

Article

Facial Beauty Prediction Using a Generative Adversarial Network for Dataset Augmentation

Junying Gan ^{1,*}, Zhen Chen ¹, Hantian Chen ¹, Wenchao Xu ¹, Zhenxin Zhuang ¹ and Junling Xiong ²

¹ School of Electronics and Information Engineering, Wuyi University, Jiangmen 529020, China; 2112304039@wyu.edu.cn (Z.C.); chenlantian0925@gmail.com (H.C.); 2112304067@wyu.edu.cn (W.X.); 2112304008@wyu.edu.cn (Z.Z.)

² School of Electronic Information and Control Engineering, Software Engineering Institute of Guangzhou, Guangzhou 510990, China; xjl@mail.seig.edu.cn

* Correspondence: junyinggan@163.com

Abstract

Facial beauty prediction (FBP) is a significant research direction in the field of computer vision; however, the performance of models developed for this task is often constrained due to the scarcity of high-quality annotated data. Generative adversarial networks (GANs) are efficient image generation networks that are capable of rapidly generating facial images. This study proposes an FBP method—named FBP-GAN—which aims to address this shortage of data by generating high-quality synthetic facial images. First, we construct a facial image generation network based on StyleGAN2-ADA to generate diverse and realistic facial images. Second, we combine transfer learning and data augmentation techniques to utilize the synthesized images for training set augmentation while optimizing the category distribution to enhance the generalization ability and prediction accuracy of the classification network. The experimental results demonstrate that, when using MobileViT or ResNeXt as the classification network, our proposed approach achieves prediction accuracies of 76.38% and 77.94% on the SCUT-FBP5500 dataset, respectively, representing improvements of 0.55% and 1.65% over the baseline models' 75.83% and 76.29%. The proposed approach effectively improves the accuracy of FBP under data-scarce scenarios and opens new avenues for the application of GANs in computer vision tasks.

Keywords: facial beauty prediction; generative adversarial networks; data augmentation

Academic Editor: Silvia
Liberata Ullo

Received: 23 December 2025

Revised: 15 January 2026

Accepted: 28 January 2026

Published: 30 January 2026

Copyright: © 2026 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 cutting-edge topic in the field of artificial intelligence that relates to human cognitive processes and patterns [1]; in particular, it explores the endowment of computers with the ability to judge or predict facial attractiveness in humans. Investigating how to better interpret, quantify, and predict beauty can facilitate a more scientific and objective understanding and description of beauty, thereby advancing FBP as an interdisciplinary research direction. For millennia, human comprehension of beauty has remained largely intuitive, with perceptions being highly subjective and difficult to quantify. Since Plato introduced the concept of aesthetics, studies on the nature of beauty and its evaluation criteria have been ongoing across philosophy, psychology,

medicine, and other domains without reaching any scientific consensus. In recent years, the emergence of deep learning technology has revitalized FBP research. Leveraging the highly non-linear fitting capabilities of neural networks, advanced semantic features capturing facial beauty can be extracted, thereby endowing machines with human esthetic reasoning abilities. One study [2] has proposed a method for understanding facial beauty through deep facial feature analysis, in which convolutional neural networks (CNNs) [3] are employed for facial feature extraction and random forests are used to evaluate facial attractiveness, demonstrating the importance of deep features in FBP. Additionally, an adaptive attribute-aware convolutional neural network [4] has been introduced, in which the network filters are adjusted to incorporate attribute perception as an additional input to the model, forming a pseudo-attribute-aware convolutional neural network [5]. In this line, a lightweight pseudo-attribute distiller can learn input pseudo-attribute perception, effectively improving FBP performance.

However, training a CNN typically requires large amounts of labeled data, which are scarce and difficult to obtain in reality. When insufficient training samples are available, determining how to train a robust network becomes an urgent challenge. In this context, transfer learning offers a viable solution. Another study [6] first transferred deep features from pre-trained models into Bayesian ridge regression algorithms for FBP, integrating multi-scale CNNs, transfer learning, and maximum feature maps as activation functions to achieve better results by aggregating features at different scales [7].

At present, FBP necessitates extensive labeled data; however, the effort, time, and resources required for annotating facial beauty data are extremely high. Without sufficient labeled data to train models, there is a significant risk of overfitting, negatively impacting model performance.

In recent years, several large-scale facial recognition datasets have been released successively, such as Labeled Faces in the Wild [8], Large-scale CelebFaces Attributes [9], and MS-Celeb-1M [10]. The publication of these datasets has significantly propelled advancements in facial recognition technology. However, these publicly available datasets still have limitations. First, their sample distribution is uneven, potentially leading to model bias on specific attributes. Second, the required facial data for FBP must meet certain environmental and contextual conditions, which public datasets often fail to satisfy. Additionally, real-world data are frequently subject to privacy issues, raising ethical concerns.

To address these challenges, data augmentation techniques have emerged as a solution. Traditional data augmentation methods include image rotation, scaling, translation, and color adjustments, which can increase the diversity of a dataset through applying simple transformations to the original images. While traditional data augmentation approaches can alleviate the issue of insufficient data to a certain extent, the generated samples are confined with respect to the variations in the original data, thus failing to create entirely new instances or significantly enhance the diversity of the dataset.

Since their introduction by Goodfellow et al. in 2014 [11], generative adversarial networks (GANs) have achieved remarkable progress in tasks such as image generation and data augmentation. GANs leverage adversarial training between generator and discriminator networks to learn the underlying data distribution, producing samples that closely resemble real data. This capability underpins the immense potential of GANs for application in tasks such as facial image generation and data augmentation.

This study proposes a novel FBP method, which we call FBP-GAN. Through introducing synthetic data generated using the GAN into FBP datasets, face beauty prediction models can be trained jointly with the synthetic and original data, effectively enhancing their generalization ability and prediction accuracy. This approach not only mitigates the challenges of data scarcity and category distribution imbalance that are prevalent in the context of FBP but also offers a feasible data augmentation strategy for other deep learning tasks

that rely on large-scale data. In terms of model architecture, this study employs high-performance MobileViT [12] or ResNeXt [13] for the development of a classification network to achieve efficient facial beauty prediction. The proposed method is tested on the SCUT-FBP5500 dataset [14] and a large-scale dataset of female Asian faces [15] (Large Scale Asia Facial Beauty Database, LSAFBD), allowing for exploration of the performance enhancements brought by FBP-GAN. The experimental results demonstrate that FBP-GAN achieves satisfactory outcomes when using both Transformer [16] and CNN architectures.

The primary contributions of this study are as follows:

- Through investigation of the class distribution imbalance issue in FBP datasets, a new data category distribution that is suitable for FBP is proposed.
- Based on the feature space of the SCUT-FBP5500 dataset, images in LSAFBD are reconstructed to alleviate the class imbalance issue in SCUT-FBP5500, thereby enhancing the classification accuracy of the MobileViT and ResNeXt networks on this dataset.
- An effective solution is provided for FBP in scenarios with scarce data, offering new insights regarding the application of generative adversarial networks (GANs) in computer vision tasks.

The rest of this paper is organized as follows: Section 2 briefly reviews the related works. Section 3 introduces the FBP-GAN architecture and details its implementation. Section 4 details the experimental procedures and analyzes the results. Section 5 introduces the conclusion and future work directions.

2. Related Works

2.1. Generative Adversarial Networks

GANs are based on a dual-network structure consisting of a generator and a discriminator, which undergo adversarial training to enable the generator to synthesize realistic image data. Following its introduction, researchers have made numerous improvements to the GAN architecture and associated training methods in order to address issues such as training instability and mode collapse, which were inherent to the original version. One early advancement was the Deep Convolution Generative Adversarial Network (DCGAN) proposed by Radford et al. [17], in which convolutional layers are incorporated into both the generator and discriminator networks, significantly enhancing the visual quality of the generated images. This advance laid the groundwork for subsequent CNN-based image generation studies.

Subsequently, the Wasserstein GAN (WGAN) [18] was introduced, in which the Wasserstein distance is employed to measure discrepancies between generated and real samples. This approach effectively addresses the gradient vanishing problem, which commonly occurs during training. The WGAN variant with Gradient Penalty (WGAN-GP) further improved the training stability through the incorporation of a gradient penalty mechanism [19]. These advancements enabled GANs to generate high-quality images in more complex contexts, thereby advancing their application in facial image generation, image super-resolution, and image restoration tasks.

To generate higher-resolution images, the Progressive Growing of GANs (ProGAN) [20] method was introduced. ProGAN begins training with low-resolution image generation and progressively increases the network complexity, enabling the generator to learn how to produce higher-resolution images in a more stable state. This incremental training approach effectively avoids the common issue of training instability faced when generating high-resolution images. ProGAN's outstanding performance in facial image generation tasks has been demonstrated, producing images with high quality and relatively natural details.

