Facial Beauty Prediction Fusing Transfer Learning and Broad Learning System

Junying Gan (junyinggan@163.com)
Wuyi University https://orcid.org/0000-0002-6418-7316

Xiaoshan Xie
Wuyi University

Yikui Zhai
Wuyi University

Guohui He
Wuyi University

Chaoyun Mai
Wuyi University

Heng Luo
Wuyi University

Research Article

Keywords: Facial Beauty Prediction, Transfer Learning, Broad Learning System.

Posted Date: February 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1349480/v1

Facial Beauty Prediction Fusing Transfer Learning and Broad Learning System

Junying Gan Xiaoshan Xie Yikui Zhai Guohui He Chaoyun Mai Heng Luo

Corresponding author: Junying Gan, email: junyinggan@163.com

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

Abstract

Facial Beauty Prediction (FBP) is an important and challenging problem in the field of computer vision and machine learning. Not only it is easily prone to over-fitting due to the lack of large-scale and effective data, but also difficult to quickly build robust and effective face beauty evaluation models because of the variability of facial appearance and the complexity of human perception. Transfer learning can be able to reduce the dependence on large amounts of data as well as avoid overfitting problems. Broad Learning System (BLS) can be capable of quickly completing models building and training. For this purpose, transfer learning was fused with BLS for facial beauty prediction in this paper. Firstly, a feature extractor is constructed by way of CNN model based on transfer learning for facial feature extraction, in which EfficientNet is used in this paper, and the facial features extracted are transferred to BLS for facial beauty prediction, called E-BLS. Then, on the basis of E-BLS, a connection layer is designed to connect the feature extractor and BLS, called ER-BLS. Experimental results show that, compared with the previous BLS and CNN methods existed, the accuracy of facial beauty prediction was improved by E-BLS and ER-BLS, indicating the effectiveness of the method presented, which can also be widely used in pattern recognition, object detection and image classification, etc.

Keywords Facial Beauty Prediction; Transfer Learning; Broad Learning System.

1 Introduction

Facial Beauty Prediction (FBP) is an important means that can be used to decipher human perception of facial beauty. It automatically evaluates the beauty attractiveness of human faces through judgments based on human perception, such as makeup evaluation (Lin et al. 2019), makeup transfer (Wan et al. 2022), personalization Recommendation (Lin et al. 2019) and cosmetic surgery planning (Xie et al. 2015), etc. However, overfitting easily occurs for the lack of large-scale valid data. And large facial appearance variance and the complexity of human perception make it difficult to construct an effective and robust beauty assessment model. The CNN model based on transfer learning aims at improving the performance of learners in the target domains by transferring the prior knowledge contained in different but related source domains (Zhuang et al. 2021), thereby reducing the dependence on data and avoiding overfitting in this way. Meanwhile, Broad Learning System (BLS) provides a new learning strategy different from deep

convolutional neural networks (Gong et al. 2021), which can quickly complete model building and training. The advantages of the methods presented provide opportunities to solve the FBP problems.

In recent years, facial beauty prediction based on transfer learning has been widely studied. Xu et al. (2018) first proposed to transfer the rich deep features in the pre-trained model to the Bayesian Ridge Regression algorithm for facial beauty prediction. Our group utilized Multiscale CNN, transfer learning and max-feature-map as activation function to solve the FBP problem, through integrating the different scales features to get good results (Zhai et al. 2019). In order to consider the correlation between tasks, our group proposed a multi-task transfer learning facial beauty prediction (2M BeautyNet), which took gender recognition as an auxiliary task and facial beauty prediction as the main task, by way of information sharing between multiple tasks to achieve fine results (Gan et al. 2020). Before long, Vahdati et al. (2020) adopted transfer and multi-task learning for facial beauty prediction, which used gender recognition and ethnicity recognition as auxiliary tasks to improve the performance of FBP. Although these methods have achieved better results, it relies heavily on high-performance hardware devices and networks have low efficiencies. In addition, in order to improve the generalization ability and accuracy of models, it needs to be retrained on the newly added data so that a lot of computer resources and time are cost.

To address these problems, Broad Learning System was proposed (Chen et al. 2018), which is an efficient incremental learning system without deep architecture. In recent years, the emergence of BLS is moving towards establishing more efficient and effective machine learning methods. Zhang et al. (2019) proposed a face recognition method based on Broad Learning System with feature block, demonstrating how face recognition with the help of BLS is not affected by the number of facial features in strong illumination and occlusion, and maintaining high accuracy. The superiority of BLS and its variants were revealed to several existing learning algorithms in time series prediction and performance regression of face recognition databases (Chen et al. 2019). Furthermore, a new method was designed for facial emotion recognition in human robot interaction based on enhanced broad Siamese network (Li et al. 2021). The fusion features of the face are extracted by Siamese network, and these features are input into the BLS for facial emotion recognition. It efficiently reduced consumption of computing time and memory resources. But the accuracy of facial beauty prediction is much lower than that of methods based on deep convolutional neural networks, because BLS is insensitive to the features of face images.

Transfer learning can be capable of reducing the data dependence of FBP and avoiding overfitting phenomenon while extracting effective face features, but it requires a lot of training time. The FBP model can be quickly constructed and trained by BLS. But it is not sensitive to facial features, resulting in low accuracy of FBP.

To solve these problems, we integrate transfer learning and BLS to solve the FBP problem in this paper, which can improve the training speed of the model and ensure the accuracy of FBP. First of all, the CNN model based on transfer learning is used as a feature extractor to extract facial features that would be transferred to BLS for facial beauty prediction, called E-BLS. Secondly, on the basis of the E-BLS, a connection layer is designed to connect the feature extractor and BLS, called ER-BLS. In the connection layer, facial features were performed by global average pooling, normalization and regularization operations, and were activated by the Radial Basis Function (RBF).

We conducted extensive experiments on the SCUT-FBP5500 (Liang et al. 2018) database and the Large Scale Asian Female Beauty Dataset (LSAFBD) (Zhai et al. 2016) to explore the properties of E-BLS and ER-BLS. Experimental results show that the E-BLS and ER-BLS presented achieve better results and outperform previous BLS methods. Meanwhile, our methods were compared with the state-of-the-art related methods on SCUT-FBP5500 and LSAFBD databases, respectively.

