

Article

Facial Beauty Prediction Combined with Multi-Task Learning of Adaptive Sharing Policy and Attentional Feature Fusion

Junying Gan ^{*}, Heng Luo, Junling Xiong, Xiaoshan Xie , Huicong Li and Jianqiang Liu

Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China; 2112104068@wyu.edu.cn (H.L.); 2112204092@wyu.edu.cn (J.X.); xiaoshanxie.xsx@gmail.com (X.X.); 2112204010@wyu.edu.cn (H.L.); liu.geigei0203@gmail.com (J.L.)

* Correspondence: junyinggan@163.com

Abstract: Facial beauty prediction (FBP) is a leading research subject in the field of artificial intelligence (AI), in which computers make facial beauty judgments and predictions similar to those of humans. At present, the methods are mainly based on deep neural networks. However, there still exist some problems such as insufficient label information and overfitting. Multi-task learning uses label information from multiple databases, which increases the utilization of label information and enhances the feature extraction ability of the network. Attentional feature fusion (AFF) combines semantic information and introduces an attention mechanism to reduce the risk of overfitting. In this study, the multi-task learning of an adaptive sharing policy combined with AFF is presented based on the adaptive sharing (AdaShare) network in FBP. First, an adaptive sharing policy is added to multi-task learning with ResNet18 as the backbone network. Second, the AFF is introduced at the short skip connections of the network. The proposed method improves the accuracy of FBP by solving the problems of insufficient label information and overfitting issues. The experimental results based on the large-scale Asia facial beauty database (LSAFBD) and SCUT-FBP5500 databases show that the proposed method outperforms the single-database single-task baseline and can be applied extensively in image classification and other fields.

Keywords: attentional feature fusion; facial beauty prediction; image classification; multi-task learning



Citation: Gan, J.; Luo, H.; Xiong, J.; Xie, X.; Li, H.; Liu, J. Facial Beauty Prediction Combined with Multi-Task Learning of Adaptive Sharing Policy and Attentional Feature Fusion.

Electronics **2024**, *13*, 179.

<https://doi.org/10.3390/electronics13010179>

electronics13010179

Academic Editor: Eva Cernadas

Received: 21 November 2023

Revised: 23 December 2023

Accepted: 26 December 2023

Published: 30 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Facial beauty prediction (FBP) is a leading research subject in the field of artificial intelligence (AI), in which computers make facial beauty judgments and predictions similar to those of humans. With the development of AI, the applications of FBP are constantly expanding, including virtual makeup, plastic surgery, portrait photography and other fields. Research on FBP not only helps people understand and interpret beauty more scientifically and objectively but also promotes the development of AI, which has important significance. Currently, deep learning methods are generally used in FBP, which requires large amounts of label information. Existing facial beauty databases have certain issues, such as insufficient label information. Solving the aforementioned issue has become a popular subject in the field of FBP research. At present, some progress has been made in FBP research [1–7]. In [1], a novel personalized FBP approach based on meta-learning was designed to apply in some small databases. In [2], a self-correcting noise labels method was proposed. It can automatically select clean samples for learning and can make full use of all data to reduce the negative impact of noise labels. In [3], a fusion model of pseudolabel and cross-network was applied to solve the problems of weak generalization ability and insufficient label information in FBP. In [4], an innovative method of broad learning fused with transfer learning was applied in FBP, which received better performance in prediction accuracy and training speed. In [5], an adaptive transformer with global and local multihead self-attention was proposed for FBP, which achieved better performance

on several datasets of different scales. In [6], a dynamic convolution vision transformer named FBFormer was proposed which aims to focus on both local and global facial beauty features. Furthermore, an instance-level dynamic exponential loss function was designed to adjust the optimization objectives of the model dynamically. In [7], a novel method was proposed to improve the facial beauty feature extraction ability of CNNs, in which generative adversarial networks (GAN) were used to generate facial data.

Although the research above improved the accuracy of FBP, it did not efficiently solve the problems of insufficient label information and overfitting. Multi-task learning improves the generalization ability of a network by training related tasks containing domain-specific information. In the era of deep learning, multi-task learning has been transformed into designing networks that can learn shared representations from the label information of multiple tasks. Compared with a single-task learning network, a multi-task learning network has greater advantages. For example, related tasks can share complementary information or act as regularizers, thereby improving the network performance. FBP based on multi-task learning has been extensively studied in recent years [8,9]. In [8], a neural architecture search (NAS) was applied to FBP to automatically determine the backbone network for multi-task learning. Moreover, a new preprocessing method was introduced to enhance the diversity of data and a nonlocal spatial attention module was proposed which further improved the performance of the network on the FBP task. By combining ResNeXt-50 and Inception-v3, the dual-branch network can extract more facial beauty features and balance performance and parameter quantity [9]. Simultaneously, adaptive and dynamic loss functions are introduced.

At present, multi-task learning networks can be generally divided into hard parameter sharing and soft parameter sharing networks. In the hard parameter sharing network, the parameters are divided into a shared parameter and a task-specific parameter, and a hard parameter sharing network usually consists of several shared network layers and a task-specific network layer. In the soft parameter sharing network, each task has a separate feature extraction layer, with L_2 -norm or trace norms to constrain the parameters of the shared feature extraction layer. However, these two kinds of multi-task learning mostly set up the network layer statically. The method for adaptively training the network has become one of the key issues of multi-task learning.

Adaptive training can be divided into three key methods. First, the optimal backbone network is adaptively obtained according to different tasks through the NAS. Auto-multi-task learning (AutoMTL) proposed an automatic and efficient multi-task learning network framework for vision tasks, which takes a backbone network and a set of tasks to be learned as input and automatically generates a high-precision multi-task learning model [10]. Although NAS can automatically generate a high-precision multi-task learning network, it requires high computing equipment and a long calculation time. Second, in the process of parameter backpropagation for network optimization, adaptive task weights are learned based on the importance of different tasks. A Bayesian task weight learner is used to adjust the task weights and back-propagate the joint loss of different tasks [11]. The adaptive weight learning method based on the verification loss trend can measure the importance of different tasks and adjust the weights of different tasks [12]. In [13], a new training algorithm was proposed to utilize the similarity between tasks to learn the task relationship coefficients and neural network parameters. Although the optimization algorithm consumed fewer resources, the improvement in the network performance was limited. Finally, the network layer parameters that can be shared across tasks are determined. In adaptive sharing (AdaShare) networks, researchers have investigated sharing policies between tasks to achieve the highest accuracy and improve resource efficiency [14]. However, in an AdaShare network, the atrous spatial pyramid pooling (ASPP) method loses local information when extracting multiscale information from images. Attentional feature fusion (AFF) combines local and global information with semantic information at different levels [15].

There were differences in the distributions of the means and variances in the different databases. Shared batch normalization (BN) layers tend to exhibit poor network

performance. A dataset-aware block (DAB) was applied to capture the homogeneous convolutional representations and heterogeneous statistics across different datasets, where the dataset alternation training (DAT) mechanism was utilized to facilitate the optimization process [16]. Reference [17] proposed a different training mechanism. Each batch of training data consisted of data randomly selected from all the datasets for batch loss calculation.