In recent years, research on GANs has gradually deepened in the field of image generation, with researchers exploring how to control the style of generated images. For

example, StyleGAN, proposed by Karras et al. [21] from NVIDIA Corporation, Santa Clara, CA, USA, is an improved GAN architecture that leverages style transfer techniques to control different features in the generated images, enabling the creation of highly realistic and controllable facial images. The introduction of StyleGAN has advanced facial image generation to a new stage, particularly excelling in terms of diversity and fine image detail control.

StyleGAN2-ADA is a GAN that is specifically designed for high-quality image generation in scenarios characterized by limited data [22]. Building upon StyleGAN2 [23], it introduces Adaptive Discriminator Augmentation (ADA)—an innovative solution to address the overfitting issues that arise when training discriminators with limited samples. The ADA mechanism dynamically enhances the discriminator’s training data through techniques such as color transformation, geometric transformation, and noise perturbation. It further adjusts the intensity of these augmentations based on the training dynamics, effectively preventing overfitting. Remarkably, StyleGAN2-ADA was shown to achieve generation quality comparable to that of models trained on 50,000 samples using only 1000 samples, thereby significantly reducing data requirements and computational costs.

2.2. Lightweight General Vision Transformer

MobileViT [12] is a lightweight vision Transformer model that successfully addresses the challenge of deploying traditional Vision Transformers (ViTs) on mobile devices. Proposed by Apple Inc., Cupertino, CA, USA, this model combines the spatial inductive bias of CNNs with the global modeling capabilities of ViTs through a lightweight design, achieving efficient feature extraction. MobileViT introduces a novel MobileViT block structure, replacing standard convolutional local processing with Transformer-based global processing. This involves locally encoding spatial information using convolutions, followed by token expansion and interaction with Transformer layers to learn cross-regional global dependencies, and finally restoring the spatial structure through folding operations. This design retains pixel spatial order while achieving full-field global modeling, significantly reducing computational complexity. Due to its compact architecture, MobileViT has been practically deployed in edge devices and mobile phone vision perception tasks, effectively balancing model accuracy and runtime efficiency.

2.3. Improved Residual Network

ResNeXt [13], proposed by Xie et al. in 2017, represents an improved residual network structure that enhances the resulting model’s expressiveness through the introduction of the “grouped convolution” concept, focusing not solely on network depth or width. Unlike traditional ResNet [24], each residual block in ResNeXt employs multiple parallel low-dimensional convolution operations, forming a unified and efficient feature extraction approach following the “Split–Transform–Merge” design principle. This structural innovation reduces computational complexity while enhancing the model’s adaptability to diverse feature patterns. Under similar computational costs, ResNeXt has been shown to possess superior classification performance and generalization capabilities when compared to ResNet. Consequently, it has been widely applied across various visual tasks such as image classification, object detection, and semantic segmentation, and remains an important reference model for subsequent efficient network design.

3. Methods

3.1. Overall Framework

Figure 1 illustrates the framework of FBP-GAN, where Data-1 refers to the SCUT-FBP5500 dataset [14], which serves as the primary dataset for the training and validation of

the classification network in this study. Data-2 corresponds to the LSAFBD dataset [15], which acts as a secondary dataset and is primarily used for feature extraction to enable the generation network based on StyleGAN2-ADA to produce more realistic images. The initial resolutions of images in Data-1 and Data-2 are 350×350 and 144×144 pixels, respectively. The proposed method consists of three stages: data preprocessing, generation, and classification.

In the data preprocessing stage, the primary task is to resize the resolutions of both Data-1 and Data-2 to make them consistent. Specifically, the Data-2 dataset is upscaled using the waifu2x image super-resolution system [25] to a resolution of 288×288 pixels and then further resized to 256×256 pixels via rescaling. For Data-1, the dataset is directly downsized to a resolution of 256×256 pixels, employing Lanczos resampling during resizing to minimize any image distortion caused by scaling.

In the generation stage, the first step involves training the generation network based on the training set of Data-1. Subsequently, a fixed number of samples are randomly extracted from the training set of Data-2 according to their categories. These samples are converted into high-dimensional feature representations via a feature extractor and then input into the generation network to synthesize a new dataset, Data-3, with a resolution of 256×256 pixels. This stage is the most critical phase of the model framework because it not only performs GAN-based data augmentation for the primary dataset (Data-1) but also optimizes the class distribution of Data-1.

Finally, in the classification stage, the training set for the classification network (denoted as Data-4) is constructed by combining the images from both Data-1 and Data-3 (with a resolution of 256×256 pixels). This hybrid dataset is then used to train the classification network, ultimately achieving the image classification task based on the primary dataset, Data-1. This phase represents the implementation stage of the FBP task, where MobileViT and ResNeXt are employed as efficient classification architectures to enhance the overall performance of the FBP-GAN framework.

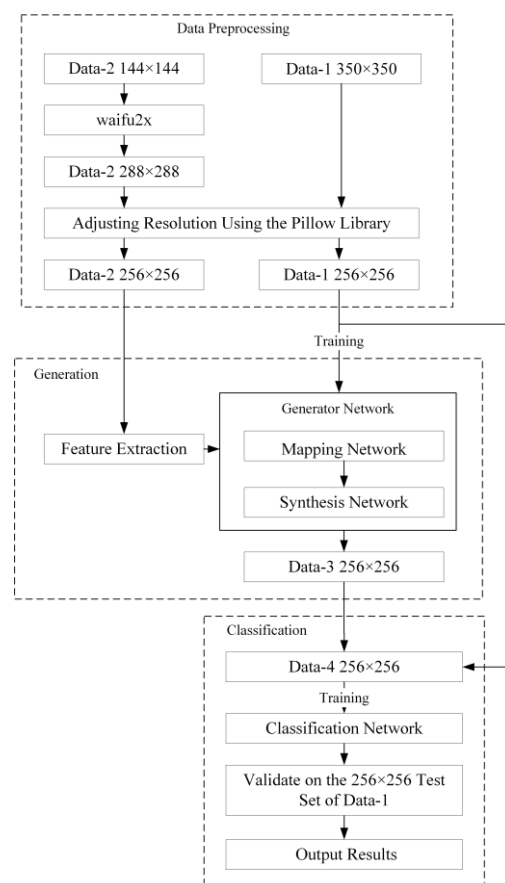


Figure 1. The framework of the proposed FBP-GAN method in this study.

3.2. Style Information Transfer

To enhance the performance of the generator network, FBP-GAN employs a pruned StyleGAN2-ADA variant named StyleGAN2-ADA-s, with approximately 12.2 M total parameters and relatively fast training speed. Figure 2 depicts the StyleGAN2-ADA-s framework, which is primarily composed of a discriminator and a generator that supports conditional generation [26].

First, unlike traditional GANs, the generator requires not only random noise as an input, but also inputs from the latent space w , which represents information controlling the style of the generated images. This w has a more significant impact on the generation process than random noise. The w vector is generated by the mapping network from the latent vector z , which is then divided into multiple control vectors. After weight demodulation, these are fed into each stage of the generator to influence the overall generation process, ultimately controlling the style of the generated images (e.g., faces). In addition, the random noise used as a direct input to the generator is scaled and introduced into every generation stage alongside w . This allows both the style and the fine details of the generated images to be controlled through w and the scaled random noise, respectively, ensuring stability in image generation. Real images from the source dataset are utilized for pre-training of the entire generator network. The adversarial loss expressions for generator G and discriminator D are represented as follows

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))], \quad (1)$$

where $p_{data}(x)$ denotes the true data distribution, $p_z(z)$ represents the noise distribution in the latent space, and $D(\cdot)$ outputs the probability of a sample being real, with values closer to 1 indicating higher confidence in its genuineness. In particular, G aims to minimize $V(D, G)$, which means that it seeks $D(G(z))$ to be as close to 1 as possible; on the other hand, D aims to maximize $V(D, G)$, striving for $\log D(x)$ to approach 1 and $D(G(z))$ to trend towards 0. This setup encourages the generator to continuously refine its output, producing increasingly realistic samples until it achieves an effect where the generated data are indistinguishable from the real data.