The contributions of this paper are summarized as follows:

1. The fusion of transfer learning and Broad Learning System is presented to solve the FBP problem for the first time, improving the model accuracy and training speed.
2. Two fusion strategies of transfer learning and Broad Learning System are presented, namely E-BLS and ER-BLS, and the facial features extracted by the feature extractor based on transfer learning are transferred to BLS to complete the facial beauty prediction.
3. Extensive experimental results illustrate the superiority of E-BLS and ER-BLS to traditional BLS and the state-of-the-art related methods. Moreover, our methods can be widely applied to the fields of pattern recognition, object detection and image classification, etc.

The content of this paper is arranged as follows. Section 2 outlines the related work of this paper and section 3 describes the fusion strategy of transfer learning and BLS. Then, we present the experimental platform and implemental details for our paper and analyze experimental results in Section 4. Finally, Section 5 will conclude the research work of this paper.

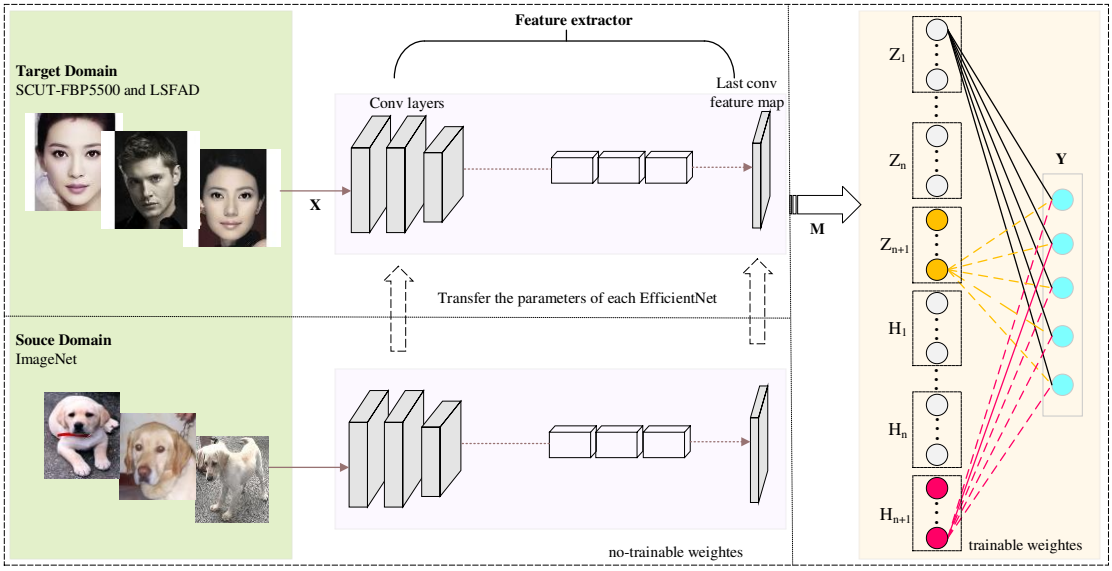

Figure 1: Architecture of E-BLS. It consists of a feature extractor and Broad Learning System. The feature extractor can be used to extract facial features as the input of BLS. And the BLS predicts facial beauty.

2 Related works

2.1 Transfer Learning

Transfer learning improves the learning effect of the learners in the target domain by transferring the prior knowledge of the relevant source domain, thereby reducing the target learner's

dependence on a large amount of data and avoiding the phenomenon of overfitting. EfficientNet based on transfer learning is used to build a feature extractor, and is a family of models that optimize floating point operations and parameter efficiency (Tan et al. 2019). It optimizes the computational complexity and parameters of the model by balancing the scaling multipliers（d, r, w） of the three dimensions of the network: depth, width, and resolution of the image. The scaling criterion is

$$\begin{aligned} &depth: d = \alpha^{\lambda} \\ &width: w = \beta^{\lambda} \\ &resolution: r = \gamma^{\lambda} \\ &s.t. \quad \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2; \\ &\qquad \alpha \geq 1,\ \beta \geq 1,\ \gamma \geq 1 \end{aligned} \tag{1}$$

where α, β, γ are constants determined by a small grid search, which specify how to assign these extra resources to network width, depth, and resolution respectively. Intuitively, λ is a user-specified coefficient that controls how many resources are available for model scaling and the speed of model operation is proportional to d, w^2, r^2. Compared with the other deep convolutional neural networks, EfficientNet has a better trade-off in terms of computing speed and accuracy.

2.2 Broad Learning System

Broad Learning System is a high-speed learning system without deep architecture, containing three essential parts such as mapping feature nodes, enhancement feature nodes and output matrix. Above all, the data can be mapped to feature nodes with random weights. Then, feature nodes are mapped to enhancement feature nodes with random weights. Finally, outputs of the model are computed through mapping feature nodes and enhancement feature nodes. In the basic BLS, assume that the input sample is expressed by X and X is mapped by feature nodes Z_i, that is

$$Z_i = \phi_i\left(XW_{ei} + b_{ei}\right), i = 1, 2, \cdots, n \tag{2}$$

Where Z_i is the *i*-th mapping feature node, W_{ei} and b_{ei} are random weights and bias. Denote $Z^n = [Z_1, Z_2, \cdots, Z_n]$, which is the concatenation of all the *n*-th groups of mapping features.

Similarly, Z^n is mapped to enhancement feature nodes H_k by

$$H_k = \xi_k\left(Z^n W_{hk} + b_{hk}\right), k = 1, 2, \cdots, m \tag{3}$$

Where H_k is the *k*-th enhancement feature node, W_{hk} and b_{hk} are random weights and biases. Denote $H^m = [H_1, H_2, \cdots, H_m]$, which is the concatenation of all the *m*-th groups of enhancement nodes.

