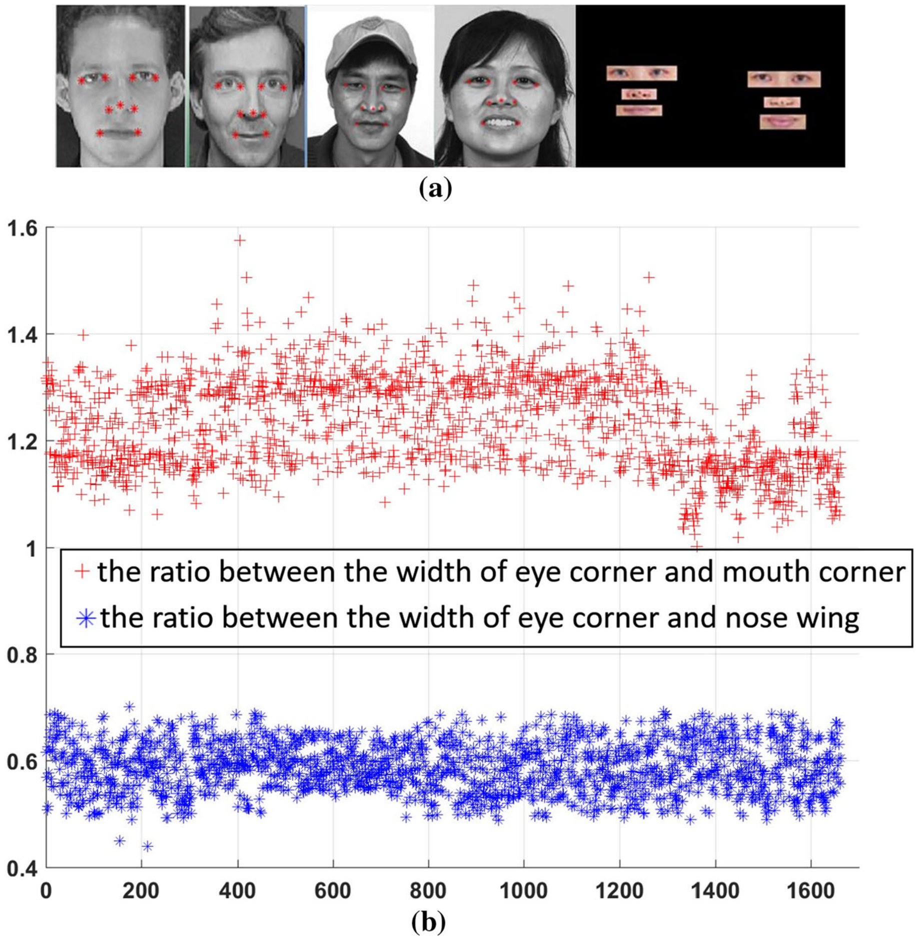

Fig. 1 The ratio of facial features. **a** Facial features (from the face database of ORL, Our lab, Second-generation ID). **b** The ratio between the width of eye corner and mouth corner $\doteq\doteq(*)$ and the ratio between the width of eyes corner and nosewings $(+)$

face recognition process, any one of the facial features can be identified as the unique feature [4]. However, the recognition and alignment accuracy of the facial features will play a decisive role in an individuals whole recognition process, and the recognition performance will decrease sharply when the illumination and pose changes.

2.2 FTFA-DFL Module

As shown in Fig. 2, the proposed face recognition method consists of two FTFA-DFL feature learning branches with same structure and parameters. Each FTFA-DFL feature learning branch includes two sub-modules, facial texture feature extraction module and ResNet [32] based deep learning feature learning module, as shown in Fig. 3. The two sub-modules are detailed introduced as follows.