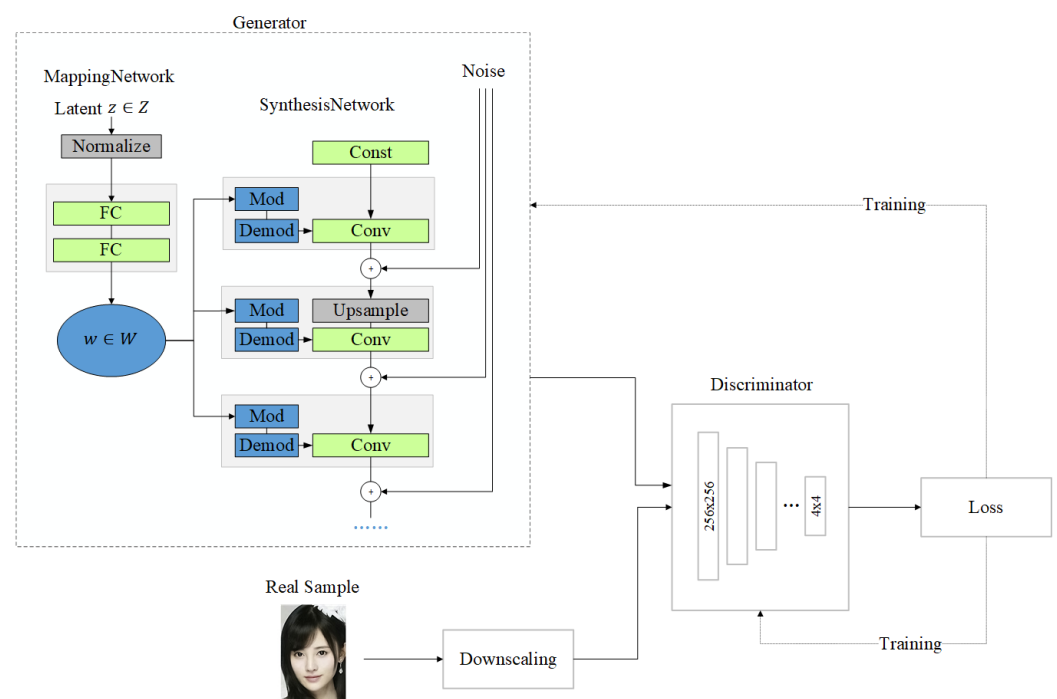
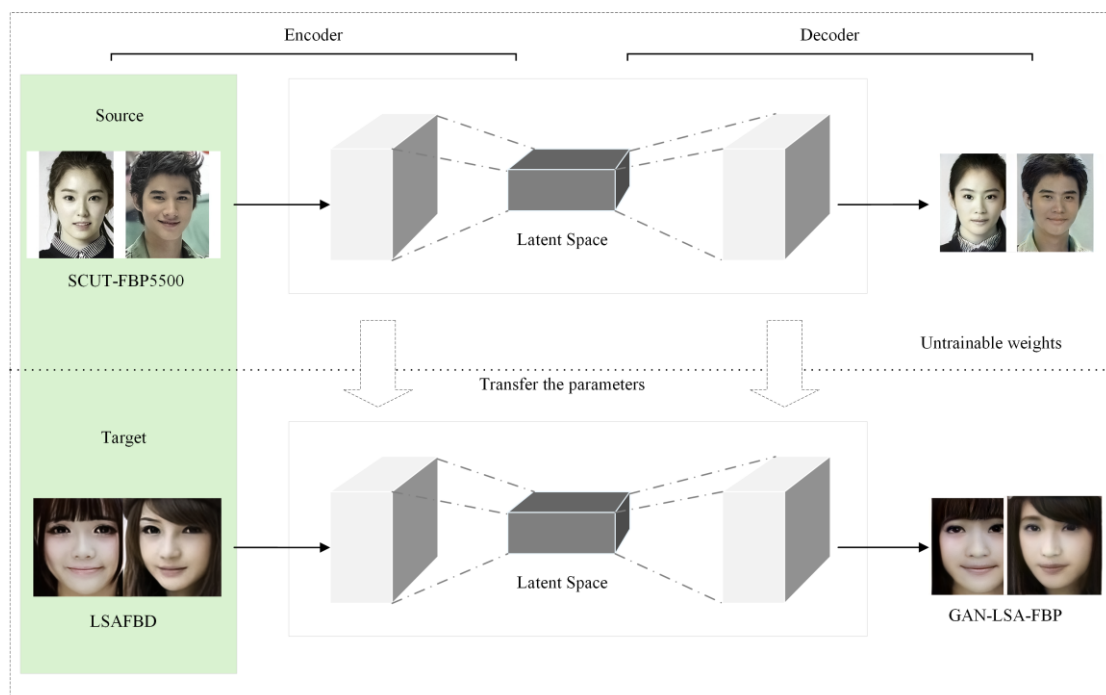


Figure 2. The framework of StyleGAN2-ADA-s.

To augment the source dataset, transfer learning can be employed to migrate style information from the source data to the target facial dataset, thereby expanding the source dataset using the target data. Without transfer learning, the generation network achieves feature fusion directly on the source dataset for augmentation. Figure 3 illustrates the style transfer process between the SCUT-FBP5500 [14] and LSAFBD [15] datasets.

First, StyleGAN2-ADA-s is trained on the SCUT-FBP5500 dataset without pretrained weights, yielding SCUT-FBP5500-optimized StyleGAN2-ADA-s weights. Second, feature information from the LSAFBD dataset is used as input into this SCUT-FBP5500-weighted StyleGAN2-ADA-s, synthesizing new data (GAN-LSA-FBP) that approximates the feature distribution of the SCUT-FBP5500 dataset.

**Figure 3.** Style information transfer process.

To input the facial features from the LSAFBD dataset [15] into StyleGAN2-ADA-s, this study employs the VGG16 model [27] to extract features from facial images and store them in the latent space W . Subsequently, the generator decodes and reconstructs these encoded facial features from the latent space back into their original images. During the face generation process, a specific latent space vector w is sought such that, after decoding, the target face can be reconstructed. The process of reconstructing the target face in the latent space is illustrated in Algorithm 1.

Algorithm aims to minimize the discrepancy between the synthesized image and the target image in the VGG feature space. It adopts a composite loss function that combines perceptual loss with noise-regularization loss. During optimization, stochastic gradient descent is used to update both the latent variables and the injected noise, subject to normalization constraints. Learning-rate cosine annealing and regularized noise smoothing are further introduced to ensure stability. The algorithm runs for 600 iterations by default, but can be stopped earlier when the loss change $\Delta loss$ remains below 10^{-3} for consecutive steps; no explicit stopping threshold is otherwise enforced.

Algorithm 1. Face Generation in Latent SpaceInput: Input vector X , generator G , feature extraction network N , real face image F_R Output: Synthetic face image F_S

```

1: while  $Error > \varepsilon$  or  $steps \leq T$  do
2:    $F_S = G \rightarrow generate\_face(X)$  ▷ Generate synthetic face
3:    $L_S = N(F_S)$  ▷ Embedding of synthetic face in latent space
4:    $L_R = N(F_R)$  ▷ Embedding of real face in latent space
5:    $Error = diff(L_S, L_R)$  ▷ Difference between synthetic and real face
6: end while

```

First, the generator produces an initial synthetic image based on an initial latent space vector. Second, through a pre-trained VGG16 network, the generated image is mapped to the latent space to extract high-dimensional feature representations. Next, the same VGG16 network is used to encode and map real face images from the LSAFBD dataset [15] to the same latent space. Then, the feature differences between the synthetic face and the target real face in the latent space are calculated, serving as the optimization objective. Finally, the Adam optimizer is employed to optimize the latent space vector iteratively, refining the generation results and progressively reducing the feature errors between the synthetic and real images. VGG16 serves as a perceptual encoder in latent space for computing feature-similarity loss, guiding the optimization of the generator.

3.3. FBP-GAN Method

The essence of FBP-GAN lies in combining StyleGAN2-ADA's latent-space feature-transport mechanism with the Omega distribution for category re-balancing, providing a brand-new GAN-augmentation paradigm for facial beauty prediction. Unlike conventional methods that merely generate images, FBP-GAN embeds constraints in the latent space to reinforce feature fidelity, achieving more accurate data balance and richer sample diversity.

While training a model with a mix of real and synthetic data poses significant challenges, due to the uniqueness of the FBP task and advancements in transfer learning techniques, special generative networks such as CycleGAN [28] can easily transform an image of one form into another. In the process of FBP-GAN data augmentation, new features are not directly introduced from other datasets; instead, existing features are used to reconstruct new ones, resulting in relatively stable augmented data that closely resemble the original data, minimally impacting the classification model's convergence process and enhancing robustness. FBP is unique when compared to other image classification tasks, as people's perceptions of facial beauty tend to be subjective and emotional, making it difficult for models to converge during training due to prediction accuracy varying among individuals. Drawing from experiments in other scientific domains, such as studies on facial attractiveness [29], data augmentation and fusion can be conducted based on prior theory without significantly disrupting the original data distribution. The LSAFBD [15] and SCUT-FBP5500 [14] datasets share consistent category counts and both focus exclusively on FBP research. While the LSAFBD dataset consists solely of lower-resolution images of Asian women, in this study, it serves as a supplementary resource to the SCUT-FBP5500 dataset, according to prior theory that standards for judging male facial attractiveness are statistically comparable to those for female facial attractiveness [30].