Denote $A_m^n = [Z^n \mid H^m]$, the output of BLS is given by

$$Y = A_m^n W_n^m \tag{4}$$

where $W^m = (A_m^n)^+ Y$. W^m are the connecting weights for the broad structure and can be easily computed through the ridge regression approximation of $(A_m^n)^+$, that is the pseudo-inverse matrix of A_m^n by formula (5), where λ is the regularization parameter.

$$A_m^{n+} = \lim_{\lambda \to 0}\left(\lambda I + A_m^n A_m^{n\mathrm{T}}\right)^{-1} A_m^{n\mathrm{T}} \tag{5}$$

In incremental BLS, assume that the initial structure A_n^m consists of *n* groups feature mapping nodes Z^n and *m* groups enhancement nodes H^m. Consider that the (*n*+1)-th feature mapping group nodes are added and denoted as

$$Z_{n+1} = f\left(XW_{e_{n+1}} + b_{e_{n+1}}\right) \tag{6}$$

where $W_{e_{n+1}}$ and $b_{e_{n+1}}$ are randomly generated.

The corresponding enhancement nodes are randomly generated, that is

$$H_{e_{xm}} = \left[\xi\left(Z_{n+1}W_{ex1} + b_{ex1}\right), \ldots, \xi\left(Z_{n+1}W_{exm} + b_{exm}\right)\right] \tag{7}$$

where W_{ex} and b_{ex} are randomly generated weights and bias.

The $(m+1)$-th enhancement node added to the structure is

$$H_{m+1} = \xi(Z^{n+1}W_{h_{m+1}} + b_{h_{m+1}}) \tag{8}$$

where W_h and b_h are the random weights and bias.

The updated combined matrix and its pseudo-inverse matrix are given by

$$A_{n+1}^{m+1} = [A_n^m \mid Z_{n+1} \mid H_{m+1} \mid H_{e_{xm}}] \tag{9}$$

$$\left(A_{n+1}^{m+1}\right)^+ = \begin{bmatrix} (A_n^m)^+ - DB^{\mathrm{T}} \\ B^{\mathrm{T}} \end{bmatrix} \tag{10}$$

where the matrices D and B are computed by

$$D = (A_n^m)^+[Z_{n+1} \mid H_{m+1} \mid H_{e_{xm}}] \tag{11}$$

$$B^{\mathrm{T}} = \begin{cases} C^+ & , \ \mathrm{C} \neq 0 \\ (1 + D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}(A_n^m)^+ & , \ \mathrm{C} = 0 \end{cases} \tag{12}$$

$$C = [Z_{n+1} \mid H_{m+1} \mid H_{e_{xm}}] - AD \tag{13}$$

The output of BLS and updated weights are readily obtained by

$$Y = A_{m+1}^{n+1}W_{n+1}^{m+1} \tag{14}$$

$$\left(W_{n+1}^{m+1}\right) = \left(A_{n+1}^{m+1}\right)^+ Y = \begin{bmatrix} (W_n^m) - DB^{\mathrm{T}}Y \\ B^{\mathrm{T}}Y \end{bmatrix} \tag{15}$$

In the incremental learning process, BLS merely needs to calculate the pseudo-inverse matrix of the added feature nodes and the enhanced nodes without retraining the whole model, which greatly improves the learning efficiency of the system.

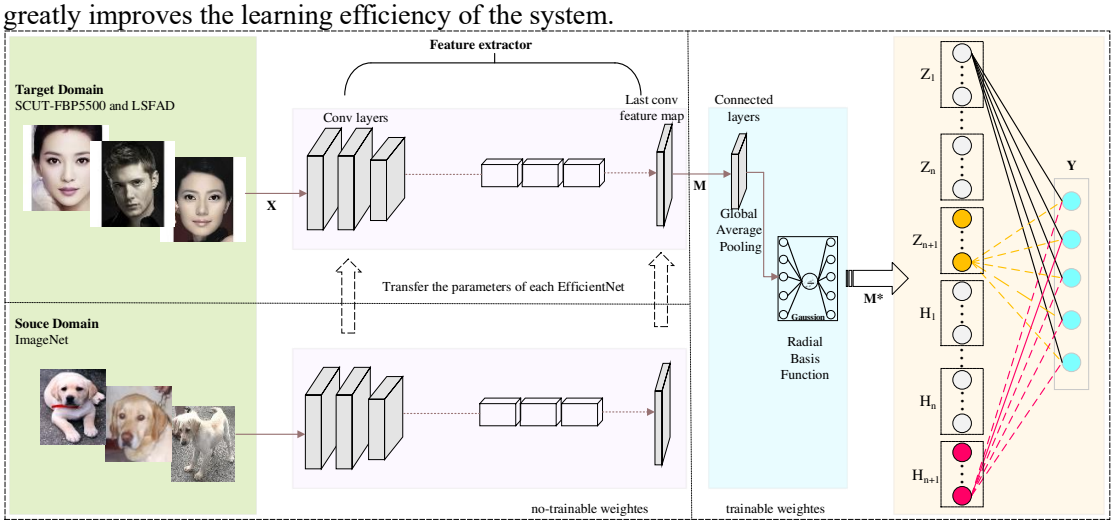

Figure 2: Architecture of ER-BLS. It consists of a feature extractor, connecting layer and Broad Learning System. The feature extractor can be used to extract facial features as the input of connected layer. In the connection layer, facial features were performed by global average pooling, normalization and regularization operations, and were activated by RBF. BLS predicts facial beauty.

3 Strategies fusing transfer learning and BLS

3.1 E-BLS

The architecture of E-BLS is shown in Figure 1. Among them, the network backbone contains the feature extractor and the BLS. The feature extractor is used to extract the features of the face, which can be implemented by various existing CNN models. And we apply EfficientNet in

experiments. BLS is used for FBP model training and testing. The details of E-BLS are shown in Algorithm 1.

Face image is fed into a feature extractor, which is built on EfficientNet with transfer learning. The fusion features are output from the last convolutional layer of EfficientNet by the following equation

$$M = \varphi\left(XW_c + b_c\right) \tag{16}$$

where X represents the input image and φ represents the feature extraction function. W_c and b_c are the fixed weights and biases of EfficientNet.

