2M BeautyNet: Facial Beauty Prediction Based on Multi-Task Transfer Learning

JUNYING GAN[1], (Member, IEEE), **LI XIANG**[1], **YIKUI ZHAI**[1,2], (Member, IEEE),
CHAOYUN MAI[1], **GUOHUI HE**[1], **JUNYING ZENG**[1], **ZHENFENG BAI**[1],
RUGGERO DONIDA LABATI[2], (Member, IEEE), **VINCENZO PIURI**[2], (Fellow, IEEE),
AND FABIO SCOTTI[2], (Member, IEEE)

[1]Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China
[2]Departimento di Information, Università Degli Studi di Milano, 20133 Milano, Italy

Corresponding author: Yikui Zhai (yikuizhai@163.com)

ABSTRACT Facial beauty prediction (FBP) has become an emerging area in the field of artificial intelligence. However, the lacks of data and accurate face representation hinder the development of FBP. Multi-task transfer learning can effectively avoid over-fitting, and utilize auxiliary information of related tasks to optimize the main task. In this paper, we present a network named Multi-input Multi-task Beauty Network (2M BeautyNet) and use transfer learning to predict facial beauty. In the experiment, beauty prediction is the main task, and gender recognition is the auxiliary. For multi-task training, we employ multi-task loss weights automatic learning strategy to improve the performance of FBP. Finally, we replace the softmax classifier with a random forest. We conduct experiments on the Large Scale Facial Beauty Database (LSFBD) and SCUT-FBP5500 database. Results show that our method has achieved good results on LSFBD, the accuracy of FBP is up to 68.23%. Our 2M BeautyNet structure is suitable for multiple inputs of different databases.

INDEX TERMS Facial beauty prediction, multi-task transfer learning, multi-input multi-output network, multi-task loss weight automatic learning strategy.

I. INTRODUCTION

At present, the basic research of facial beauty prediction (FBP) has promoted the rapid development of plastic surgery and cosmetics industry, like cosmetic recommendation [1], esthetic surgery planning [2], face-based pose analysis [3] and facial beautification [4]. It has become an essential role in our daily life, while it also suffers from factors like appearance, social status, personal feelings and others. Consequently, FBP has become an emerging area in the field of artificial intelligence.

In recent years, most studies of FBP are based on deep learning [5]–[7]. Although these methods have achieved good results, there are still some challenges. One of the main difficulties is the lack of training data. Currently, the study of FBP is based on a small database with only a few thousand

images. SCUT-FBP5500 [6] is a facial beauty database of 5500 images, constructed by South China University of Technology. Our group has built a Large Scale Facial Beauty Database (LSFBD) [7], including 20,000 labeled images (10000 male images and 10000 female images) and 80000 unlabeled images. However, training a deep convolutional neural network by a small database is prone to over-fitting. Hence our group constructed a multi-scale network structure with transfer learning to solve the problem of insufficient data, and the accuracy on LSFBD is 67.4% [8].

Currently, most researchers consider single-task learning (STL) in FBP, ignoring the correlation between tasks. Unlike the single task, multi-task learning (MTL) utilizes the additional useful information from the auxiliary, improving the generalization performance and learning efficiency [9], [10]. Besides, the multi-task network has other merits. For example, it contains shared layers that avoid recalculating the characteristics of each task, thus the memory usage is

---

The associate editor coordinating the review of this manuscript and approving it for publication was Fan-Hsun Tseng.
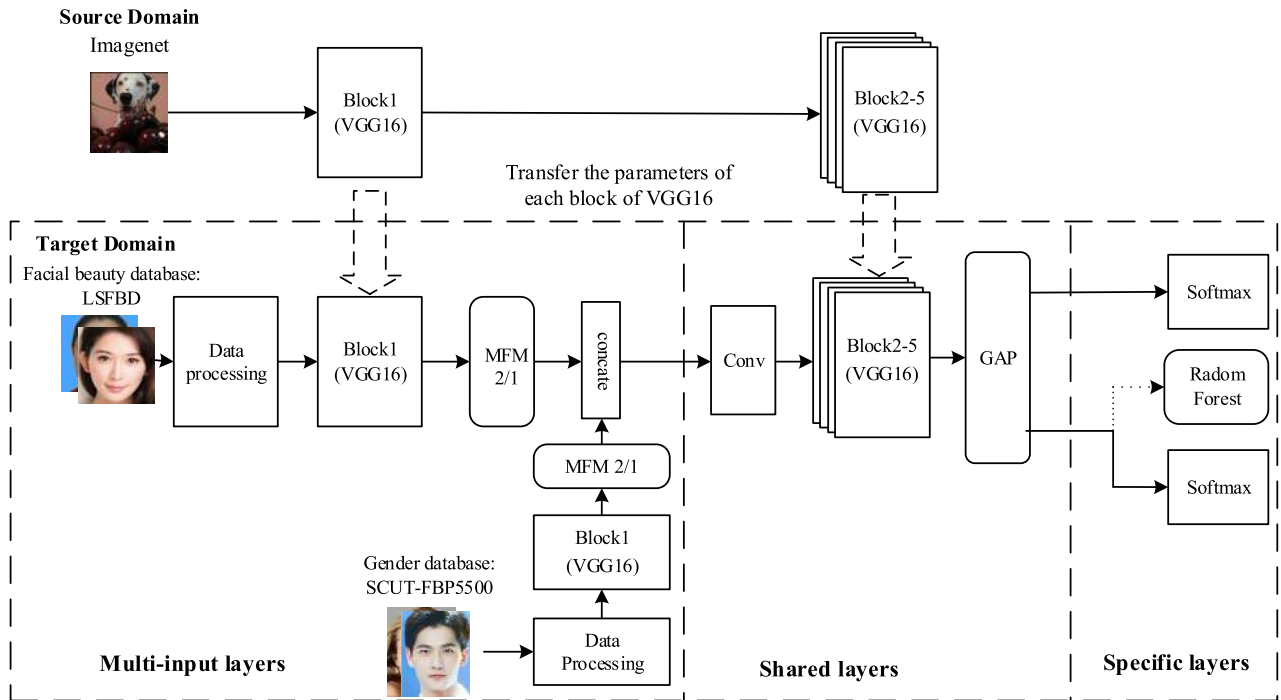
FIGURE 1. Framework of the proposed method.

reduced [11], and the computation speed is improved [12]. Gao *et al.* [13] proposed a novel multi-task network consisting of facial beauty prediction and landmark detection. The best correlation score is up to 0.92 on the SCUT-FBP benchmark. However, the related task of FBP in their framework is landmark detection, ignoring the influence of other facial attributes, such as gender, emotion, and age, *etc.*

It is worth noting that some studies show smiling faces are more attractive [14], [15]. Therefore, we consider using other face attributes as auxiliary tasks of FBP for MTL. More recently, multi-task network structures are mostly single input and multiple outputs. It is suitable for inputting a dataset with multiple labels, which is called multi-label learning. Unfortunately, LSFBD has a facial beauty label simply but does not include other facial attribute labels. Hence, our MTL researching on FBP and other facial attribute tasks cannot use multi-label learning, but we can perform multi-task researching on different facial attribute tasks from different databases. For this reason, combined related tasks of FBP with transfer learning, we propose a multi-input and multi-output network to improve the classification accuracy. As is displayed in Fig.1, it manifests the whole framework of our proposed method. Our major contributions are as follows:

(i). We propose a novel network for MTL. 2M BeautyNet was designed for FBP and its auxiliary tasks. We research FBP and gender recognition based on multi-task learning. Meanwhile, transfer learning is combined to extract the depth sharing features of faces. Thereby, the shallow features of main task are enriched, and the classification accuracy is improved.

