ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096554

CUSTOMIZED AUTOMATIC FACE BEAUTIFICATION

Wang Chen[1], *Peizhen Chen*[1], *Weijie Chen*[2], *Luojun Lin*[1✉]

[1]Fuzhou University, Fuzhou, China
[2]Zhejiang University, Hangzhou, China

ABSTRACT

In the age of social media, posting attractive mugshots is commonplace, leading to an urgent need for automatic facial beautification techniques. To better meet the esthetic preferences of users, we devise a customized automatic face beautification task that can retouch the face adaptively to match the user-entered target score whilst preserving the ID information as much as possible. To accomplish this task, we propose a Human Esthetics Guided StyleGAN Inversion method to retouch each face in the embedding space using StyleGAN inversion. This process is guided by a pre-trained facial beauty prediction model that measures the difference between the target score and the predicted score of the retouched face. We conduct extensive experiments on various faces with different attributes, where the experimental results show that our method achieves the competitive performance, both in terms of visual effect and the proposed criterion.

Index Terms— Face beautification, Facial beauty prediction, GAN inversion, StyleGAN

1. INTRODUCTION

Sociologists and psychologists have proved that attractive faces tend to bring more advantages and privileges in human activities, such as social acceptance, career development, interpersonal relationships, and social status [1]. The attractive face plays such an important role in our social life that people are generally obsessed with beautifying their facial images (*e.g.*, selfies). Especially with the advent of mobile Internet, various applications have emerged, such as dating websites, live broadcasts, short videos, and so on. There is an urgent need for face beautification techniques.

Recent advances in facial beautification algorithms have focused on retouching facial skin texture by developing smoothing filters [2, 3], or transferring makeup from a reference image [4–6]. Another line of works merely aim to ameliorate facial geometric ratios via matching face geometry to an average facial template [7], or using deep models to guide the modification of facial landmarks [8]. However, most of these beautification methods usually do not take into account the user aesthetic preferences when editing facial images. Actually, users tend to customize their face beautification requirements. For example, some people prefer to preserve more identity information, while others prefer a more obvious beautification effect. For this purpose, we develop a new task, namely, *Customized Automatic Face Beautification* (see Fig. 1), which aims to retouch facial images based on the user aesthetic preferences (*i.e.*, user-entered target score), whilst preserving the face ID as much as possible by making minimal changes to the face geometry, expression, and pose.

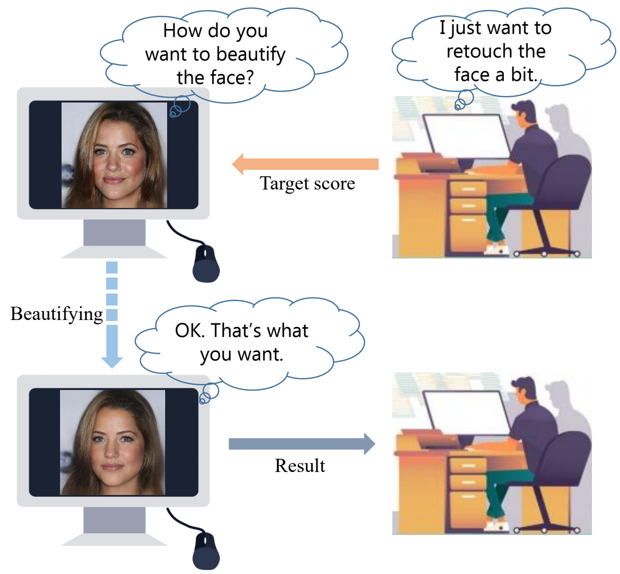


Fig. 1. Customized automatic face beautification task.