The chief contributions of this study are summarized as follows:

- We extend the AdaShare network, introduce DAB to solve the issue of distribution differences between different databases in multi-task learning on FBP and apply the network in various databases.
- We propose multi-task learning of an adaptive sharing policy combined with AFF to solve the issue of insufficient label information and overfitting for FBP, in which the receptive field is expanded, and more semantic information is obtained from the images.
- The experimental results show that multi-task learning of the adaptive sharing policy combined with AFF outperforms the baseline model and the other method on FBP.

2. Methods

2.1. Network Model

A schematic diagram of the network model structure is shown in Figure 1, including the pre-trained network module, multi-task learning of the adaptive sharing policy combined with the AFF and classification module. The pre-training network module transfers the parameters to multi-task learning of the adaptive sharing policy combined with the AFF module through the ResNet18 [18] network trained by the ImageNet datasets. Multi-task learning of adaptive sharing policy combined with the AFF module contains multi-task learning of adaptive sharing policy with ResNet18 as the backbone network plus the AFF introduced at the short-skip connection. It primarily performs sharing policy learning, image feature extraction, and fusion. The classification module includes an average pooling layer, a fully connected layer, a Dropout [19] layer and a softmax classifier. Database1 and database2 are two different databases, task1 and task2 are two different tasks. First, the parameters of the pretrained network module are transferred to the multi-task learning of the adaptive sharing policy combined with the AFF module. Second, the images from database1 and database2 are simultaneously entered into the multi-task learning of the adaptive sharing policy combined with the AFF module. Meanwhile, the sharing policies and image features of task1 and task2 are learned from the module. Finally, the features are entered into the classification module, in which the categories of task1 and task2 are the outputs.

2.2. Multi-Task Learning of Adaptive Sharing Policy Combined with AFF Module

The multi-task learning of the adaptive sharing policy combined with the AFF module was extended with ResNet18 as the backbone network; its schematic is shown in Figure 2. Among them, the multi-task learning of the adaptive sharing policy combined with the AFF module contains four-layer blocks, and each layer block is composed of four BasicBlock structures. First, the images from database1 and database2 are entered into the network simultaneously. In the convolution, ReLU, and max-pooling layers, the network parameters of the two tasks are shared. Second, the image features must pass through four-layer blocks. The BasicBlock structure of each layer block includes an adaptive sharing policy and AFF. A schematic of the layered block structure is shown in Figure 3, and the BasicBlock structure is shown in Figure 4. Finally, the features produced by layer4 are entered into the classification module and the classification results of task1 and task2 are produced.

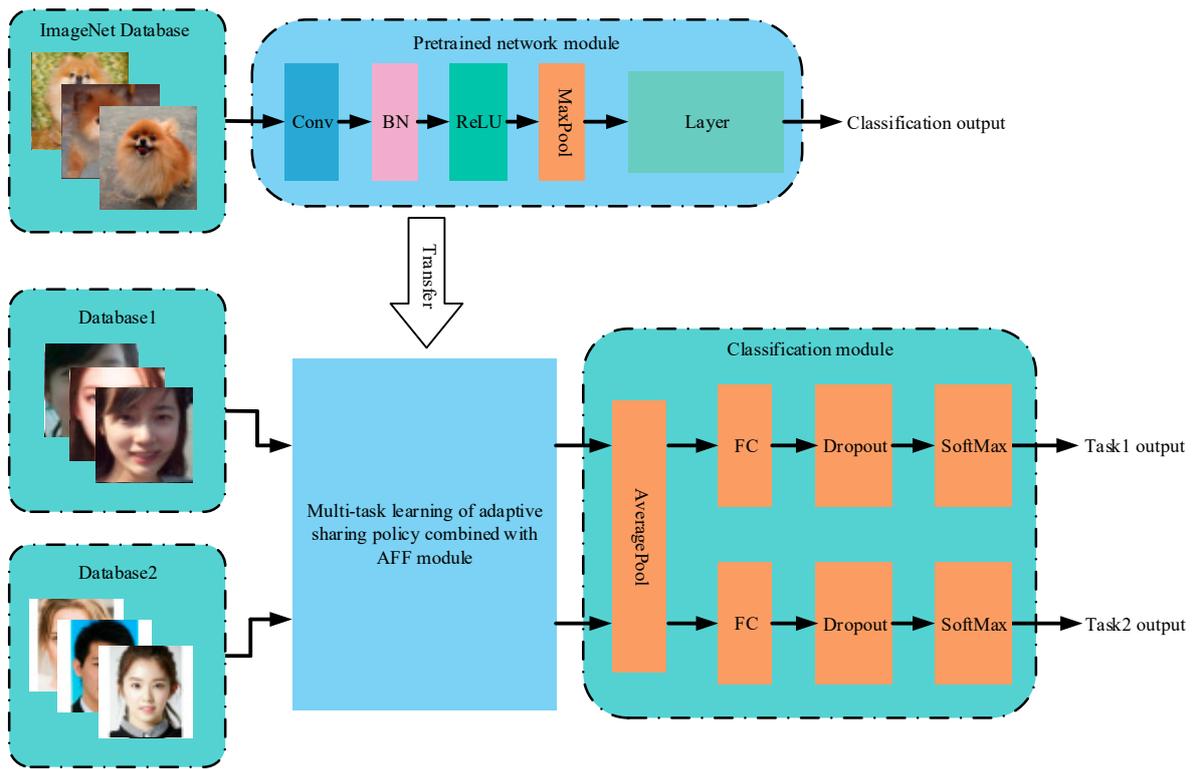


Figure 1. Schematic diagram of network model structure.

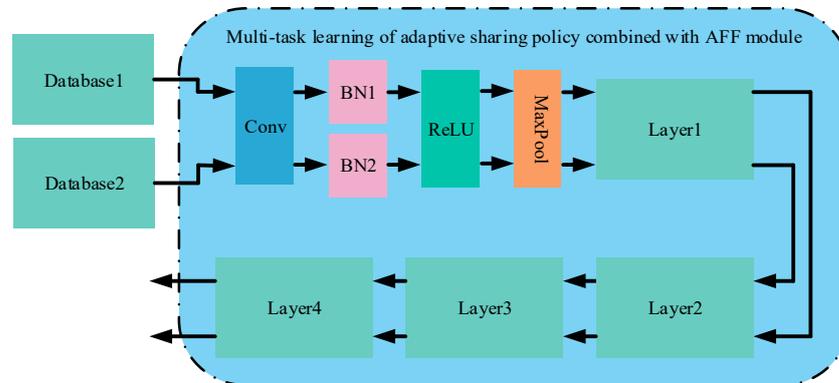


Figure 2. Schematic structure of multi-task learning of adaptive sharing policy combined with AFF module.