(ii). We utilize a multi-task loss weights automatic learning strategy. Combined with multi-task loss weights automatic learning strategy, 2M BeautyNet avoids the phenomenon, in which the model has one task that dominates the entire loss, while the other tasks cannot influence the learning process of the shared layer.

(iii). We combine the traditional method with deep learning method. After multi-task training, the softmax classifier is replaced by a random forest for classification. Finally, the result is improved by about 1%.

The remainder of this paper here is as follows: section 2, reviews some related work of FBP and multi-task transfer learning. Section 3 describes 2M BeautyNet architecture, and multi-task loss weights automatic learning strategy. Then, we present the experimental platform and implemental details for our paper and analyze experimental results in Section 4. Section 5 will conclude our works.

II. RELATED WORKS
A. FACIAL BEAUTY PREDICTION
The research on FBP has gone through three stages, facial beauty attractive hypothesis verification, handcrafted feature classification, and depth apparent feature classification. At first, people's understanding of facial beauty mainly focused on a series of facial beauty attractive hypotheses and aesthetic research elements proposed by psychology and biology. The facial beauty attractive hypothesis includes averageness hypothesis, evolutionary hypothesis, symmetry hypothesis, sexual dimorphism, etc. Aesthetic research elements include

three courts and five eyes, golden ratio, etc [16]–[18]. Later, people began to extract handcrafted features of face images for FBP, such as geometric features and texture features. In recent years, using machine learning and computer vision technology to analyze facial beauty will empower machines to judge beauty, which has become another emerging research topic in the field of artificial intelligence [19], [20].

In general, FBP usually was regarded as a classification or regression problem. This process is usually divided into feature extraction and feature classification/regression. In early studies, researchers pay great efforts to use different machine learning algorithms and design various features. Eisenthal *et al.* [21] measured 92 frontal face images for geometric feature point distance and ratio, using SVM and KNN algorithm for facial beauty evaluation. Gunes *et al.* [22] used a 15-dimensional distance vector and 13-dimensional distance ratio vector for geometric features and training the classifier with C4.5 decision tree. Gray *et al.* [23] were the first to use texture features to analyze facial beauty. The face image is filtered by 48 filters, then different face image features are extracted by changing the image resolution. Although these methods can achieve certain results, handcrafted features lack universality and are low-level features that contain deeper levels of facial beauty information.

With the spread of up-to-data deep learning methods, higher-level features learned from CNNs have been applied for facial beauty computation tasks. Our group utilized a deep self-learning and Convolution Restricted Boltzmann Machines (CRBM) to predict facial beauty [24]. A set of works used the pre-trained VGG-16 to extract the depth features of facial beauty prediction [25], [26]. However, simply using the trained network as the feature extractor cannot achieve the superior effect. Xu *et al.* [27] used a new deep cascaded fine-tuning scheme, which is using various face image inputs to fine-tune the network respectively. These works indicate that the hidden layers of the deep model can learn useful facial features. Before long, a psychologically inspired convolutional neural network (PI-CNN) for FBP was proposed [28]. To solve the problem of lack of labeled data and distinguishing features in FBP, Liu *et al.* [29] fused depth features and geometric features for the first time, then an end-to-end label distribution learning (LDL) framework are constructed. Shi *et al.* [30] used a co-attention learning mechanism to improve the importance of different regions and different facial components, and pixel-wise labeling masks were seen as the meta information of the face. Whether based on features or other means, these works are single-task learning, ignoring the correlation between tasks.

B. MULTI-TASK TRANSFER LEARNING

Multi-task learning can be seen as a type of transfer learning, which was first proposed by [31]. It makes use of shared information between complementary tasks to enhance the generalization ability and the learning and recognition ability. In deep multi-task learning (DMTL), MTL typically shares

parameters between hidden layers. The sharing mechanism is divided into hard sharing and soft sharing. Hard sharing is the most common mechanism in deep multi-task learning. All tasks share the structure and parameters, but the output layer is respective. Baxter *et al.* [32] proved that hard sharing mechanism can effectively reduce the risk of over-fitting. However, in the soft sharing mechanism, each task has its structure and parameters, which guarantees the similarity of parameters through regularizing the distance of parameters between models [33], [34]. In traditional multi-task learning, soft sharing is largely influenced by regularization technology. Recently, there is a growing wave of research in multi-task learning, by which diverse computer vision tasks are solved. Lu *et al.* [35] proposed a hierarchical multi-task network (HMTNet), which can achieve the effect of simultaneously identifying a person's gender, race, and facial attractiveness from a given portrait image. A DMTL model for keypoint detection, face detection, posture estimation was proposed by [36]. In MTL, the challenge of each task is different, and the loss weight of each task in these methods is a fixed value set according to experience. Guo *et al.* [37] proposed a dynamic task priority method, utilizing the indicators of performance to determine the difficulty level of each task. If the loss value of task was accumulated directly, a task will converge but the other will be screw up. Therefore, Alex *et al.* [10] used uncertainly to measure the loss in multi-task learning. To a certain extent, these efforts can avoid the phenomenon of one task dominating the whole loss.

The purpose of transfer learning is to improve the learning efficiency of the target domain. It transfers the source domain knowledge to the target domain [38]. The easiest way for transfer learning is that target domain uses the weight of source domain. Because the pre-trained model already contains a lot of basic information, it enriches the low-level features of the target task and improves the model learning performance. For the problem of heterogeneous unsupervised domain adaptation, a shared fuzzy equivalence relation (SFER) method was proposed [39]. Based on distribution adaptation, Jiang *et al.* [40] proposed a multi-label metric transfer learning (MLMTL). Recently, Tan *et al.* [41] studied a novel transfer learning problem termed Distant Domain Transfer Learning (DDTL). The differences can exist between target domain and source domain. Xin *et al.* [42] used transfer learning for smile detection and fine-tuned the face recognition model with different inputs, which effectively improved the performance of smile detection. Also, Amira *et al.* [43] applied transfer learning to medical research and proposed a segmentation recommender for skin lesion extraction. Lv *et al.* [44] proposed an unsupervised incremental learning algorithm, which is realized by transferring the pedestrian's spatio-temporal pattern in the target domain.