Since StyleGAN [9] has achieved considerable success in image editing tasks, we consider applying it to tackle customized automatic face beautification as well. Based on StyleGAN, we propose a *Human Esthetics Guided StyleGAN Inversion (HEGS inversion)* method that introduces a Facial Beauty Prediction (FBP) model to guide the inversion of StyleGAN to manipulate face retouching in the embedding space rather than in the image space. The pipeline of our method is shown in Fig. 2, where an advanced StyleGAN model is used to convert each input face into a content code

Wang Chen and Peizhen Chen contribute equally.
Luojun Lin is the corresponding author (ljlin@fzu.edu.cn).
This work was supported by the Fujian Provincial Natural Science Foundation (No. 83322030); and the Fujian Provincial Youth Education and Scientific Research Project (No. 650722).

and an appearance code through specific encoders firstly, and then integrate these codes into the generative model to produce an edited face. The FBP model calculates the distance loss between the target score and the predicted score of the edited face, and the loss is back-propagated to iteratively refine the content code and appearance code until the FBP score of the edited face matches the target score. Under the guidance of FBP model, the content and appearance codes will be optimized in the direction that most closely matches the user preferences.

To better quantify the effect of face beautification, we also propose a new evaluation criterion for such a customized face beautification task, which measures the trade-off between ID preservation (i.e., the distance between the original face and edited face) and beautification effect (i.e., the distance between the target score and the predicted FBP score of edited face). Our method was evaluated on face images from different populations, e.g., different ages and ethnic groups. The experimental results have shown that compared to other face beautification methods, our method can achieve an adaptive beautification effect with the user-entered target score and preserve more ID information, which stresses the significance of our method. Our contributions are summarized as follows:

- We define a new task, namely Customized Automatic Face Beautification, which aims to beautify facial images according to the user aesthetic preferences.
- We propose a Human Esthetics Guided StyleGAN Inversion method to achieve the above new task through the guidance of FBP model, which is the first work of Face Beautification based on GAN inversion.
- We design a new evaluation criterion to quantify the effect of face beautification algorithms that measures ID preservation and beautification effect simultaneously.
- In FBP community, there is a debate over whether FBP can be learned by machine learning since it is a subjective task. Our work answers this question from a model inversion aspect and will bring inspirations to the FBP community and the aesthetic cognitive psychology.

2. RELATED WORKS

Face Beautification. There are two common directions to solve face beautification. The first one is skin texture beautification, which aims to reduce blemishes and increase skin luster by using a perceptual filter [2, 10], or transferring makeup from a reference face [11–13]. The second one is face geometry beautification, which morphes the mask of the input face into a more attractive one, *e.g.*, by deforming the face mask to a reference mask [14–16], matching to an average face template [7], or using a deep model to regress landmark coordinates for morphing [8]. Most of them do not take into account individual aesthetic preferences. So they may not produce the most pleasing result due to different user preferences.

Facial Beauty Prediction. Psychologists have found that human perceptions of facial beauty are highly consistent across different races, ages, genders, and cultures [17], which encourages the development of machine-based facial beauty prediction (FBP). Traditional FBP methods extract handcrafted features [18, 19], and more recent FBP methods use deep networks [20–22] to achieve superior performance. It suggests that the FBP model has captured human esthetic information so that it can be used for face beautification.

GAN Inversion. Given an image and a pre-trained model, GAN inversion aims to obtain the latent code that can be traced back to the given image. Existing GAN inversion methods can be divided into the following three categories. (1) Optimizing the latent code by minimizing the reconstruction error between the given image and the reconstructed image [9, 23, 24]. (2) Using an encoder to map the given image to the latent space directly [25]. (3) A hybrid approach combining both optimization and encoder techniques [26,27]. To speed up the algorithm, we develop the encoder-based inversion method in this paper.

3. PROPOSED METHOD

As shown in Fig. 2, our method consists of two parts: 1) Image reconstruction, which generates fake face from content code; 2) StyleGAN Inversion, which regresses content code from the fake face. These two parts run alternatively until convergence. Let us describe the details of each part below.

3.1. Preliminaries: start from generating code

Since we found that the original StyleGAN [9] has excellent editability but lower reconstruction quality, we decided to employ HyperInverter [28], an extended StyleGAN that introduces hypernetworks to improve the reconstruction quality. Before reconstruction, we encode each input face x into the latent code $\omega_c \in \mathcal{W}$ via an encoder E with weights of θ_e:

$$\omega_c = E(x|\theta_e), \tag{1}$$

where $\omega_c(\omega_c \in \mathbb{R}^{512})$ is a content code that encodes coarse information about the facial content, including facial structure, expression, skin tone, etc.