Algorithm 1 Facial Beauty Prediction with E-BLS

Input: training samples set X;
Output: output matrix set Y and W;
Calculate $M = \varphi\left(XW_c + b_c\right)$;
For $i = 0$; $i \le n$ **do**
  Random W_{ei}, b_{ei}
  Calculate $Z_i = \phi_i\left(MW_{ei} + b_{ei}\right)$;
end
Set the feature mapping group $Z^n = [Z_1, Z_2, \cdots, Z_n]$;
For $k = 0$; $k \le m$ **do**
  Random W_{hk}, b_{hk};
  Calculate $H_k = \xi_k\left(Z^n W_{hk} + b_{hk}\right)$;
end
Set the enhancement nodes group $H^m = [H_1, H_2, \cdots, H_m]$;
Calculate $Y = [Z^n \mid H^m]W_n^m$;
Calculate $W^m = [Z^n \mid H^m]^+ Y$;
While the training accurary threshold is not satisfied **do**
  Random W_{en+1}, b_{en+1}
  Calculate $Z_{n+1} = \phi(MW_{e_{n+1}} + b_{e_{n+1}})$;
  Random W_{hk}, b_{hk};
  Calculate $H_{m+1} = \xi(Z^{n+1}W_{h_{m+1}} + b_{h_{m+1}})$;
  Set $A_{n+1}^{m+1} = [A_n^m \mid Z_{n+1} \mid H_{m+1} \mid H_{e_{xm}}]$;
  Calculate $\left(A_{n+1}^{m+1}\right)^+$ by Eq. (10);
  Calculate $Y = [Z^{n+1} \mid H^{m+1}]W_{n+1}^{m+1}$;
  Calculate $\left(W_{n+1}^{m+1}\right)$ by Eq. (11,12,13,14);
  $n = n + 1$;
  $m = m + 1$;
end

We map M to the feature nodes Z with random weights by formula (2). Then we map Z to the enhancement nodes H with random weights by formula (3). Finally, we compute the output matrix for Z and H by formula (4) to assess facial beauty. If the training accuracy threshold does not satisfy our expectation, we expand the feature nodes and enhancement nodes of the model to improve accuracy and training speed by formulas (6, 7, 8, 9).

3.2 ER-BLS

The architecture of ER-BLS is shown in Figure 2, which consists of feature extractor, connection layer and BLS. We designed a connected layer to connect the feature extractor and BLS. The facial features are processed by the connected layer and then input to BLS for facial beauty prediction. In the connected layer, global average pooling, normalization and regularization are

performed to facial features, which are activated with RBF. At the same time, the connected layer can heighten the training speed of the model and avoid overfitting. The details of ER-BLS are shown in Algorithm 2.

Images X are fed into feature extractor for features extraction and output facial features M. After processing through the connected layer, we output new facial features M^*, which are fed into BLS for facial beauty prediction by equation

$$M^* = \psi\left(MW_r + b_r\right) \tag{17}$$

where ψ denotes the RBF of connected layer, W_r and b_r are the fixed weights and biases of connected layer.

Algorithm 2 Facial Beauty Prediction with ER-BLS

Input: training samples set X;
Output: output matrix set Y and W;
Calculate $M = \varphi\left(XW_c + b_c\right)$;
Calculate $M^* = \psi\left(MW_r + b_r\right)$;
For $i = 0$; $i \le n$ **do**
  Random W_{ei}, b_{ei}
  Calculate $Z_i = \phi_i\left(M^* W_{ei} + b_{ei}\right)$;
end
Set the feature mapping group $Z^n = [Z_1, Z_2, \cdots, Z_n]$;
For $k = 0$; $k \le m$ **do**
  Random W_{hk}, b_{hk};
  Calculate $H_k = \xi_k\left(Z^n W_{hk} + b_{hk}\right)$;
end
Set the enhancement nodes group $H^m = [H_1, H_2, \cdots, H_m]$;
Calculate $Y = [Z^n \mid H^m] W_n^m$;
Calculate $W^m = [Z^n \mid H^m]^+ Y$;
While the training accurary threshold is not satisfied **do**
  Random W_{en+1}, b_{en+1}
  Calculate $Z_{n+1} = \phi(MW_{e_{n+1}} + b_{e_{n+1}})$;
  Random W_{hk}, b_{hk};
  Calculate $H_{m+1} = \xi(Z^{n+1} W_{h_{m+1}} + b_{h_{m+1}})$;
  Set $A_{n+1}^{m+1} = [A_n^m \mid Z_{n+1} \mid H_{m+1} \mid H_{e_{xm}}]$;
  Calculate $\left(A_{n+1}^{m+1}\right)^+$ by Eq. (10);
  Calculate $Y = [Z^{n+1} \mid H^{m+1}] W_{n+1}^{m+1}$;
  Calculate $\left(W_{n+1}^{m+1}\right)$ by Eq. (11,12,13,14);
  $n = n + 1$;
  $m = m + 1$;
end

We map M^* to the feature nodes Z with random weights by formula (2). Then we map Z to the enhancement nodes H with random weights by formula (3). Finally, we compute the output matrix for Z and H by formula (4) to assess facial beauty. If the training accurary threshold does not meet our expectation, we extend the feature nodes and enhancement nodes of the model to improve accuracy and training speed with formulas (6, 7, 8, 9).

4 Experiments and results analysis

To explore the characteristics of E-BLS and ER-BLS, firstly, the method in this paper is compared with the deep convolutional neural network method based on transfer learning. Then, the effectiveness of E-BLS and ER-BLS has been demonstrated through numerous trials. Finally, our methods were compared with related methods separately. All experiments were implemented on a Python software platform with an Intel-i7 3.6 GHz CPU and 64 GB RAM desktop computer.

4.1 Experimental databases

4.1.1 SCUT-FBP5500 database

SCUT-FBP5500 is a facial beauty prediction database established by South China University of Technology. It contains 5,500 frontal face images at 350×350 resolution with different races, gender and age. Each image is rated by 60 volunteers and is labeled with a beauty score ranging from 1 to 5. The larger the score, the more attractive. Among them, there were 76 images in score "1", 821 images in score "2", 3278 images in score "3", 1226 images in score "4" and 99 images in score "5". Figure 3 shows the beautiful scores distribution and some image samples of SCUT-FBP5500.