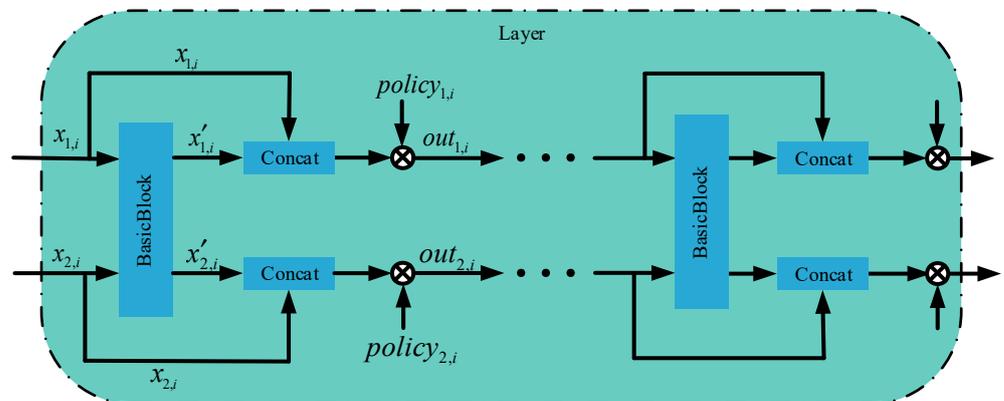


Figure 3. Schematic diagram of layer block structure.

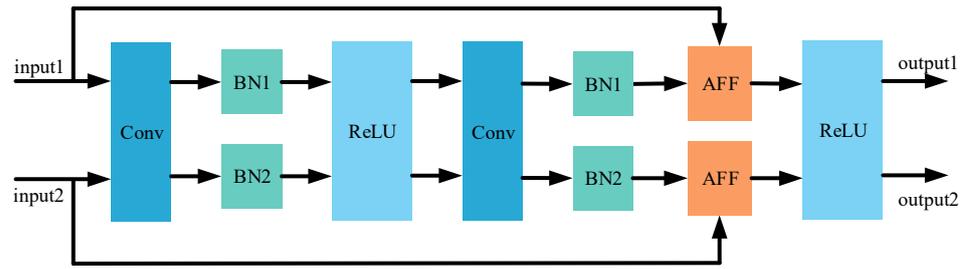


Figure 4. Schematic diagram of BasicBlock structure.

In Figure 3, $x_{1,i}, x'_{1,i}, out_{1,i}, policy_{1,i}$ are the variables of task1 in the i – th ($2 < i < 17$) layer network. Among them, $x_{1,i}, x'_{1,i}$ are feature maps with a resolution of $m \times n$ and a channel number of l , that is $x_{1,i}, x'_{1,i} \in \mathbb{R}^{m \times n \times l}$. $policy_{1,i}$ represents the sharing policy of task1 on the i – th layer network. In Figure 3, the specific process can be described as follows:

First, $x_{1,i}$ is changed into $x'_{1,i}$ through the BasicBlock structure. At the same time, $x_{1,i}$ concatenates $x'_{1,i}$ and its result is multiplied by $policy_{1,i}$. Finally, $out_{1,i}$ is obtained. $out_{1,i}$ can be expressed as follows:

$$out_{1,i} = [x'_{1,i} \ x_{1,i}] \cdot policy_{1,i} = \begin{cases} x'_{1,i} \cdot 0 + x_{1,i} \cdot 1, & \text{when } policy_{1,i} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ x'_{1,i} \cdot 1 + x_{1,i} \cdot 0, & \text{when } policy_{1,i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{cases} \quad (1)$$

where $policy_{1,i} = [0 \ 1]^T$ indicates that the sharing policy of task1 in the i – th layer network is skipped. $policy_{1,i} = [1 \ 0]^T$ indicates that the sharing policy of task1 in the i – th layer network is implemented. The multi-task learning of adaptive sharing policy aims to learn the sharing policy and network weights from the loss function through backpropagation. But each $policy_{1,i}$ is discrete and non-differentiable so the gradient of the entire network cannot be backpropagated. Therefore, the Gumbel Softmax [20] function is applied to solve this non-differentiable problem to complete backpropagation and update the parameters. $x_{2,i}, x'_{2,i}, x''_{2,i}, out_{2,i}, policy_{2,i}$ represent the variables of task2 in the i -th layer network, which are the same as the variables of task1 in the network. The details of the adaptive sharing policy are expressed in Algorithm 1.

Algorithm 1 Facial beauty prediction via adaptive sharing policy

- Input:** sample set x
Output: output set out
- 1: n is the number of layers in the backbone;
 - 2: m is the number of blocks in each layer;
 - 3: $policy$ is the adaptive policy of the current layer;
 - 4: φ indicates the BasicBlock structure;
 - 5: ϕ indicates the concatenation and multiplication;
 - 6: for $i, i \leq n$ do
 - 7: for $j, j \leq m$ do
 - 8: Calculate $x' = \varphi(x)$
 - 9: Calculate $out = \phi(x, x', policy)$
 - 10: end
 - 11: end
-

To improve the convergence speed of the network, a multi-stage training method is adopted during the training phase. Initially, the multi-task learning network shares all the parameters. As the number of training epochs increases, the deep network adopts a shared policy for training. In a deep convolutional network, the BN layer can be understood as a simplified whitening operation on the input value of each layer of the deep network. This whitening operation is significantly affected by the distribution of the databases. Therefore, to apply the label information of multiple databases and to solve the issue of distribution differences caused by different databases, DAB is introduced, which implies that different tasks will use different BN layers. Figure 4 shows a schematic of the improved BasicBlock structure proposed in this study, where input1 and input2 represent the Image features of database1 and database2 in Figure 2, and the tasks of each database use different BN layers. The AFF aims to extract features that are more relevant to the current task and fuse channel features at different scales.

2.3. Attentional Feature Fusion

The AFF module was introduced to fuse the semantic information of different network layers and generate the fusion weights for the mapping and residuals of the network [15]. Figure 5 shows a schematic of the AFF structure, where $b, r \in \mathbb{R}^{C \times H \times W}$ and C is the channels, H is the height, and W is the width. In ResNet [18], b is the mapping and r is the residual. Based on the multiscale channel attention module (MS-CAM), the AFF can be expressed as follows:

$$z = c' \otimes b \oplus (1 - c') \otimes r \tag{2}$$

where $z \in \mathbb{R}^{C \times H \times W}$ is the fusion feature of the i -th layer network, the \otimes operation is the multiplication of each element, the \oplus operation is the sum of each element, and $c' = MS(c)$, $c = b \oplus r$, and $1 - c'$ is obtained from c' by passing it through the Diff operation. The output c' after the MS-CAM structure is a real number between 0 and 1, and $1 - c'$ is also a real number between 0 and 1. Therefore, the network can improve the feature extraction ability by learning the fusion weight of the mapping and residual, thereby improving the accuracy of the target task. Figure 6 shows a schematic of the MS-CAM structure.

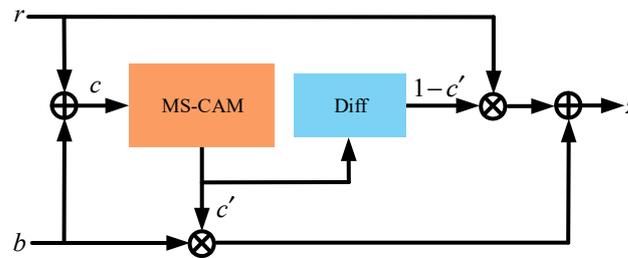


Figure 5. Schematic diagram of AFF structure.