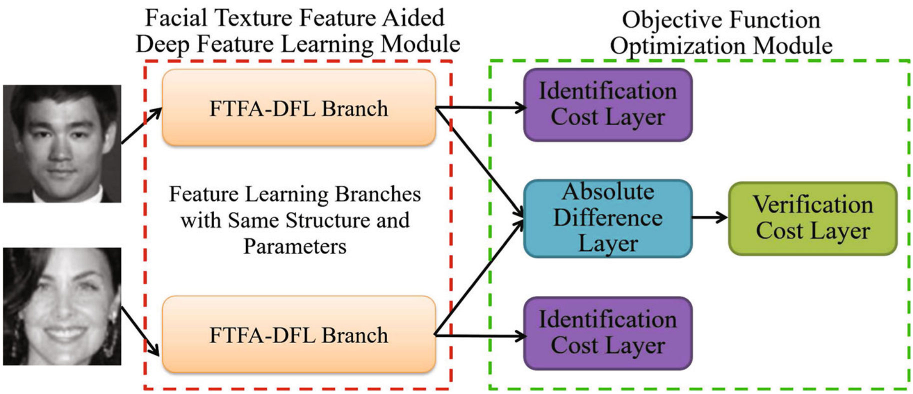

Fig. 2 The framework of the proposed face recognition method

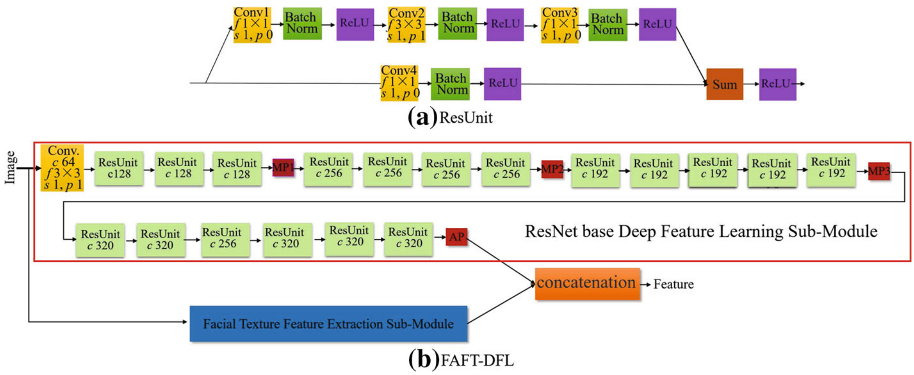

Fig. 3 The FTFA-DFL feature learning branch

2.2.1 Facial Texture Feature Extraction Module

Local binary pattern (LBP) and its variants are most commonly used features in the face recognition field. In this paper, the multi block local binary pattern (MBLBP) [6] is applied, because MBLBP has better light illumination robustness than LBP. Considering that facial regions, such as eyes, nose, and mouth, are the most discriminating regions and in order to reduce the complexity of MBLBP feature extraction, this paper only extract MBLBP features from facial regions rather than the whole face region.

To extract MBLBP features from facial regions, we need to locate facial regions first, this is related to the face alignment technology [25]. In this paper, we adopt the CASIA-WebFace [24] database as the training set. The CASIA-WebFace database provides aligned face images. See Fig. 4 for some aligned face samples from the CASIA-WebFace database. Since the face images from the CASIA-WebFace are aligned, we can directly locate facial regions in some fixed coordinates. As shown in Fig. 5, each fixed facial region is represented by 1×1 and 3×3 uniform MBLBP histogram with $59 \times 2 = 118$ dimensions. The features of eyes, nose and mouth regions are concatenated together to form the final hand-craft facial features and these calculations are implemented by the facial texture feature extraction module shown in Fig. 3b.

Fig. 4 Landmark based aligned face image from the CASIA-WebFace [24]] database

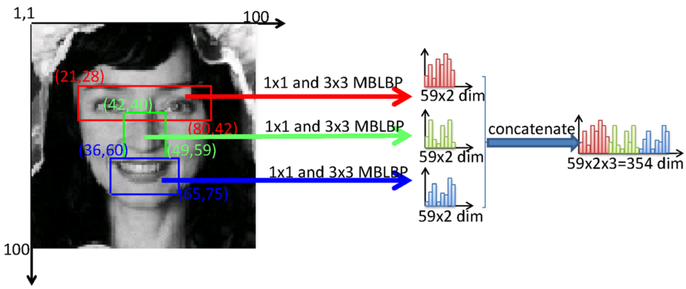

Fig. 5 MBLBP feature extraction from facial regions

2.2.2 ResNet based Deep Feature Learning Module

From Fig. 2a, one can see that there are four convolutional layer (i.e. Conv1, Conv2, Conv3 and Con4). The 11 sized filters applied in Conv1, Conv3 and Conv4, while the 3×3 sized filters applied in Conv2. For all convolutional layers zero padding and 1-pixel step stride is applied. The MP1, MP2 and MP3 in Fig. 2 represent three max pooling layers and the AP layer means an average pooling layer. For these three MPs, the 2×2 sized pooling sub-windows and 2-pixel step stride are used. In this paper, all face images are unified resized as 100×100 sizes, thus the pooling sub-window used in the AP layer is 7×7.