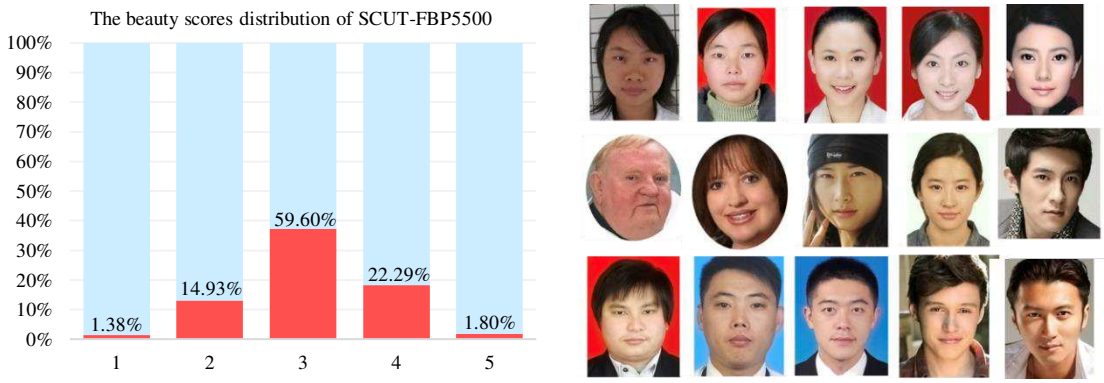

Figure 3: The distribution of SCUT-FBP5500 with beautiful score and faces with different properties

4.1.1 LSAFBD database

LSAFBD is a facial beauty prediction database established by our group, which consists of 20,000 labeled images and 80,000 unlabeled images with the resolution of 144×144. Most facial images include variations in background, pose, and age. Each image is rated by 75 volunteers and all images were divided into five categories, labeled as "0", "1", "2", "3" and "4", in increasing order of beauty. Among them, there were 948 images in category "0", 1148 images in category "1", 3846 images in category "2", 2718 images in category "3" and 1333 images in category "4". This paper focuses on the prediction of female beauty, and only 10,000 LSFBD female images were used to verify the effectiveness of our methods for the FBP problems. Figure 4 shows the beautiful score distribution and some image samples of LSAFBD.

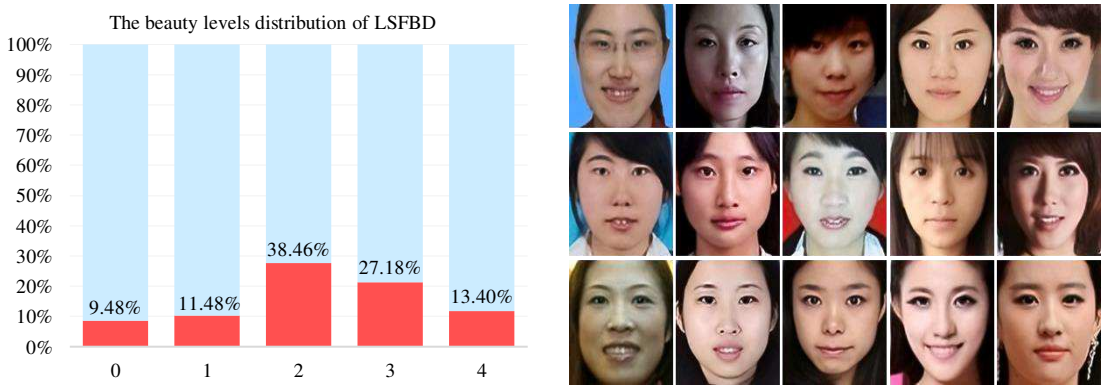

Figure 4: The distribution of LSAFBD with beauty levels and faces with different properties.

4.2 Model Training and Testing

These two databases are randomly divided into the training set and the testing set in the ratio of 8: 2 in our experiments, respectively. The E-BLS and ER-BLS presented contain a very small number of hyperparameters. For the BLS, the hyperparameters include the number of feature windows (*N1*), the number of nodes in each feature window (*N2*), and the number of enhancement nodes (*N3*). In this section, Hyperopt was used to optimize the optimal value of *N1*, *N2* and *N3*. To quantify the improvement effect of the presented methods, the plain BLS model was trained with the same hyperparameters. In addition, some classical deep CNNs, such as ResNet50 (He et al. 2016), InceptionV3 (Szegedy et al. 2016), DensNet121 (Huang et al. 2017), InceptionResNetV2 (Szegedy et al. 2017), EfficientNetB7 (Tan et al. 2019), MobileNetV2 (Sandler et al. 2018), NASNet (Zoph et al. 2018), and Xception (Chollet et al. 2017) based on transfer learning were also trained with the same hyperparameters. The initial learning rate is 0.001. When the training accuracy does not improve for more than 3 epochs, the multiplicative factor of learning rate decay is 0.5. The batch size and the number of epochs is 16 and 50. The coefficient of regularization is 0.3 and activation function utilizes linear rectification function. Meanwhile, the initial weights of these networks are from the ImageNet. Because deep CNNs were trained by transfer learning, all the layers except for around the final layer were frozen and then trained only around the final layer. When images are fed into these networks for training, the original resolution is maintained.

4.3 Experiments and results analysis by E-BLS algorithm

4.3.1 Experiments on SCUT-FBP5500

In this section, we conducted extensive experiments on the SCUT-FBP5500 database to explore the properties of E-BLS. In E-BLS, Hyperopt was used to optimize the hyperparameters: $N1 = 12$, $N2 = 54$ and $N3 = 2966$ for the training of BLS and E-BLS. Experimental results are shown in Table 1. The prediction accuracy of facial beauty prediction by BLS directly is only 65.85%, which is the lowest among all methods. Nevertheless, the testing accuracy of E-BLS is 73.13%, which is 7.28% better than BLS, 0.43% better than EfficientNetB7, and only 0.03% lower than InceptionV3's 73.16%. Furthermore, the training time for deep convolutional neural networks based on transfer learning was between 4224.4 2 and 2589 6.82 seconds, while E-BLS only needs 1399.88 seconds. As a result, E-BLS improved the network efficiency by several times to more than a dozen times while maintaining the accuracy of the model.