Multi-task transfer learning method was used in our works. By the way, not only the low-level information from in the source domain but also the shared information of auxiliary tasks can be used. A host of experiments demonstrated that multi-task transfer learning can improve the classification
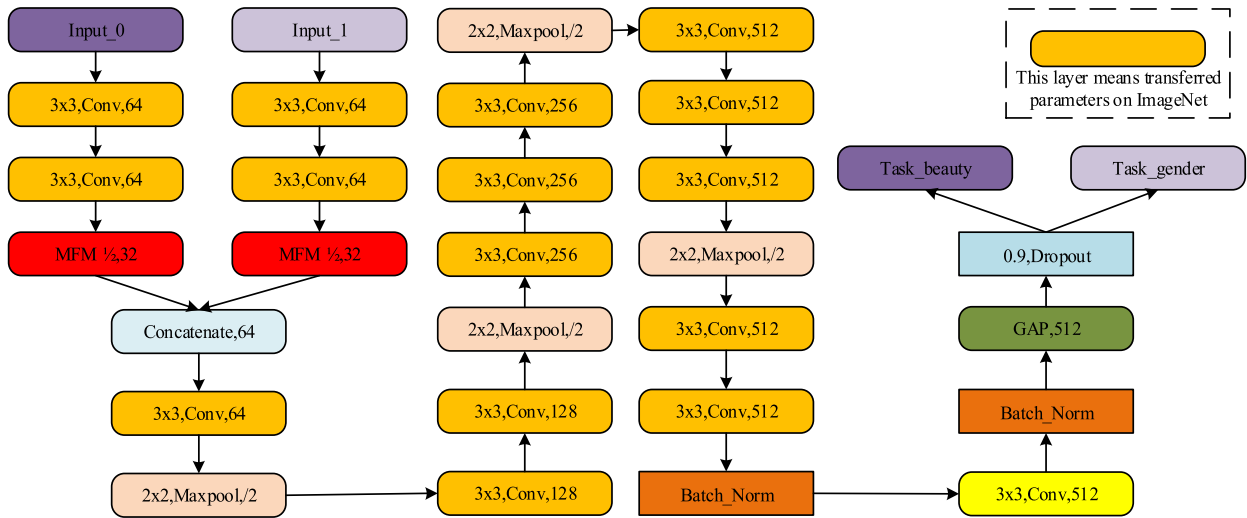
FIGURE 2. The architecture of 2M BeautyNet.

accuracy of FBP, and alleviate the over-fitting phenomenon to some extent.

III. PROPOSED METHODS

A. 2M BEAUTYNET ARCHITECTURE

Most of the existing MTL network structures used multiple labels of one image to jointly train the network with a single input and multiple outputs. At this stage, the basic multi-task network is usually a single input. What's different is that we combine the related tasks with FBP to construct a network structure with multiple inputs. Here we will give two reasons for designing this network structure. On the one hand, we study facial beauty prediction on LSFBD with only a single beautiful degree label. The commonly used single-input and multi-output networks are not suitable for our multi-task research. On the other hand, few facial beauty sample training depth networks are prone to over-fitting. If the network structure with multi-input multi-task is used, the input data of related tasks can make up for the defect of insufficient training samples to some extent.

Fig.2 illustrates the architecture of 2M BeautyNet, and the parameter setting is shown in Table.1. Our model consists of multi-input layers, shared layers and specific layers. Among them, multi-input layers include the first block of VGG16 (2 convolution layers) and Max Feature Map (MFM) layer. It can transfer existing CNN parameters and learn tasks from different databases. MFM reduces the dimensions of each branch, and finally recombines the feature maps to retraining them. The shared layer is composed of the other block of VGG16, and the specific layers consist of Global Average Pooling (GAP) layers and two classifiers. The input image of the network is RGB images from SCUT-FBP5500 and LSFBD. Firstly, the images of Input_0 and Input_1 are input to the multi-input layers for training respectively. Secondly, the respective feature images are input into the shared layer

TABLE 1. The structural details of 2M BeautyNet.

Name	Filter Size/Stride	Output Shape	Parameters
Input_0	-	128×128×3	-
Block0_conv1	3×3/1	128×128×64	1792
Block0_conv2	3×3/1	128×128×64	36928
MFM0	-	128×128×32	-
Input_1	-	128×128×3	-
Block1_conv1	3×3/1	128×128×64	1792
Block1_conv2	3×3/1	128×128×64	36928
MFM1	-	128×128×32	-
Conv1	3×3/1	128×128×64	36928
Maxpool1	2×2/2	64×64×64	-
Block2_conv1	3×3/1	64×64×128	73856
Block2_conv2	3×3/1	64×64×128	147584
Block2_pool	2×2/2	32×32×128	-
Block3_conv1	3×3/1	32×32×256	295168
Block3_conv2	3×3/1	32×32×256	590080
Block3_conv3	3×3/1	32×32×256	590080
Block3_pool	2×2/2	16×16×256	-
Block4_conv1	3×3/1	16×16×512	1180160
Block4_conv2	3×3/1	16×16×512	2359808
Block4_conv3	3×3/1	16×16×512	2359808
Block4_pool	2×2/2	8×8×512	-
Block5_conv1	3×3/1	8×8×512	2359808
Block5_conv2	3×3/1	8×8×512	2359808
BN1	-	8×8×512	2048
Block5_conv3	3×3/1	8×8×512	2359808
BN2	-	8×8×512	2048
GAP	-	1×1×512	-
Task_beauty	-	1×1×5	2565
Task_gender	-	1×1×2	1026
Total			14798023

for training at the same time. Finally, the two classifiers are used to classify FBP and gender recognition. Among them, Input_0 and Input_1 represent facial beauty prediction images and gender images, respectively. Task_beauty is the facial beauty prediction classifier, while Task_gender is the gender recognition classifier.
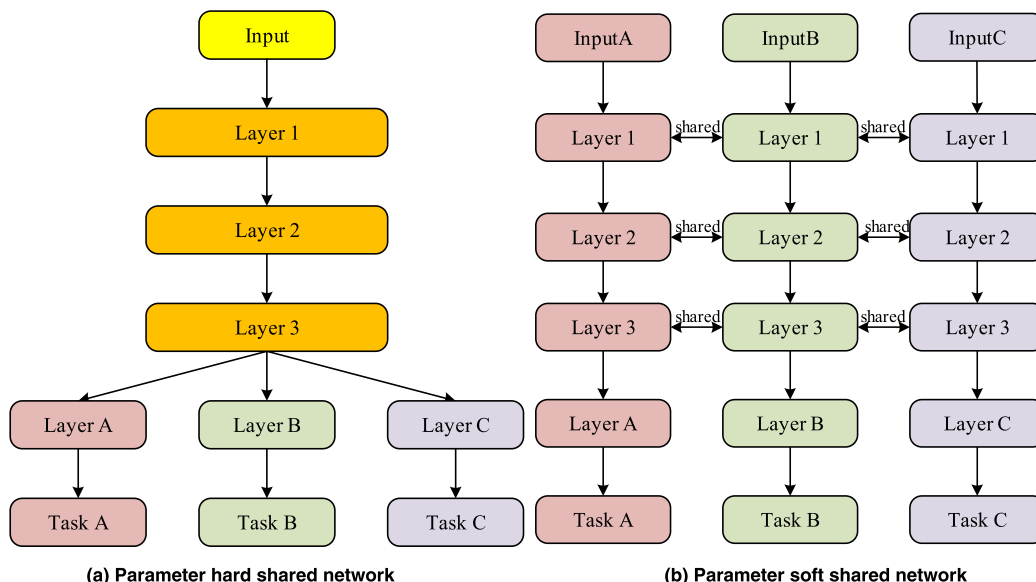
(a) Parameter hard shared network   (b) Parameter soft shared network

