

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.Doi Number

Facial Beauty Prediction via Local Feature Fusion and Broad Learning System

YIKUI ZHAI¹, IEEE MEMBER, CUILIN YU¹, CHUANBO QIN¹, WENLUE ZHOU¹, QIRUI KE¹, JUNYING GAN¹, RUGGERO DONIDA LABATI², IEEE MEMBER, VINCENZO PIURI², IEEE FELLOW, FABIO SCOTTI², IEEE SENIOR MEMBER

¹Department of Intelligent Manufacturing, Wuyi University, Jiangmen, China, 529020

²Dipartimento di Information, Università, Degli Studi di Milano, via Celoria 18, 20133 Milano (MI), Italy

Corresponding author: CHUANBO QIN (e-mail: tenround@163.com).

This work is supported by National Natural Science Foundation (No.61771347); The Open Project of Guizhou Province Public Big Data Key Laboratory (No.2019BDKFJJ015); Basic Research and Applied Basic Research Key Project in General Colleges and Universities of Guangdong Province (No.2018KZDXM073, 2019KZDZX1017, 2020ZDZX3031); Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515010716).

ABSTRACT Facial beauty prediction (FBP), as a frontier topic in the domain of artificial intelligence regarding anthropology, has witnessed some good results as deep learning technology progressively develops. However, it is still limited by the complexity of the deep structure network in need of a large number of parameters and high dimensions, easily leading to a great consumption of time. To solve this problem, this paper proposes a fast training FBP method based on local feature fusion and broad learning system (BLS). Firstly, two-dimensional principal component analysis (2DPCA) is employed to reduce the dimension of the local texture image so as to lessen its redundancy. Secondly, local feature fusion method is adopted to extract more advanced features through avoiding the effects from unstable illumination, individual differences, and various postures. Finally, extensional feature eigenvectors are input to the broad learning network to train an efficient FBP model, which effectively shortens operational time and improve its preciseness. Extensive experiments with the proposed method on large scale Asian female beauty database (LSAFBD) can be conducted within 13.33s while sustaining an accuracy of 58.97%, impressively outstripping other state-of-the-art methods in training speed.

INDEX TERMS Facial beauty prediction (FBP); local feature fusion; broad learning system (BLS);

I. INTRODUCTION

Facial beauty prediction (FBP) [1] is a significant scientific research direction in machine vision and artificial intelligence, favored by many domestic and foreign researcher. Exploring how to interpret, quantify, and predict beauty better will offer people a more scientific perspective on beauty, thus promoting the development of the cross-research direction of FBP. There will be different answers to the evaluation of human appearance based on different aesthetics. Therefore, assisted with artificial intelligence in anthropology, psychology, and public aesthetics to predict human facial beauty, a relatively scientific and objective judgment standard can be obtained. In the past decade, a series of research outcomes have been achieved in the field [2-4]. The theoretical results of FBP research have been widely adopted in practice, such as social applications, cosmetic surgery, and beauty product shopping guides. However, FBP still suffers from issues including unclear

evaluation indicators, massive time consumption of deep networks training, and lack of high-performance equipment.

Deep learning has fully displayed its strengths in the field of FBP in the past decade [5-7]. Compared with conventional machine learning approaches, deep learning networks can automatically extract higher-level features from facial data [8]. Although deep learning has realized excellent prediction results in FBP, its imperfections [9] can never be underestimated: (1) time-consuming and costly training process due to a great deal of parameters; (2) poor model generalization [10] arisen from deep learning's greater sensitivity to image texture features than to image shape features; (3) huge calculation needed resulting from the complex structure of deep learning network. It requires highly configured equipment to complete model training, further increasing the economic burden.

To overcome the issues of deep learning above, this paper proposes a fast training method based on local feature

fusion and broad learning system (BLS) [11]. Above all, different local texture feature methods are employed to extract features of face images. Because of the need of feature fusion and it will lead to significantly grow of data dimension, a two-dimensional principal component analysis (2DPCA) [12] is designated to reduce each kinds of texture images dimension. Compared with other dimensional reduction methods, 2DPCA excels in extracting more advanced facial features, while improving the computer's operating efficiency and model recognition accuracy.

Texture invariance is acquired by fusing different kinds of local texture features to constrain the bias of the texture. Geirhos R et al. [10] proved that the neural network is strongly biased in recognizing the texture of the images rather than the shape of the images. Hence, enhancing the ability of the neural network to learn the texture of the images can strengthen the robustness of the network. Simultaneously, local feature fusion can make up for the disadvantage of single local features to improve the robustness of image features.

Broad learning system (BLS) is adopted as a rapid and simple classifier algorithm for FBP task. It is based on the random vector functional-link neural network (RVFLNN) [13-15] structure. Compared with the "depth" structure, the "width" structure is simpler because there is no layer-to-layer coupling. When the network's accuracy cannot meet the requirements, it can be improved by broadening the "width" of the network. Similarly, because of no multi-layer connection, the broad learning network can rapidly be remodeled in an incremental way without complete retraining process. Besides, compared with the time-consuming training demanding high-performance equipment in deep learning, BLS can be established in rapid way with a normal PC. So, in the case of limited equipment resources, BLS is still suitable for FBP tasks.

The primary contributions in this paper can be summarized as follows:

- This is the first time for broad learning to be incorporated in overcoming the FBP training problems, which has ensured a promising performance for model learning speed and accuracy.
- A texture-based local feature fusion method combined with the 2DPCA is proposed to improve the model's semantic representation and transform the bias from texture to shape.
- An efficient and simple FBP network is introduced in this paper. Experimental results prove that this network can be easily constructed in an expeditious way even with general computer.

The layout of this paper is arranged as follows. Section 2 provides an overview of the related works of this paper while Section 3 elaborates the method and algorithm process. Experimental results are explained in Section 4, followed by discussions and conclusions of this paper in Section 5.

II. RELATED WORKS

Since the proposed method is based on local feature fusion and broad learning to solve the problem of FBP, this section will introduce the related works in FBP as well as in local feature fusion and broad learning.

A. Facial Beauty Prediction

Conventional FBP largely focuses on geometric features, enabling it to collect key points on the face. It operates with the calculation of the Euclidean distance of the key point and the ratio vector between them and then inputs the obtained distance data into the classifier for learning. In 2006, Eisenthal et al. [16] utilized support vector machine and the K nearest neighbor algorithm to predict the facial beauty on 100 frontal face images, while based on which Fan et al. [17] adopted higher-dimensional geometric features for traditional supervised classification. Compared with geometric features, texture features have more robust characterization capabilities and fewer restrictions on face poses. Geometric features are only suitable for face beauty research under general conditions. Hence, Yan et al. [18] proposed a cost-sensitive sequence regression (CSSR) FBP method, through extracting and testing the original pixels and texture features, successfully counter the problem of unbalanced data set categories, reaching a classification accuracy of 52.12%. Furthermore, an extensive database named large scale Asian female beauty database (LSAFBD) was established by Zhai et al. [19] to settle the problem of insufficient facial data via applying apparent multiscale features to FBP. The above experiments reveal that if the texture feature uses the part of the whole face as the feature extraction object, it can substantially reduce manual intervention and improve model calculation efficiency.

In recent years, some new methods for studying facial beauty [20] have sprung up in neural networks and deep learning. Gan et al. [21] employed the deep convolutional network model for feature learning with a correlation of 0.739 while Jie et al. [22] established a psychologically stimulated convolutional neural network model to predict facial beauty, achieving a correlation of 0.87. Gan et al. [23] introduced a research on FBP based on deep convolutional features of dual activation layers which vastly outperformed the previous methods both in classification and regression prediction. It possessed better real-time performance and accuracy in mainstream CNN models. After that, Gan et al. [24] adopted a network named Multi-input Multi-task Beauty Network (2M BeautyNet) and use transfer learning to predict facial beauty, which replace the SoftMax classifier with a random forest achieving better results than before.