Table 1. Results of facial beauty prediction on SCUT-FBP5500

Model	Training time (s)	Testing accuracy (%)	Training accuracy (%)
E-BLS (ours)	**1399.88**	**73.13**	**75.38**
BLS	640.5	65.85	68.83
NASNet	5493.06	68.11	73.92
MobileNetV2	4224.42	69.03	79.3
DensNet121	9862.52	70.77	75.97
ResNet50	8211.02	71.23	76.63
InceptionResNetV2	25896.82	71.69	74.28
EfficientNetB7	17589.68	72.7	76.68
Xception	6770.64	72.98	75.01
ER-BLS (ours)	**1291.83**	**74.69**	**76.76**
InceptionV3	11630.7	73.16	78.02

The training loss curve and validation loss curve of these deep CNNs have been shown in Figure 5 and 6. Each network nearly tends to converge after about 50 epochs. When we increase epochs with little performance improvement but greatly increasing the training time. In addition, the loss curve of EfficientNetB7 training and validation is more stable than the other networks.

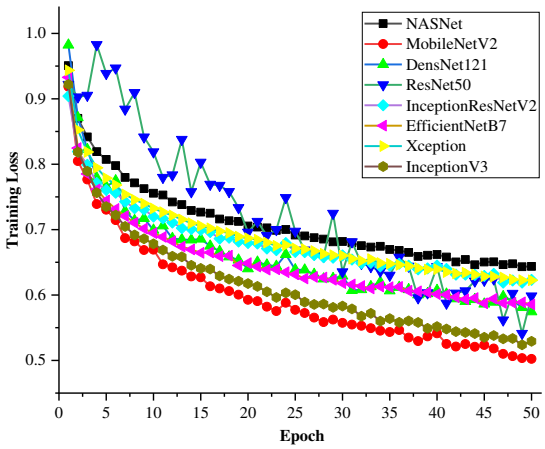

Figure 5 Training loss curves of each algorithm

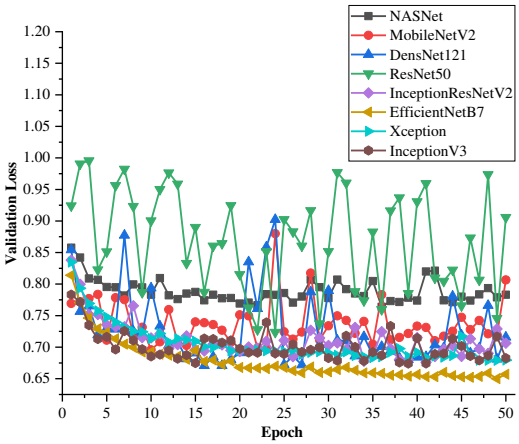

Figure 6 Validation loss curves of each algorithm

4.3.2 Experiments on LSAFBD

In this section, we continue to conduct extensive experiments to evaluate the performance and explore the properties of E-BLS on the LSAFBD database. In E-BLS and BLS, the hyperparameters are as follows: *N1* is 25, *N2* is 72 and *N3* is 3088. Experimental results are shown in Table 2. The testing accuracy of facial beauty prediction by BLS directly is 52.96%, which is the lowest among all the methods. Nevertheless, the testing accuracy of E-BLS is 60.82%, which is 9.17% better than BLS, 2.91% better than EfficientNetB7. Its testing accuracy outperformed the convolutional neural networks. Furthermore, the training time for deep convolutional neural networks based on transfer learning was between 7830.47 and 36853.53 seconds, while E-BLS merely needs 2300.48 seconds. Therefore, the training efficiency of FBP with E-BLS is significantly improved.

Table 2. Results of facial beauty prediction on LSAFBD

Model	Training time (s)	Testing accuracy (%)	Training accuracy (%)
ER-BLS (ours)	**2286.54**	**62.13**	**72.34**
E-BLS (ours)	**2300.48**	**60.82**	**71.58**
BLS	1446.37	52.96	67.89
NASNet	12043.95	53.12	60.72
MobileNetV2	7830.47	54.74	66.66
Xception	16852.60	57.01	62.91
InceptionResNetV2	36853.53	57.31	60.91
EfficientNetB7	21165.61	57.91	61.32
ResNet50	20014.05	58.37	65.10
InceptionV3	25201.29	58.42	66.47
DensNet121	19520.28	59.93	63.54

4.4 Experiments and results analysis by ER-BLS algorithm

4.4.1 Experiments on SCUT-FBP5500

In order to solve the overfitting problem of E-BLS, ER-BLS is presented in this paper in combination with E-BLS. We use the same hyperparameters of E-BLS to train ER-BLS: *N1* = 12, *N2* = 54 and *N3* = 2966. Experimental results are shown in Table 1. The performance of ER-BLS is better than E-BLS. Its training accuracy is 76.76%, testing accuracy is 74.69%, and the training time is 1291.83 seconds, which is the best among all the algorithms. Improved quantification of ER-BLS is shown in Table 3. Compared with E-BLS, the testing accuracy of ER-BLS is improved by 1.56% and the training time is reduced by 108.05 seconds. On the other hand, compared with the other algorithms, the accuracy of ER-BLS is improved between 1.53% and 8.84%, and the training time is shortened between 2932.59 and 24604.99 seconds.

Table 3 Quantitative results of performance improvement on SCUT-FBP5500

Model	Decreased time(s)	Improved accuracy (%)
InceptionResNetV2	24604.99	3.00
EfficientNetB7	**16297.85**	**1.99**
InceptionV3	10338.87	1.53
DensNet121	8570.69	3.92
ResNet50	6919.19	3.46
Xception	5478.81	1.71
NASNet	4201.23	6.58
MobileNetV2	2932.59	5.66
E-BLS (ours)	**108.05**	**1.56**
BLS	-651.33	8.84

ER-BLS is able to get fine classification accuracy with a few parameters. ER-BLS is trained with 12×54 feature nodes and 2966 enhancement nodes. These nodes directly determine the learning effect of ER-BLS. To further explore the classification performance of ER-BLS, studies with different mapping feature nodes and enhancement nodes were conducted. From Table 4, we can see that the more the total number of nodes, the longer the training time, but the testing accuracy of the model decreases after initial improvement. This implies that the choice of the number of nodes will become a key factor affecting the accuracy of the model. Thus, it is a reliable scheme to select parameters by Hyperopt.