FIGURE 3. Basic multi-task network of soft sharing and hard sharing.
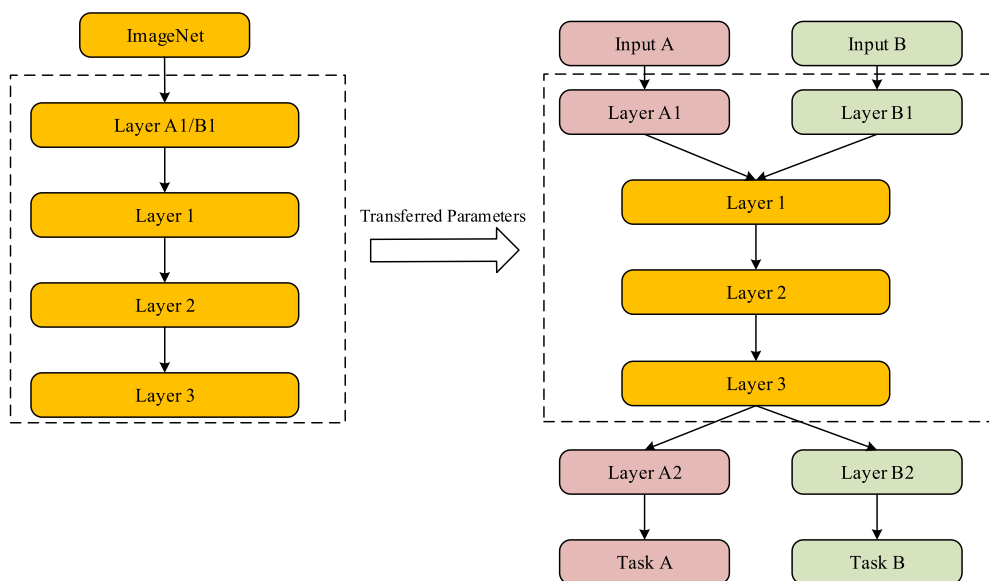

FIGURE 4. The proposed multi-input multi-output network.

B. MULTI-INPUT MULTI-OUTPUT NETWORK

In this section, we will reveal two basic multi-task networks that are based on hard sharing and soft sharing. Fig.3 (a) is a simple parameter hard shared multi-task network, and parameter soft shared network is shown in Fig.3 (b). Based on parameter hard shared multi-task network, we designed multi-input multi-task network for tasks from different databases to ensure transferred parameters during multiple tasks training. Fig.4 shows multi-input multi-task network. All tasks of parameter hard shared network share a bottom layer. The shared part learns the shared representation of multiple tasks, which have strong abstraction capabilities.

It is adaptable to several different but related target tasks, often enabling the better generalization of the main tasks in MTL. However, as shown in Fig.3 (b), each task of parameter soft shared model has its own underlying. All parameters or partial parameters can be shared by designing a shared structure between the underlying layers. Based on parameter hard shared model, this paper designs 2M BeautyNet in combination with related tasks from different databases.

Researches have shown that transfer learning can effectively solve the problem of insufficient training samples [9], [24]. For multi-task transfer learning, we transferred the trained parameters of VGG16 from the ImageNet database

to different layers of our multi-input multi-task network for retraining. The details are described in Section 3.3. Next, we will illustrate why we design multi-input multi-task network.

Since the output size of the transfer network convolution layer is fixed, parameter transfer can only be performed if the input data dimension is the same as transfer layer. However, in the case of multiple inputs, the dimensions of the input data are usually not guaranteed to be the same. To solve this problem, a multi-input multi-task network that contains a multiple inputs layer is designed. We use VGG16 as the basic transfer network, where input data shape is (128, 128, 3). If the two input data of beauty and gender are directly concatenated, the input shape of the first convolution layer will be changed to (128, 128, 6), and it will not be able to transfer the trained parameters. Therefore, we have designed a multi-input layer with two branches. The output feature map type of the two branches after passing through the convolutional layer is (128, 128, 64). If you want to continue to transfer the parameters of the layers behind VGG16, you must ensure that the data type, which is input to the next convolutional layer, is (128, 128, 64). But the data type of the two branches directly concatenated is (128, 128, 128). Therefore, we reduce the dimension through MFM activation function instead of 1×1 convolution kernel. The first reason is that Relu function will cause the collapse of low-dimensional data. The probability of a low-dimensional feature distribution on Relu function activation band is small, and the information is damaged severely after the feature map passes through Relu layer. Therefore, we use MFM to process low-dimensional feature maps. MFM activation function suppresses the low-activation neurons in each layer through competitive relationships and has advantages of compact features and reduced parameters. If convolution kernel is used for convolution, not only the amount of model parameters is increased, but also subsequent parameter transfer is hindered. Another reason is that MFM adopts a split-polymerization method, which can achieve compact feature representation and variable selection, dimensionality reduction and sparse gradient function. We assume that the convolutional layer of the output is $C \in R^{h \times w \times 2n}$, dividing the convolutional layer into two. MFM 1/2 activation function is expressed as

$$f_{i,j}^k = \max_{1 < k < n} \left(C_{i,j}^k, C_{i,j}^{k+n} \right) \qquad (1)$$

where the channel of the input convolution layer is $2n$, $1 < i < h$, $1 < j < w$. The gradient of (1) takes the following form,

$$\frac{\partial f}{\partial C^{k'}} = \begin{cases} 1, & C_{ij}^k \geq C_{ij}^{k+n} \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

Among them, $1 \leq k' \leq 2n$, and

$$k = \begin{cases} k', & 1 \leq k' \leq n \\ k' - n, & n + 1 \leq k' \leq 2n \end{cases} \qquad (3)$$

Usually, the activation amount of MFM 1/2 is 50%. Similar network structures can be used if the network input

comes from three different databases. We just change the activation function to MFM 2/3. Some studies have proved that the activation performance of MFM 2/3 is better than MFM 1/2 [10], [45].

C. AUTOMATIC LEARN WEIGHTS OF MULTI-TASK LOSS

In MTL, joint optimization between multiple tasks may make a simple single task converge due to data imbalances and diverse difficulty in tasks. If we add multiple losses directly, one task tends to be better and the other tasks are bad. To solve this problem, Alex *et al.* [10] propose that the weight of multiple losses can be obtained by way of uncertainty. The effectiveness of the method was verified on multiple regression tasks, but it was not verified on the classification task. We verify this method in two classification tasks.

