Computer Methods and Programs in Biomedicine 197 (2020) 105622

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition

Jialin Tang[a,b], Qinglang Su[b], Binghua Su[a], Simon Fong[c,*], Wei Cao[a], Xueyuan Gong[a,*]

[a] Beijing Institute of Technology, Zhuhai 519088, China
[b] City University of Macau, Macau, China
[c] Department of Computer and Information Science, University of Macau, Macau, China

ARTICLE INFO

Article history:
Received 28 April 2020
Accepted 15 June 2020

Keywords:
Convolutional Neural Networks (CNN)
Local Binary Patterns (LBP)
Ensemble learning
Face recognition

ABSTRACT

Background and Objective: Face recognition success rate is influenced by illumination, expression, posture change, and other factors, which is due to the low generalization ability of a single convolutional neural network. A new face recognition method based on parallel ensemble learning of convolutional neural networks (CNN) and local binary patterns (LBP) is proposed to solve this problem. It also helps to improve the low pedestrian detection rate caused by occlusion.

Methods: First, the LBP operator is employed to extract features of the face texture. After that, 10 convolutional neural networks with 5 different network structures are adopted to further extract features for training, to improve the network parameters and get classification result by using the Softmax function after the layer is fully connected. Finally, the method of parallel ensemble learning is used to generate the final result of face recognition using majority voting.

Results: By this method, the recognition rates in the ORL and Yale-B face datasets increase to 100% and 97.51%, respectively. In the experiments, the proposed approach is illustrated not only enhances its tolerance to illumination, expression, and posture but also improves the accuracy of face recognition and the poor generalization performance of the model, which is normally caused by the learning algorithm being trapped in a local minimum. Moreover, the proposed method is combined with a pedestrian detection model as a hybrid model for improving the detection rate, which shows in the result that the detection rate is improved by 11.2%.

Conclusion: In summary, the proposed approach greatly outperforms other competitive methods.

1. Introduction

Since face information processing has many advantages, such as high safety coefficient, convenient data acquisition and easy popularization in biometric feature recognition. Face recognition technology is useful in many finance and security applications, such as video surveillance and intelligent payment. It is one of the most popular research directions in machine learning and computer vision at present.

Face recognition is often used in unconstrained condition. The collected images are affected by several environmental factors such as illumination, expression, posture and so on. Therefore, the traditional feature extraction approach is not an ideal face recognition solution. Since Histogram of Oriented Gradient (HOG) [13] maintains good invariance to optical and geometric deformation of image, and local binary patterns (LBP) [8] has the advantages of grayscale invariance, insensitivity to illumination and so on, they are employed to get facial features for face recognition.

In this era, a large number of face training data sets have emerged, and a huge number of attentions has been attracted to deep learning. The neural network model can process two-dimensional images, learn the characteristics of the images from a large number of samples, and classify the two-dimensional images according to the learned characteristics, so as to realize the face recognition. The feature extraction by machines can avoid excessive intervention of subjective factors. At the same time, convolutional neural network is widely adopted in face recognition applications because of its self-learning ability, parallel processing ability, good fault tolerance and generalization performance.

[*] Corresponding authors.
E-mail addresses: thong03@qq.com (J. Tang), sonnysu@cityu.mo (Q. Su), 01004@bitzh.edu.cn (B. Su), ccfong@um.edu.mo (S. Fong), 17303@bitzh.edu.cn (W. Cao), 18202@bitzh.edu.cn (X. Gong).

https://doi.org/10.1016/j.cmpb.2020.105622

The convolutional neural network can detect the facial features from image directly, but it will learn the noise of the image in the meantime. Both HOG and LBP can process the face images and reduce the noise interference, thus making the features of images more obvious. Since HOG is more focused on extracting the presentational and shape features of the target, while LBP can extract the texture features of the target, therefore, LBP has better effect on extracting facial features. In this paper, LBP is employed to detect facial texture features to reduce the influence of illumination and expression; and CNN and skip connection are used for parallel convolution processing, so as to reduce the training time and improve the accuracy of classification. Meanwhile, by introducing the parallel ensemble learning method and parallel connection of two or more convolutional neural networks with different structures for face recognition, the diversity of the ensemble individuals is improved and the generalization ability of the network is enhanced. One face recognition method based on parallel ensemble learning of LBP and CNN is proposed in this paper, which effectively improves the face recognition accuracy.

In the experiment part, two different experiments are conducted. First, ORL [8] and Yale-B [13] face data sets are adopted to test the accuracy of our proposed CNN model in face recognition problem. In detail, PCA, HOG-CNN, CNN and the proposed method are compared in the experiment, the result of which demonstrates that our method outperforms the others. Second, our method for face recognition is combined with a CNN model for pedestrian detection, we call it hybrid model. This hybrid model improves the accuracy of pedestrian detection caused by occlusion with our proposed face recognition model. Intuitively, human face is a small part of a whole human in an image. Therefore, even half of the human is occluded, the pedestrian is available to be detected as long as his/her face is clear.