In AFF, MS-CAM fuses local and global features in the attention mechanism, which not only assigns different weights to each channel but also gathers multi-scale feature context information. Thus, it improves the network’s ability to extract the target task features. By aggregating multiscale contextual information along the channel dimension, MS-CAM can simultaneously emphasize global and local information [15]. Therefore, the MS-CAM was utilized as a multiscale feature extractor. The local information extractor can be computed as follows:

$$L(c) = B(PWConv_2(\delta(B(PWConv_1(c)))))) \tag{3}$$

where $PWConv_1$ indicates that the channels of the input feature $c \in \mathbb{R}^{C \times H \times W}$ are reduced to the original $1/r$ by the point convolution of 1×1 , B indicates the BN layer, δ indicates the ReLU activation layer, $PWConv_2$ indicates that the channels are restored to the original

input channel by the point convolution of 1×1 , and r is the channel-scaling ratio. The global information extractor can be represented as follows:

$$G(c) = B(PWConv_2(\delta(B(PWConv_1(A(c)))))) \tag{4}$$

where $A(c)$ denotes the global average pooling layer. The final output c' can be calculated as follows:

$$c' = c \otimes \sigma(L(c) \oplus G(c)) \tag{5}$$

where σ is the Sigmoid activation function. Therefore, a network with AFF not only fuses different semantic information but also introduces an attention mechanism that improves the feature extraction ability of the network, reduces the risk of overfitting, and improves the accuracy of the target task.

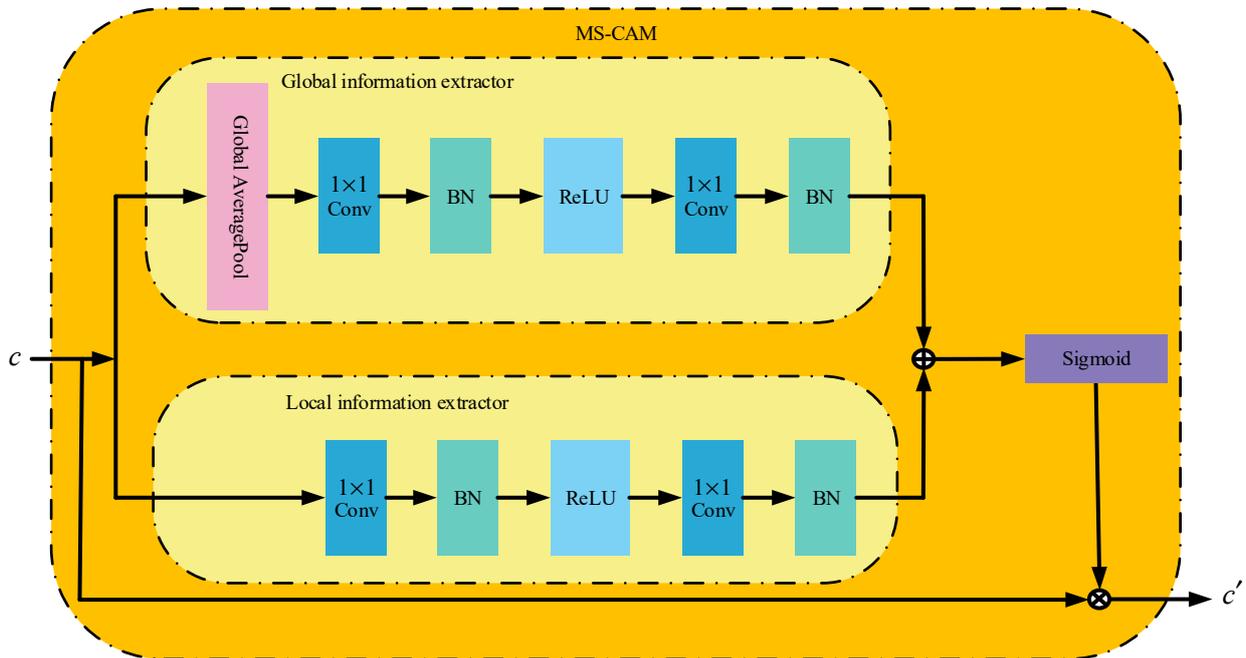


Figure 6. Schematic diagram of MS-CAM structure.

2.4. Loss Function

In this study, cross-entropy was adopted as the loss function for task1 and task2, which can be defined as follows:

$$L = \sum_{i=1}^N y_i * \log(p_i) \tag{6}$$

where N is the number of categories in task1 or task2, y_i is the label value of the i – th category of the image, and p_i is the probability value of the image being predicted as the i – th category. The total task loss function can be formalized as follows:

$$L_{\text{task}} = \lambda_1 L_1 + \lambda_2 L_2 \tag{7}$$

where L_1 and L_2 represent the loss function value of task1 and task2, respectively; λ_1 and λ_2 represent the weight coefficient of task1 and task2, respectively. In multi-task learning, the weight ratio $\lambda_1 : \lambda_2$ of different tasks affects the accuracy of the target task.

3. Experiments and Analysis

3.1. Experimental Databases

3.1.1. LSAFBD Database

The authors established the LSAFBD database with 20,000 labeled facial images (including 10,000 male and 10,000 female facial images) and 80,000 unlabeled facial images, with a resolution of 144×144 . It is divided into five categories, including “0”, “1”, “2”, “3” and “4”, which correspond to five attractiveness levels of facial beauty, with “0” being the lowest level and “4” being the highest level. This study primarily focused on experiments with 10,000 labeled female facial images from the LSAFBD database. The distribution of facial beauty labels in the LSAFBD database and some image samples in the LSAFBD database are shown in Figures 7 and 8, respectively.

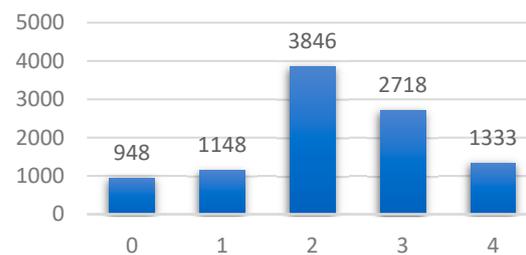


Figure 7. Distribution of facial beauty labels on LSAFBD.

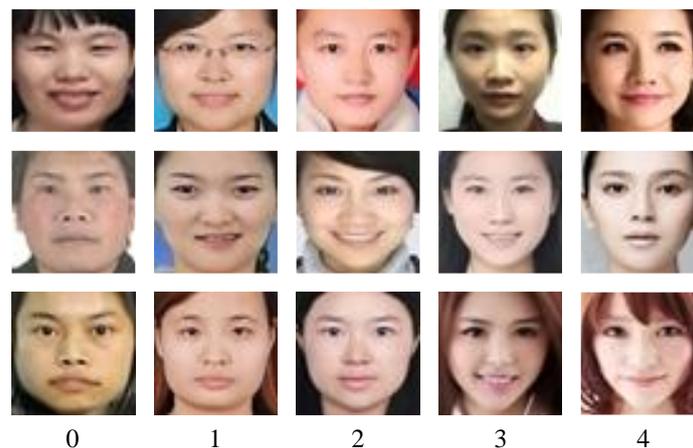


Figure 8. Facial images with different properties of LSAFBD.