We assume that multi-task prediction error obeys Gaussian distribution, that is, the mean is 0 and the variance is σ^2. Based on the maximizing Gaussian likelihood function with homoscedastic uncertainty, the multi-task loss function is adjusted. Let $f^{\mathbf{W}}(\mathbf{x})$ be the output of a neural network with weights $\mathbf{W}$ on input $\mathbf{x}$. For the classification task, we usually pass the model output through a Softmax classification function. And the final output is expressed as

$$p\left(\mathbf{y}|f^{\mathbf{W}}(\mathbf{x}), \sigma\right) = \text{Soft max}\left(\frac{1}{\sigma^2} f^{\mathbf{W}}(\mathbf{x})\right) \qquad (4)$$

where $\mathbf{y}$ represents the output of the model and σ is a positive scalar. This scalar is either fixed or can be learned, and the Softmax input is scaled by σ^2. The log-likelihood for this output can be written as

$$\log p\left(\mathbf{y} = c|f^{\mathbf{W}}(\mathbf{x}), \sigma\right) = \frac{1}{\sigma^2} f_c^{\mathbf{W}}(\mathbf{x}) - \log \sum_{c'} \exp\left(\frac{1}{\sigma^2} f_{c'}^{\mathbf{W}}(\mathbf{x})\right) \qquad (5)$$

In the case of multiple outputs, we assume that each task is independently and equally distributed. The maximum likelihood estimation of multiple tasks can be expressed as

$$p\left(\mathbf{y}_1, \ldots, \mathbf{y}_j|f^{\mathbf{W}}(\mathbf{x}), \sigma\right) = p\left(\mathbf{y}_1|f^{\mathbf{W}}(\mathbf{x}), \sigma\right) \ldots p\left(\mathbf{y}_j|f^{\mathbf{W}}(\mathbf{x}), \sigma\right) \qquad (6)$$

where $\mathbf{y}_1, \ldots, \mathbf{y}_j$ means that the model has j outputs, which means that there are j tasks. In this paper, FBP is the main task, and the auxiliary is gender recognition. We take the log for maximum likelihood estimation of the two tasks and get the minimization objective function of our multi-output model, represented by

$$L\left(\mathbf{W}, \sigma_1, \sigma_2\right)$$
$$= -\log p\left(y_1, y_2 = c|f^{\mathbf{w}}(\mathbf{x})\right)$$
$$= -\log(\text{Softmax}(y_1 = c_1; f^{\mathbf{w}}(\mathbf{x}), \sigma_1)$$
$$\cdot \text{Softmax}(y_2 = c_2; f^{\mathbf{w}}(\mathbf{x}), \sigma_2))$$
$$= -\left[\log p\left(\mathbf{y}_1 = c_1|f^{\mathbf{W}}(\mathbf{x}), \sigma_1\right) + \log p\left(\mathbf{y}_2 = c_2|f^{\mathbf{W}}(\mathbf{x}), \sigma_2\right)\right]$$

FIGURE 5. Some female examples of LSFBD.

$$= - \left[\frac{1}{\sigma_1^2} f_{c_1}^{\mathbf{w}}(\boldsymbol{x}) - \log \sum_{c'} \exp \left(\frac{1}{\sigma_1^2} f_{c'}^{\mathbf{W}}(\boldsymbol{x}) \right) + \frac{1}{\sigma_2^2} f_{c_2}^{\mathbf{w}}(\boldsymbol{x}) \right.$$

$$\left. - \log \sum_{c'} \exp \left(\frac{1}{\sigma_2^2} f_{c'}^{\mathbf{W}}(\boldsymbol{x}) \right) \right]$$

$$\approx \frac{1}{\sigma_1^2} L_1(\mathbf{W}) + \frac{1}{\sigma_2^2} L_2(\mathbf{W}) + \log \sigma_1 + \log \sigma_2 \quad (7)$$

where $L_1(\mathbf{W}) = f_{c_1}^{\mathbf{w}}(\boldsymbol{x})$, $L_2(\mathbf{W}) = f_{c_2}^{\mathbf{w}}(\boldsymbol{x})$. The weight of each task is $1/\sigma_1^2$ and $1/\sigma_2^2$. If σ is larger, the contribution of the corresponding task in the training process is greater.

D. RANDOM FOREST CLASSIFIER

Leo Breiman proposed a classifier called random forest [46]. Its principle is to use an army of decision trees to train and predict samples. It has the advantage of giving an importance score for each variable and evaluating the role of each variable in the classification when classifying the data. It's worth noting that bagging is an important idea in random forests, and the main principle of implementation is bootstrap sampling. That is to say, each decision tree in random forests is a classifier, and if there are n trees there will be n classification results. The final classification results are obtained by considering all the classifiers. By voting on all the results of the classification in random forests, the output with the most votes is selected. The implementation process of random forest can be summarized as follows:

Step 1: Suppose the number of original training sets is S. Then, we can randomly select k new bootstrap sample sets from S, and construct k classification trees. Finally, each time the sample that has not been extracted constitutes k out-of-bag data.

Step 2: Assuming that the feature dimension of each sample is M, specify a constant m, and satisfy the condition $m \ll M$. Then, the feature subset of m-dimension is randomly selected from the M-dimension feature. Finally, each time the tree is split, the best feature is selected from m feature subsets.

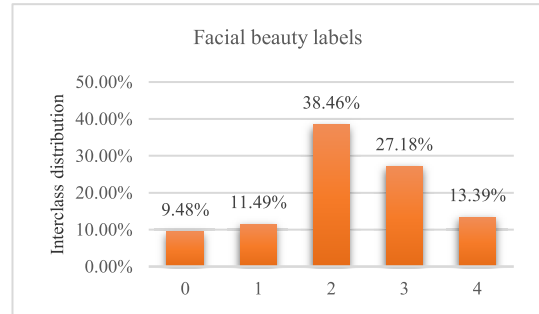*Step 3:* Every tree grows to its fullest extent and there is no pruning process.

FIGURE 6. The distribution of LSFBD.

Step 4: The generated multiple classification trees form a random forest, and we can use this random forest classifier to recognize and classify the new.

IV. EXPERIMENTS ANS ANALYSIS

We implement our method with Keras on an Ubuntu server with GTX1080 GPU, i3-7350k CPU, and 48G memory. In this paper, we use different classifiers in the skict_learn toolkit to achieve the best classification by adjusting some of the parameters in the function. Because 2M BeautyNet structure is improved based on VGG16. Then all the single-task experiments are based on VGG16, and the multi-task experiments are based on 2M BeautyNet.

A. DATABASES

1) LSFBD

LSFBD consists of 20,000 labeled images (10000 male images and 10000 female images) and 80000 unlabeled images with the resolution of 144×144. Moreover, because of the diversity and inhomogeneity of image quality, it's difficult to predict facial beauty. The images were divided into five categories, labeled as "0", "1", "2", "3" and "4". These labels correspond to the words "extremely unattractive", "unattractive", "averages", "attractive" and "most attractive". This paper focuses on the prediction of female beauty, and only 10,000 LSFBD female images were used to verify the effectiveness of our overall framework for the FBP task. 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". Fig.5 shows some examples of LSFBD, and the distribution histogram of LSFBD is illustrated in Fig.6.

2) SCUT-FBP5500

SCUT-FBP5500 database contains 5,500 high-resolution frontal face images with different races, gender and age. Each image is labeled with a beauty score ranging from 1 to 5, and the larger the score, the more attractive it is. These scores were assessed by 60 volunteers. They named facial images of male and female, Asians and Caucasians in different ways. So we can analyze the naming of each image to get a gender label, with "0" for males and "1" for females. The face image quality of this database is excellent, which is beneficial
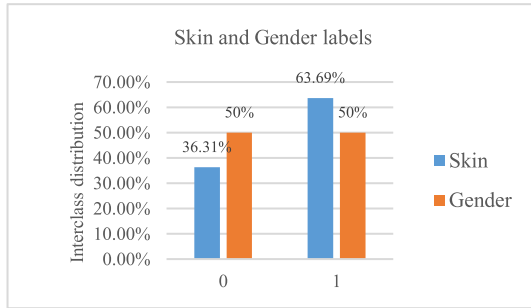
FIGURE 7. The distribution of skin and gender.