2. Literature review

Tahira et al. [1] used PCA to reduce face dimensions so as to realize face recognition. Face image acquisition is influenced by illumination, expression and posture, which results in a large difference of the same individual, and the reduction of the face recognition rate [2]. The face recognition approach based on local binary pattern introduced by Ahonen et al. [3] can divide a face image into several regions for face recognition. The face recognition approach based on multi-direction local binary pattern proposed by Liu et al. [4] acquires the feature vector by replacing a single pixel with the regional mean value, which not only introduces the whole spatial image information, but also reduces the dimensions of the image. By this way, the face recognition rate under complex illuminations is significantly improved. The face recognition method based on a hybrid model of CNN and LBP proposed by Wang et al. [5] can effectively overcome the disadvantage of poor grayscale stability of CNN and reduce the influence of illumination, expression and posture change. The face recognition approach based on a hybrid model of HOG and CNN proposed by Ahamed et al. [6] makes face recognition by inputting CNN with the shape of the target. There are still some problems in the above methods. For example, with the change of illumination intensity and posture, the recognition rate of PCA will be greatly reduced. The HOG is more focused on extracting the presentational and shape features of the target, instead of specific facial feature points. A large number of experiments prove that the increasement of depth and width of the network improves the accuracy. However, with the deepening of the convolutional neural network, problems like the increasing of parameters and the surging of computations will occur. A deeper network can lead to a rapid saturation of accuracy, and after saturation, a higher error rate will occur as the number of training times increases [7].

Therefore, the parallel ensemble learning based on LBP and CNN is proposed in this paper to extract face image features, which reduces the depth of the convolutional neural network, improves the accuracy of classification and reduces the influence of illumination, expression and posture change, etc. The main difference of our approach and the existing approaches are that our approach utilizes LBP to extract facial features firstly. Afterwards, they are fed into ResNet and classified, where ResNet is a cutting-edge CNN model for classifying images.

3. FACE recognition based on parallel ensemble learning of LBP and CNN

This paper mainly studies the face recognition based on parallel ensemble learning of LBP and CNN. In this paper, LBP is first employed to analyze the texture of the input images, then CNN is utilized to get the facial features of the images processed by LBP, finally, the parallel ensemble learning method is utilized to improve the poor generalization performance of the CNN caused by the learning algorithm being trapped in a local minimum, so as to improve the effect of distinguishing different faces. The implementation flow of the proposed approach is given in Fig. 1, and the detailed implementation way is described as follows.

4. Local Binary Pattern (LBP)

The local binary pattern, with its simple principle, low computational complexity, grayscale invariance, and illumination insensitivity, can extract the texture features of images and fuse the overall features of an image.

In this paper, the LBP model improved by Ojala [8] and other researchers can obtain the texture features of images by changing the radius of circle and the number of pixels. Bilinear interpolation is used to obtain the point gray value that is not in the center of the pixel box, which makes the algorithm more robust. Using the improved LBP, the radius and pixels are expressed by R and P, respectively. The LBP algorithm for different radii and pixels is demonstrated in Fig. 2.

The basic idea of LBP is to compare the gray value of every neighboring pixel with that of the center pixel. Taking the radius of 1 and the number of pixels of 8 as an example, taking the center point as the base point, the gray value of the central point is compared with the gray values of 8 pixels in neighborhood. If the gray value of the neighboring pixels is greater than that of the center pixel, the gray values of all neighboring pixels are set to 1; on the contrary, the gray values of all neighboring pixels are set to 0. The LBP-based image feature extraction procedure is shown in Fig. 3.

The formula of LBP can be expressed as:

$$LBP_{P,R}(x,y) = \sum_{n=0}^{P-1} 2^n s(i_n - i_{x,y}) \tag{1}$$

$$S(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

In which, $LBP_{P,R}(x,\ y)$ represents the LBP texture feature with the center pixel $(x,\ y)$, the radius R, and the neighboring pixels P.i_n represents the gray value of the nth neighboring pixel, and $i_{x,y}$ represents the gray value of the center pixel.