Table 4 ER-BLS with different feature nodes and enhancement nodes on SCUT-FBP5500

Feature nodes	Enhancement nodes	Training time (s)	Testing accuracy (%)
12×54	1000	1292.1	73.77
12×54	2000	1294.65	74.226
12×54	3000	1294.71	74.59
12×54	4000	1295.97	74.317
12×54	5000	1295.95	74.317
12×54	6000	1296.36	73.678
12×54	7000	1295.75	73.133
12×54	8000	1297.60	72.951
14×54	8000	1300.92	72.222
16×54	8000	1303.64	72.86
16×57	8000	1306.09	72.86
16×60	8000	1305.21	72.678

4.4.2 Experiments on LSAFBD

In this section, we explore the properties of ER-BLS on LSAFBD. The parameters of ER-BLS are set as $N1$ = 25, $N2$ = 72 and $N3$ = 3088, which are consistent with E-BLS. The results are presented in Table 5. The performance of ER-BLS is better than E-BLS. Its training accuracy is 72.34%, testing accuracy is 62.13%, and the training time is 2286.54 seconds, which is the best among all the algorithms. Improved quantification of ER-BLS is shown in Table 5. Compared with E-BLS, the testing accuracy of ER-BLS is improved by 1.31% and the training time is reduced by 13.94 seconds. On the other hand, compared with the other algorithms, the accuracy of ER-BLS is improved between 2.20% and 9.17%, and the training time is shortened between 5543.93 and 34566.99 seconds. The accuracy and training speed of FBP are significantly improved with ER-BLS.

Table 5 Quantitative results of performance improvement on LSAFBD

Model	Decreased time(s)	Improved accuracy (%)

InceptionResNetV2	34566.99	4.82
InceptionV3	22914.75	3.71
EfficientNetB7	18879.07	4.22
ResNet50	17727.51	3.76
DensNet121	17233.74	2.20
Xception	14566.06	5.12
NASNet	9757.41	9.01
MobileNetV2	5543.93	7.39
E-BLSNet(ours)	**13.94**	**1.31**
BLSNet	-840.17	9.17

Both the number and distribution of samples may affect the accuracy of ER-BLS. To further investigate this issue, we modified the number of training samples in LSAFBD. LSAFBD contains 10000 images, in which 8000 images are used as the training set and 2000 images as the testing set. The number of images in the testing set is not changed, and the number of images in the training set is modified to 4000~8000, as shown in Table 6. As can be seen from the table, the testing accuracy increases greatly when the number of training samples increases. Thus, we believe that the number of training samples has a large influence on the accuracy of ER-BLS.

Table 6 ER-BLS with different number of training samples on LSAFBD

Training sample numbers	Training time (s)	Testing accuracy (%)	Training accuracy (%)
4000	2280.87	51.30	93.45
5000	2282.06	51.40	89.10
6000	2283.64	55.11	81.17
7000	2285.04	58.72	77.19
8000	2286.54	62.13	72.34

4.5 Incremental learning by ER-BLS

ER-BLS can complete model construction and training by incremental learning. It is a dynamic system whose performance would be improved by adding additional feature nodes, enhancement nodes or new data. At the same time, the structure of ER-BLS would be quickly updated and trained, which greatly improves the efficiency of ER-BLS. To verify the incremental learning effect of the presented methods, we conducted experiments on SCUT-FBP5500 and LSAFBD by adding feature nodes and enhancement nodes. For SCUT-FBP5500, the initial feature nodes are 548 and enhancement nodes are 500. For LSAFBD, the ER-BLS were initialized with 1700 feature nodes and 1000 enhancement nodes. Then, the incremental algorithm was adopted to add dynamically 20 feature nodes and 500 enhancement nodes each time. The testing results of incremental learning are presented in Tables 7 and 8. For SCUT-FBP5500, ER-BLS is reconstructed and trained within 12 seconds. For LSAFBD, ER-BLS is reconstructed and trained within 37 seconds. Compared with deep convolutional neural networks that consume a lot of time for retraining, ER-BLS greatly improved the efficiency of the model.

Table 7 Incremental learning of ER-BLS on SCUT-FBP5500

Feature nodes	Enhancement nodes	Training time (s)	Testing accuracy (%)
548	500	5.18	73.32
548-568	500-1000	8.39	73.50
568-588	1000-1500	9.49	73.68
588-608	1500-2000	10.12	73.59
608-628	2000-2500	11.19	74.14
628-648	2500-3000	11.98	73.86

Table 8 Incremental learning of ER-BLS on LSAFBD

Feature nodes	Enhancement nodes	Training time (s)	Testing accuracy (%)
1700	1000	19.54	60.62
1700-1720	1000-1500	26.52	61.12
1720-1740	1500-2000	28.49	60.87
1740-1760	2000-2500	30.99	61.17%
1760-1780	2500-3000	32.94	61.32%
1780-1800	3000-3500	36.45	61.42%

4.6 Comparison of different methods

To further verify the effectiveness of E-BLS and ER-BLS, the methods presented were compared with the related methods, respectively. The results are shown in Table 9 and 10. For SCUT-FBP5500, the Pearson Correlation Coefficient of E-BLS is 0.9104 and ER-BLS is 0.9303, which is the best among the various methods. For LSAFBD, the testing accuracy of E-BLS is 60.82% and ER-BLS is 62.13%, which is the best among the various methods. The E-BLS and ER-BLS presentd can be extended to the fields of pattern recognition, object detection and image classification.