3.2. Image Reconstruction: from code to image

The content code ω_c is fed into the generator G with weights of θ_g to reconstruct the face:

$$x_\omega = G(\omega_c|\theta_g). \tag{2}$$

Actually, the output x_ω is reconstructed at a coarse level due to the low dimension of the content code, which may cause most of fine-grained information to be lost. Hence, we introduce a new stage to refine x_ω subsequently, using an encoder T with weights of θ_t to encode the image x and x_ω jointly into an appearance code ω_a:

$$\omega_a = T(x, x_\omega|\theta_t), \tag{3}$$

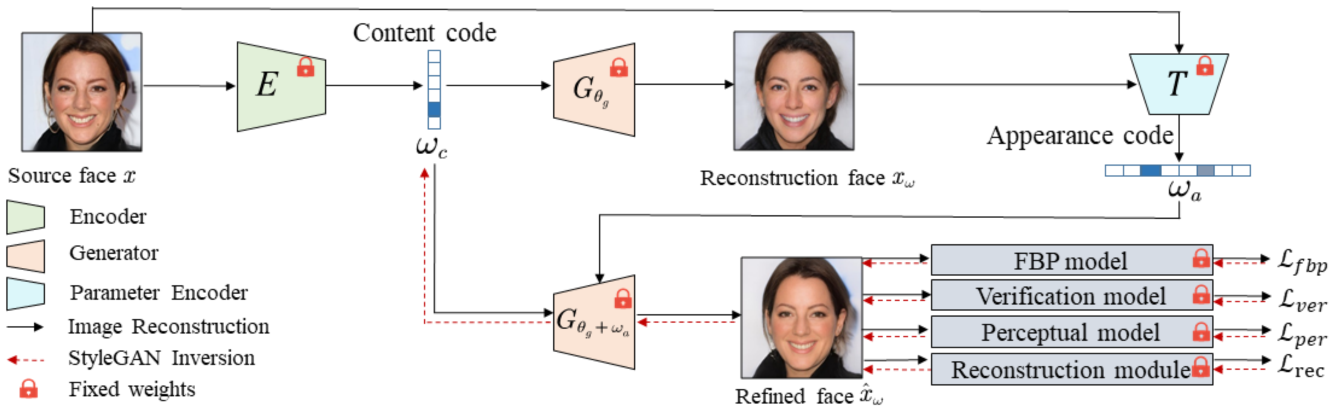


Fig. 2. The pipeline of our method. The face is first encoded into the content code, which is reconstructed into the fake face by the StyleGAN (including the generator G and encoder T), and then the reconstructed face is fed into the prediction models (*e.g.*, FBP model) to compute the loss that drives the StyleGAN inversion. The two steps are alternated until convergence.

where ω_a represents the granular information about the face x, *e.g.*, hair strands, wrinkles, freckles, pockmarks, etc. Then, the content code ω_c and appearance code ω_a are integrated together to generate the refined image $\hat{x}_\omega$:

$$\hat{x}_\omega = G(\omega_c, \omega_a|\theta_g) \rightarrow G(\omega_c|\theta_g + \omega_a), \tag{4}$$

In fact, ω_a serves as dynamic residual weights added to the generator weights θ_g, which is a more effective way to integrate the content code and the appearance code. In this way, the refined image $\hat{x}_\omega$ retains more information than the coarse version x_ω. Note that the encoders E, T and the generator G are pre-trained on a large face dataset, and their weights retain fixed in our method.