The texture features after being extracted by LBP are insensitive to light change, and they are less affected by illumination variations than that of the original image. Facial features of the same individual with various illumination intensities are extracted by LBP and the effect is shown in Fig. 4. In addition, as shown in Fig. 4, not only is LBP able to capture the facial features with only a little affection of illumination, also that LBP preserves adequate details

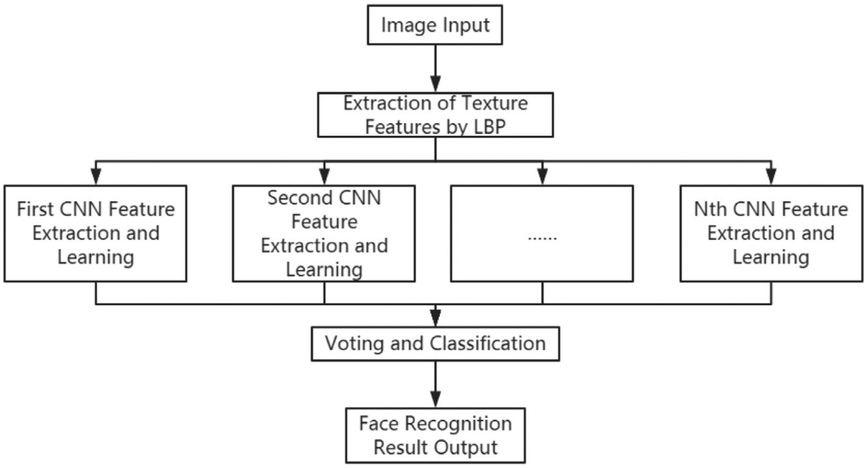


Image Input
Extraction of Texture Features by LBP
First CNN Feature Extraction and Learning
Second CNN Feature Extraction and Learning
......
Nth CNN Feature Extraction and Learning
Voting and Classification
Face Recognition Result Output

Fig. 1. Flow Chart for the Implementation Method Introduced in the Paper.

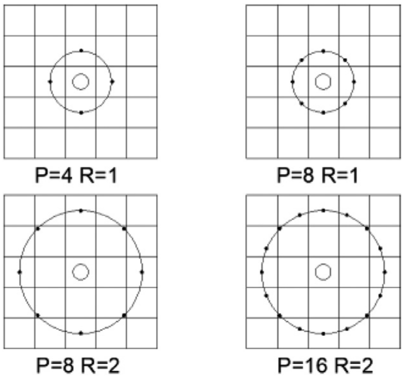


Fig. 2. Schematic Diagram of the LBP Algorithm.

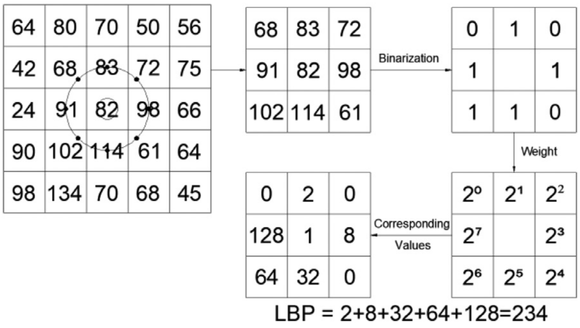


Fig. 3. Calculation Process with the LBP Algorithm.

of facial features (e.g. hair) in order to avoid misclassifying two different people as same ones just because of their same expressions. In other words, CNNs and other classifiers may be misguided by the same expressions between two different people without giving more detailed facial features.

In order to preserve more facial features for the CNNs, in this paper, the texture feature images extracted by LBP are directly used as inputs to the CNNs for learning to improve the accuracy.

5. Convolutional Neural Network (CNN)

In the paper, after the texture extraction of LBP, the facial features are further extracted by CNN. In order to increase the network width, accelerate the training time and improve the accuracy of classification, this paper adopts Inception module [9] and the famous Batch Normalization [10] for skip connection [7] via CNN. The network structure in Fig. 5 is one of the CNN models used in this paper, which consists of four convolutional layers, five maximum pooling layers and one stacking model of Inception modules.

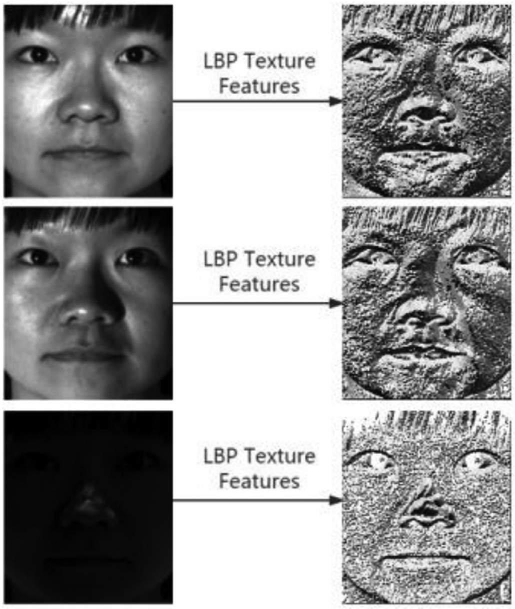


Fig. 4. Texture Features Extracted from Face Images.