At present, although methods based on deep learning can theoretically extract the in-depth features of face images, limited results in this regard gained by scholars have told a different story. For one thing, the main reason belongs to not enough exploration about the prediction model training

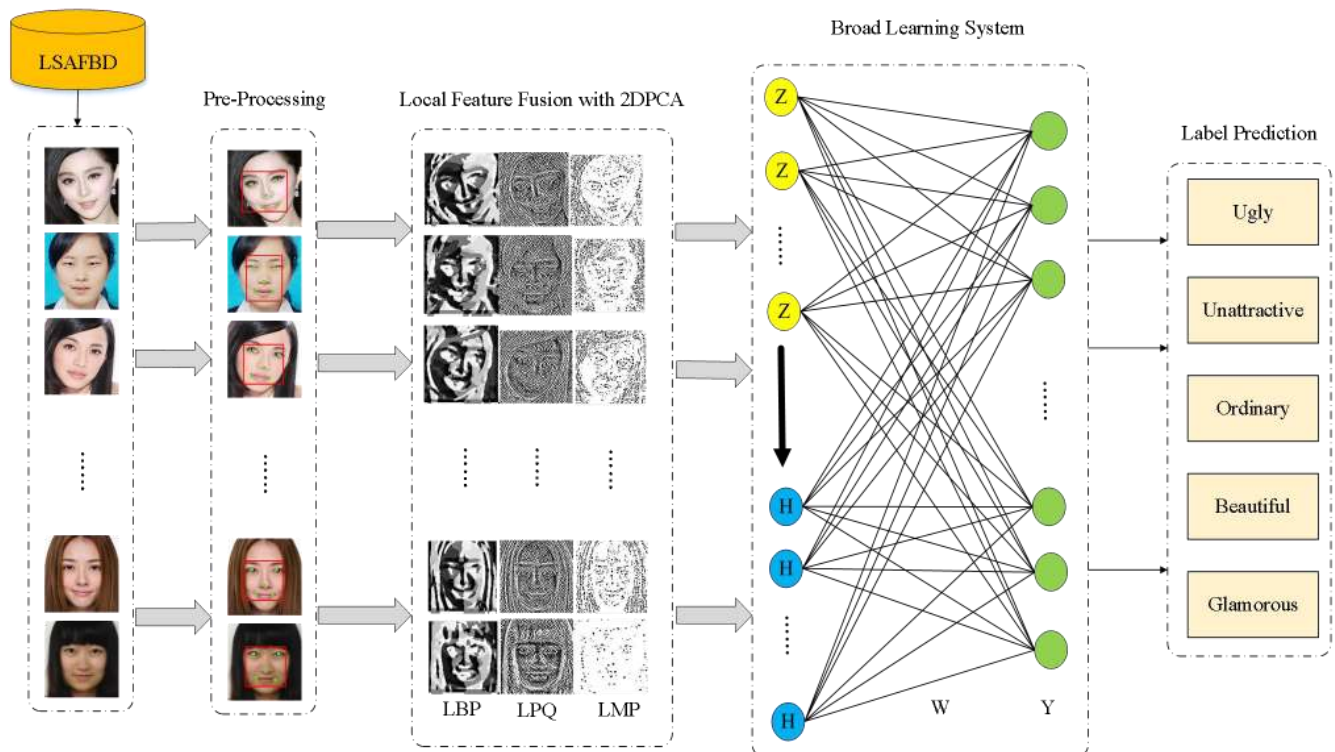


FIGURE 1. Facial beauty prediction network via local feature fusion and broad learning system.

for the optimal hierarchical structure. For another, deep learning with numerous parameter designs involved can lower operating efficiency. Therefore, reducing the network model's scale and improving training efficiency are the core of the research with important scientific significance and research value.

B. Local Feature Fusion

Local feature fusion [25] is a classic computer vision concept, which has been widely used in biometrics identification [26]. The traditional method is to use single local feature to train neural network to realize biometric recognition. The difference between the traditional method and local feature fusion lies in the existence of complementary advantages between different features in the latter one, which can make up for the shortcomings of a single feature and enhance the robustness of image features. Yu et al. [27] realized the recognition of human ears by extracting the texture feature and edge feature of the image separately by histogram of oriented gradient (HOG) for local feature fusion. Min et al. [28] verified the effectiveness of the multi-feature fusion algorithm in enhancing the model performance in the details of the VR panoramic image more than the single-feature fusion algorithm. Yu et al. [29] suggested an emotion recognition model algorithm focusing on long short-term memory (LSTM) with the fusion of different local facial features, which acquired a 24.38% higher accuracy compared with the typical method.

Since both texture information and semantic information play crucial roles in FBP tasks, the proposed

method in this paper using local feature fusion. Experimental results verify the effectiveness of the proposed method in favorably fulfilling FBP tasks.

C. Broad Learning System

BLS is an expeditious model without the need of deep network. This system first transfers the original images and places them in the feature node as a mapping feature, followed by the extensive expansion of the structure in the enhanced node. If the network deems to be expanded, incremental learning algorithms are developed for fast remodeling in broad expansion without a retraining process. It can improve the training speed while ensuring the accuracy of the model.

Many scholars have verified the effectiveness of broad learning in the field of image recognition through experiments. For example, Zhang et al. [30] provided evidence about how face recognition with the help of BLS could remain unaffected by the number of facial features with intense illumination and occlusion and meanwhile maintaining a prominent accuracy. Chen et al. [31] unraveled the superiority of BLS and its variants to several existing learning algorithms in time series prediction and performance regression on face recognition database. Compared with other classic structures, the efficiency and effectiveness of the BLS variants have been fully testified. What is more, Xu et al. [32] raised a new adaptive neural control framework based on broad learning, giving birth to a better human neuromotor system than conventional adaptive neural control did.

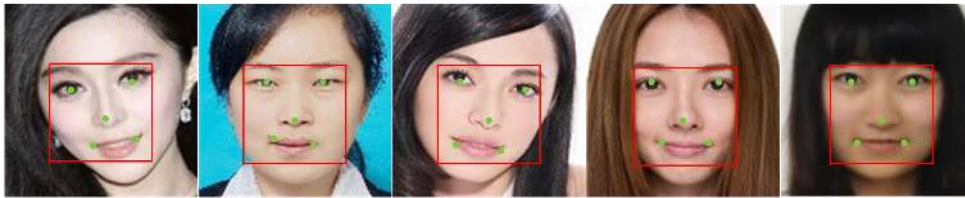


FIGURE 2. Face detection and key point detection example.

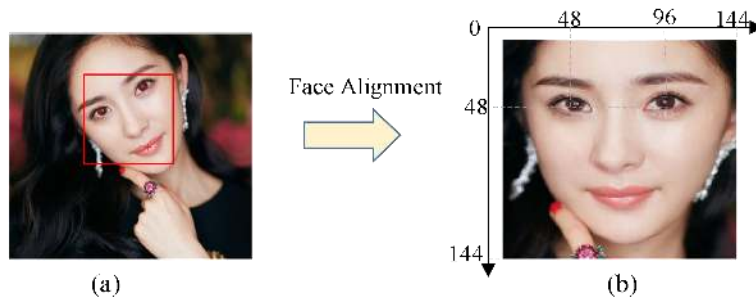


FIGURE 3. Face horizontal alignment example. (a) Detection image; (b) Alignment image.

This is the initial attempt for the BLS to serve as an FBP classifier in FBP tasks. Through substantial experiments, it is certified that BLS performs outstandingly in optimizing FBP training efficiency and the final prediction accuracy.

III. PROPOSED METHOD

In this paper, we proposed an effective and simple FBP system with the participation of local feature fusion and BLS. To begin with, pre-process the images in the LSAFBD, followed by the utilization of different local texture feature including LBP [33], [34] LPQ [35], [36], and LMP [37] methods to extract facial features. Secondly, 2DPCA [12] is adopted to achieve dimensional reduction of texture images. After that, the texture images are processed with local feature fusion method. At last, input the fused feature eigenvectors into BLS to predict facial beauty. Fig. 1 displays the whole processing of the proposed method in this paper.

A. Facial Image Pre-processing

Facial detection and key point detection are both performed on the face images. The key point method of the face is estimated by the three-level convolutional neural network, while the key point prediction equation of the face is based on the multi-level regression:

$$\sum X = \frac{x_1^{(1)} + \dots + x_i^{(l)}}{l_1} + \sum_{k=2}^n \frac{\Delta x_1^{(k)} + \dots + \Delta x_{l_k}^{(k)}}{l_k} \quad (1)$$

where for n -level cascade, x represents the input value of the key points of human face. There is a prediction at level k , among which the first level prediction refers to the absolute value prediction bit and the next level prediction serves as an adjustment to realize the detection of key points of the face. A red frame indicates the detected face area, and green dots locate the five key points of the detected eyes, nose, and mouth corners, as

shown in Fig. 2.

Owing to the detrimental factors such as deflection and tilt in the face images of the original database, it is of urgent necessity to perform operations such as alignment on the detected face. Firstly, use the key points at the center of the left and right eyes to calculate the connecting line and horizontal line angle so that the images can be aligned horizontally in case of posture deflection. Then, fix the pixels of the center point of the eyes and the center point of the mouth to 48 pixels with the normalization of scale following. Finally, image with a size of 144×144 can be obtained as the input of the network shown in Fig. 3.