3.1.2. SCUT-FBP5500 Database

The SCUT-FBP5500 database was established by the South China University of Technology and contains a total of 5,500 facial images with a resolution of 350×350 . Each facial image contained various label information, including gender (male or female), race (Asian or White), and facial beauty. The facial beauty level of the SCUT-FBP5500 database is divided into five levels, namely “0”, “1”, “2”, “3” and “4”, which correspond to the five attractiveness levels of facial beauty, with “0” as the lowest grade, and “4” as the highest grade. The facial beauty grade of each image was given by 60 volunteers; therefore, this study takes the grade with the largest number of volunteers as the facial beauty grade of the image. The distribution of the facial beauty labels in the SCUT-FBP5500 database and some image samples of the SCUT-FBP5500 database are shown in Figures 9 and 10, respectively.

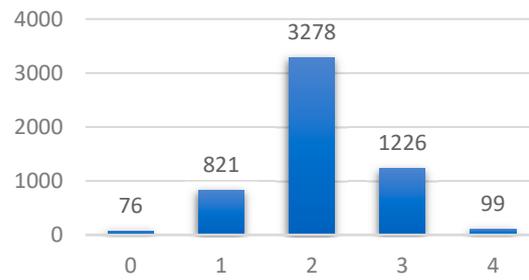


Figure 9. Distribution of facial beauty labels on SCUT-FBP5500.



Figure 10. Facial images with different properties of SCUT-FBP5500.

3.2. Experimental Environment

Table 1 describes the experimental environment. In Figure 3, the variables $x_{1,i}$, $x'_{1,i}$, and $x''_{1,i}$ have a resolution of 56×56 , the number of channels is 64, the task weight ratio $\lambda_1 : \lambda_2$ of the training phase is 1:0.6, the data batch size is 32, the initial learning rate is 0.001, and the optimizer is AdamW [21]. In this study, the accuracy (ACC) and F1 score were applied as the performance evaluation metric.

Table 1. Experimental environment configuration.

Environment	Parameters
Deep learning framework	Pytorch1.12.1
Operating system	Ubuntu20.04
Memory	64 G
Resolution $m \times n$	56×56
Channels l	64
Task weight ratio $\lambda_1 : \lambda_2$	1:0.6
Learning rate	0.001
Batch size	32
Optimizer	AdamW

The experimental setting in Figure 1, is shown in Table 2. Database1 and database2 represent the LSAFBD database and SCUT-FBP5500 database, respectively. Task1 and task2 represent the facial beauty prediction (FBP) and gender recognition (GR), respectively.

Table 2. Explanation of the experimental setting.

Experiment Settings	Explanation
Database1	LSAFBD
Database2	SCUT-FBP5500
Task1	FBP
Task2	GR

3.3. Comparison Experiment between the Proposed Method and the Baseline

3.3.1. Experiments Based on Different Databases

The experimental results of the proposed method and the baseline based on the LSAFBD database are shown in Tables 3–5. The ratio of the training, verification and testing set was 6:2:2, and the experiments included the training and testing phases. In the training phase, the facial beauty-labeled data from the LSAFBD database and gender-labeled data from the SCUT-FBP5500 database are used as an input of the network for the proposed method. In the testing phase, the proposed method was based on the testing set of the LSAFBD database for FBP. The baseline based on transfer learning was a single-database, single-task method with ResNet18 as the backbone network. In the training phase, the facial beauty-labeled data from the LSAFBD database are used as an input of the network of baseline. During the testing phase, the baseline was based on the testing set of the LSAFBD database for FBP.

Table 3. Experimental results based on LSAFBD (ACC(%), F1 score(%)).

Batch Size	Task	Method	Baseline without AFF		Baseline with AFF		Ours without AFF		Ours with AFF	
			ACC	F1 Score	ACC	F1 Score	ACC	F1 Score	ACC	F1 Score
32		FBP	58.01	56.51	59.52	57.70	59.12	57.72	61.37	59.72
16		FBP	58.02	56.53	59.77	57.76	59.02	57.62	61.12	59.53

Note: The bold is the optimal value.

Table 4. Experimental results of the proposed method based on LSAFBD (ACC(%), time (s), difference of ACC(%)).

Batch Size	Task	Method	Ours without AFF				Ours with AFF			
			Training Time	Training ACC	Testing ACC	Difference of ACC	Training Time	Training ACC	Testing ACC	Difference of ACC
32		FBP	2976.04	63.42	59.12	4.30%	3867.85	63.49	61.37	2.12%
16		FBP	3555.86	63.12	59.02	4.10%	4587.81	63.31	61.12	2.19%

Note: The bold is the optimal value.

Table 5. Experimental results of the baseline based on LSAFBD (ACC(%), time (s), difference of ACC(%)).

Batch Size	Task	Method	Baseline without AFF				Baseline with AFF			
			Training Time	Training ACC	Testing ACC	Difference of ACC	Training Time	Training ACC	Testing ACC	Difference of ACC
32		FBP	637.23	65.03	58.01	7.02%	853.19	61.61	59.52	2.09%
16		FBP	800.91	65.28	58.02	7.26%	1154.07	61.72	59.77	1.95%

It can be observed from Table 3 that with AFF, the accuracy of the FBP in the proposed method was 61.37%, which was 1.85% higher than the baseline accuracy of 59.52% and the F1 score of the FBP in the proposed method was 59.72%, which was 2.02% higher than the baseline F1 score of 57.70. Without AFF when the batch size was 32, the accuracy of the FBP in the proposed method was 59.12%, which was 1.11% higher than the baseline accuracy of 58.01% and the F1 score of the FBP in the proposed method was 57.72%, which was 1.21% higher than the baseline F1 score of 56.51%. The experimental results showed

that the proposed multi-task learning of the adaptive sharing policy outperformed the baseline. In the current method, the accuracy of FBP with AFF was 61.37%, which was 2.25% higher than that without AFF (59.12%) and the F1 score of FBP with AFF was 59.72%, which was 2.00% higher than that without AFF (57.72%). At baseline, the accuracy of FBP with AFF was 59.52%, which was 1.51% higher than the accuracy of 58.01% and the F1 score of FBP with AFF was 57.70%, which was 1.19% higher than the F1 score of 56.51%. The experimental results showed that AFF can improve the network's ability to extract facial beauty features, thereby improving the accuracy of FBP. When the batch size was 16, the proposed method also achieved better performance than the baseline.

It can be observed from Table 4 to Table 5 that when the batch size was 32, the difference between the training accuracy and testing accuracy of the proposed method with AFF was 2.12%, which is 2.18% lower than that of the proposed method without AFF of 4.30%. The difference between the training accuracy and testing accuracy of the baseline with AFF was 2.09%, which was 4.93% lower than that of the baseline without AFF of 7.02%. The experimental results showed that the AFF truly reduces the risk of overfitting and improves the feature extraction capability of the network. The improvement also can be seen when the batch size was 16.