5.1. Convolutional layer

The convolutional layer is composed of one or more convolutional feature maps, and each convolution surface is calculated from the input, the convolution kernel and the activation function. The input of convolutional layer is assumed to be a $M \times M$ matrix x, the size of convolution kernel is $K \times K$ matrix w, the bias is b, and the convolutional feature map is matrix h. Then, the convolution formula is:

$$\boldsymbol{h} = f(conv(\boldsymbol{x}, \boldsymbol{w}) + b) \tag{3}$$

In which, *f*(•) is an activation function, *conv*(•) is a convolution.

The literature [10] shows that the receptive field of a series of small convolution kernels is consistent with that of a large convolution kernel, the network parameters can be reduced by de-

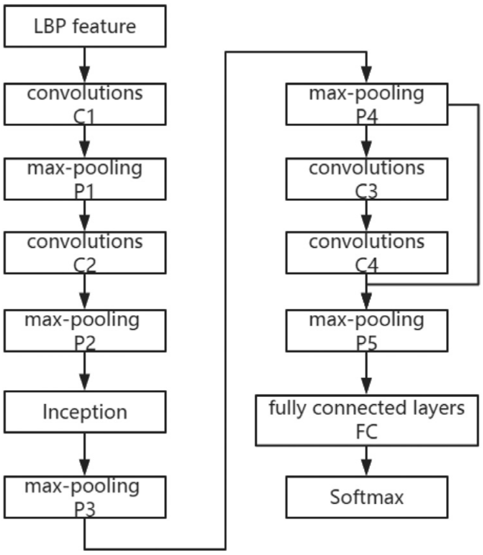


Fig. 5. CNN Structure Diagram.

composing the convolution kernel [11], and the deeper network is able to improve the accuracy of the model. So, reducing the convolution kernel size and decomposing the convolution kernel can make the network structure deeper, the computation amount reduced and the accuracy improved. Therefore, the maximum size of the convolutional kernel employed in the paper is 5 × 5, and 3 × 1 and 1 × 3 convolution kernels are used to replace the 3 × 3 convolution kernel in order to reduce the parameters and calculation amount, and improve the accuracy.

5.2. Pooling layer

Pooling layer is also called subsampled layer. The input of the pooling layer can be either the output of a convolutional layer or that of a pooling layer. The pooling layer can capture features and reduce the size of the input image, and it is normally put after the convolutional layer. The pooling layer includes several types, such as Max Pooling and Average Pooling. These two types of pooling layers are mathematically expressed as follows:

$$\boldsymbol{F} = \boldsymbol{D}_{max}^{l,l}(\boldsymbol{P}) \tag{4}$$

$$\boldsymbol{F} = \boldsymbol{D}_{avg}^{l,l}(\boldsymbol{P}) \tag{5}$$

In which, $\boldsymbol{P}$ is the input vector, $\boldsymbol{D}_{max}^{l,l}$ is the maximum pooling of the $l \times l$ block size, $\boldsymbol{D}_{avg}^{l,l}$ is the average pooling of the $l \times l$ block size, and $\boldsymbol{F}$ is the output pooling face.

5.3. Fully-connected layer

In the convolutional neural network, a fully-connected layer is often added after all convolutional and pooling layers in order to integrate the category discrimination information in multiple convolutional and pooling layers. The fully-connected layer is mathematically expressed as follows:

$$\boldsymbol{J} = f(\boldsymbol{r} \times \boldsymbol{a} + u) \tag{6}$$

In which, $\boldsymbol{r}$ is the input vector, $\boldsymbol{a}$ and u are the weight and bias of the fully-connected layer respectively, $\boldsymbol{J}$ is the output of the fully-connected layer, and $f(\bullet)$ is the activation function.

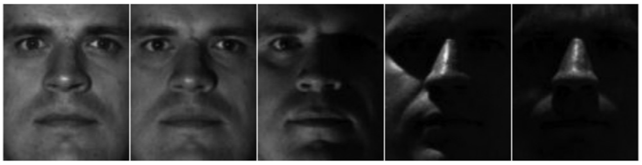


Fig. 6. Yale-B Dataset Samples.

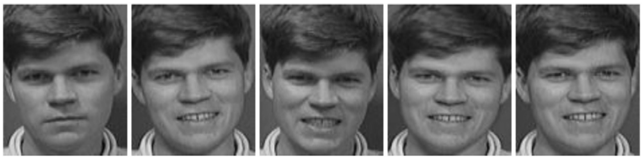


Fig. 7. ORL Dataset Samples.