B Local Feature Fusion

1) Local Descriptors

Compared with global feature fusion, local feature fusion contains richer features and lower correlation between features. In the case of occlusion, the detection and matching of other features will not be affected by the disappearance of some features. In recent years, local features [38] have been widely used in face recognition, three-dimensional reconstruction, target recognition, and panoramic image stitching. In this paper, we incorporate three efficient local texture feature methods to complete the local feature fusion operation.

Local Binary Patterns (LBP) has long been applied to face image analysis, including face detection and recognition, and gradually demonstrates its value in image feature extraction. In [33], it has been proved that LBP is a simple and effective method to depict local structure, attracting more and more attention in the field of machine vision.

Local Phase Quantization (LPQ), whose principle share similarities with LBP, mainly obtains image features by using the local discrete fourier transform (DFT) of the image. LPQ with the use of the image frequency domain feature information engenders a better

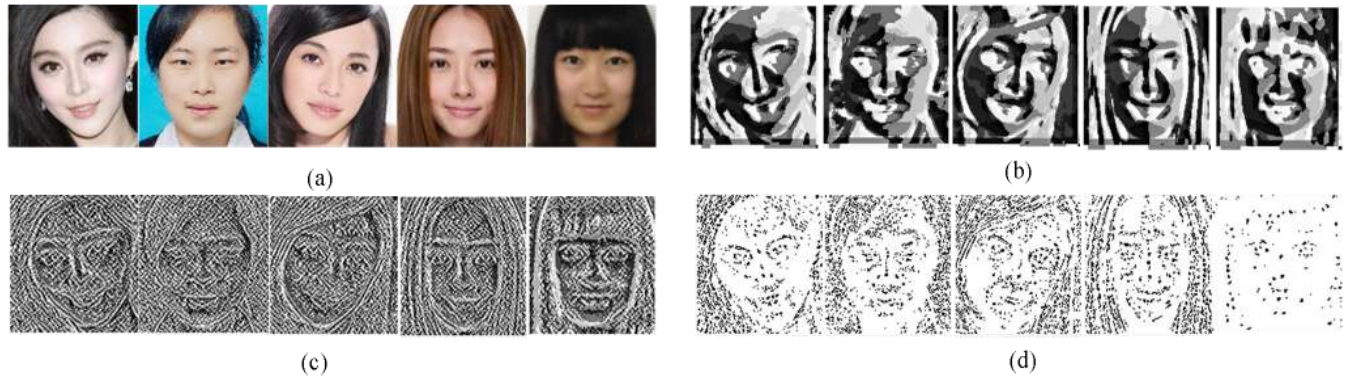


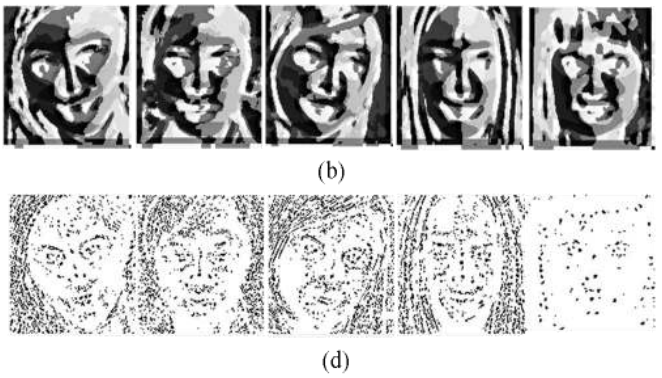
FIGURE 4. The visualization of local texture feature extraction. (a) Image of original pixels; (b) Features from LBP descriptor; (c) Features from LPQ descriptor; (d) Features from LMP descriptor.

feature invariance effect on the image interfered by blur. It means that LPQ features have good blur invariance and consequently, when using LPQ to extract the blurred image's texture information, it produces superior results. Compared with LBP, despite the LPQ feature image tends to lose more details, it still owns the competitive edge of being insensitive to blur. Hence when indefinite blurred and disturbing images are to be encountered in practice, LPQ can be helpful in enhancing the robustness of the algorithm.

Local Monotonic Pattern (LMP) mainly extracts specific micro-pattern facial features from face images with its operator applied to the image pixels. This method converts by finding the monotonic intensity of adjacent pixels with different radius, tiling the image and obtaining each tile histogram. Then through spatial information enhancement to find smaller texture features, the final feature vector is obtained from the histogram. The discriminate power of the LMP operator assists the successful integration of multiple adjacent pixel radius in the micro-pattern and the histogram of spatial encoding. The psychological experiment of Bassili [39] shows that facial features can be recognized more accurately from sequence images than from a single image. The performance of local texture feature extraction is given in Fig.4.

2) Local Feature Fusion Method with 2DPCA Dimensional Reduction

Feature fusion is basically an intermediate level process of extracting feature information from original information for comprehensive analysis and processing. The extracted feature information is a sufficient representation or statistics of the original data fusion initial information, and the multi-source information is classified and integrated accordingly. Simultaneously, multi-feature extraction can provide more feature information of the recognition target than single feature extraction, which can better build up the dimension of feature space. This paper mainly focuses on local texture feature fusion for LBP, LPQ and LMP, three of which will extract image feature information from different angles, and then fuse them to obtain a new eigenvector as the feature vector for subsequent recognition.



Before local feature fusion, 2DPCA is implemented to reduce the dimension of local texture images. If $\{B_1, B_2, \dots, B_p\}$ are the training texture images, compute the covariance matrix of the training images

$$Q = \sum_{i=1}^p (B_i - \bar{B})^T (B_i - \bar{B}) \quad (2)$$

where $i = 1, 2, \dots, p$, and $\bar{B} = \frac{1}{p} \sum_{i=1}^p B_i$.

Then, select the eigenvectors relevant to the kl biggest positive eigenvalues to form the transformation matrix V_{row} .

If B_i represents training sample, its feature matrix will be

$$G_i = (B_i - \bar{B})V_{row} \quad (3)$$

If T represents testing sample, its feature matrix will be

$$G = (T - \bar{B})V_{row} \quad (4)$$

In 2DPCA, the covariance matrix can be directly established by using the original matrix. Its main advantages are: (1) the correlation of the row eigenvectors or column eigenvectors of the image removed; (2) less time spent in calculating the eigenvalue eigenvector.

Assume that different kinds of local texture feature matrix after dimension reduction are A and B . When performing feature fusion, first, obtain a d -dimensional eigenvector F_A and an x -dimensional eigenvector F_B representing its characteristics. Then, normalize the feature eigenvectors to F'_A and F'_B according to the principle of maximum and minimum for F'_A and F'_B .

$$F'_A = \frac{F_A - \min(F_A)}{\max(F_A) - \min(F_A)} \quad (5)$$

$$F'_B = \frac{F_B - \min(F_B)}{\max(F_B) - \min(F_B)} \quad (6)$$

Where $F_A = \{F_{p1}, F_{p2}, \dots, F_{px}\}$, and $F_B = \{F_{Q1}, F_{Q2}, \dots, F_{Qd}\}$.

Finally, perform a weighted cascade fusion to the two features after normalization above mentioned

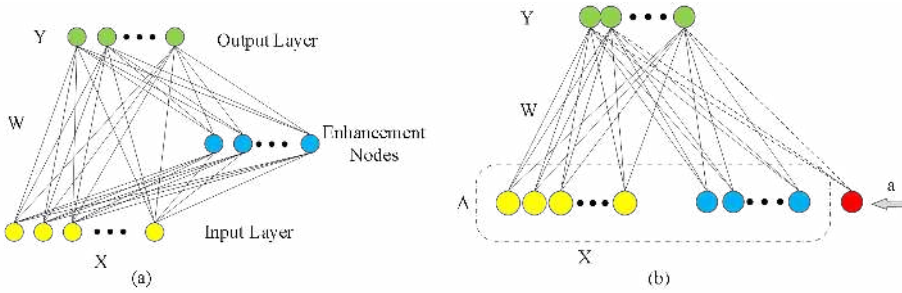


FIGURE 5. Different structure of random vector functional-link neural network. (a) The original model; (b) The update model.