3.4. Ethical Considerations

The concept of beauty is inherently subjective and varies across cultures, historical periods, and individuals. Models trained on biased datasets—such as those dominated by a single gender or ethnicity—may unintentionally reinforce stereotypes or discriminatory

aesthetic standards. This bias risks marginalizing underrepresented groups and perpetuating harmful social norms. Furthermore, as FBP systems rely on facial imagery, they involve sensitive biometric data that must be collected and processed under strict privacy and consent regulations. Misuse of such data for profiling or commercial purposes could result in serious ethical violations. Moreover, automated attractiveness assessment can distort human self-perception and intensify social pressures related to appearance, particularly when deployed in recruitment, media, or social networking applications. To mitigate these risks, researchers should adopt fairness-aware learning, ensure diverse and balanced datasets, and limit FBP deployment to transparent, research-driven contexts. Ultimately, the development of FBP systems must balance technical innovation with respect for cultural diversity, individual dignity, and ethical responsibility.

4. Experiments and Analysis

4.1. Experimental Subjects

The SCUT-FBP5500 dataset [14] was established by South China University of Technology and contains 5500 front-face images with a resolution of 350×350 pixels, covering various ethnicities, genders, and ages. Each image is scored by 60 volunteers on a beauty scale ranging from 1 to 5. Based on the mode criterion, each image is categorized into one of five levels: 0, 1, 2, 3, or 4, corresponding to “very unattractive,” “unattractive,” “average,” “attractive,” and “very attractive,” which include 76, 821, 3278, 1226, and 99 images, respectively. Figure 4 displays some sample images from the SCUT-FBP5500 dataset along with the percentage distribution of images across each category, showing a data distribution that closely resembles a normal distribution. In this study, the primary focus is on conducting enhancement experiments using the SCUT-FBP5500 dataset, with the training and testing sets divided in an 8:2 ratio.

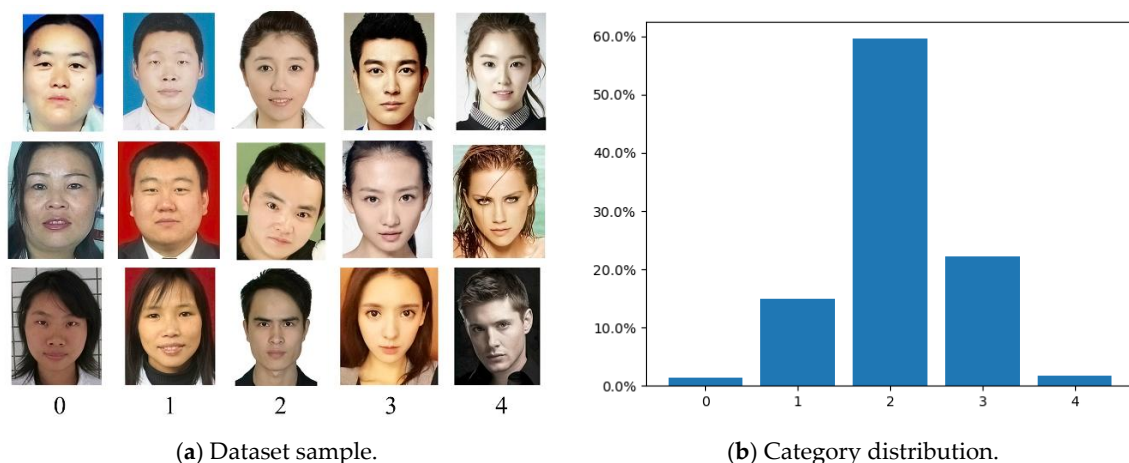


Figure 4. SCUT-FBP5500 dataset sample and category distribution.

The LSAFBD dataset [15] is an FBP dataset established by our team, which focuses exclusively on female faces, with 20,000 annotated images and 80,000 unannotated images, each with a resolution of 144×144 pixels. Most images exhibit diversity in terms of background, pose, and age. Each image was scored by 75 volunteers and categorized into one of five levels: 0, 1, 2, 3, or 4, containing 948, 1148, 3846, 2718, and 1333 images, respectively. Figure 5 shows sample images and the category distribution for the LSAFBD dataset. Specifically, Figure 5b shows the category distribution for the LSAFBD dataset, which approximates a normal distribution. Figure 5a displays some sample images from the LSAFBD dataset, demonstrating that the images have been meticulously cropped with pixels concentrated on the facial area, thus minimizing the influence of irrelevant features

and facilitating feature extraction for transfer learning. In this study, the LSAFBD dataset was primarily utilized for the extraction of feature information, which serves as the reconstruction target when employing StyleGAN2-ADA-s to generate synthetic images.

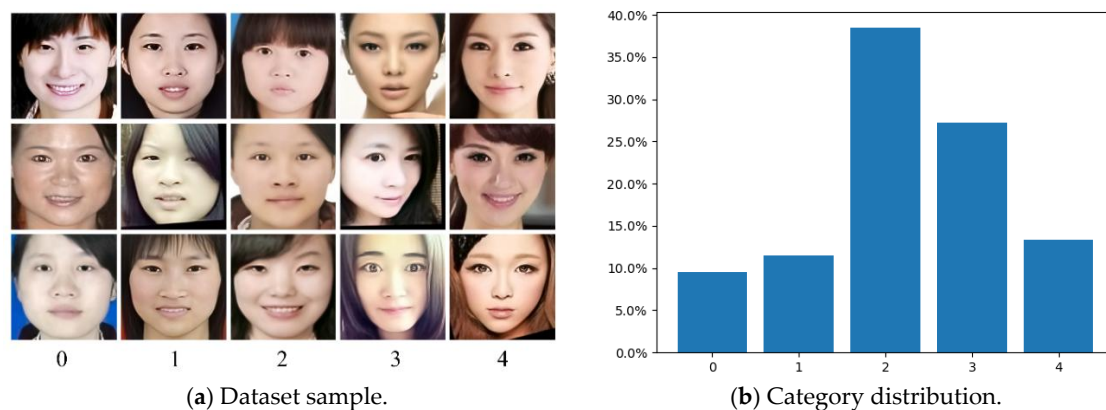


Figure 5. LSAFBD dataset sample and category distribution.

MEBeauty (Multi-ethnic Facial Beauty Dataset in-the-wild) [31] is a multi-ethnic FBP dataset designed for real-world application scenarios. The dataset contains 2334 valid facial images, covering male and female samples from multiple ethnic groups, including Black, Asian, Caucasian, Hispanic, Indian, and Middle Eastern. MEBeauty is primarily used for regression tasks in FBP, with performance evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient. Since this study focuses on the classification task of FBP, the facial images in MEBeauty are categorized into three classes: “0,” “1,” and “2,” corresponding to “unattractive,” “average,” and “attractive,” respectively. Figure 6 presents sample images and the class distribution of facial beauty scores for the MEBeauty dataset. By establishing 3.5 and 6.5 as critical thresholds, we divided the scores into three categories, resulting in an approximately normal distribution. Due to its high ethnic diversity and limited training samples, this may lead to latent-space drift during feature transfer, the dataset poses significant challenges to the generalization ability of models [32].

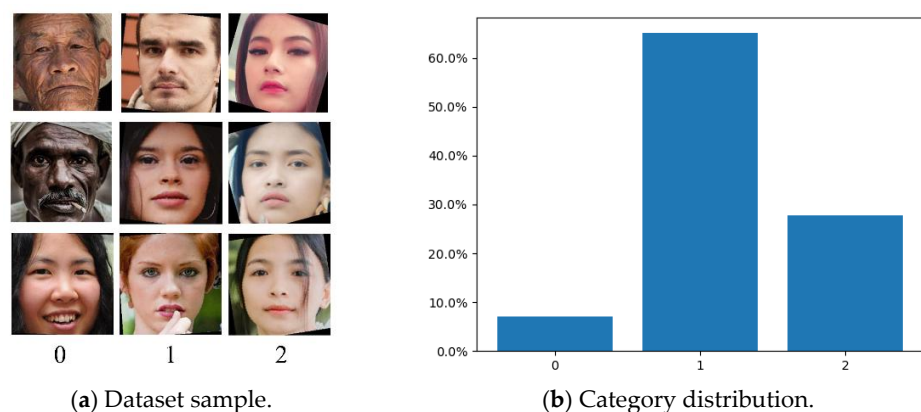


Figure 6. MEBeauty dataset sample and category distribution.