FIGURE 8. Examples of MTCNN for face detection (The top row is original images, the bottom row is cropped faces using MTCNN).

to multi-task shared feature learning, so we take this database as the input data of gender recognition. Large-scale Celeb Faces Attributes (CelebA) Dataset [47] is a large scale face attributes dataset with 40 attribute annotations. We use skin labels in the large scale face attribute, with "0" for white and "1" for black. Among them, class "0" contains 400 male and female images, class "1" contains 7017 images. We calculated the distribution of gender and skin labels, and the distribution between categories is shown in Fig. 7.

B. DATA PREPROCESSING AND AUGMENTATION

For our training data, we use different data augmentation methods to enlarge the sample size of the training set. We used three main methods to extend the two databases, such as randomly cropping, randomly flipping images from left to right and randomly rotating images from $-45°$ to $25°$. Since the SCUT-FBP5500 database has fewer samples than LSFBD, we use more types of data enhancement methods. To perform multi-task training normally, we require the training and verification samples of two databases with the same order of magnitude. In our experiment, we increased the number of samples of two training sets to 36,000. The verification set does not perform data enhancement, and the number of samples is 996. Our network only works with face, so we use Multi-task Cascaded Convolutional Networks (MTCNN) [48] to detect faces in each image. Some examples are shown in Fig. 8.

C. IMPLEMENTATION DETAIL

Our experiment combines two tasks: facial beauty prediction and gender recognition, gender recognition and skin classification. We have LSFBD as facial beauty prediction, SCUT-FBP5500 as gender recognition and CelebA as skin classification. Our 2M BeautyNet is implemented via Keras. During the training process, the initial weights of the network structure are transferred from the pre-trained weights of VGG16 on the Imagenet database. The Block1 weights of two branches in the single-task layer are the Block1 weights of VGG16. Then, after the MFM activation function, concatenate layer, and convolution operations, the feature map for each task is jointly trained. The convolutional layer weights of Block2-Block5 have corresponded to that of VGG16. In all of our tables, TL represents transfer learning and CA represents classification accuracy.

In the experiments, we use Adam optimizer to train the network for 50 epochs with batch size of 64, the initialized learning rate of 0.001, and momentum of (0.9, 0.999). Besides, this experiment uses the monitor in Keras to monitor the verification loss in the FBP task. When the verification loss does not decrease within two epochs, the current learning rate is changed to 0.1 times. Finally, we use the skict_learn toolkit to verify the performance of different classifiers.

D. COMPARISON BETWEEN SINGLE TASK AND MULTIPLE TASKS

Many studies have shown that MTL is better than STL. On the one hand, multi-task network structure contains multiple tasks shared layers. By learning the shared representation of multiple tasks, the network has strong abstraction ability. The main task can achieve better generalization capabilities when it is trained by combining several different but related target tasks. On the other hand, a specific independent layer structure is designed for each different task to learn how to use a shared representation to improve the performance of each specific task. In this paper, gender recognition is regarded as an auxiliary task for FBP. And Table. 2 compares the performance of single-task and multi-task with multi-task loss weights automatic learning strategy and softmax classifier. Meanwhile, we also compare the classification accuracy of using transfer learning. Furthermore, in order to illustrate the effectiveness of the proposed method, we conducted another experiment, which is joint gender recognition and skin classification for multi-task training. The classification accuracy of FBP, gender recognition, and skin classification tasks during training alone are showed in Table. 2. In addition, we also showed the classification accuracy of multi-task training in two experiments: FBP and gender recognition, gender recognition and skin classification.

In experiments, the network framework of STL uses VGG16, while the network framework of MTL adopts our proposed 2M BeautyNet. Whether for single-task or multi-task training, the trained VGG16 parameters on ImageNet of each layer are transferred to the corresponding layers of our network. From Table. 2, we can see that the classification accuracy of STL is slightly worse than MTL, and transferring ImageNet weights is better than direct training. To evaluate the generalization performance of our 2M BeatyNet,

TABLE 2. Classification accuracy Of single task and multiple tasks.

Task number	Task	TL	CA (%)
Single	Beauty	Yes	64.36
		No	39.00
	Gender	Yes	99.00
		No	98.09
	Skin	Yes	89.66
		No	89.96
Multiple	Gender	Yes	98.59
	Skin	Yes	90.26
	Gender	Yes	96.35
	Beauty	Yes	**66.82**

TABLE 3. Classification accuracy by combining multi-Task loss weights automatic learning strategy.

Task Set	Loss of weight	CA (%)
Multi-task-Beauty	Auto	68.23
	Fixed	66.98
Multi-task-Gender	Auto	97.43
	Fixed	97.10

we change the multiple tasks of FBP and gender recognition to skin detection and gender recognition. Experiments show that our method is better than a single task, the classification accuracy of facial beauty prediction is up to 66.82%.

E. THE IMPACT OF MULTI-TASK LOSS AUTOMATIC LEARNING STRATEGY

In MTL, joint optimization between multiple tasks may make simple single task converge due to data imbalances and diverse difficulty in tasks. If we add multiple task losses directly, one task tends to be better and the others are bad. We also compared the effects of different loss on multi-task performance, especially the manually setting fixed value and automatically learning value. Multi-task loss weight automatic learning can derive a reasonable multi-task loss function that can learn to balance various losses. The impact of multi-task loss automatic learning strategy on task performance is shown in Table. 3.

If the experiment does not adopt the method of multi-task loss weights automatic learning strategy, the default task weight ratio is 1:1. Through experiments, if the losses of multiple tasks are directly added and optimized, the model will converge to tasks with fewer classification categories. For example, if five-class FBP and two-class gender recognition are jointly optimized, the model converges on gender recognition. Five-class FBP and seven-class emotion recognition are jointly optimized, the model converges on FBP. However, if multi-task loss automatic learning strategy is adopted, both tasks will slowly converge. It can be seen from Table. 3 that FBP classification accuracy rate with multi-task loss automatic learning strategy is increased by 2%-3% compared with the fixed ratio 1:1. For FBP, the best classification accuracy is 68.23%.

TABLE 4. Classification accuracy of different classifiers.

Task Set	Classifier	CA (%)
Single-Beauty	Random Forest	65.36
	Softmax	64.52
Multi-task-Beauty	Random Forest	68.23
	Softmax	66.82

TABLE 5. Comparison Of FBP classification accuracy on SCUT-FBP5500.

Method	PC
AlexNet[8]	0.8634
PI-CNN [28]	0.86
ResNet-18 based AaNet[5]	0.9055
ResNet-18 based P-AaNet[5]	0.8965
2M BeautyNet(ours)	0.8996

F. THE IMPACT OF RANDOM FOREST CLASSIFIER

The existing FBP methods usually regard a deep neural network as a feature extractor to extract features from the final output layer, and then use softmax for classification. However, random forests are a supervised integrated learning model that aggregates multiple machine learning models to make overall performance better. It has the advantage of being parallelizable and reducing over-fitting. Thus, we also compare the influence of different classifiers on the classification accuracy of FBP. The FBP classification accuracy of the softmax classifier, random forest classifier, are shown in Table. 4.

In the experiment, we first use 2M BeautyNet and softmax classifier to train multiple tasks, the feature expression with higher accuracy can be extracted, and finally softmax is replaced by random forests for classification. As shown in Table. 4, we found that the classification accuracy of replacing the Softmax classifier by random forest is increased by about 1% to 2%.