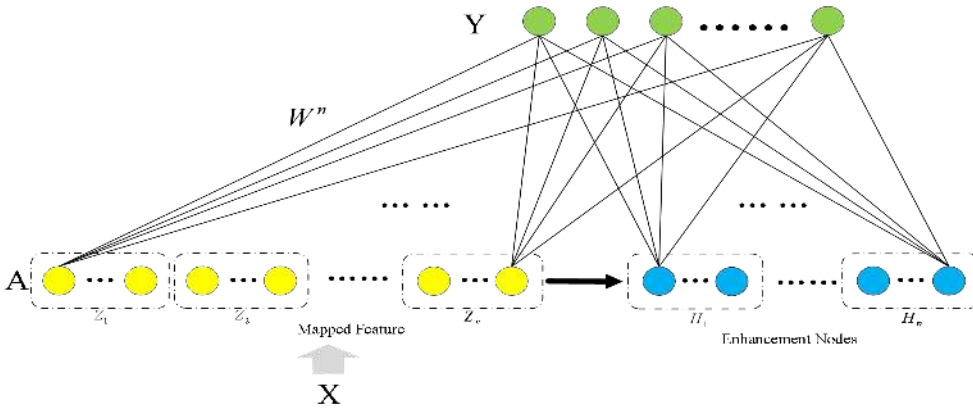


FIGURE 6. Broad learning system.

$$F = \begin{bmatrix} w_1 F'_A \\ w_2 F'_B \end{bmatrix} \quad (7)$$

In which, w_1 and w_2 are obtained through training, and $w_1 + w_2 = 1$ while the fusion ratio is directly assigned to 1:1. The fused feature F will be the final face feature.

C. Broad Learning System

1) Basic Theory of Broad Learning

This part we will introduce the basic theory of broad Learning. BLS is based on the traditional random vector functional-link neural network (RVFLNN) [41], [42]. For the basic network in general classification tasks shown in Fig. 5 (a). It is simpler and easier to update this network than the traditional model.

Assume that A_n represents an $n \times m$ pattern matrix. It is the same as adding a new column to the input matrix A^+_{n+1} , represented as

$$A_{n+1} \Delta [A_n | a] \quad (8)$$

The novel pseudo-inverse A^+_{n+1} is equals to

$$\begin{bmatrix} A^+_n - db^T \\ b^T \end{bmatrix} \quad (9)$$

where $d = A^+_n a$

$$b^T = \begin{cases} (c)^+ & \text{if } c \neq 0 \\ (1 + d^T d)^{-1} d^T A^+_n & \text{if } c = 0 \end{cases} \quad (10)$$

and $c = a - A_n d$. So, we can get the new weights below:

$$W_{n+1} = \begin{bmatrix} W_n - db^T Y_n \\ b^T Y_n \end{bmatrix} \quad (11)$$

where the weights are W_{n+1} and W_n . They represent the weights before and after adding new enhanced nodes, respectively. In this way, by calculating only the pseudo inverse of the corresponding added node, the new weight can be readily updated.

If A_n is full rank and $c=0$, the pseudo-inverse A^+_n and the weight W_n will be quickly updated. A novel node was added to the network is given at Fig. 5 (b). Where A represents the extended input matrix composed of feature nodes and enhancement nodes. A dynamic system is designed to update the weight of the model immediately for the newly added model and enhancement node.

Based on pseudoinverse and ridge regression learning algorithms, we have

$$W^m = (\lambda I + AA^T)^{-1} A^T Y \quad (12)$$

Specifically, we have that

$$A^+ = \lim(\lambda I + AA^T)^{-1} A^T \quad (13)$$

where value λ denotes the further constraints by summarizing the squared weights, W .

2) Broad Learning System Model

The BLS effectively reduces the training time of classifier, and possesses generalization ability in background of modeling and control. In this section, we mainly

introduce the structure of the broad learning network. This system needs to map the inputs to construct a set of mapped features at first. And then combine incremental learning algorithms that can update the system dynamically. Fig. 6 shows the basic structure of BLS.

Assume that input feature eigenvectors as $F \in R$. For n mapping features, m enhancement nodes can be generated. The input F is projected and the group i of feature nodes is generated; then the first mapping feature can be expressed as

$$Z_i = \varphi(FW_{ei} + \beta_{ei}) \quad (14)$$

where Z_i , $i=1,2,\dots,n$. W_{ei} and β_{ei} stand for random matrix and bias respectively. All feature nodes can be represented as $Z^n \equiv [Z_1, \dots, Z_n]$. $\varphi(\cdot)$ is an activation function. So, the m group of enhancement nodes can be denoted as

$$H_j \equiv \xi(Z^n W_{hj} + \beta_{hj}) \quad (15)$$

where, W_{hj} and β_{hj} represent random matrix and bias, respectively. All enhancement nodes can be represented as $H^m \equiv [H_1, \dots, H_m]$. $\xi(\cdot)$ is an activation function.

The combined matrix obtained by connecting the enhancement node and feature node is used as BLS's the new input. And assuming that the output matrix is $Y \in R$, the broad learning model can be abbreviated as

$$Y = [Z^n | H^m] W^m \quad (16)$$

Combining with the $W^m = (\lambda I + AA^T)^{-1} A^T Y$, and according to the pseudo-inverse and ridge regression learning algorithm, we have

$$A^+ = \lim(\lambda I + AA^T)^{-1} A^T \quad (17)$$

where $A^+ = [Z^n | H^m]^+$. This formula can be derived from the above derivation

$$W^m = [Z^n | H^m]^+ Y \quad (18)$$

which is the connection weight of the broad learning network model. The model overall construction and learning procedure of FBP via local feature fusion and broad learning is listed in Algorithm 1.

V. EXPERIMENTAL RESULTS & ANALYSIS

Experimental results are given to confirm the effectiveness of FBP network. The database we used in this experiment certifies this method's effectiveness in the LSAFBD containing 10,000 face images. There are 5 labels in LSAFBD, which are ugly, unattractive, ordinary, beautiful and glamorous. We have performed equalization processing based on this database. Fig. 7 presents the data distribution of each category in the LSAFBD. The horizontal axis denotes database label, while the vertical axis as the percentage of each category of data. Fig. 8 refers to a sample of each category labels in the database. We divided the samples in LSAFBD into training set and test

Algorithm 1: Facial Beauty Prediction via Local Feature Fusion and Broad Learning System

Input: training samples set X , Y ; Y is the output matrix; x represents one of the samples; (x_c, y_c) and g_c represents the center pixel; i_c is grayscale value; i_n represents gray value of adjacent pixels; s represents a symbolic function; $q_j(X)$ are binary coding integer values; g_{p1} and g_{p2} respectively correspond to the intensity values P equally spaced pixels on the circle with radius $R1$ and $R2$.

Output: W

for $k=0$; $k \leq n$;

Facial detection and key point detection;
Calculate $\sum X = \frac{x_1^{(1)} + \dots + x_k^{(l)}}{l_1} + \sum_{k=2}^n \frac{\Delta x_1^{(k)} + \dots + \Delta x_k^{(k)}}{l_k}$;
Cropping and gray-scale transformation;

end

Calculate $LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c)$;

Calculate $LPO(X) = \sum_{j=1}^8 q_j(X) 2^{j-1}$;

Calculate $LMP_{p,R1,R2}(x_c, y_c) = \sum_{p=0}^{p-1} s(g_{p1} - g_c) \wedge s(g_{p2} - g_{p1}) * 2^p$;

Calculate the covariance matrix $Q = \sum_{i=1}^n (B_i - \bar{B})^T (B_i - \bar{B})$;

Calculate $G_i = (B_i - \bar{B}) V_{row}$ and $G = (T - \bar{B}) V_{row}$;

Assume that local feature fusion methods are A and B ;

Calculate $F_A = \frac{F_A - \min(F_A)}{\max(F_A) - \min(F_A)}$, $F_B = \frac{F_B - \min(F_B)}{\max(F_B) - \min(F_B)}$;

Calculate $F = \begin{bmatrix} w_1 F_A \\ w_2 F_B \end{bmatrix}$;

for $i=0$; $i \leq n$;

Random W_{hj}, β_{hj} ;
Calculate $Z_i = \varphi(FW_{ei} + \beta_{ei})$;

end

Set $Z^n \equiv [Z_1, \dots, Z_n]$;

for $j=1$; $j \leq m$;

Random W_{hj}, β_{hj} ;
Calculate $H_j \equiv \xi(Z^n W_{hj} + \beta_{hj})$;

end

Set $H^m \equiv [H_1, \dots, H_m]$;

Calculate A^m and get $(A^m)^+$ with $A^+ = \lim(\lambda I + AA^T)^{-1} A^T$;

Calculate $W^m = [Z^n | H^m]^+ Y$;

if training error \geq threshold ;

Add p enhancement nodes or $n+1$ feature mapping;
Update parameters;

end

Set W ;

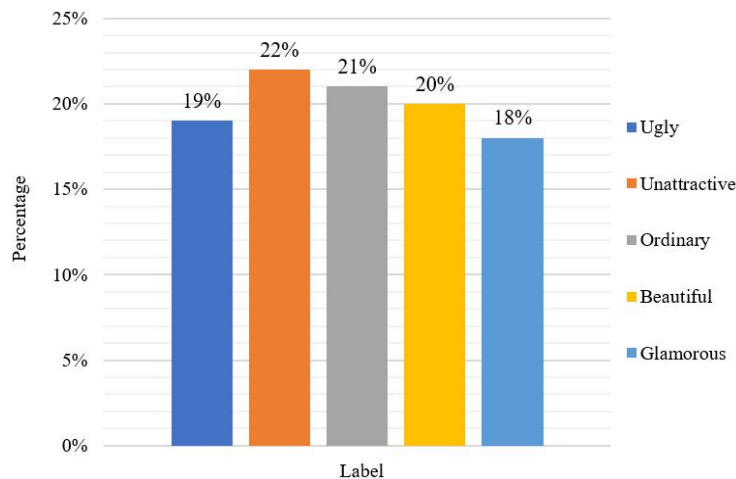


FIGURE 7. LSAFBD data distribution histogram.