6. Classification and parallel ensemble learning

In this paper, several convolutional neural networks based on different network structures of LBP are used to learn and classify the training datasets. After training, the final results are outputted by vote to construct a face classifier. When calculating the accuracy of the test dataset, each convolutional neural network outputs a classification result for each face image in the test dataset. The final classification result is obtained by majority vote, which is mathematically expressed as follows:

$$o = G_{max}(o_1, o_2, o_3, \cdots, o_n) \tag{7}$$

In which, o is the final output result,o_1,o_2,o_3,$\cdots$, o_n are the output results of n convolutional neural networks, G_{max} is a function of the output result with the most votes.

7. Experimental results and analysis

7.1. Experimental dataset and environment

For the purpose of verifying the effectiveness of the proposed approach, two public face image datasets, Yale-B and ORL, are selected in this paper. The Yale-B dataset consists of 38 individuals with 576 face images per person in 9 postures and 64 illumination conditions, while the ORL dataset consists of 40 individuals with 10 face images per person. In addition, CBCL [12] and INRIA [13] pedestrian datasets are selected to test the effectiveness of our proposed method to improve the detection rate of pedestrian.

The face images of these two datasets contain the changes in illumination, posture and expression. In comparison, the illumination variation of Yale-B is larger than that of ORL, while posture and expression variations of ORL are larger than that of Yale-B. In this experiment, the two datasets are separated into two datasets, respectively, that is, training set and test set. For each face category in the Yale-B dataset, the first 32 images are selected and put into the training subset and the last 32 images into the test subset. For each face category in the ORL dataset, the first 5 images are selected and put into the training subset and the last 5 images into the test subset. Some face data samples are shown in Figs. 6 and 7.

The main performance indicators of computer hardware and software environment are as follows: CPU: Intel(R) Core(TM) i5-6300HQ 2.3GHZ; RAM: 8G DDR4; Video Card: NVIDIA GeForce GTX 960M; Operating System: Windows 10 64bits; Development Environment: Python 3.5; Tensorflow-gpu 1.10.0.

In this paper, recognition rate and average elapsed time are used as the evaluation indexes. The recognition rate is defined as the ratio of the number of face images correctly recognized to the total number of face images in the test set; the average elapsed

Table 1
Convolutional Neural Network Structure.

Number of Layers	Structure A		Structure B		Structure C		Structure D		Structure E	
	Type	Kernel Size	Type	Kernel Size	Type	Kernel Size	Type	Kernel Size	Type	Kernel Size
1	C1	3 × 3	C1	3 × 3	C1	3 × 3	C1	3 × 3	C1	3 × 3
2	P1	2 × 2	P1	2 × 2	P1	2 × 2	P1	2 × 2	P1	2 × 2
3	C2	3 × 3	C2	3 × 3	C2	3 × 3	C2	3 × 3	C2	3 × 3
4	P2	2 × 2	P2	2 × 2	P2	2 × 2	P2	2 × 2	P2	2 × 2
5	I_C1	1 × 1	I_C1	1 × 1	I_C1	1 × 1	I_C1	1 × 1	I_C1	1 × 1
	I_C2	3 × 3	I_C2	3 × 3	I_C2	3 × 3	I_C2	3 × 3	I_C2	3 × 3
	I_C3	5 × 5	I_C3	5 × 5	I_C3	5 × 5	I_C3	5 × 5	I_C3	3 × 3
	I_P1	3 × 3	I_P1	3 × 3	I_P1	3 × 3	I_P1	3 × 3	I_P1	3 × 3
6	P3	2 × 2	P3	2 × 2	P3	2 × 2	P3	2 × 2	P3	2 × 2
7	C3	1 × 3	C3	1 × 3	C3	1 × 3	P4	2 × 2	P4	2 × 2
8	C4	3 × 1	C4	3 × 1	C4	3 × 1	C3	1 × 3	C3	1 × 3
9	P4	2 × 2	P4	2 × 2	P4	2 × 2	C4	3 × 1	C4	3 × 1
10	FC	–	FC	–	C5	1 × 3	P5	2 × 2	P5	2 × 2
11	S	–	S	–	C6	3 × 1	FC	–	FC	–
12					P5	2 × 2	S	–	S	–
13					FC	–				
14					S	–				

time is defined as the average time to complete the recognition of one face image.