Table 6 shows the experimental results of the proposed method and the baseline method based on the SCUT-FBP5500 database, respectively. In the training phase, the facial beauty-labeled and gender-labeled data from the training set are used as an input of the network of the proposed method simultaneously. In the testing phase, the proposed method was based on a testing set for FBP and GR. In the training phase, the facial beauty-labeled data and gender-labeled data from the training set are used as an input of the network of baseline, respectively. In the testing phase, the baseline implemented FBP and GR, respectively, based on the testing set.

Table 6. Experimental results based on SCUT-FBP5500 (ACC(%), F1 score(%)).

Batch Size	Task	Method	Baseline without AFF		Baseline with AFF		Ours without AFF		Ours with AFF	
			ACC	F1 Score	ACC	F1 Score	ACC	F1 Score	ACC	F1 Score
32	FBP		73.41	70.64	74.50	71.72	74.23	72.02	75.41	73.82
	GR		98.27	98.26	98.55	98.55	96.52	96.52	97.09	97.09
16	FBP		73.67	70.13	74.61	71.91	73.95	71.91	75.13	73.27
	GR		98.45	98.45	98.73	98.73	96.43	96.40	96.89	96.88

It can be observed from Table 6 that with AFF, the FBP accuracy of the proposed method was 75.41%, which was 0.91% higher than the baseline accuracy of 74.50% and the F1 score was 73.82%, which was 1.1% higher than the baseline F1 score of 71.72%. Without AFF when the batch size was 32, the accuracy of the FBP in the proposed method was 74.23%, which was 0.82% higher than the baseline accuracy of 73.41% and the F1 score was 72.02%, which was 1.38% higher than the baseline F1 score of 70.64. The experimental results showed that the proposed multi-task learning of the adaptive sharing policy outperformed the baseline. In the current method, the accuracy of FBP with AFF was 75.41%, which was 1.18% higher than the accuracy of 74.23% without AFF and the F1 score of FBP with AFF was 73.82%, which was 1.8% higher than the F1 score of 72.02% without AFF. The accuracy of the GR with AFF was 97.09%, which was 0.57% higher than the accuracy of 96.52% and the F1 score of the GR with AFF was 97.09%, which was 0.57% higher than the F1 score of 96.52%. At baseline, the accuracy of FBP with AFF was 74.50%, which was 1.09% higher than the accuracy of 73.41% without AFF, and the F1 score of FBP with AFF was 71.72%, which was 1.08% higher than the F1 score of 70.64. The accuracy of GR with AFF was 98.55%, which was 0.28% higher than the accuracy of 98.27% without AFF and the F1 score of GR with AFF was 98.55%, which was 0.29% higher than the F1 score of 98.26%. The experimental results showed that AFF can improve the network's ability to extract facial beauty and gender features, thereby improving the accuracy of FBP in both the proposed

method and the baseline. When the batch size was 16, the proposed method also achieved better performance than the baseline.

It can be observed from Table 7 to Table 8 that when the batch size was 32, the difference between FBP training accuracy and FBP testing accuracy of the proposed method with AFF was 1.43%, which was 2.02% lower than that of the proposed method without AFF of 3.45%. The difference between FBP training accuracy and FBP testing accuracy of the baseline with AFF was 0.61%, which was 6.45% lower than that of the baseline without AFF of 7.06%. The experimental results showed that the AFF truly reduces the risk of overfitting and improves the feature extraction capability of the network. The improvement also can be seen when the batch size was 16.

Table 7. Experimental results of the proposed method based on SCUT-FBP5500 (ACC(%), time (s), difference of ACC(%)).

Batch Size	Task	Method	Ours without AFF				Ours with AFF			
			Training Time	Training ACC	Testing ACC	Difference of ACC	Training Time	Training ACC	Testing ACC	Difference of ACC
32	FBP			77.68	74.23	3.45		76.84	75.41	1.43
	GR		1587.81	97.36	96.52	0.84	2073.85	97.63	97.09	0.54
16	FBP			77.49	73.95	3.54		76.44	75.13	1.31
	GR		1894.58	97.13	96.43	0.70	3076.09	97.50	96.89	0.61

Table 8. Experimental results of the baseline based on SCUT-FBP5500 (ACC(%), time (s), difference of ACC(%)).

Batch Size	Task	Method	Baseline without AFF				Baseline with AFF			
			Training Time	Training ACC	Testing ACC	Difference of ACC	Training Time	Training ACC	Testing ACC	Difference of ACC
32	FBP		372.69	80.47	73.41	7.06	482.48	75.11	74.50	0.61
	GR		362.23	98.49	98.27	0.22	484.57	98.64	98.55	0.09
16	FBP		442.85	79.49	73.67	5.82	653.17	75.24	74.61	0.63
	GR		443.31	98.61	98.45	0.16	619.89	98.86	98.73	0.13

From the experimental results, it is observed that in FBP multi-task learning of the adaptive sharing policy combined with AFF achieved the best results on the two different databases. The proposed method not only effectively utilizes GR to improve the network's ability to extract facial beauty features, but also reduces the risk of overfitting through attention feature fusion, thereby improving the accuracy of FBP. From the experimental results in Table 6, it can be observed that the accuracy of the proposed method is lower than that of the baseline method in terms of the GR. This is because the proposed method improves the accuracy of FBP through GR. Therefore, when the task weight ratio $\lambda_1 : \lambda_2$ was 1:0.6, the network learned more facial beauty features, resulting in a lower extraction ability for gender features than the baseline. From Tables 4, 5, 7 and 8, it can be observed that the proposed method requires more time for training. This is because the proposed method has a more complex network and more data to calculate.

3.3.2. Experiments with Different Weight Ratios Based on Different Databases

To study the effect of the weight ratio $\lambda_1 : \lambda_2$ on different tasks in FBP, the weight ratios of the three groups of FBP and GR were explored. The experimental results for different weight ratios based on the LSAFBD database are shown in Table 9. At a weight ratio of 1:0.6, FBP achieved an accuracy of 61.37%, surpassing the 58.91% by 2.46% at a weight ratio of 1:0.7, and exceeding the 58.62% by 2.75% at a weight ratio of 1:0.5. At a weight ratio of 1:0.6, FBP achieved an F1 score of 59.72%, surpassing the 56.94% by 2.78% at a weight ratio of 1:0.7, and exceeding the 56.70% by 3.02% at a weight ratio of 1:0.5. The experimental results showed that different weight ratios have a significant influence on the

FBP. When the weight ratio was 1:0.6, the proposed method achieved the best performance on the LSAFBD database.

Table 9. Different weight ratios of the experimental results based on LSAFBD (ACC(%), F1 score(%)).

Batch Size	$\lambda_1:\lambda_2$		1:0.7		1:0.6		1:0.5	
			ACC	F1 Score	ACC	F1 Score	ACC	F1 Score
32	Task	FBP	58.91	56.94	61.37	59.72	58.62	56.70
16		FBP	58.88	56.86	61.12	59.53	58.53	56.37

Note: The bold is the optimal value.