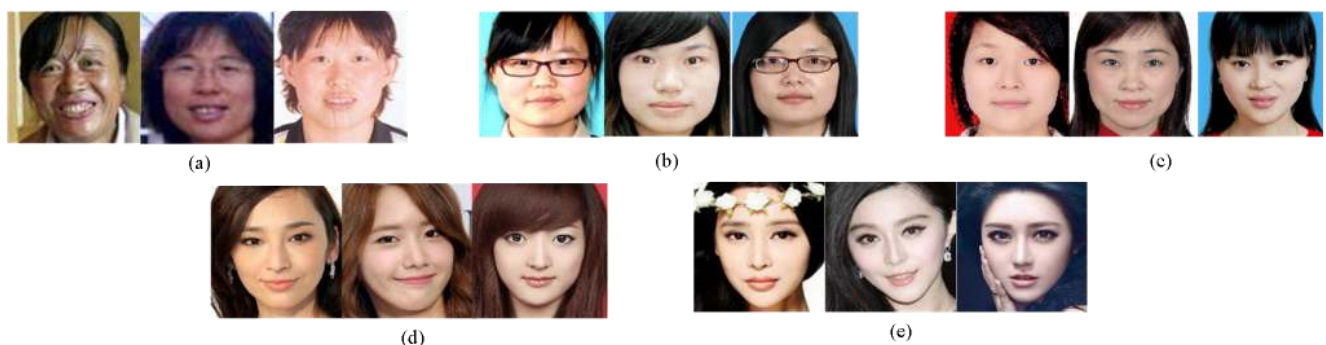


FIGURE 8. Examples of each type of label in the facial beauty database. (a) Ugly; (b) Unattractive; (c) Ordinary; (d) Beautiful; (e) Glamorous

set at a ratio of 9:1.

Comparisons are conducted between the prediction ability of our FBP network with those of the existing mainstream approaches. The experiments are equipped with a 32GB memory, Inter-i5 3.6 GHz CPU laptop on MATLAB software platform.

A. Experiments on Single Local Feature & BLS

This section will combine the broad learning network with the single local feature extraction method to build a robust model that can achieve beautiful face prediction. Table 1-4 are the FBP results of LBP, LPQ, LMP and the original pixels under different feature nodes and enhancement nodes in the BLS separately while in table 5 a comparison among different dimensional reduction methods can be found. We reduced the dimension of PCA [43] to a one-dimensional feature eigenvector 784, whereas 2DPCA calculated the eigenvectors of a 48×48 matrix.

Concluding from repeated experiments and demonstrations, we set the attenuation coefficient of the enhanced node to 0.3 and the regularization parameter of the sparse area to $2e-20$ in this experiment, which were conducive to the birth of the best test accuracy. The test accuracy result was the average value of 50 training times. Training time and testing time corresponded to the values when the accuracy was average.

It can be seen from the above experimental results that when LBP, LPQ, and LMP are used for feature extraction and original pixels as input, the accuracies of FBP are 43.09%, 52.98%, 50.19% and 54.88%, respectively. Among these subjects, LBP has the worst effect when used as feature extraction, and the prediction accuracy of original pixels remains the highest. The experiment also shows that single local feature cannot achieve robust facial feature extraction. Simultaneously, the original image as training data retains more advanced facial features, such as skin fineness and facial details leading to better prediction.

Known from the tables, based on 2DPCA the training time of LBP, LPQ, LMP and original pixels are 7.80s, 9.38s, 9.54s, and 16.24s respectively. The experiment result shows that the training speed of broad learning witnesses a geometric increase different from deep network. This corroborates the effectiveness of broad learning as a classification network.

In addition, we also compared the impact of different dimensional reduction methods on prediction performance. Table 5 proves the influence of 2DPCA, PCA and no operation on the predicted network performance. Performance results prove that 2DPCA exceeds PCA and no-operation in FBP task. After researches conducted on these two algorithms, the following conclusion can be

TABLE 1
PERFORMANCE OF LBP WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	43.75	9.10	0.87
100	900	43.87	6.17	1.02
100	1500	44.81	5.10	0.84
100	3000	45.96	9.04	0.99
100	3500	45.69	7.91	0.82
300	3000	45.76	8.55	0.91
500	3000	46.33	7.80	0.86
700	3000	45.92	9.91	1.22
800	3000	45.20	10.84	1.28

TABLE 2
PERFORMANCE OF LPQ WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	52.68	9.36	1.04
100	900	52.31	8.88	1.08
100	1500	52.06	7.88	1.06
100	3000	52.10	8.64	0.50
100	3500	53.78	7.07	0.91
300	3500	53.73	10.49	0.96
500	3500	53.67	8.99	1.57
700	3500	54.32	9.38	0.95
800	3500	54.07	11.77	1.56

TABLE 3
PERFORMANCE OF LMP WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	50.51	6.54	0.65
100	900	50.04	7.16	1.6
100	1500	51.44	8.61	1.56
100	2000	51.98	9.41	0.96
100	3000	51.51	10.08	0.95
300	2000	51.77	11.38	1.66
500	2000	52.00	10.50	1.89
700	2000	52.02	9.54	0.92
800	2000	51.87	9.23	1.82

TABLE 4
PERFORMANCE OF ORIGINAL PIXELS WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	52.64	11.34	0.65
100	900	52.68	14.65	1.96
100	1500	53.08	13.98	0.97
100	3000	53.96	15.67	1.90
100	3500	53.87	14.95	2.94
300	3000	54.10	15.34	1.33
500	3000	54.97	15.76	1.86
700	3000	55.13	16.27	1.41
800	3000	55.28	16.24	1.14

TABLE 5
PERFORMANCE COMPARISON OF DIFFERENT DIMENSIONAL REDUCTION METHOD BASED ON SINGLE LOCAL FEATURE

Method	Dimensional reduction method	Accuracy (%)	Training time (s)	Testing time (s)
LBP [34]		43.09	20.75	1.17
LPQ [35]	--	46.98	19.30	1.63
LMP [37]		47.19	18.32	1.96
Original Pixels		51.88	40.66	5.98
LBP [34]	PCA [43]	44.78	7.23	1.01
LPQ [35]		52.66	7.23	1.23
LMP [37]		51.36	6.24	0.98
Original Pixels		54.73	10.22	0.99
LBP [34]	2DPCA [12]	46.33	7.80	0.86
LPQ [35]		54.32	9.38	0.95
LMP [37]		52.02	9.54	0.92
Original Pixels		55.28	16.24	1.14

TABLE 6
PERFORMANCE OF LMP+LBP FEATURE FUSION WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	54.46	9.38	0.90
100	900	54.78	8.48	1.80
100	1500	55.84	8.65	1.20
100	3000	55.62	9.42	1.00
100	5000	55.20	10.34	1.00
300	1500	56.34	12.65	1.05
500	1500	56.83	13.85	1.15
700	1500	57.26	13.96	0.98
800	1500	56.81	12.70	1.36

TABLE 7
PERFORMANCE OF LMP+LPQ FEATURE FUSION WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	55.31	9.82	0.83
100	900	55.89	10.71	0.85
100	1500	56.42	13.53	1.71
100	3000	56.46	10.24	0.99
100	5000	56.92	12.41	1.78
300	5000	57.38	11.56	1.38
500	5000	58.86	12.94	1.43
700	5000	57.09	13.36	1.42
800	5000	56.88	13.35	1.57