7.2. Experimental results of face recognition

According to the above experimental dataset and the method proposed in this paper, the experiment integrates 10 convolutional neural networks in parallel with 5 different structures, in which CNN1, CNN4, and CNN5 have the same structure (hereinafter referred to as Structure A), CNN2 and CNN3 have the same structure (hereinafter referred to as Structure B), CNN6, CNN7, and CNN8 have the same structure (hereinafter referred to as Structure C), the other two structures are for CNN9 (hereinafter referred to as Structure D) and CNN10 (hereinafter referred to as Structural E) respectively. Note that the combination of different CNN structures is employed based on our testing results that compared the performances of various combinations of CNN structures. For each combination, we had executed around 30 times and recorded the averaged accuracy. After several combinations are tested, the one with the best accuracy is chosen. The Inception in this method is made up of three convolutional layers and one maximum pooling layer in parallel. The detailed network structure is given in Table 1. C denotes the convolutional layer with a step size of 1; P denotes the maximum pooling layer with a step size of 2; I_C denotes the convolutional layer in the Inception with a step size of 1; I_P represents the pooling layer in the Inception with a step size of 1; FC denotes the fully connected layer, and S denotes the Softmax logic regression function. Among them, Structure B, Structure D and Structure E all contain a skip connection, as shown in Fig. 8. Structure B jumps from the output of Max Pooling P3 to the input of Max Pooling P4, while Structure D and Structure E jump from the output of Max Pooling P4 to the input of Max Pooling P5.

The recognition results of the iterations and accuracy of each network in the ORL dataset are demonstrated in Fig. 9. Because of the difference of CNN structure, the features learned are different, and the accuracy of verification is also deviated. The data with better learning effect in each structure are compared in Fig. 10. Obviously, the accuracy of the integrated data is about 98%. After 2800 trainings, the accuracy of the test set is 100%, which is much higher than that of single convolutional neural network. Despite the obvious change of expression and posture in the ORL face dataset, the proposed method can still achieve high accuracy, which proves that it is highly robust to expression and posture change.

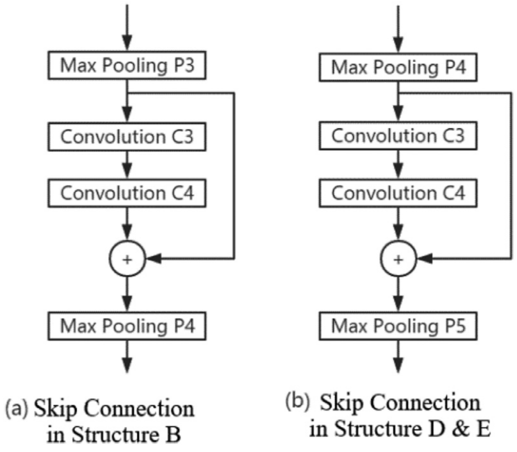


Fig. 8. Schematic Diagram of Network Skip Connection.

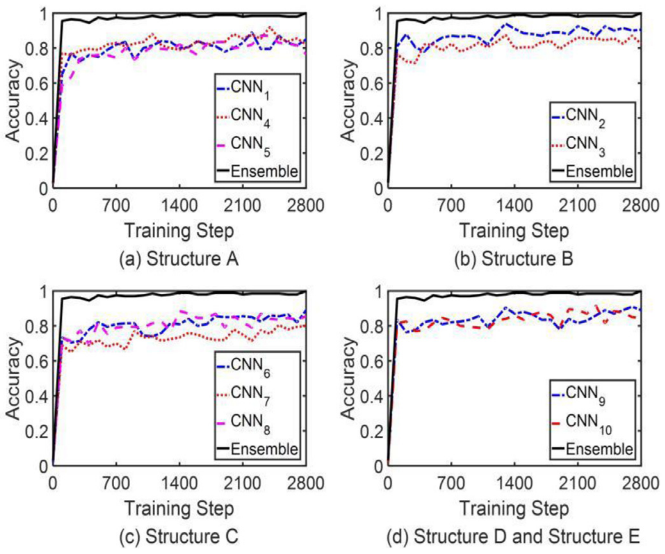
The recognition results of Yale-B dataset are demonstrated in Fig. 11. The accuracy of single network is about 85%, and that of the integrated network is up to 97.51%, which is much higher than that of single convolutional neural network.

In summary, the parallel ensemble learning is beneficial to strengthen the robustness of the algorithm and improve the final recognition rate.

7.3. Experimental results of pedestrian detection

CBCL [12] and INRIA [13] pedestrian datasets are selected to test the effectiveness of our proposed method to improve the detection rate of pedestrian in this subsection. The CBCL data set contains 924 images, among which the size of each image is 64 by 128. Besides, the INRIA data set consists of 2416 images, where include 614 positives and 1218 negatives. Comparing to images in the CBCL data set, images in the INRIA are of higher resolution.