Based on the ResUnit, the ResNet based deep feature learning module is constructed, as shown in Fig. 2b. More specially, along the four MP layers, there are 3, 4, 5 and 6 ResUnits are arranged, with 128, 256, 192 and 320 channels, respectively.

As shown in Fig. 2, both the facial texture features and deep learning features are concatenate together. This concatenated feature are then fed into the objective function optimization module for adaptively adjusting the deep learning features to cooperate with the facial texture features, by adjusting the parameters, W_i and W_v in the identification (i.e. Eq. 2) and verification (i.e. Eq. 3) cost functions, respectively.

2.3 Objective Function Optimization Module

As shown in Fig. 3, the fusion feature is obtained by concatenating the facial texture and CNN learning features together. Then, the fusion feature is inserted into the objective function optimization model for the training of face recognition, which adaptively adjusts the CNN learning features so that it can better cooperate with the facial features, thus is benefit for face recognition. Similar with DeepID2 [16], in our objective function optimization model, the identification and verification cost are jointly optimized, which is formulated as follows:

$$J(W_i W_v) = I(W_i) + \lambda V(W_v), \tag{1}$$

where $I(W_i)$ and $V(W_v)$ represent identification and verification cost, respectively; W_i and $W_v$are the weight coefficients used to project the fusion feature; $\lambda \geq 0$ is a balance coefficient used to control the contribution of the verification cost $V(W_v)$ and in the following experiment, it is set as 0.0001. The identification cost $I(W_i)$ function is based on the softmax function, which is formulated as follows:

$$I(W_i) = -\sum\nolimits_{n=1}^{N} \sum_{j=1}^{M} 1\{y_n = j\} \log \frac{\exp(W_{i,j}^T \cdot x_n)}{\sum\nolimits_{m=1}^{M} \exp(W_{i,m}^T \cdot x_n)}, \tag{2}$$

where N and M are the number of samples and the number of subjects; x is an image represented by the fusion feature; y is the subject identification of x; $W_i = [\mathrm{W}_{i,1}, \mathrm{W}_{i,2}, \mathrm{W}_{i,3}, \ldots, \mathrm{W}_{i,M}]$ is the learned coefficient.

The verification cost function is based on a Log-logistic function, which is defined as follows:

$$V(W_v) = \frac{-1}{P \times G} \sum_{p=1}^{P} \sum_{q=1}^{Q} \log(1 + \exp(-l_{p,q} \cdot W_v^T \cdot |x_p - x_q|)), \tag{3}$$

where P and G are the image number of probe set and the image number of gallery set, respectively; x is an image represented by the fusion feature; l_{pq} is indicator function, if x_p and x_q are belong to the same subject, then $l_{pq} = 1$, else $l_{pq} = -1$. Back propagation (BP) [28] is used to learn the parameters of Eq. (1) and there are many public CNN tools, such as matconvnet [29] and caffe [30].

2.4 Experiment and Analysis

We evaluate the proposed face recognition method on CASIA-WebFace [24] and LFW [22] databases to validate the superiority of the proposed method. The CASIA-WebFace database is released by Institute of Automation, Chinese Academy of Sciences, which includes 494,414 images of 10,575 subjects. The average image number of each subject is 47 and each subject has 15 images at least. LFW is the most commonly used face dataset, which includes 13,233 images of 5,749 subjects. We apply cross-dataset evaluation protocol in our experiment, that is, for training, we train the face recognition model on the CASIA-WebFace database and do not use the images of LFW for model fine-tuning; for testing, we use the standard evaluation protocol on the LFW to evaluate our face recognition model.

2.4.1 Training Configuration

We initialize the weights in each layer based on a normal distribution N(0, 0.01), and the biases are initialized to 0. The balance coefficient $\lambda \geq 0$ in Eq. (1) it is set as 0.0001. The size of mini-batch is 128 including 128 images which are randomly selected from the whole dataset. The momentums are set to 0.9. The learning rates start with 0.01 and gradually decreased along the training progress. That is, if the objective function is convergent at a stage, the learning rates are reduced to 1/10 of the current values, and the minimum learning rates are 0.0001. The max iteration number is 100,000.

2.4.2 Comparison with State-of-the-art Method

The comparison of our proposed method and the state of the art is shown in Table 1. As shown in Table 1, our proposed FTFA-CNN method beats the high-dim LBP [31] with 1.85% and obtains a better result than DR [24] under the same training and testing configurations. Since we only train single face recognition model, we report the result of DeepID2 [16] with single face recognition model to ensure the comparison fairness. Compared with DeepID2, our proposed FTFA-CNN method obtains a higher accuracy.