TABLE 8
PERFORMANCE OF LBP+LPQ FEATURE FUSION WITH 2DPCA

Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Testing time (s)
100	500	56.10	9.32	0.93
100	900	56.29	10.71	0.87
100	1500	57.12	12.33	1.45
100	3000	57.46	12.64	1.29
100	3500	57.90	12.17	1.78
300	3500	58.78	11.56	1.38
500	3500	58.56	10.94	2.03
700	3500	58.97	13.33	1.89
800	3500	57.88	12.75	1.96

TABLE 9
PERFORMANCE COMPARISON OF DIFFERENT DIMENSIONAL REDUCTION METHOD BASED ON LOCAL FEATURE FUSION

Method	Dimensional reduction method	Accuracy (%)	Training time (s)	Testing time (s)
LMP+LBP		54.68	41.56	1.75
LMP+LPQ	--	53.68	35.10	1.33
LBP+LPQ		55.68	45.59	1.62
LMP+LBP	PCA [43]	56.32	10.46	1.01
LMP+LPQ		57.28	12.78	0.85
LBP+LPQ		57.32	11.12	0.78
LMP+LBP	2DPCA [12]	57.26	13.96	0.98
LMP+LPQ		58.86	12.94	1.43
LBP+LPQ		58.97	13.33	1.89

drawn. Because the PCA algorithm needs to convert the image into a one-dimensional eigenvector, it cannot make good use of the structural information of the image. However, the 2DPCA algorithm can directly transform the image matrix, which can effectively utilize the two-dimensional structure information of the image and enrich the feature information. The experiments hereafter will perform local feature fusion to realize a higher precision on the FBP model.

B. Experiments on Local Feature Fusion & BLS

The following experiments will perform local feature fusion on the acquired data to further improve the prediction accuracy. Local feature fusion mainly conducts more advanced feature extraction by cascading the generated feature eigenvectors of the face image.

Due to the large size of data dimension after cascade, inputting it directly into the broad learning network will raise the training time and reduce the classification accuracy affected by background noise, light and other factors. Therefore, after local texture feature extraction pro-

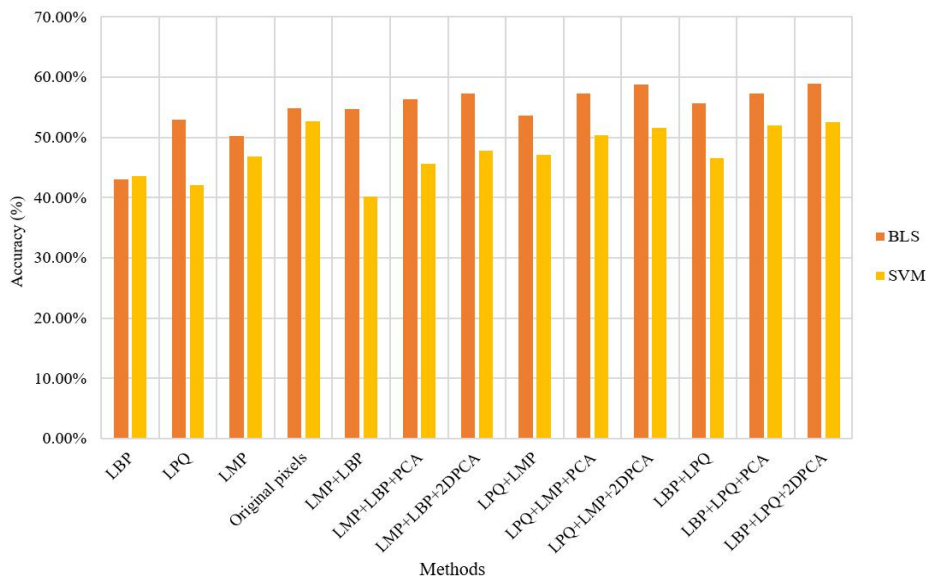


FIGURE 9. Comparison of accuracy between BLS and SVM using different feature extraction methods.

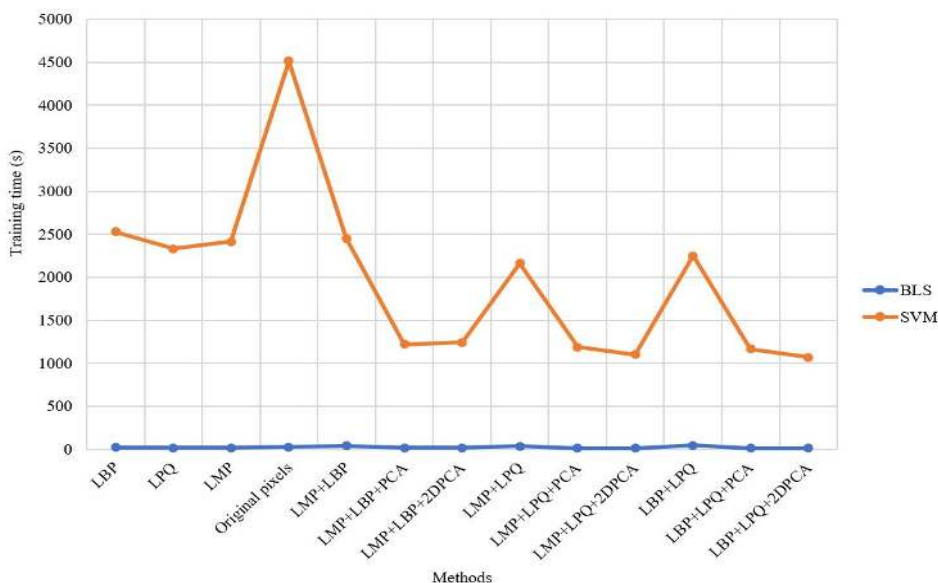


FIGURE 10. Comparison of training time between BLS and SVM using different feature extraction methods.

cess, 2DPCA need to invite again to reduce high-dimensional data. Table 6-8 are the prediction results of different numbers of feature nodes and local texture features under enhanced nodes after dimensional reduction and fusion. Table 9 also compares the experiment results among dimensional reduction methods. The dimensional reduction of PCA and 2DPCA remained the same as experiment A. But the attenuation coefficient of the enhanced node and the regularization parameter of the sparse area are set with different values.

In this experiment, the attenuation coefficient of the enhanced node set to 0.3, and the regularization parameter of the sparse area set to $2e-7$, which could optimize the experimental results. The test accuracy result was the

average value of 50 training times. Training time and testing time corresponded to the value when the accuracy was the average.

Compared with the prediction results in Table 5, the prediction performance of FBP based on local feature fusion is much higher than that of single local feature. This method can reach the prediction accuracy at 58.97% in 13.33s. Through local feature fusion, texture invariance is achieved. Also, the influence of background noise is avoided and redundant features are removed, preserving more advanced facial features. From the comparison results in Table 9, 2DPCA proves its effectiveness again in improving FBP accuracy and robustness. 2DPCA is capable of projecting the original features to the dimension with as

TABLE 10
COMPARISON OF FBP ON SCUT-FBP5500

Method [24]	PC	Training time (s)
AlexNet	86.34	94.89
PI-CNN	86.00	107.43
ResNet18 based AaNet	90.55	92.09
ResNet18 based P-Net	89.65	109.53
2M beautyNet	89.96	123.78
Our method	86.93	10.23

much information as possible through reducing the dimension of the original features.

In order to further verify the effectiveness of our method, we test on a smaller dataset SCUT-FBP5500 [44] for experiments. This is a facial beauty dataset with a beauty score ranging from 1 to 5, containing 5500 face images, which involves different races, genders and ages. The image quality of this dataset is wonderful, which is helpful for the verification of the experimental method in this paper.

We use Pearson correlation (PC) [45] to evaluate our method for FBP. The experimental results of deep learning method in the database SCUT-FBP5500 in Table 10 are from paper [24]. Experimental results show that our method can still achieve good results similar to deep learning in a very short time.

C. Performance Comparison with SVM Classifier

In this section, the traditional methods support vector machine (SVM), a generalized linear classifier, is taken as classifier performance comparison to ensure fair conclusion in this experiment, because broad learning is also based on linear features mapping. The results of performance comparison between BLS and SVM can be referred to Fig. 9 and Fig.10.

Fig.9 and Fig.10 compare the training time and prediction accuracy of BLS and SVM, showing that BLS greatly shortens the training time while improving the prediction accuracy compared with SVM. This outcome consolidates the effective status of BLS as a classifier. In addition, we can also find that whether it is with SVM or BLS, local feature fusion with 2DPCA can obtain the best results in FBP accuracy.