G. COMPARISON OF DIFFERENT METHODS

In order to verify the validity of the 2M BeautyNet, we compare its performance with other existing algorithms on LSFBD, such as one traditional method with four basic CNN models. The detailed comparison is shown in Table. 6. The first part shows the classification accuracy of LSFBD by the traditional method of FBP [9]. The second part compares the classification accuracy of different networks designed by deep learning, including NIN [49], GoogleNet [50], VGG16 [51] and BeautyNet [10]. As we can see, the classification accuracy of deep learning methods is higher than traditional methods. At the same time, we also compare the performances of these four networks after the transfer, in which the accuracy of facial beauty classification is generally increased by 2%-5%. The above experiments are all about FBP single task, and the highest accuracy was 67.48%. However, the last part compares the accuracy of 2M BeautyNet combined with different strategies, including transfer learning,

TABLE 6. Comparison Of FBP pearson correlation among other methods.

Method	TL	CA (%)
CRBM[9]	No	51.62
K-means[9]	No	52.54
Multiscale K-means[9]	No	55.07
NIN [50]	Yes	58.30
GoogleNet[51]	Yes	62.59
VGG16 [52]	Yes	62.29
BeautyNet[10]	Yes	67.48
2M BeautyNet+Radom Forest	No	66.34
2M BeautyNet+Radom Forest	Yes	66.98
2M BeautyNet+Auto loss	No	65.59
2M BeautyNet+Auto loss	Yes	66.82
2M BeautyNet+Radom Foresst+Auto loss	No	67.14
2M BeautyNet+Radom Foresst+Auto loss	Yes	**68.23**

multi-task loss automatic learning strategy, and random forest. In summary, the classification accuracy of our method is better than others, and the highest classification accuracy is 68.23%. In addition, we verified our proposed network on SCUT-FBP5500. As shown in Table 5, our network has achieved similar results. However, 2M BeautyNet is based on VGG16 with relatively few parameters.

V. CONCLUSION

In this work, we have proposed an effective multi-input multi-task 2M BeautyNet to jointly learn two tasks including facial beauty prediction and gender recognition. This network can properly transfer pre-trained network parameters when inputting from different databases simultaneously. Different from the single-task network structure in FBP, 2M BeautyNet combines task-assisted information from the other databases related to FBP to improve the performance of the main task. Besides, transfer learning is combined to extract the depth sharing features of faces. Multi-task transfer learning can enrich the low-level features of the main task and improve classification accuracy. Then we utilize multi-task loss automatic learning strategy to avoid the phenomenon, in which the model has one task that dominates the entire loss, while the other tasks cannot influence the learning process of multi-task shared layer. Finally, we combine the traditional method with deep learning method. After multi-task training, the softmax classifier is replaced by a random forest for classification. The extensive experiments on LSFBD and SCUT-FBP5000 show that our method achieves superior accuracy over the other methods in facial beauty prediction task, which is up to 68.23%.

In the future, we will find the tasks more relevant to FBP for multi-task training by measuring the correlation of the tasks. Considering how to design a more versatile and effective multi-input multi-task network, and combining the local information and the other factors that influence facial
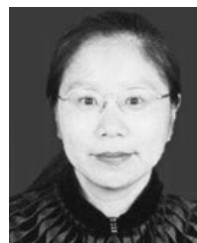
beauty in multi-task learning, we believe that this area will advance the progress of aesthetic surgery planning, cosmetic recommendation, and facial beautification, *etc.*

REFERENCES

[1] T. Alashkar, S. Jiang, and Y. Fu, "Rule-based facial makeup recommendation system," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 325–330.

[2] A. Laurentini and A. Bottino, "Computer analysis of face beauty: A survey," *Comput. Vis. Image Understand.*, vol. 125, pp. 184–199, Aug. 2014.

[3] Y. Liu, Z. Xie, X. Yuan, J. Chen, and W. Song, "Multi-level structured hybrid forest for joint head detection and pose estimation," *Neurocomputing*, vol. 266, pp. 206–215, Nov. 2017.

[4] L. Liang, L. Jin, and X. Li, "Facial skin beautification using adaptive region-aware masks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2600–2612, Dec. 2014.

[5] L. Lin, L. Liang, L. Jin, and W. Chen, "Attribute-aware convolutional neural networks for facial beauty prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 847–853.

[6] 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.

[7] Y. Zhai, Y. Huang, Y. Xu, J. Zeng, F. Yu, and J. Gan, "Benchmark of a large scale database for facial beauty prediction," in *Proc. Int. Conf. Intell. Inf. Process. (ICIIP)*, 2016, pp. 131–135.

[8] Y. Zhai, H. Cao, W. Deng, J. Gan, V. Piuri, and J. Zeng, "BeautyNet: Joint multiscale CNN and transfer learning method for unconstrained facial beauty prediction," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–14, Jan. 2019.

[9] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 2–12.

[10] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491.

[11] I. Kokkinos, "UberNet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6129–6138.

[12] D. Neven, B. De Brabandere, and S. Georgoulis, "Fast scene understanding for autonomous driving," 2017, *arXiv:1708.02550*. [Online]. Available: https://arxiv.org/abs/1708.02550

[13] L. Gao, W. Li, Z. Huang, D. Huang, and Y. Wang, "Automatic facial attractiveness prediction by deep multi-task learning," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3592–3597.

[14] X. Liu and T. Li, "Understanding beauty via deep facial features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1–11.

[15] R. Hassin and Y. Trope, "Facing faces: Studies on the cognitive aspects of physiognomy," *J. Pers. Social Psychol.*, vol. 78, no. 5, pp. 837–852, Sep. 2005.

[16] J. H. Langlois and L. A. Roggman, "Attractive faces are only average," *Psychol. Sci.*, vol. 1, no. 2, pp. 115–121, Mar. 1990.

[17] J. P. Swaddle and I. C. Cuthill, "Asymmetry and human facial attractiveness: Symmetry may not always be beautiful," *Proc. Roy. Soc. London*, vol. 261, no. 1360, pp. 111–116, 1995.

[18] D. I. Perrett, K. J. Lee, I. Penton-Voak, D. Rowland, S. Yoshikawa, D. M. Burt, S. P. Henzi, D. L. Castles, and S. Akamatsu, "Effects of sexual dimorphism on facial attractiveness," *Nature*, vol. 394, no. 6696, pp. 884–887, Aug. 1998.

[19] Y. Ren and X. Geng, "Sense beauty by label distribution learning," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2648–2654.

[20] L. Lin, L. Liang, and L. Jin, "Regression guided by relative ranking using convolutional neural network (R3CNN) for facial beauty prediction," *IEEE Trans. Affective Comput.*, to be published.

[21] Y. Eisenthal, G. Dror, and E. Ruppin, "Facial attractiveness: Beauty and the machine," *Neural Comput.*, vol. 18, no. 1, pp. 119–142, Jan. 2006.

[22] H. Gunes and M. Piccardi, "Assessing facial beauty through proportion analysis by image processing and supervised learning," *Int. J. Hum.-Comput. Stud.*, vol. 64, no. 12, pp. 1184–1199, Dec. 2006.

[23] D. Gray, K. Yu, and W. Xu, "Predicting facial beauty without landmarks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 434–447.