The GAN-LSA-FBP data refer to synthetic images generated using StyleGAN2-ADA-s. These images were obtained by reconstructing samples from the LSAFBD dataset [15] using the feature space mapping learned from the SCUT-FBP5500 dataset [14]. To ensure that the categories of synthetic images align with those of the original dataset, StyleGAN2-ADA-s requires category information as input during image generation. The reconstruction results are visualized in Figure 7. The reconstruction quality was evaluated using the

Mean Opinion Score (MOS): a perceptual metric ranging from 1 to 5, with higher scores indicating greater photorealism. Although SCUT-FBP5500's limited data volume compromised the ability to realistically reconstruct environmental lighting conditions, the model achieved exceptional facial texture recovery. The overall image clarity consistently exceeded that of the original LSAFBD samples.

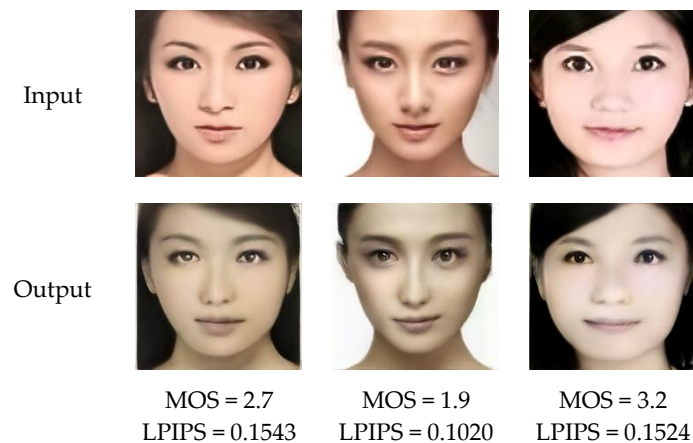


Figure 7. Image reconstruction performance of StyleGAN2-ADA-s.

4.2. Experimental Environment

The experiments were conducted on a computer equipped with an NVIDIA GeForce RTX 4060 Ti GPU, an Intel Core i5-10505 CPU, and 64G DDR4 memory, operating under Microsoft Windows 10. The Python version utilized was 3.9.18, employing PyTorch 2.4.1 as the deep learning framework with CUDA 12.4. Table 1 presents the pre-training phase of StyleGAN2-ADA-s. During the pre-training phase of StyleGAN2-ADA-s, the mapping depth was set to 2, the learning rate to 0.0025, the batch size to 16, the total_kimg to 3500, the snapshot interval to 50, and the R1 regularization strength to 0.8192, utilizing the Adam optimizer.

For the training phase of the classification network, the data were augmented according to specific category ratios, followed by traditional data augmentation methods. The final images were resized to an actual resolution of 224×224 , enabling them to serve as inputs to the model. The traditional data augmentation processes included random cropping, random horizontal flipping, and normalization, and the image order was randomized during training. During testing, the augmentation processes consisted of center cropping and normalization, which were sequentially applied on the test set. In the evaluation stage, data augmentation involved center cropping and normalization, with the test images shuffled based on a fixed seed to ensure more accurate model prediction assessment while maintaining the stability of the experimental results. During training, pre-trained weights from ImageNet-1K were incorporated [33], with a batch size of 32 and iteration count of 45, employing an initial learning rate of 0.0125 and the SGD optimizer.

Table 1. The pre-training phase of StyleGAN2-ADA-s.

Parameters	Recommended Values
mapping depth	2
lr	0.0025
batch size	16
kimg	3500
snapshot	50
r1	0.8192

optimizer	Adam
Seed count	1

4.3. Evaluation Metrics and Augmentation Parameters

In this study, Accuracy (ACC), Macro-F1, and Evaluate Accuracy (*Evaluate ACC*) were employed as the primary evaluation metrics. Evaluation AC is calculated as ACC on a shuffled test set, and is computed in the same way as ACC . The relevant formulas are as follows:

$$ACC = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)} \times 100\%, \quad (2)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \times 100\%, \quad (3)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \times 100\%, \quad (4)$$

$$Precision_{Macro} = \frac{1}{n} \sum_{i=1}^n Precision_i, \quad (5)$$

$$Recall_{Macro} = \frac{1}{n} \sum_{i=1}^n Recall_i, \quad (6)$$

$$F1_{Macro} = \frac{2 \times Precision_{Macro} \times Recall_{Macro}}{Precision_{Macro} + Recall_{Macro}} \times 100\%. \quad (7)$$

The total synthetic sample count S_{total} and the total original sample count T_{total} are defined as

$$S_{total} = \sum_{i=0}^4 S_i, \quad (8)$$

$$T_{total} = \sum_{i=0}^4 T_i, \quad (9)$$

where i represents the level of facial attractiveness, T_i denotes the number of real samples in the training set for category i , and S_i indicates the number of synthetic samples for category i .

The expansion ratio ER in the experiment is defined as the ratio between the number of synthetic samples S_{total} and the expansion reference amount N_{refer} . The definitions of the expansion ratio ER and the total number of samples EP are, respectively,

$$ER = \left(\frac{S_{total}}{N_{refer}} \times 100 \right) \%, \quad (10)$$

$$EP = \left(\frac{T_{total} + S_{total}}{T_{total}} \times 100 \right) \%, \quad (11)$$

where ER denotes how much the current dataset's class distribution is shifted toward another distribution, while EP reflects the strength of the data augmentation process. The definitions of the actual number of augmented samples per category S_i is

$$S_i = \begin{cases} R_i N_i (R_i = ER), p_{target} \in \{p_{SCUT-FBP5500}, p_{LSAFBD}\} \\ R_i N_i (R_i = V_i^\alpha), p_{target} = V \end{cases} \quad \forall i \in \{0, 1, 2, 3, 4\}, \quad (12)$$

where R_i represents the expansion ratio of samples for the i th category; N_i denotes the reference quantity for expanding the i th category's samples; p_{target} is the target class distribution, which serves as the distribution that the current dataset's class distribution is to approach; V stands for Omega class distribution; and $p_{SCUT-FBP5500}$ and p_{LSAFBD} represent the SCUT-FBP5500 and LSAFBD class distributions, respectively, which are the

target class distributions for expansion. In this study, we primarily investigate how to enhance the classification network's performance using a limited amount of synthetic data; hence, P was set to a maximum of 150%.

4.4. Training and Evaluation Based on the Transformer Model

MobileViT is a lightweight visual Transformer model that integrates the local bias of CNNs with the ViT's global modeling capabilities, achieving efficiency akin to a CNN and performance comparable to ViT. To enhance the operational efficiency of FBP-GAN, for experiments based on the Transformer model, this study employs MobileViT as the classification network for FBP-GAN (designated as FBP-GAN-M). In particular, the utilized MobileViT variant was MobileViT-S, with the benchmark network also being MobileViT-S for comparison. In the experiments, the suffixes MA and MB indicate that the augmented label distributions are respectively aligned with those of LSAFBD (Figure 4b) and SCUT-FBP5500 (Figure 5b), whereas MO denotes alignment with the proposed Omega distribution (Figure 8).

When augmenting data based on the category distribution of the auxiliary LSAFBD dataset, the proposed method is termed FBP-GAN-MA, where N_{refer} denotes the total sample count of the LSAFBD training set, N_i represents the quantity of samples in class i in the LSAFBD training set, the maximum EP is set to 136%, p_{target} indicates the LSAFBD category distribution, and the maximum ER is set to 20%. The prediction results obtained with FBP-GAN-MA models are presented in Table 2, from which it is evident that, as the EP value increases, all metrics of the FBP-GAN-MA models improve: ACC reaches a maximum of 76.75%, representing a 0.92% improvement over the baseline network's 75.83%; Evaluate ACC peaks at 76.96%, a 2.37% increase compared to the baseline network's 74.59%; and Macro-F1 reaches 50.78%, marking a 5.68% enhancement over the baseline network's 45.10%.

Table 2. Comparison experiment results of FBP-GAN-MA on the SCUT-FBP5500 dataset.

Method	ER	EP	Macro-F1	ACC	Evaluate ACC
Baseline	0	100%	45.10	75.83	74.59
FBP-GAN-MA	10%	118%	45.15	75.92	74.86
FBP-GAN-MA	20%	136%	50.78	76.75	76.96

Note: Bolded numbers indicate the best results in each column.

When performing data augmentation based on the category distribution of the primary SCUT-FBP5500 dataset, the proposed method is denoted as FBP-GAN-MB, N_{refer} signifies the total sample count of the SCUT-FBP5500 training set, N_i denotes the quantity of the sample in class i in the SCUT-FBP5500 training set, the maximum EP is configured at 150%, p_{target} indicates the SCUT-FBP5500 category distribution, and the maximum ER is set to 50%. The prediction outcomes for FBP-GAN-MB models are detailed in Table 3, with the experimental results demonstrating that a greater sample volume EP correlated with higher Macro-F1, ACC, and Evaluate ACC for FBP-GAN-MB models: ACC peaks at 76.84% versus the baseline network's 75.83%, reflecting a 1.01% improvement; Evaluate ACC reaches 77.32% versus 74.59%, marking a 2.73% increase; Macro-F1 achieves 50.42% versus 45.10%, representing a 5.32% enhancement; and performance gains across the metrics show negligible variance between the LSAFBD and SCUT-FBP5500 category distributions.