D. Performance Comparison with Deep learning Methods

To corroborate the effectiveness of our method, we choose five kinds of well-performed deep networks to constitute comparison with our method. These five state-of-the-art deep learning methods [46-54] can even produce satisfying outcomes even under the ImageNet challenge, such as Resnet-50, Inception-v3, and GoogLeNet; however, they demand deep structure [55] and high-performance computer in their operations.

In the deep learning experiments, related experiments about CNN needs to be tested on a VARM 5G, Intel-i7 64GB memory computer with 8G GPU on Pytorch framework. Simultaneously, we used the optimizer Adam to train 50 times on the batch size of 64. The initial learning

TABLE 11
PERFORMANCE COMPARISON OF DEEP LEARNING METHODS

Method	Accuracy (%)	Training time (s)
NIN-Imagenet [46]	55.60	1542.90
DeepID2 [47]	55.90	1342.02
GoogLeNet [48]	57.20	1243.02
VGG-CNN-S [49]	57.30	1557.89
Deep Cascaded Forest [50]	54.29	5029.40
Resnet-18 [51]	53.93	75.55
Resnet-50 [52]	54.25	165.24
ResneXt-50 [53]	55.65	187.79
Inception-v3 [54]	43.68	240.98
Our method	58.97	13.33

rate set at 0.03.

According to table 11, we compared more than a dozen deep learning methods in terms of prediction accuracy and training time. Although broad learning only performs linear feature extraction, it still obtains better prediction results and consumes less time than some popular deep learning methods did. Besides, compared with hours of training time equipped with high-performance PC in deep network, our methods enable the establishment of a high accuracy FBP model in a normal laptop within a few minutes.

VI. CONCLUSION

In this paper, a novel method is proposed based on local feature fusion and broad learning for FBP task. Unlike other models for FBP, BLS as a simple and fast network is adopted to predict for improving the training speed while ensuring prediction accuracy, successfully providing an alternative for deep learning structure. Experiments verify that broad learning can efficiently update and reshape the model on a regular computer. Furthermore, local feature fusion with 2DPCA dimensional reduction is designed to extract highly discriminative features specific to facial images, improving the model's semantic representation and transforming the bias from texture to shape.

In this work, comparisons have been conducted between the accuracy and training time of the proposed method and other state-of-the-art approaches so as to demonstrate its validity. The experiments on LSAFBD confirm the effectiveness and the efficiency of the proposed FBP model, which obviously outperforms the existing deep neural network in terms of training speed. However, in view of the serious labeling errors on FBP task, efforts will be further devoted into the exploration of new methods that can satisfactorily solve the cost-sensitive problem of facial beauty in the future.

REFERENCES

- [1] F. Chen, X. Xiao and D. Zhang, "Data-Driven Facial Beauty Analysis: Prediction, Retrieval and Manipulation," in *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 205-216, 1 April-June 2018.
- [2] L. Liu, J. Xing, S. Liu, et al, "Wow! You Are So Beautiful Today," *ACM Transactions on Multimedia Computing*,

- Communications, and Applications*, pp. 11(1): 20, 2014.
- [3] L. Marchesotti, N. Murray, F. Perronnin, "Discovering Beautiful Attributes for Aesthetic Image Analysis," *International Journal of Computer Vision*, pp. 113(3): 246-266, 2015.
- [4] F. Chen, X. Xiao and D. Zhang, "Data-Driven Facial Beauty Analysis: Prediction, Retrieval and Manipulation," in *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 205-216, 1 April-June 2018.
- [5] J. Xu, L. Jin, L. Liang, Z. Feng, D. Xie and H. Mao, "Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN)," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, LA, 2017, pp. 1657-1661.
- [6] F. Chen, D. Zhang, C. Wang et al., "Comparison and fusion of multiple types of features for image-based facial beauty prediction," in *Proceedings of Chinese Conference on Biometric Recognition*, pp. 23-30, Springer, 2017.
- [7] J. Gan et al., "2M BeautyNet: Facial Beauty Prediction Based on Multi-Task Transfer Learning," in *IEEE Access*, vol. 8, pp. 20245-20256, 2020.
- [8] Y. Zhai et al., "Asian Female Facial Beauty Prediction Using Deep Neural Networks via Transfer Learning and Multi-Channel Feature Fusion," in *IEEE Access*, vol. 8, pp. 56892-56907, 2020.
- [9] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," in *IEEE Access*, vol. 7, pp. 53040-53065, 2019.
- [10] Geirhos R, Rubisch P, Michaelis C, et al. (2018). "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," [Online]. Available: <https://arxiv.org/abs/1811.12231>
- [11] C. L. P. Chen and Z. Liu, "Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 10-24, 2018.
- [12] J. Yang, D. Zhang, A. F. Frangi and Jing-yu Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.1, pp.131-137, 2004.
- [13] Y.-H. Pao and Y. Takefuji, "Functional-link net computing: Theory, system architecture, and functionalities," *Computer*, vol. 25, no. 5, pp. 76-79, 1992.
- [14] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163-180, 1994.
- [15] B. Igel'nik and Y.-H. Pao, "Stochastic choice of basic functions in adaptive function approximation and the functional-link net," *IEEE Transaction. Neural Netw.*, vol. 6, no. 6, pp. 1320-1329, 1995.
- [16] Eysenck S, Dror G, Ruppin E. "Facial attractiveness: beauty and the machine," *Neural Computation*, pp.18 (1): 119-142. 2006.
- [17] F.J, Chau K P, Wan X, et al., "Prediction of facial attractiveness from facial proportion," *Pattern Recognition*, pp. 2326-2334. 2012.
- [18] Y.H. "Cost-sensitive ordinal regression for fully automatic facial beauty assessment," *Neurocomputing*, pp. 129(5):334-342, 2014.
- [19] Y. Zhai, Y. Huang, Y. Xu et al., "Benchmark of a large-scale database for facial beauty prediction," in *Proceedings of International Conference on Intelligent Information Processing*, 2016, pp. 1-5.
- [20] Y. Zhai et al., "Asian Female Facial Beauty Prediction Using Deep Neural Networks via Transfer Learning and Multi-Channel Feature Fusion," in *IEEE Access*, vol. 8, pp. 56892-56907, 2020.
- [21] J. Gan, L. Li, Y. Zhai, Liu Y H. "Deep Self-taught Learning for Facial Beauty Prediction," *Neurocomputing*, pp.144: 295-303, 2014.
- [22] J. Xu, L. Jin, L. Liang, Z. Feng, D. Xie. "A New Humanlike Facial Attractiveness Predictor with Cascaded Fine-tuning Deep Learning Model," *Computer Vision and Pattern Recognition*, pp.70(1): 45-79, 2015.
- [23] J. Gan, Y. Zhai, Huang Y, et al., "Research on facial beauty prediction based on dual activation layer deep convolution features," *Chinese Journal of Electronics*, pp.047(003):636-642, 2019.
- [24] J. Gan et al., "2M BeautyNet: Facial Beauty Prediction Based on Multi-Task Transfer Learning," in *IEEE Access*, vol. 8, pp. 20245-20256, 2020.
- [25] Y. Zhou, Y. Liu, G. Han and Z. Zhang, "Face Recognition Based on Global and Local Feature Fusion," *2019 IEEE Symposium Series on Computational Intelligence*, pp. 2771-2775, 2019.
- [26] D. Yu, L. X. Yu, D. Shuzhan and Y. Lei, "Group Emotion Recognition Based on Global and Local Features," in *IEEE Access*, vol. 7, pp. 111617-111624, 2019.
- [27] Y. W, Qian G, W. T, et al., "Ear Recognition Algorithm Based on Haar Local Feature Fusion," *Computer Engineering and Applications*, pp.340-352, 2019.
- [28] M. Zhu and X. Yu, "Multi-Feature Fusion Algorithm in VR Panoramic Image Detail Enhancement Processing," in *IEEE Access*, vol. 8, pp. 142055-142064, 2020.
- [29] D. Yu, L. Xingyu, D. Shuzhan and Y. Lei, "Group Emotion Recognition Based on Global and Local Features," in *IEEE Access*, vol. 7, pp. 111617-111624, 2019.
- [30] D. Zhang, H. Yang, P. Chen and T. Li, "A Face Recognition Method Based on Broad Learning of Feature Block," *2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems*, Suzhou, China, 2019, pp. 307-310
- [31] C. L. P. Chen, Z. Liu and S. Feng, "Universal Approximation Capability of Broad Learning System and Its Structural Variations," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 4, pp. 1191-1204, April 2019.
- [32] H. Huang, T. Zhang, C. Yang and C. L. P. Chen, "Motor Learning and Generalization Using Broad Learning Adaptive