2.4.3 Analysis of the Proposed Method

In order to evaluate the role of the facial features, we make a comparison using (i.e. FTFA-DFL) and not using (i.e. ResNet) the aid of facial texture features. As shown in Table 2,

Table 1 The performance comparison of the proposed method and multiple state-of-the-art methods

Algorithm	Accuracy (%) + std (%)	Evaluation condition
FTFA-DFL	97.02 + 0.26	Cross-dataset
High-dim LBP [31]	95. 17 + 1.13	Cross-dataset
Deep Representation (DR) [24]	96.13 + 0.30	Cross-dataset
DeepID2 [16]	95.43 + N/A	Cross-dataset

Table 2 Comparison between using and not using the aid of facial textures

Algorithm	Accuracy (%)+std (%)	Evaluation condition
FTFA-DFL	97.02 + 0.26	Cross-dataset
ResNet	96.58 + 0.36	Cross-dataset

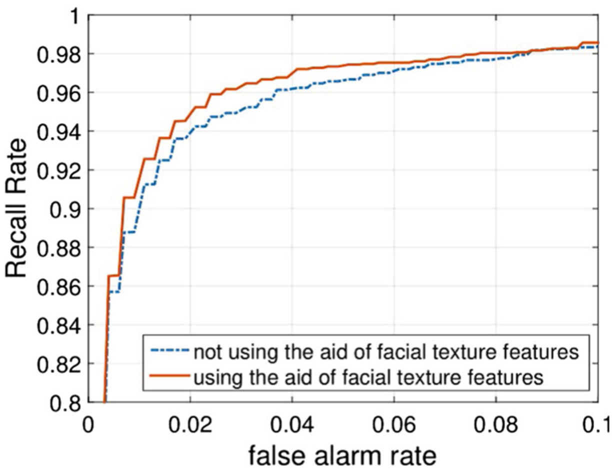

Fig. 6 ROC comparison between using and not using the aid of facial texture features

using the aid of facial features gets a higher accuracy. Moreover, as shown in Fig. 6, the ROC of using the aid of facial features is obvious better than that of not using facial features. These results illuminate the effectiveness of the aiding of facial features.

3 Conclusion

A Facial texture feature aided deep feature learning feature learning method (FTFA-DFL) is proposed face recognition algorithm. The FTFA-DFL is able to combine both advantages of hand-crafted facial texture features and deep learned features. In the proposed FTFA-DFL, the hand-craft features are texture features extracted from the eyes, nose, and mouth regions. Then, the hand-craft features are used to aid deep learning features by adding both deep learning and hand-craft features into the objective function module, which adaptively adjusts the deep learning features so that it can better cooperate with the hand-craft features and obtain a better face recognition performance. The experimental results show that the proposed face recognition algorithm (with one model) on the LFW face database to achieve the accuracy rate of 97.02%.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under the Grants 61502364, in part by the Fundamental Research Funds for the Central Universities under the Grant K50510010007, and in part by Weinan normal university Funds 16YKS001.

References

1. Dix, A. (2009). Human-computer interaction. In *Encyclopedia of database systems* (pp. 1327–1331). New York: Springer.
2. Arigbabu, O. A., et al. (2015). Integration of multiple soft biometrics for human identification. *Pattern Recognition Letters*, *68*, 278–287.
3. Husain, I., et al. (2016). Attendance system based on face recognition by using raspberry Pi. *International Journal of Research*, *3*(3), 22–26.
4. Feng, D., Siu, W.-C., & Zhang, H. J. (2013). *Multimedia information retrieval and management: Technological fundamentals and applications*. Dordrecht: Springer.
5. Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(12), 2037–2041.
6. Zhang, L., et al. (2007). Face detection based on multi-block LBP representation. In *Advances in biometrics* (pp. 11–18). London: Springer.
7. Albiol, A., et al. (2008). Face recognition using HOGCEBGM. *Pattern Recognition Letters*, *29*(10), 1537–1543.
8. Dniz, O., et al. (2011). Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, *32*(12), 1598–1603.
9. Lei, Z., et al. (2008). Gabor volume based local binary pattern for face representation and recognition. In *8th IEEE International Conference on Automatic Face & Gesture Recognition, FG'08*, 2008.
10. Luo, J., et al. (2007). Person-specific SIFT features for face recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.
11. Liao, S., Jain, A. K., & Li, S. Z. (2013). Partial face recognition: Alignment-free approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(5), 1193–1205.
12. Huang, G. B., Lee, H., Learned-Miller, E. (2012). Learning hierarchical representations for face verification with convolutional deep belief networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
13. Sun, Y., Wang, X., Tang, X. (2013). Hybrid deep learning for face verification. In *Proceedings of the IEEE International Conference on Computer Vision*.
14. Deepface, Wolf, L. (2014). Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