The experimental results for different weight ratios based on the SCUT-FBP5500 database of the proposed method are shown in Table 10. When the weight ratio was 1:0.6, the accuracy of the FBP was 75.41%, which was 1.76% higher than 73.65% when the weight ratio was 1:0.7, and 1.83% higher than 73.58% when the weight ratio was 1:0.5. When the weight ratio was 1:0.7, the accuracy of the GR was 97.18%, which was 0.09% higher than 97.09% when the weight ratio was 1:0.6, and 0.73% higher than 96.45% when the weight ratio was 1:0.5. When the weight ratio was 1:0.6, the F1 score of the FBP was 73.82, which was 2.40% higher than 71.42 when the weight ratio was 1:0.7, and 2.42% higher than 71.40% when the weight ratio was 1:0.5. When the weight ratio was 1:0.7, the F1 score of the GR was 97.18%, which was 0.09% higher than 97.09% when the weight ratio was 1:0.6, and 0.77% higher than 96.41% when the weight ratio was 1:0.5. The experimental results showed that different weight ratios have a significant influence on the FBP. When the weight ratio was 1:0.6, the proposed method achieved the best performance on the SCUT-FBP5500 database.

Table 10. Different weight ratios of the experimental results based on SCUT-FBP5500 (ACC(%), F1 score(%)).

Batch Size	$\lambda_1:\lambda_2$		1:0.7		1:0.6		1:0.5	
			ACC	F1 Score	ACC	F1 Score	ACC	F1 Score
32	Task	FBP	73.65	71.42	75.41	73.82	73.58	71.40
		GR	97.18	97.18	97.09	97.09	96.45	96.44
16	Task	FBP	73.27	71.67	75.13	73.27	73.41	70.67
		GR	97.15	97.14	96.89	96.88	96.35	96.31

From the experimental results, it is observed that when the weight ratio is 1:0.6, the FBP accuracy of the proposed method based on two different databases reaches the highest value in the existing experiments. When the weight ratio was 1:0.7, the gender features learned by the network increased and the facial beauty features decreased, resulting in a slight improvement in the accuracy of the FBP compared with the single-task network. When the weight ratio was 1:0.5, the gender features learned by the network were insufficient, and compared with the single-task network, it could only slightly improve the accuracy of the FBP.

3.4. Comparison Experiments between the Proposed Method and Other Models

In this section, the proposed method is compared with other models based on the LSAFBD database and SCUT-FBP5500 databases. The experimental results for the proposed method and other models are shown in Table 11. Based on the LSAFBD database, during the training phase, the facial beauty-labeled data from the LSAFBD database and the gender-labeled data from the SCUT-FBP5500 database are used as an input of the network of the proposed method simultaneously. In the testing phase, the proposed method was based on the testing set of the LSAFBD database for FBP. In the training phase, the facial beauty-labeled data from the LSAFBD database are used as an input of the other models. In the testing phase, the other models were based on the testing set of the LSAFBD database for the FBP.

Table 11. Experimental results compared with other models (ACC%).

Method	Task	LSAFBD	SCUT-FBP5500
		FBP	FBP
GoogleNet [22]		56.06	72.77
MobileNetV2 [23]		50.70	72.13
MobileNetV3 [24]		52.35	72.31
ShuffleNetV2 [25]		59.97	75.14
DenseNet [26]		59.32	73.95
EfficientNet [27]		59.02	75.04
RegNet [28]		59.02	74.68
ConvNeXt [29]		60.67	75.32
Proposed method		61.37	75.41

Note: The bold is the optimal value.

Based on the SCUT-FBP5500 database, in the training phase, facial beauty labeled data and gender labeled data of the SCUT-FBP5500 database are used as an input of the network of the proposed method simultaneously. During the testing phase, the proposed method was based on the testing set of the SCUT-FBP5500 database for FBP. In the training phase, the facial beauty-labeled data from the SCUT-FBP5500 database are used as an input for the other models. During the testing phase, the other models were based on the testing set of the SCUT-FBP5500 database for FBP.

In Table 1, GoogleNet [22] improved the utilization of computing resources inside the network through the inception structure. MobileNetV2 [23] introduced a residual structure that ascended and descended the dimensions to enhance the propagation of gradients and significantly reduce the memory footprint required during inference. MobileNetV3 [24] added a lightweight attention model Squeeze Excitation (SE) structure based on MobileNetV2. ShuffleNetV2 [25] proposed that the ratio of the input feature matrix channel to the output matrix channel should be equal to or close to one. The input to each network layer in DenseNet [26] is a concatenation of all previous network outputs. EfficientNet [27] was proposed to keep the channels of features, depth of the network model, and image resolution small, which can create a competitive and computationally efficient CNN model. RegNet [28] aims to determine the optimal search space. Using the search space, a series of design criteria for the model can be obtained and extended to other scenarios. In ConvNeXt [29], better CNN structures and parameter settings were determined through numerous experiments.

Based on the LSAFBD database, the FBP accuracy of the proposed method was 61.37%, which was 5.31% higher than 56.06% on GoogleNet, 11.3% higher than 50.70% on MobileNetV2, 9.02% higher than 52.35% on MobileNetV3, 1.4% higher than 59.97% on ShuffleNetV2, 2.05% higher than 59.32% on DenseNet, 2.35% higher than 59.02% on EfficientNet, 2.35% higher than 59.02% on RegNet, and 0.7% higher than 60.67% on ConvNeXt. The experimental results showed that the proposed method can effectively utilize GR to improve the accuracy of FBP, which is better than other single-task network models.

Based on the SCUT-FBP5500 database, the FBP accuracy of the proposed method was 75.41%, which was 2.64% higher than 72.77% on GoogleNet, 3.28% higher than 72.13% on MobileNetV2, 3.1% higher than 72.31% on MobileNetV3, 0.27% higher than 75.14% on ShuffleNetV2, 1.46% higher than 73.95% on DenseNet, 0.37% higher than 75.04% on EfficientNet, 0.73% higher than 74.68% on RegNet, and 0.09% higher than 75.32% on ConvNeXt. The experimental results showed that the proposed method can effectively apply GR to improve the accuracy of FBP, which is superior to other models.

3.5. Comparison Experiments between the Proposed Method and Other Methods

To further validate the effectiveness of the proposed method, we also compared the proposed method with other methods based on the LSAFBD and SCUT-FBP5500. The results are listed in Table 12. In [2], a self-correcting noise labels method was proposed,

which can make full use of all data to reduce the negative impact of noise labels. In [3], a fusion model of pseudolabel and cross-stitch network was applied to solve the problems of weak generalization ability and insufficient label information in FBP. In [4], a network named E-BLS fusing EfficientNet and a broad learning system was applied in FBP. In [5], a tiny network named TransBLS-T fusing transformer and broad learning system was proposed to improve FBP. The performance of the proposed method surpasses that of the other method. Based on the LSAFBD database, the method by way of self-correcting noise labels achieves poor results. This is because the method is based on single-task deep neural networks (DNNs) and does not utilize label information from multiple databases. The experimental results of the cross-network based on multi-task learning illustrate the superiority of multi-task learning. The methods by way of E-BLS and TransBLS-T are better than those of DNNs based on the LSAFBD database, which is attributed to the attention mechanism of the transformer. The proposed method in this paper combines the advantages of multi-task learning and attention feature fusion to achieve the best results.

Table 12. Experimental results compared with other methods (ACC(%)).