Table 9 Comparison of FBP Pearson Correlation among the other methods on SCUT-FBP5500

Model	Pearson Correlation Coefficient
P-AaNet (Lin et al. 2019)	0.8965
2M BeautyNet (Gan et al. 2020)	0.8996
AestheticNetG (Danner et al. 2021)	0.9011
AaNet (Lin et al. 2019)	0.9055
E-BLS (ours)	**0.9104**
R^3CNN (Lin et al. 2019)	0.9142
ER-BLS (ours)	**0.9303**

Table 10 Comparison of FBP testing accuracy among the other methods on LSAFBD

Model	Testing accuracy (%)
Deep Cascaded Forest (Zhou et al. 2017)	54.29
NIN (Szegedy et al. 2015)	58.30
NetA+DAL (Gan et al. 2019)	59.90
DeepID2 (Zhai et al. 2019)	60.25
E-BLS (ours)	**60.82**
LDCNN (Gan et al. 2020)	62.00
ER-BLS (ours)	**62.13**

5 Conclusion

In this work, we have presented effective fusion algorithms of transfer learning and Broad Learning System to solve FBP problems, including E-BLS and ER-BLS. Among them, E-BLS transfers the facial features extracted by the feature extractor with transfer learning to BLS for facial beauty prediction. On the basis of E-BLS, ER-BLS adds a connection layer to connect the feature extractor and BLS, which has solved the overfitting problem of E-BLS and further improves the performance of FBP. Extensive experimental results show that our methods are effective. 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 PC within 40 minutes. In the future, we will consider how to design a more versatile and effective

fusion algorithm, and combine the local features and information such as gender and race that influence facial beauty.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61771347, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515010716 and in part by the Basic Research and Applied Basic Research Key Project in General Colleges and Universities of Guangdong Province under Grant 2018KZDXM073.

Compliance with ethical standards

Conflict of interest All authors declare that he/she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Lin L, Liang L, Jin L, Chen W (2019) Attribute-aware convolutional neural networks for facial beauty prediction. In: Proc. 28th Int. Joint Conf. Artif. Intell, pp 847–853

Wan Z, Chen H, An J, Jiang W, Yao C, Luo J (2022) Facial attribute transformers for precise and robust makeup transfer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 1717-1726

Lin L, Liang L, Jin L (2019) Regression Guided by Relative Ranking Using Convolutional Neural Network (R^3CNN) for Facial Beauty Prediction. In: IEEE Transactions on Affective Computing, pp 1-14

Xie D, Liang L, Jin L, Xu J, Li M (2015) SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception. In: IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, pp 1821-1826

Zhuang F, Qi Z, Duan K et al (2019) A Comprehensive Survey on Transfer Learning. In: Proceedings of the IEEE, pp 43-76

Gong X, Zhang T, Chen C, Liu Z (2021) Research Review for Broad Learning System: Algorithms, Theory, and Applications. In: IEEE Transactions on Cybernetics, pp 1-29

Xu L, Xiang J, Yuan X (2018) Transferring rich deep features for facial beauty prediction. arXiv preprint arXiv:1803.07253

Zhai Y, Cao H, Deng W, Gan J, Piuri V, Zeng J (2019) BeautyNet: Joint multiscale CNN and transfer learning method for unconstrained facial beauty prediction. In: Computational intelligence and neuroscience, pp 1–14

Gan J, Xiang L, Zhai Y, Mai C, He G, Zeng J, Bai Z, Labati R, Piuri V, Scotti F (2020) 2M BeautyNet: Facial beauty prediction based on multi-task transfer learning. In: IEEE Access, pp 20245–20256

Vahdati E, Suen C (2020) Facial Beauty Prediction Using Transfer and Multi-task Learning Techniques. In: International Conference on Pattern Recognition and Artificial Intelligence, pp 441-452

Chen C, Liu Z (2018) 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, pp 10-24

Zhang D, Yang H, Chen P, Li T (2019) A face recognition method based on broad learning of feature block. In: Proc. IEEE 9th Annu. Int. Conf. CYBER Technol. Automat., Control, Intell. Syst. (CYBER), pp 307–310

Chen C, Liu Z, Feng S (2019) Universal approximation capability of broad learning system and its structural variations. In: IEEE Trans. Neural Netw. Learn. Syst, pp 1191–1204

Li Y, Zhang T, Chen C (2021) Enhanced Broad Siamese Network for Facial Emotion Recognition in Human–Robot Interaction. In: IEEE Transactions on Artificial Intelligence, pp 413-423

Liang L, Lin L, Jin L, Xie D, Li M (2018) SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In: Proc. 24th Int. Conf. Pattern Recognit. (ICPR),

pp 1598–1603
Zhai Y, Huang Y, Xu Y, Zeng J, Yu F, Gan J (2016) Benchmark of a Large Scale Database for Facial Beauty Prediction. In: Proc. Int. Conf. Intell. Inf. Process. (ICIIP), pp 131–135
Tan M, Le Q (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc. 36th Int. Conf. Mach. Learn., pp 6105-6114
He K, Zhang X, Ren S, Sun Jian (2016) Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 770-778
Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2818-2826
Huang G, Liu Z, Maaten L, Kilian Q, Weinberger (2017) Densely Connected Convolutional Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 4700-4708
Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-V4 inception-ResNet and the impact of residual connections on learning. In: Proc. AAAI, pp 1-3
Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 4510-4520
Zoph B, Vasudevan V, Shlens J, Le Q (2018) Learning Transfer- able Architectures for Scalable Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pat- tern Recognition (CVPR), pp 8697-8710
Chollet F (2017) Xception: Deep Learning With Depthwise Separable Convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1251-1258
Danner M, Weber T, Peng L, Gerlach T, Su X, Rätsch M (2021) Ethically aligned deep learning: unbiased facial aesthetic prediction. arXiv preprint arXiv: 211 1.05149
Zhou Z, Feng J (2017) Deep forest: Towards an alternative to deep neural networks. In: Proc. 26th Int. Joint Conf. Artif. Intell., Melbourne,VIC, Australia, pp 3553–3559
Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, "Going deeper with convolutions. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp 1–9
Gan J, Zhai Y, Huang Y, Zeng J, et al. (2019) Research of facial beauty prediction based on deep convolutional features using double activation layer. In: ACTA ELECTONICA SINICA, pp 636-643
Gan J, Jiang K, Tan H, He G (2020) Facial beauty prediction based on lighted deep convolution neural network with feature extraction strengthened. In: China. J. Electron, pp 312-321