15. Sun, Y., Wang, X., Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
16. Sun, Y., et al. (2014). Deep learning face representation by joint identification-verification. In *Advances in Neural Information Processing Systems*.
17. Silapachote, P., Karuppiah, D. R., & Hanson, A. R. (2005). Feature selection using AdaBoost for face expression recognition. Massachusetts Univ Amherst Dept Of Computer Science.
18. Qin, J., He, Z.-S. (2005). A SVM face recognition method based on Gabor-featured key points. In *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, 2005. IEEE.
19. Sharma, A., Jacobs, D.W. (2011). Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. IEEE.
20. Chen, D., Cao, X., Wang, L., Wen, F., & Sun, J. (2012). Bayesian face revisited: A joint formulation. In *European Conference on Computer Vision, 2012, Florence, Italy* (pp. 566–579). Springer.
21. Davis, J. V., et al. (2007). Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning ACM*.
22. Huang, G. B. et al. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments., *Technical Report 07–49*. Amherst: University of Massachusetts.
23. Wolf, L., Hassner, T., Maoz, I. (2011). Face recognition in unconstrained videos with matched background similarity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. IEEE.
24. Zhu, X., Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. IEEE.
25. Yi, D., et al. (2014). Learning face representation from scratch. arXiv preprint arXiv:1411.7923. Accessed 28 Nov 2014.
26. Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167. Accessed 2 Mar 2015.
27. Chen, D., Xudong, C., Fang, W., & Jian, S. (2013). Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *IEEE Conference on Computer vision and pattern recognition (CVPR), 2013* (pp. 3025–3032). IEEE.
28. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 22782324.
29. A. Vedaldi, Lenc, K. (2015). Matconvnet: Convolutional neural networks for matlab, In *Proceedings of the 23rd ACM international conference on Multimedia, ACM*, pp. 689C692.
30. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T. Caffe: Convolutional architecture for fast feature embedding, arXiv preprint arXiv:1408.5093. Accessed 20 Jun 2014.
31. Chen, D., Cao, X., Wen, F., Sun, J. (2013). Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 30253032.
32. K. He, Zhang, X., Ren, S., et al. Deep Residual Learning for Image Recognition, arXiv preprint arXiv:1512.03385. Accessed 10 Dec 2015.
33. Wen, Y., Zhang, K., Li, Z., Yu, Q. (2016). A discriminative feature learning approach for deep face recognition, In *European Conference on Computer Vision*, pp. 499–515.
34. Ding, C., Choi, J., Tan, D., & Davis, L. (2016). Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *38*(3), 518.
35. Ding, H., Zhou, S., Chellappa, R. (2017). FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition, In *IEEE International Conference on Automatic Face & Gesture Recognition IEEE*, pp. 118–126.
36. Wu, Y., Liu, H., Li, J., & Fu, Y. (2017). Deep Face Recognition with Center Invariant Loss. In *ACM Multimedia-The matic Workshops*. CA, USA ACM: Mountain View.
37. Waisy, A., Qahwaji, R., Ipson, S., Fahdawi, S. (2017). A multimodal deep learning framework using local feature representations for face recognition, In Machine Vision & Applications pp. 1–20.

Yunfei Li received his BS degree in Northwest Normal University China, in 1998 and master degree in Xidian University China, in 2007. He is currently pursuing the PhD degree in School of Telecommunications Engineering, Xidian University, Xi'an 710071, China. His research interests include pattern recognition, and image processing.

Zhaoyang Lu is a Ph.D. professor and Ph.D. supervisor with School of Telecommunications Engineering, Xidian University, Xi'an 710071, China. His research interests include machine learning, pattern recognition, image and video coding, intelligent transportation system and natural language analysis.

Jing Li is a Ph.D. associate professor with School of Telecommunications Engineering, Xidian University, Xi'an 710071, China. Her research interests include machine learning, computer vision, pattern recognition, intelligent video surveillance, augmented reality.

Yanzi Deng is currently pursuing the Ph.D. degree in School of Telecommunications Engineering, Xidian University, Xi'an 710071, China. Her research interests include machine learning, computer vision, pattern recognition, intelligent video surveillance, augmented reality.