Method	Task	LSAFBD	SCUT-FBP5500
		FBP	FBP
Noise Labels [2]		60.80	75.30
Cross Network [3]		61.29	-
E-BLS [4]		60.82	73.13
TransBLS-T [5]		61.27	75.23
Proposed method		61.37	75.41

Note: The bold is the optimal value.

In summary, multi-task learning of the adaptive sharing policy combined with AFF utilizes the label information of two different databases, solves the problem of insufficient label information on the single-task network for FBP, and improves the network's ability to extract facial beauty features. Simultaneously, the network combines AFF to reduce the risk of overfitting, thereby improving the accuracy of the FBP.

4. Conclusions

To address the issue of insufficient label information and easy overfitting in FBP, multi-task learning of an adaptive sharing policy combined with AFF based on the AdaShare network is proposed. Among them, multi-task learning of the adaptive sharing policy utilizes the label information of two different databases to improve the accuracy of FBP by solving the insufficient label information issue. The AFF reduces the risk of overfitting and improves the feature extraction capability of the network by adding a feature fusion and attention mechanism at the short skip connections of ResNet. The experimental results based on the LSAFBD database and SCUT-FBP5500 databases showed that the multi-task learning of the adaptive sharing policy combined with AFF outperforms the single-task baseline method. Future studies will be focused on label information from multiple databases, how to set the weight ratio of different tasks adaptively, how to balance the category of databases, and on continuously optimizing the current method to obtain greater improvement.

Author Contributions: Conceptualization, J.G., H.L. (Heng Luo); methodology, H.L. (Heng Luo), J.X.; software, H.L. (Heng Luo), J.X., X.X.; validation, H.L. (Heng Luo), X.X., H.L. (Huicong Li); formal analysis, J.G., X.X.; investigation, J.G., X.X., H.L. (Huicong Li), J.L.; resources, J.G.; data curation, H.L. (Heng Luo), J.X., X.X.; writing—original draft preparation, J.G., H.L. (Heng Luo), J.X.; writing—review and editing, J.G., H.L. (Heng Luo), J.X., H.L. (Huicong Li), J.L.; supervision, J.G.; project administration, J.X., X.X., H.L. (Huicong Li), J.L.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 61771347).

Data Availability Statement: SCUT-FBP5500: Dataset utilized in this research is publicly available: <https://github.com/HCIILAB/SCUT-FBP5500-Database-Release> (accessed on 20 November 2023). LSAFBD: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lebedeva, I.; Ying, F.; Guo, Y. Personalized facial beauty assessment: A meta-learning approach. *Vis. Comput.* **2023**, *39*, 1095–1107. [[CrossRef](#)]
2. Gan, J.; Wu, B.; Zhai, Y.; He, G.; Mai, C.; Bai, Z. Self-correcting noise labels for facial beauty prediction. *Chin. J. Image Graph.* **2022**, *27*, 2487–2495.
3. Gan, J.; Wu, B.; Zou, Q.; Zheng, Z.; Mai, C.; Zhai, Y.; He, G.; Bai, Z. Application Research for Fusion Model of Pseudolabel and Cross Network. *Comput. Intell. Neurosci.* **2022**, *2022*, 1–10. [[CrossRef](#)] [[PubMed](#)]
4. Gan, J.; Xie, X.; Zhai, Y.; He, G.; Mai, C.; Luo, H. Facial beauty prediction fusing transfer learning and broad learning system. *Soft Comput.* **2023**, *27*, 13391–13404. [[CrossRef](#)]
5. Gan, J.; Xie, X.; He, G.; Luo, H. TransBLS: Transformer combined with broad learning system for facial beauty prediction. *Appl. Intell.* **2023**, *53*, 26110–26125. [[CrossRef](#)]
6. Liu, Q.; Lin, L.; Shen, Z.; Yu, Y. FBPFormer: Dynamic Convolutional Transformer for Global-Local-Contextual Facial Beauty Prediction. In Proceedings of the Artificial Neural Networks and Machine Learning (ICANN), Heraklion, Greece, 26–29 September 2023; pp. 223–235. [[CrossRef](#)]
7. Laurinavičius, D.; Maskeliūnas, R.; Damaševičius, R. Improvement of Facial Beauty Prediction Using Artificial Human Faces Generated by Generative Adversarial Network. *Cogn. Comput.* **2023**, *15*, 998–1015. [[CrossRef](#)]
8. Zhang, P.; Liu, Y. NAS4FBP: Facial Beauty Prediction Based on Neural Architecture Search. In Proceedings of the Artificial Neural Networks and Machine Learning (ICANN), Bristol, UK, 6–9 September 2022; pp. 225–236.
9. Bougourzi, F.; Dornaika, F.; Taleb-Ahmed, A. Deep learning based face beauty prediction via dynamic robust losses and ensemble regression. *Knowl.-Based Syst.* **2022**, *242*, 108246–108251. [[CrossRef](#)]
10. Zhang, L.; Liu, X.; Guan, H. AutoMTL: A Programming Framework for Automating Efficient Multi-task Learning. In Proceedings of the Advances in Neural Information Processing Systems (NeuralIPS), New Orleans, LA, USA, 28 November–9 December 2022; pp. 34216–34228.
11. Li, H.; Wang, Y.; Lyu, Z.; Shi, J. Multi-task learning for recommendation over heterogeneous information network. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 789–802. [[CrossRef](#)]
12. Fan, X.; Wang, H.; Zhao, Y.; Li, Y.; Tsui, K.L. An adaptive weight learning-based multi-task deep network for continuous blood pressure estimation using electrocardiogram signals. *Sensors* **2021**, *21*, 1595. [[CrossRef](#)] [[PubMed](#)]
13. Zhou, F.; Shui, C.; Abbasi, M.; Robitaille, L.-E.; Wang, B.; Gagne, C. Task similarity estimation through adversarial multi-task neural network. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 466–480. [[CrossRef](#)] [[PubMed](#)]
14. Sun, X.; Panda, R.; Feris, R.; Saenko, K. AdaShare: Learning What to Share for Efficient Deep Multi-task Learning. In Proceedings of the Advances in Neural Information Processing Systems (NeuralIPS), Virtual, 6–12 December 2020; pp. 8728–8740.
15. Dai, Y.; Gieseke, F.; Oehmcke, S.; Wu, Y.; Barnard, K. Attentional Feature Fusion. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Virtual, 5–9 January 2021; pp. 3559–3568.
16. Wang, L.; Li, D.; Liu, H.; Peng, J.; Tian, L.; Shan, Y. Cross-dataset collaborative learning for semantic segmentation in autonomous driving. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; pp. 2487–2494.
17. Kapidis, G.; Poppe, R.; Veltkamp, R.C. Multi-Dataset, Multi-task Learning of Egocentric Vision Tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 6618–6630. [[CrossRef](#)] [[PubMed](#)]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
19. Srivastava, N.; Hinton, G.E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
20. Jang, E.; Gu, S.; Poole, B. Categorical Reparameterization with Gumbel-Softmax. In Proceedings of the 5th International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
21. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.

23. Sandler, M.; Howard, A.G.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
24. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
25. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
26. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
27. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
28. Radosavovic, I.; Kosaraju, R.P.; Girshick, R.; He, K.; Dollár, P. Designing Network Design Spaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10428–10436.
29. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.