[24] J. Gan, L. Li, Y. Zhai, and Y. Liu, "Deep self-taught learning for facial beauty prediction," *Neurocomputing*, vol. 144, pp. 295–303, Nov. 2014.

[25] R. Rothe, R. Timofte, and L. V. Gool, "Some like it hot-visual guidance for preference prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5553–5561.

[26] L. Xu, J. Xiang, and X. Yuan, "Transferring rich deep features for facial beauty prediction," 2018, *arXiv:1803.07253*. [Online]. Available: https://arxiv.org/abs/1803.07253

[27] J. Xu, L. Jin, and L. Liang, "A new humanlike facial attractiveness predictor with cascaded fine-tuning deep learning model," 2015, *arXiv:1511.02465*. [Online]. Available: https://arxiv.org/abs/1511.02465

[28] J. Xu, L. Jin, L. Liang, Z. Feng, D. Xie, and H. Mao, "Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN)," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 1657–1661.

[29] S. Liu, B. Li, Y.-Y. Fan, Z. Quo, and A. Samal, "Facial attractiveness computation by label distribution learning with deep CNN and geometric features," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1344–1349.

[30] S. Shi, F. Gao, X. Meng, X. Xu, and J. Zhu, "Improving facial attractiveness prediction via co-attention learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4045–4049.

[31] R. Caruana, "Multitask Learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.

[32] J. Baxter, "A Bayesian/information-theoretic model of learning to learn via multiple task sampling," *Mach. Learn.*, vol. 28, no. 1, pp. 7–39, 1997.

[33] L. Duong, T. Cohn, and S. Bird, "Low resource dependency parsing cross-lingual parameter sharing in a neural network parser," in *Proc. Assoc. Comput. Linguistics (ACL)*, 2015, pp. 845–850.

[34] Y. Yang, TM. Hospedales, "Trace norm regularized deep multi-task learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–4.

[35] L. Xu, H. Fan, and J. Xiang, "Hierarchical multi-task network for race, gender and facial attractiveness recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1–4.

[36] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019.

[37] M. Guo, A. Haque, and D.-A. Huang, "Dynamic task prioritization for multitask learning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 270–287.

[38] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[39] F. Liu, J. Lu, and G. Zhang, "Unsupervised heterogeneous domain adaptation via shared fuzzy equivalence relations," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3555–3568, Dec. 2018.

[40] S. Jiang, Y. Xu, T. Wang, H. Yang, S. Qiu, H. Yu, and H. Song, "Multi-label metric transfer learning jointly considering instance space and label space distribution divergence," *IEEE Access*, vol. 7, pp. 10362–10373, 2019.

[41] B. Tan, Y. Zhang, and S. J. Pan, "Distant domain transfer learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 2604–2610.

[42] X. Guo, L. Polania, and K. Barner, "Smile detection in the wild based on transfer learning," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 679–686.

[43] A. Soudani and W. Barhoumi, "An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction," *Expert Syst. Appl.*, vol. 118, pp. 400–410, Mar. 2019.

[44] J. Lv, W. Chen, Q. Li, and C. Yang, "Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7948–7956.

[45] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

[46] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[47] Z. Liu, P. Luo, and X. Wang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Jun. 2015, pp. 3730–3738.

[48] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[49] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 14–16.

[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[51] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

JUNYING GAN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical information engineering from Beihang University, in 1987, 1992, and 2003, respectively. In 1992, she joined Wuyi University, Guangdong, China, as a Full Professor. She is currently the Executive Director of Guangdong image graphics association. She has published more than 50 journal articles. Her research interests include biometric extraction and pattern recognition. She has received several provincial technology awards.

LI XIANG received the B.S. degree in electronic and communication engineering from Anhui Xinhua University, in 2017. She is currently pursuing the master's degree with the Department of Intelligence Manufacturing, Wuyi University. Her research interests include biometric extraction and pattern recognition.

YIKUI ZHAI (Member, IEEE) received the bachelor's and master's degrees in optical electronics information and communication engineering and signal and information processing from Shantou University, Guangdong, China, in 2004 and 2007, respectively, and the Ph.D. degree in signal and information processing from Beihang University, in June 2013. Since October 2007, he has been working with the Department of Intelligence Manufacturing, Wuyi University, where he is currently an Associate Professor. He has been a Visiting Scholar with the Department of Computer Science, Università Degli Studi di Milano, since 2016. His research interests include image processing, deep learning, and pattern recognition.

CHAOYUN MAI received the Ph.D. degree in information and communication engineering from Beihang University, Beijing, China, in 2017. He is currently a Lecturer with the Department of Intelligent Manufacturing, Wuyi University. His current research interests include waveform design and signal processing.

GUOHUI HE was born in Pingxiang, Jiangxi, China. He received the B.S. degree in computer engineering from the Shenyang University of Aeronautics and Astronautics, in 1983, and the M.S. degree from BUAA, in 1991. He joined Wuyi University, in 1994. His research interest includes multimedia information systems.
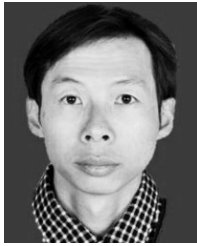
JUNYING ZENG was born in Ganzhou, Jiangxi, China, in 1977. He received the Ph.D. degree in physical electronics from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008. He has been working with the Department of Intelligence Manufacturing, Wuyi University, Guangdong, China. His research interests include intelligent signal processing and pattern recognition.

ZHENFENG BAI received the B.S. degree from the Qingdao Hengxing University of Science and Technology, in 2018. He is currently pursuing the master's degree with the Department of Intelligence Manufacturing, Wuyi University. His research interests include biometric extraction and pattern recognition.

RUGGERO DONIDA LABATI (Member, IEEE) received the Ph.D. degree in computer science from the Universitá Degli Studi di Milano, Crema, Italy, in 2013. He has been an Assistant Professor of computer science with the Universitá Degli Studi di Milano, since 2015. He has been a Visiting Researcher with Michigan State University, East Lansing, MI, USA. Original results have been published in more than 50 articles 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. He is an Associate Editor with the *Journal of Ambient Intelligence and Humanized Computing* (Springer).

VINCENZO PIURI (Fellow, IEEE) received the M.S. and Ph.D. degrees in computer engineering from the Politecnico di Milano, Italy. He has been a Full Professor with the University of Milan, Italy, since 2000, where he was a Department Chair, from 2007 to 2012. He was an Associate Professor with the Politecnico di Milano, Italy, from 1992 to 2000, a Visiting Professor with The University of Texas at Austin, USA, from summer 1996 to summer 1999, and a Visiting Researcher with George Mason University, USA, from summer 2012 to summer 2016. He founded a start-up company, Sensure srl, in the area of intelligent systems for industrial applications (leading it from 2007 to 2010). He was active in industrial research projects with several companies. His main research and industrial application interests are intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, the Internet of Things, cloud computing, fault tolerance, application-specific digital processing architectures, and arithmetic architectures. He is an ACM Fellow.

FABIO SCOTTI (Member, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003. He has been an Associate Professor of computer science with the Universitá Degli Studi di Milano, Crema, Italy, since 2015. Original results have been published in more than 100 articles 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. He 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 Co-Editor of the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

• • •