Table 3. Comparison experiment results of FBP-GAN-MB on the SCUT-FBP5500 dataset.

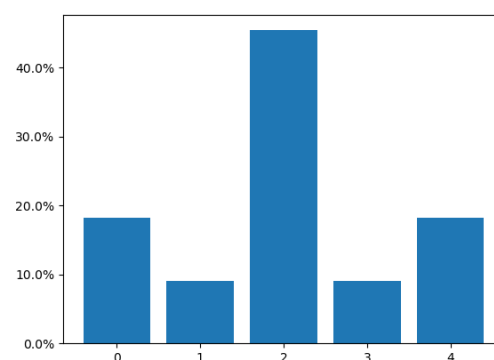
Method	ER	EP	Macro-F1	ACC	Evaluate ACC
Baseline	0	100%	45.10	75.83	74.59
FBP-GAN-MB	25%	125%	46.38	76.01	75.96
FBP-GAN-MB	50%	150%	50.42	76.84	77.32

Note: Bolded numbers indicate the best results in each column.

In the SCUT-FBP5500 dataset [14], there is a significant difference in the number of face samples across different beauty levels, with both the 0 and 4 category samples accounting for less than 5% each. To mitigate category shift-induced training bias, this study referenced classic imbalanced learning research [34] and conducted targeted expansion of categories with fewer samples to further enhance the FBP-GAN-M model's discrimination capabilities across all categories and its overall generalization performance, while simultaneously minimizing the number of synthetic samples. Additionally, regarding the targeted expansion strategy, we controlled the expansion intensity to ensure that the distribution of expanded categories remained overall consistent with the original distribution, thus avoiding issues such as overfitting or distribution drift.

To conduct targeted expansion, this study proposes an ideal category distribution, namely, the Omega category distribution, which is specifically designed for FBP and is shown in Figure 8. The design inspiration for the Omega category distribution stems from human cognitive habits. When faced with previously unseen objects, people instinctively combine rich prior knowledge about familiar items to infer the attributes of the unknown—filling in gaps through an assembly of prevalent known features [35]. This mechanism exemplifies human imagination, and Omega seeks to emulate such cognitive patterns. Its distribution shape is similar to the Chinese character 'shan', serving as the target category distribution for directed expansion.

When using the Omega category distribution to augment data, this method is referred to as FBP-GAN-MO, with N_{refer} being the total number of samples in the SCUT-FBP5500 training set and N_i representing the number of samples for the i th category in the SCUT-FBP5500 training set, with p_{target} set as the Omega category distribution.

**Figure 8.** Ideal category distribution for FBP.

The Omega expansion ratio is a targeted expansion ratio designed to approach the ideal category distribution, aiming to make the sample size in the five categories conform to the ratio 2:1:5:1:2 as closely as possible. The focus is on expanding samples with scarce quantities while keeping the number of alternative category samples unchanged, thereby bringing the sample category distribution closer to the Omega category distribution. The calculation formula for the Omega expansion ratio R_i^* is

$$R_i^* = \frac{S_i + T_i}{T_{total} + S_{total}} \approx V_i, V = \left(\frac{2}{11}, \frac{1}{11}, \frac{5}{11}, \frac{1}{11}, \frac{2}{11} \right) \quad (13)$$

In this method, the Omega expansion ratio is implemented by directly expanding the sample quantities of the 0 and 4 categories in the SCUT-FBP5500 training set to twice their original numbers, making the category distribution approximate to the ideal distribution. The expansion ratio is denoted as V^α , with the actual value of R_i being

$$R_i = V_i^\alpha, V^\alpha = (100\%, 0, 0, 0, 100\%). \quad (14)$$

The experimental results are shown in Table 4, which indicate that the FBP-GAN-MO model demonstrated excellent performance, achieving an Evaluate ACC of 77.69% with just 103% total sample quantity, comprising a 3.1% improvement over the baseline network. This surpasses FBP-GAN-MA and FBP-GAN-MB, whose sample totals far exceeded 103%. Macro-F1 and ACC presented improvements of 3.97% and 0.55%, respectively, when compared to the baseline network. Figure 9 illustrates the loss and ACC change curves obtained during training of the FBP-GAN-M-related methods, from which it can be observed that data expansion has little effect on the stability of the training process, and that FBP-GAN-M shows certain robustness with respect to the FBP task.

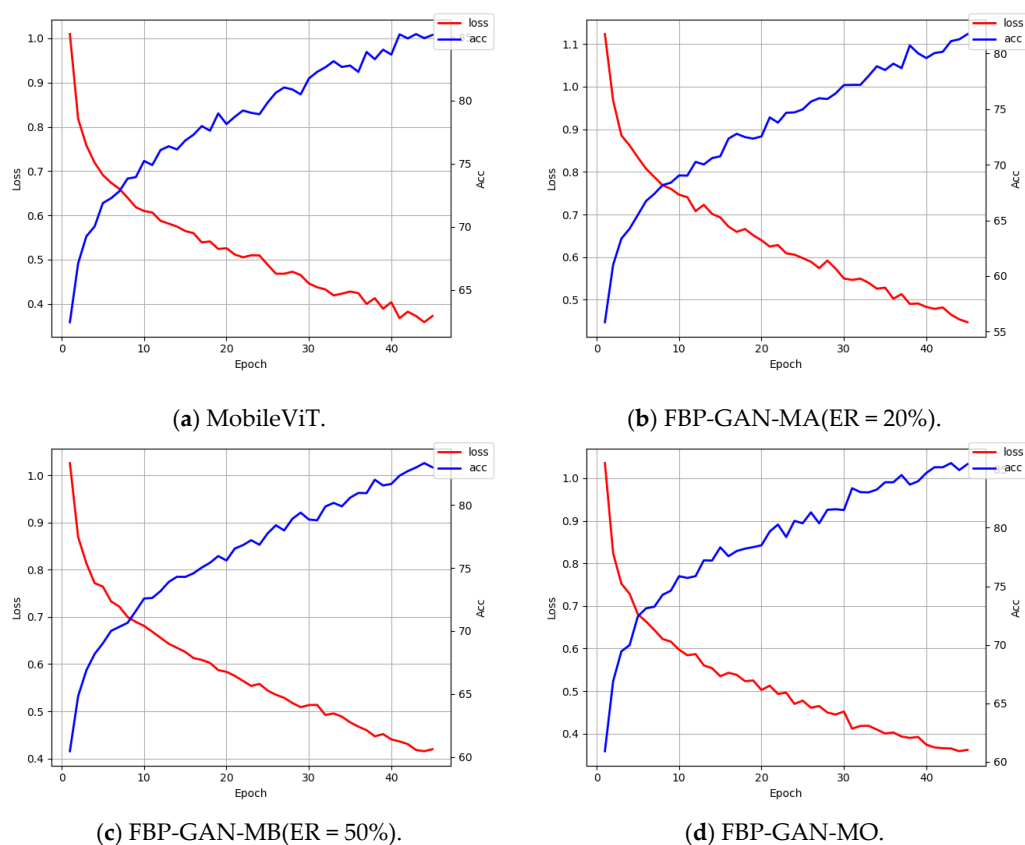


Figure 9. Accuracy and loss curves during Transformer model training.

Table 4. Comparison experiment results of FBP-GAN-MO on the SCUT-FBP5500 dataset.

Method	ER	EP	Macro-F1	ACC	Evaluate ACC
Baseline	0	100%	45.10	75.83	74.59
FBP-GAN-MO	3%	103%	49.07	76.38	77.69

Note: Bolded numbers indicate the best results in each column.

The demand for data volume and capacity in Transformer models exceeds that of CNN models [36], which is a crucial factor enabling the FBP-GAN method to function when using MobileViT. From the confusion matrix in Figure 10, it can be seen that MobileViT has low recognition accuracy for images categorized as 0; after incorporating FBP-GAN, the recognition accuracy significantly improved, with accuracy for nearby categories also increasing. This indicates that the bottleneck in MobileViT's classification performance on the SCUT-FBP5500 dataset [14] derives from the severe lack of data in the 0 category, leading to its reduced capability to judge 'very unattractive' faces. From an aesthetic perspective, as faces tend towards 'very unattractive', their feature richness and complexity far exceed those deemed 'very attractive' [37]. This is one of the key factors preventing deep learning models from achieving ultra-high prediction accuracy on datasets with category distributions approaching a normal distribution. If the dataset's distribution is further optimized, most models can be expected to show an improvement in accuracy for FBP tasks.