- Neural Control," in *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8608-8617, Oct. 2020.
- [33] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in *IEEE Transaction. Pattern Anal.Mach. Intell.* pp.24: 971-987, 2002.
- [34] W. Li, C. Chen, H. Su and Q. Du, "Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681-3693, 2015.
- [35] V. Ojansivu, J. Heikkilä, "Blur insensitive texture classification using local phase quantization," *International Conference on Image and Signal Processing, Springer*, 2008, pp. 236-243.
- [36] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using PHOG and LPQ features," in: *Automatic Face & Gesture Recognition and Workshops, IEEE International Conference on, IEEE*, 2011, pp. 878-883.
- [37] T. Mohammad, and M.L. Ali, "Robust facial expression recognition based on local monotonic pattern (LMP)," in: *Computer and Information Technology, 2011 14th International Conference on, IEEE*, 2011, pp. 572-576.
- [38] Z. Han, Z. Liu, J. Han, C. Vong, S. Bu and C. L. P. Chen, "Unsupervised Learning of 3-D Local Features from Raw Voxels Based on a Novel Permutation Voxelization Strategy," in *IEEE Transactions on Cybernetics*, vol. 49, no. 2, pp. 481-494, 2019.
- [39] J. N. Bassili, "Emotion recognition: the role of facial movement and the relative importance of upper and lower area of the face," *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2049-2058, 1979.
- [40] Israel, A. "Generalized inverses: theory and applications," New York, USA, *J. Wiley*, 1974.
- [41] C. L. P. Chen and J. Z. Wan, "A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 1, pp. 62-72, Feb. 1999.
- [42] C. L. P. Chen, "A rapid supervised learning neural network for function interpolation and approximation," *IEEE Transaction. Neural Netw.*, vol. 7, no. 5, pp. 1220-1230, 1996.
- [43] C. Chuan, C. Zheng, L. Bo and L. Ligang, "Computation of Atmospheric Optical Parameters Based on Deep Neural Network and PCA," in *IEEE Access*, vol. 8, pp. 102256-102262, 2020.
- [44] L. Liang, L. Lin, L. Jin, D. Xie, and M. Li, "SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug.2018, pp.1598-1603.
- [45] L. Lin, L. Liang and L. Jin, "Regression Guided by Relative Ranking Using Convolutional Neural Network (R3CNN) for Facial Beauty Prediction," in *IEEE Transactions on Affective Computing*, 2019.
- [46] L. M, C. Q, Y. S. "Network In Network," *Computer ence*, 2013.
- [47] Y. Sun, Y. Chen, X. Wang et al., "Deep learning face representation by joint identification-verification," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 1988-1996, 2014.
- [48] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.
- [49] Simonyan K, Zisserman A. "Very Deep Convolutional Networks for Large-Scale Image Recognition". *Computer ence*, 2014.
- [50] Z.-H. Zhou and J. Feng. "Deep forest: Towards an alternative to deep neural networks," In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, Melbourne, Australia, 2017, pp.3553-3559.
- [51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778
- [52] X. Tian and C. Chen, "Modulation Pattern Recognition Based on Resnet50 Neural Network," *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP)*, Weihai, China, 2019, pp. 34-38.
- [53] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 5987-5995.
- [54] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 2818-2826.
- [55] Dornaika, F., Wang, K., Arganda-Carreras, I., Elorza, A., & Moujahid, A. "Toward graph-based semi-supervised face beauty prediction." *Expert Systems with Applications*, pp.142, 112990, 2019.



Yikui Zhai is an Associate Professor at Wuyi University, Guangdong, China. He was a visiting scholar at the University of Milan, Department of Computer Science during June 2016 to June 2017. He received his Ph. D degree in Signal and Information Processing at Beihang University in June 2013. He received the Bachelor and Master degree in Optical Electronics Information and Communication Engineering, and Signal and Information Processing from Shantou University, Guangdong, China in 2004 and 2007 respectively. Since October 2007, he has been working at Department of Intelligence Manufacturing, Wuyi University, Guangdong, China, and his research interests include: image processing, biometric extraction, deep learning and pattern recognition.



Cuilin Yu is currently a researcher at Faculty of Intelligent Manufacturing, Wuyi University, Guangdong, China. She received her master's degree in Department of Decision Sciences, The Hang Seng University of Hong Kong in June 2020. She received her B.S. degree in Wuyi University in June 2019. Her

current research interests include: machine vision and pattern recognition.



Chuanbo Qin is currently a Lecturer at Faculty of Intelligent Manufacturing, Wuyi University, Guangdong, China. He received his Ph. D degree in pattern recognition and intelligence system at South China University of Technology, Guangdong, China, in June 2015. He

received the Bachelor and Master degree in Communication Engineering and Signal and Information Processing from Wuyi University, Guangdong, China in 2004 and 2008 respectively. Since September 2015, he has been working at Department of Intelligence Manufacturing, Wuyi University, Guangdong, China. His current research interests include image processing and embedded system.



Wenlv Zhou is currently pursuing his Master in Department of Intelligence Manufacturing, Wuyi University. He received his B.S. degree in Beijing Institute of Technology, Zhuhai in 2019. His research interests include: biometric extraction and pattern recognition.



Qirui Ke is currently pursuing his Master in Department of Intelligence Manufacturing, Wuyi University. He received his B.S. degree at Hubei University of Arts and Science in 2018. His research interests include: biometric extraction and pattern recognition.



Junying Gan is a Full Professor at Wuyi University, Guangdong, China. She received her B.S., M.S. and Ph.D. degrees in Electrical Information Engineering from Beihang University in 1987, 1992 and 2003, respectively. She joined Wuyi University in 1992. She is the executive director of Guangdong

image graphics association, and an IEEE member. She has published more than 50 journal papers, and has received several provincial technology awards. Her research interests include: biometric extraction and pattern recognition.



Ruggero Donida Labati (S'11–M'13) received the Ph.D. degree in computer science from the Università degli Studi di Milano, Crema, Italy, in 2013.

He has been an Assistant Professor in Computer Science with the Università degli Studi di Milano since 2015. He has been a Visiting Researcher with Michigan State University, EastLansing, MI, USA. Original results have been published in more than 50 papers in international journals, proceedings of international conferences, books, and book chapters. His current research interests include intelligent systems, signal and image processing, machine learning, pattern analysis and recognition, theory and industrial applications of neural networks, biometrics, and industrial applications.

Dr. Dnida Labati is an Associate Editor with the Journal of mbient Intelligence and Humanized Computing (Springer).



Vincenzo Piuri (S'86–M'96–SM'00–F'01) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 1989.

He has been a Full Professor in Computer Engineering with the Università degli Studi di Milano, Crema, Italy, since 2000. He has been an Associate Professor with Politecnico di Milano and a Visiting Professor with the University of Texas at Austin, Austin, TX, USA and George Mason University, Fairfax, VA, USA. He is an Honorary Professor with Obuda University, Budapest, Hungary, the Guangdong University of Petrochemical Technology, Maoming, China, the Muroran Institute of Technology, Muroran, Japan, and Amity University, Noida,

India. Original results have been published in more than 350 papers in international journals, proceedings of international conferences, books, book chapters, and patents. His current research interests include intelligent systems, signal and image processing, machine learning, pattern analysis and recognition, theory and industrial applications of neural networks, biometrics, intelligent measurement systems, industrial applications, fault tolerance, digital processing architectures, and cloud computing infrastructures.

Dr. Piuri is a Distinguished Scientist of ACM and a Senior Member of INNS. He is the Editor-in-Chief of the IEEE SYSTEMS JOURNAL from 2013 to 2019 and an Associate Editor of the IEEE TRANSACTIONS ON COMPUTERS, the IEEE TRANSACTIONS ON CLOUD COMPUTING, and has been an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS, and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

He has been the IEEE Vice-President for Technical Activities in 2015, the IEEE Director, the President of the IEEE Computational Intelligence Society, the Vice-President for Education of the IEEE Biometrics Council, the Vice-President for Publications of the IEEE Instrumentation and Measurement Society and the IEEE Systems Council, and the Vice-President for Membership of the IEEE Computational Intelligence Society.



Fabio Scotti (M'03–SM'12) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003.

He has been an Associate Professor in Computer Science with the Università degli Studi di Milano, Crema, Italy, since 2015. Original results have been

published in more than 100 papers in international journals, proceedings of international conferences, books, book chapters, and patents.

His current research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, three-dimensional reconstruction, industrial applications, intelligent measurement systems, and high-level system design.

Dr. Scotti is an Associate Editor with the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS AND SOFT COMPUTING (Springer). He has been an Associate Editor with the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and a Guest Coeditor for the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.