First, we trained a CNN model to recognize pedestrian straightforwardly. After that, our proposed method is combined with the CNN model as a hybrid model, which aims at improving the pedestrian detection rate. As shown in Table 2, our hybrid model improves the detection rate of CNN from 96.3% to 99.5% on CBCL data set, which claims the effectiveness of the Hybrid approach. In addition, as the INRIA data set contains images with higher resolution, the hybrid approach prefers such data set since high resolution carries more facial details to help the hybrid approach to



(a) Structure A

(b) Structure B

(c) Structure C

(d) Structure D and Structure E

Fig. 9. Recognition Results of ORL Face Image Dataset.

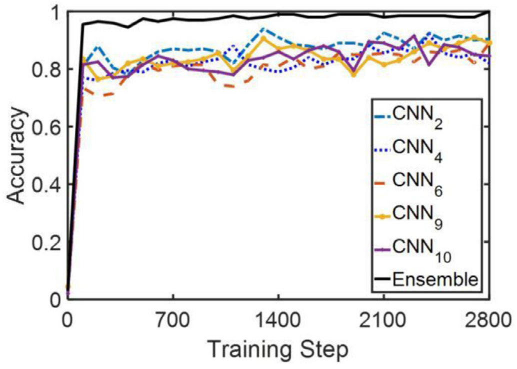


Fig. 10. Network Structure vs. Integrated Recognition Results.

Table 2
Experimental Results of Pedestrian Detection.

Method	Accuracy of CBCL/%	Accuracy of INRIA/%
CNN	96.3	89.2
Hybrid	99.5	98.1

detect pedestrians, that leads to a 98.1% accuracy for the hybrid approach on INRIA data set comparing to a 89.2% accuracy for the CNN model.

7.4. Analysis and comparison of the results with other algorithms

In order to verify the performance, the proposed approach is compared with PCA, CNN, HOG-CNN and LBP-CNN. PCA maps high-dimensional face image to low-dimensional subspace, which makes it easier to classify the face images. PCA is a popular face recognition method at present. In this paper, the PCA-based face recognition is used as Contrast Scheme 1, hereinafter referred to as Scheme 1. The face image recognition directly based on the CNN is a common solution in face recognition. According to the experimental results in previous sections, if the convolutional neural network does not use the ensemble learning scheme for classification, the accuracy will be greatly reduced. Therefore, the CNN-based parallel ensemble learning is used as Contrast Scheme 2, hereinafter referred to as Scheme 2. The convolutional neural network structure of Scheme 2 is consistent with that of Table 1, and the difference between Scheme 2 and the proposed approach in the paper is that the texture features of the image input to CNN in Scheme 2 are not extracted by LBP. The convolutional neural networks based on HOG features can also be employed for face recognition. In the paper, the parallel ensemble learning of convolutional neural network based on HOG features is used as Contrast Scheme 3, hereinafter referred to as Scheme 3. The convolutional neural network structure of Scheme 3 is consistent with that of Table 1, and the difference between Scheme 3 and the proposed method in this paper is that Scheme 3 uses HOG to replace the LBP proposed in this paper to extract the features. Note the 4th algorithm LBP-CNN is similar to our proposed algorithm, yet the backbone of CNNs in our algorithm is chosen from ResNet, which is able to go deeper due to its skip connection comparing to the basic backbone in LBP-CNN.

It can be seen from Table 3 that, in the ORL face image dataset identification, the recognition rate of the proposed approach is significantly higher than that of PCA (principal component analysis), HOG-CNN and LBP-CNN. At the same time, the recognition rate of the proposed method is a little higher than that of the traditional convolutional neural network.

It can be seen in Table 4 that, in the identification of Yale-B dataset, the accuracy of the proposed approach is 10% higher than that of Scheme 2 ranked second, is over 50% higher than

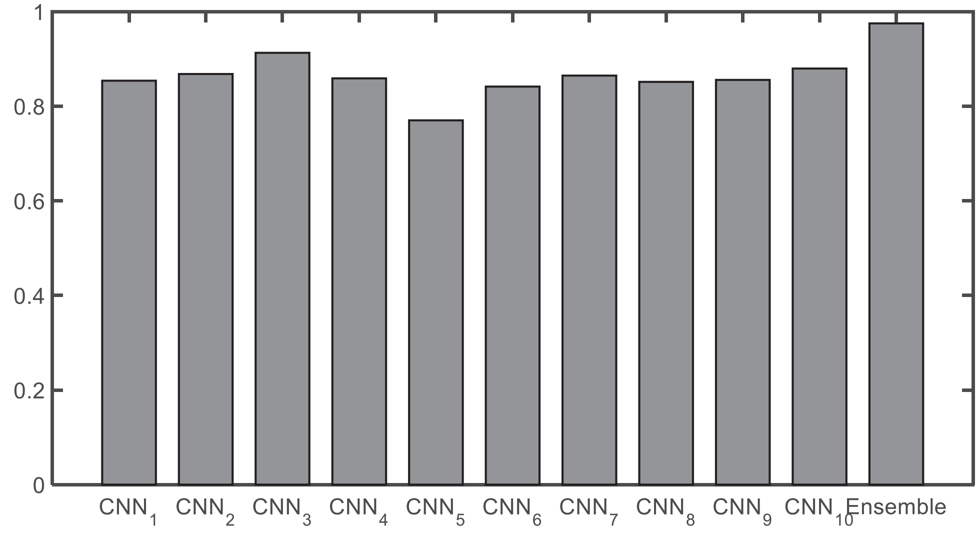


Fig. 11. Recognition Results of Yale-B Face Database.