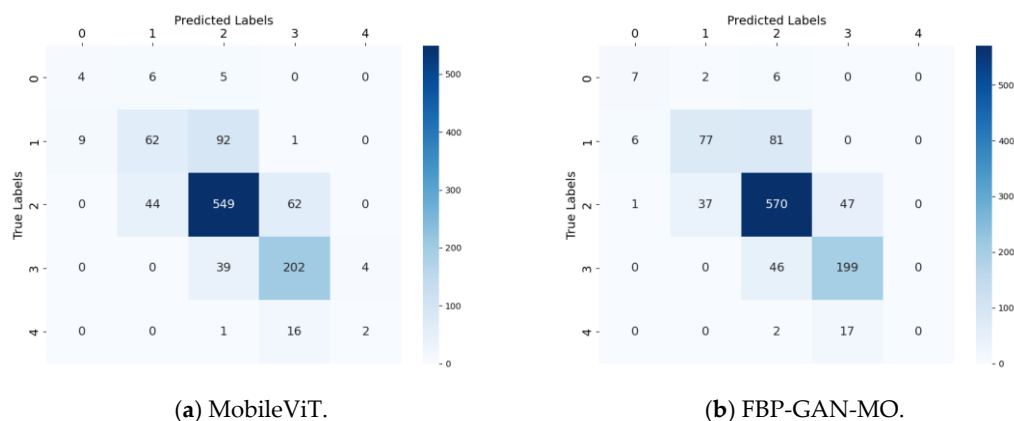


Figure 10. Confusion matrices for the Transformer-based models.

4.5. Training and Evaluation Based on CNN Models

ResNeXt is an upgraded version of ResNet. Although ResNeXt is not a lightweight model, its performance has been improved compared to that of ResNet. To better demonstrate the effect of the designed FBP-GAN, the medium-sized ResNeXt model was selected for CNN model experiments.

When the classification network of FBP-GAN is ResNeXt-50 and the Omega expansion ratio is applied, this method is referred to as FBP-GAN-RO. The same expansion method was used, in order to double the number of samples in the 0 and 4 categories in the training set.

In the experiments on Transformer models detailed in Section 4.3, we have already proven the designed expansion method to be relatively balanced. Therefore, only the Omega expansion ratio was tested for the CNN models, with the experimental results shown in Table 5. The baseline network for comparison was ResNeXt-50, while the experimental results indicate that, compared to the baseline network's ACC of 76.29% and Evaluate ACC of 75.14%, FBP-GAN-RO achieved improvements of 1.65% and 1.91%, respectively, with these results aligning with our expectations.

Table 5. Comparison experiment results of FBP-GAN-RO on the SCUT-FBP5500 dataset.

Method	ER	EP	Macro-F1	ACC	Evaluate ACC
Baseline	0	100%	51.20	76.29	75.14
FBP-GAN-RO	3%	103%	51.56	77.94	77.05

Note: Bolded numbers indicate the best results in each column.

Figure 11 shows the training accuracy and loss curves for ResNeXt-50 and FBP-GAN-RO. It can be observed that, compared to ResNeXt-50, the accuracy and loss curves of FBP-GAN-RO are slightly steeper but eventually trend toward stability as the number of iterations increases. From the confusion matrices shown in Figure 12, it is evident that FBP-GAN-RO achieved a significant improvement in recognition accuracy for images in category 0, and a slight improvement for those in category 3. The experimental results demonstrate that FBP-GAN also performs well when used with medium-sized CNN models.

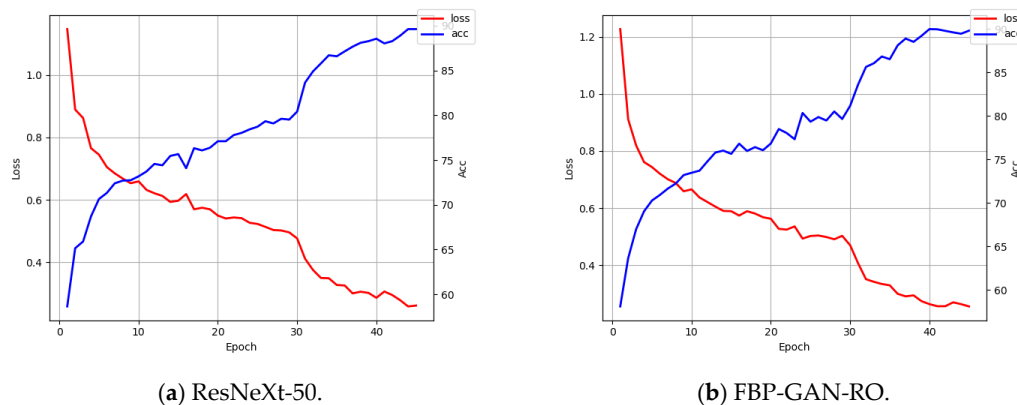


Figure 11. Accuracy and loss curves during CNN model training.

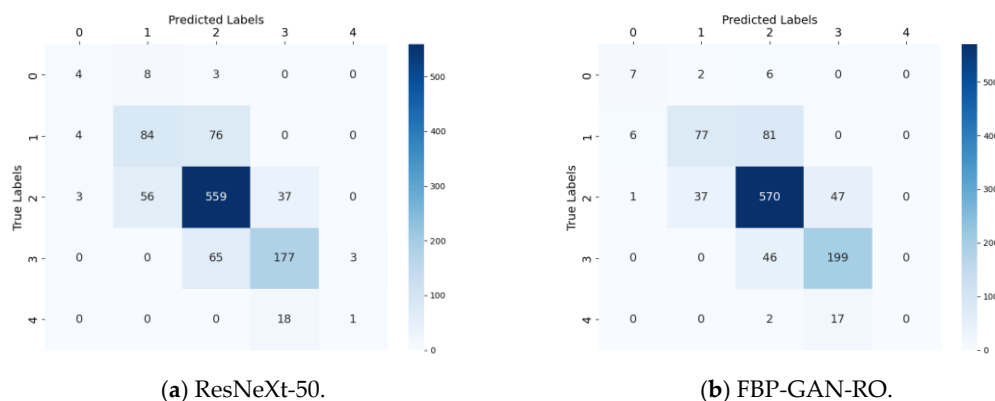


Figure 12. Confusion matrices of CNN-based models.

4.6. Migrate to Other Datasets

MEBeauty is a small-scale dataset with high complexity for FBP tasks. To evaluate the generalizability of FBP-GAN, we adapted it to the MEBeauty dataset. Given that MEBeauty primarily focuses on FBP regression tasks, we designed a simplified conversion framework to enable effective application of FBP-GAN.

As illustrated in Figure 13, the conversion process begins by transforming linear scores into nine discrete score bins. Subsequently, the number of images in the leftmost and rightmost bins is doubled to approximate an Omega distribution. These nine bins are then merged into three classes to form an augmented dataset suitable for FBP classification. Finally, FBP-GAN's classification network is trained on this three-class augmented dataset to perform the FBP classification task. This framework enables more refined data augmentation and enhances augmentation efficiency.

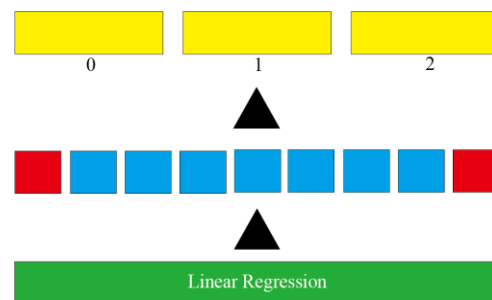


Figure 13. FBP Regression-to-Classification Conversion Framework. The green rectangle represents the dataset with linear scores. The blue rectangles represent the classes in the 9-class dataset that were not involved in the augmentation parameter calculation, while the red rectangles represent the classes that were used in the augmentation parameter calculation. The yellow rectangles denote the actual 3-class dataset used for training and testing; FBP-GAN’s generator network is also trained on this 3-class dataset.

For small-scale datasets, the CNN-based FBP-GAN-RO model is recommended. As shown in Table 6, FBP-GAN-RO achieved 72.99% accuracy, 57.34% Macro-F1 score, and 72.38% Evaluation Accuracy on the MEBeauty dataset, surpassing the ResNeXt-50 baseline by 1.56, 14.16, and 1.5 percentage points, respectively.

These results demonstrate the robust performance of FBP-GAN-RO on small-scale, complex FBP datasets and validate the transferability of the FBP-GAN framework to diverse datasets.

Table 6. Experimental Results of FBP-GAN-RO on the MEBeauty Dataset.

Method	ER	EP	Macro-F1	ACC	Evaluate ACC
Baseline	0	100%	43.18	71.43	70.88
FBP-GAN-RO	1%	101%	57.34	72.99	72.38

Note: Bolded numbers indicate the best results in each column.

4.7. Statistical Significance Test

To verify whether the performance advantage of the proposed method is statistically significant, we conducted paired-samples t-tests on the results of five independent runs. In each run, the identical training/validation split was employed to ensure pairwise comparability of ACC between the proposed method and the baseline. The results are shown in Table 7; two-tailed tests revealed that the differences in both metrics were statistically significant.

Table 7. Paired *t*-test *p*-values for ACC between the proposed method and the baseline.

Method	Baseline	SCUT-FBP5500 (<i>p</i> -Value)
FBP-GAN-MO	MobileViT-S	$p = 0.0025$
FBP-GAN-RO	ResNeXt-50	$p < 0.001$

4.8. Comparison with Other Methods

To further verify the effectiveness of the FBP-GAN method, it was compared with other methods previously used in the context of FBP. The comparison results are shown in Table 8. In one study [38], the data quality was improved through label correction to enhance the model’s classification accuracy, while another [39] utilized width learning systems to obtain a lightweight FBP model. On the other hand, [40] integrated attention mechanisms and multi-task learning into width learning for FBP improvement, and [37] proposed a multi-task learning method that combines an Adaptive Sharing Policy (ASP)

with Attentional Feature Fusion (AFF), based on an adaptive shared network framework, to enhance performance in FBP tasks. Our proposed method can be seen to demonstrate superior prediction accuracy when compared to these other methods.

Table 8. Comparative experiments between proposed method and the other FBP methods (ACC).

Method	SCUT-FBP5500	MEBeauty
ResNeXt-50 [13]	76.29	71.43
Noise Labels [38]	75.30	—
E-BLS [39]	73.13	—
ER-BLS [39]	74.69	—
TransBLS-T [40]	75.23	—
Adaptive multi-task method [41]	75.41	—
<i>FBP-GAN-MO (ours)</i>	76.38	—
<i>FBP-GAN-RO (ours)</i>	77.94	72.99

Note: The text in italics represents our method.

In summary, the method proposed in this study uses the latent features of images from the LSAFBD dataset [15] for reconstruction of the SCUT-FBP5500 dataset [14], thus achieving targeted data augmentation for the latter. This enables effective optimization of the category distribution, further unlocking the inherent potential of the original dataset. This method provides a feasible and effective approach to address data scarcity issues in the context of FBP tasks. Additionally, it was proven to improve the accuracy of MobileViT and ResNeXt models on the SCUT-FBP5500 dataset.

4.9. Methodological Limitations

Firstly, although FBP-GAN effectively enhances the performance of classification networks, the training of the generative network and the image synthesis stage require a relatively long time, which increases the overall training cost and makes it unsuitable for environments with strict computational constraints. Secondly, the modular design of the FBP-GAN framework improves flexibility but results in weak inter-module correlations, limiting the construction of a highly integrated and generalizable model while increasing computational overhead. Future research may consider integrating the core concepts of FBP-GAN into the design of classification networks to further enhance overall performance. Finally, during data synthesis, FBP-GAN has a small probability of generating unrecognizable or distorted images, which can introduce noise and potentially degrade the quality of the dataset.

5. Conclusions

The proposed FBP-GAN is an innovative data augmentation method tailored to FBP datasets, ingeniously utilizing GANs for data augmentation without compromising the dataset's quality or feature distribution. It serves as a crucial approach to address FBP data challenges and achieve significant and consistent improvements in FBP tasks. This method achieved favorable results when paired with both Transformer- and CNN-based FBP models. Furthermore, we identified category distribution issues common to FBP-related datasets, allowing for the proposal of an idealized category distribution for further research. These advances open new directions for in-depth studies of FBP and provide fresh insights into the application of GANs in complex computer vision tasks. In future work, we intend to explore the incorporation of images of faces from different ethnic groups to address data bias [42], and to conduct experiments using other generative network architectures, such as SANA [43], to improve image generation efficiency.

Author Contributions: Conceptualization, J.G. and Z.C.; methodology, Z.C.; software, Z.C.; validation, Z.C., H.C. and W.X.; formal analysis, Z.C.; investigation, Z.C., Z.Z. and J.X.; resources, J.G.; data curation, Z.C.; writing—original draft preparation, Z.C.; writing—review and editing, Z.C. and H.C.; visualization, Z.C.; supervision, J.G.; project administration, Z.C. and J.G.; 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 upon request from the corresponding author. The data are not publicly available due to privacy. MEBeauty: The dataset is publicly available at <https://github.com/fbplab/MEBeauty-database> (accessed on 3 June 2024).

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

References

1. Saeed, J.; Abdulazeez, A.M. Facial beauty prediction and analysis based on deep convolutional neural network: A review. *J. Soft Comput. Data Min.* **2021**, *2*, 1–12.
2. Chen, C.L.P.; Liu, Z.; Feng, S. Universal approximation capability of broad learning system and its structural variations. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *30*, 1191–1204.
3. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90.
4. Lin, L.; Liang, L.; Jin, L. Regression Guided by Relative Ranking Using Convolutional Neural Network (R3 CNN) for Facial Beauty Prediction. *IEEE Trans. Affect. Comput.* **2022**, *13*, 122–134.
5. Lin, L.; Liang, L.; Jin, L.; Chen, W. Attribute-Aware Convolutional Neural Networks for Facial Beauty Prediction. In Proceedings of the International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 847–853.
6. Xu, L.; Xiang, J.; Yuan, X. Transferring rich deep features for facial beauty prediction. *arXiv* **2018**, arXiv:1803.07253.
7. Zhai, Y.; Cao, H.; Deng, W.; Gan, J.; Piuri, V.; Zeng, J. BeautyNet: Joint multiscale CNN and transfer learning method for unconstrained facial beauty prediction. *Comput. Intell. Neurosci.* **2019**, *2019*, 1910624.
8. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Proceedings of the Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition, Marseille, France, 17 October 2008.
9. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 3730–3738.
10. Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 87–102.
11. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
12. Mehta, S.; Rastegari, M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.
13. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
14. Liang, L.; Lin, L.; Jin, L.; Xie, D.; Li, M. SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In *Proceedings of the 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018*; IEEE: New York, NY, USA, 2018; pp. 1598–1603.
15. Zhai, Y.; Huang, Y.; Xu, Y.; Gan, J.; Cao, H.; Deng, W.; Labati, R.D.; Piuri, V.; Scotti, F. Asian female facial beauty prediction using deep neural networks via transfer learning and multi-channel feature fusion. *IEEE Access* **2020**, *8*, 56892–56907.

16. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
17. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
18. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 214–223.
19. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein GANs. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5767–5777.
20. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
21. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4401–4410.
22. Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training generative adversarial networks with limited data. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12104–12122.
23. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8110–8119.
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
25. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307.
26. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
27. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
28. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
29. Wen, F.F.; Zuo, B. The influence of masculinity and femininity on face preference: Evidence from image processing technology and eye movement. *Acta Psychol. Sin.* **2012**, *44*, 14–29. <https://doi.org/10.3724/SP.J.1041.2012.00014>.
30. Dai, L.Q.; Jin, Z.; Sun, M.M. Facial geometric beauty score based on semi-supervised regression learning. *Comput. Appl. Softw.* **2015**, *32*, 209–211.
31. Lebedeva, I.; Guo, Y.; Ying, F. MEBeauty: A multi-ethnic facial beauty dataset in-the-wild. *Neural Comput. Appl.* **2022**, *34*, 14169–14183.
32. Sumsion, A.; Torrie, S.; Lee, D.J.; Sun, Z. Surveying racial bias in facial recognition: Balancing datasets and algorithmic enhancements. *Electronics* **2024**, *13*, 2317.
33. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: New York, NY, USA, 2009; pp. 248–255.
34. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259.
35. Pearson, J. The human imagination: The cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* **2019**, *20*, 624–634.
36. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021), Virtual Event, 3–7 May 2021.
37. Lee, P.; Li, J.; Rafiee, Y.; Jones, B.C.; Shiramizu, V.K.M. Further evidence that averageness and femininity, rather than symmetry and masculinity, predict facial attractiveness judgments. *Sci. Rep.* **2025**, *15*, 5498.
38. 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.
39. 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.
40. 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.

41. 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* **2023**, *13*, 179.
42. Yeung, M.; Teramoto, T.; Wu, S.; Fujiwara, T.; Suzuki, K.; Kojima, T. VariFace: Fair and Diverse Synthetic Dataset Generation for Face Recognition. *arXiv* **2024**, arXiv:2412.06235.
43. Xie, E.; Chen, J.; Chen, J.; Cai, H.; Tang, H.; Lin, Y.; Zhang, Z.; Li, M.; Zhu, L.; Lu, Y.; et al. SANA: Efficient high-resolution text-to-image synthesis with linear diffusion transformers. In Proceedings of the Thirteenth International Conference on Learning Representations, Singapore, 24–28 April 2025.

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.