Table 3
Recognition Results of Each Method in ORL Face Image Dataset.

Method	Recognition Rate /%
Scheme 1	92.00
Scheme 2	99.50
Scheme 3	93.50
Scheme 4	96.6
Proposed Method	100.00

Table 4
Recognition Results of Each Method in Yale-B Face Image Dataset.

Method	Recognition Rate /%
Scheme 1	38.23
Scheme 2	87.72
Scheme 3	62.74
Scheme 4	96.7
Proposed Method	97.51

Table 5
The Average Elapsed Time of the approach Proposed in the Paper.

Database	Average Elapsed Time /MS
ORL	14
Yale-B	34

that of PCA, is over 30% higher than that of HOG-CNN and is a slightly higher than LBP-CNN. Because the illumination variation of Yale-B face image dataset is relatively large, the recognition rate will be reduced when the original image features are used. However, the proposed approach in the paper adopts LBP to extract texture features, which greatly reduces the influence of illumination changes. The robustness to the change of illumination of the proposed method in this paper is stronger than that of the three methods compared, and the extraction of face detail features and recognition rate of the proposed method are much better than that of Scheme 3.

In summary, the proposed method in this paper is superior to the three contrast schemes in both datasets. Compared with Schemes 1, 2 and 3, the proposed method in this paper has stronger adaptability and higher accuracy in face recognition situations with large changes in illumination and expression.

In the identification of ORL and Yale-B face image datasets, the average elapsed time of the proposed method in this paper is shown in Table 5, and the recognition time is all less than 40 ms, which is in accordance with the real-time processing requirements.

8. Conclusion

The face recognition approach based on the parallel ensemble learning of LBP and CNN introduced in the paper adopts LBP to extract the texture features as training data for the parallel CNN and finally is applicable to face recognition. LBP features are mainly utilized to extract facial texture features, because LBP can reduce the influence of illumination on facial features and improve the face recognition accuracy. In CNN, the Inception module is utilized to increase the width of the CNN, the Batch Normalization is employed to reduce the training time, and the skip connection is employed to improve the accuracy of face recognition. The parallel ensemble learning makes the network structure no longer single, and greatly improves the accuracy and generalization ability of the proposed approach in the paper. In the experiments, we compared the proposed approach in this paper with other three methods, which are PCA, HOG-CNN and CNN respectively. The final results illustrate that the proposed approach is more effective in face recognition, and its accuracy of face recognition is promising.

Declaration of Competing Interest

None.

Acknowledgements

This work is supported by the Research and Development Fund Beijing Institute of Technology, Zhuhai (No. XK-2018-07), a grant from Innovative Youth Foundation of Guangdong Educational Committee (No. 2018KQNCX346), and Guangdong government grant, (No. EF003/FST-FSJ/2019/GSTIC).

References

[1] Z. Tahira, H.M. Asif, Effect of averaging techniques on PCA algorithm and its performance evaluation in face recognition applications, in: International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), 2018, pp. 1–6.
[2] D.N. Parmar, B.B. Mehta, Face recognition methods & applications, CoRR abs/1403.0485 (2014).
[3] T. Ahonen, A. Hadid, M. Pietikainen, Face Description with Local Binary Patterns: application to Face Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[4] J. Liu, Y. Chen, S. Sun, Face recognition based on multi-direction local binary pattern, in: 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 1606–1610.
[5] M. Wang, Z. Wang, J. Li, Deep convolutional neural network applies to face recognition in small and medium databases, in: 4th International Conference on Systems and Informatics (ICSAI), 2017, pp. 1368–1372.
[6] H. Ahamed, I. Alam, M.M. Islam, HOG-CNN based real time face recognition, in: International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), 2018, pp. 1–4.
[7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[8] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich., Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[10] S. Ioffe, C. Szegedy, Batch normalization accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning (ICML), 2015, pp. 448–456.
[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826.
[12] CBCL pedestrian database, website: http://cbcl.mit.edu/software-datasets/PedestrianData.html. Last accessed: 10/27/2019.
[13] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.