3.3. StyleGAN Inversion: from image to code

To retouch the face in $\mathcal{W}$-space, we employ a set of loss functions to guide StyleGAN inversion that the content code is optimized towards user preferences. Given a source face x, its content code is ω_c, and the inversion is formulated as:

$$\hat{\omega}_c = \omega_c - \gamma \frac{\partial(\eta_{fbp}\mathcal{L}_{fbp} + \eta_{rec}\mathcal{L}_{rec} + \eta_{per}\mathcal{L}_{per} + \eta_{ver}\mathcal{L}_{ver})}{\partial \omega_c}, \tag{5}$$

where γ is the learning rate, and $\eta_{fbp}, \eta_{rec}, \eta_{per}, \eta_{ver}$ are the weights of losses. Each loss is defined as below:

Beauty loss. We use a pre-trained FBP model $f(\cdot|\theta_f)$ to calculate the difference between the target score s and the predicted score of the refined face $\hat{x}_\omega$:

$$\mathcal{L}_{fbp} = ||f(\hat{x}_\omega|\theta_f) - s||_1. \tag{6}$$

Reconstruction loss. We also measure the pixel-level similarity between the source face x and its refined face $\hat{x}_\omega$:

$$\mathcal{L}_{rec} = ||x - \hat{x}_\omega||_2. \tag{7}$$

Perceptual loss. We ensure the feature-level consistency between x and $\hat{x}_\omega$, using a pre-trained AlexNet [29] as the perceptual feature extractor $h(\cdot|\theta_h)$:

$$\mathcal{L}_{per} = ||h(x|\theta_h) - h(\hat{x}_\omega|\theta_h)||_2. \tag{8}$$

Verification loss. To ensure the preservation of face identity, we design an verification loss developed from a pre-trained face recognition model $v(\cdot|\theta_v)$, namely ArcFace [30], to measure the cosine distance between x and $\hat{x}_\omega$:

$$\mathcal{L}_{ver} = 1 - \frac{v(x|\theta_v) \cdot v(\hat{x}_\omega|\theta_v)}{||v(x|\theta_v)||_2 \cdot ||v(\hat{x}_\omega|\theta_v||_2}. \tag{9}$$

In each iteration, the edited content code follows the sequence $\hat{\omega}_c \rightarrow x_\omega \rightarrow \omega_a \rightarrow \hat{x}_\omega$ (see Eq.2, Eq.3, and Eq.4) to generate the retouched face. This kind of coarse to fine-grained refinement is iterated until convergence.

4. EXPERIMENTS

4.1. Experimental Setup

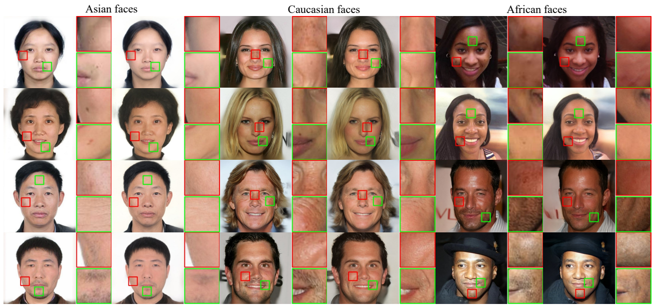
Evaluation criterion. We define a new criterion, namely beauty-to-preservation (BTP), to evaluate the beautification effect in terms of a given target score :

$$BTP = -\frac{||s - f(x_\omega|\theta_f)||_1}{\log ||x - \hat{x}_\omega||_2} \tag{10}$$

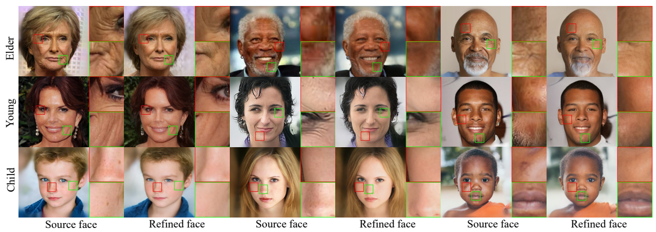
Evaluation data. Our method is evaluated on faces selected from the Internet and the CelebA-HQ dataset. These faces have diverse attributes, *e.g.*, race, gender and age, which makes the evaluation more comprehensive and challenging.

Pre-trained model. All models are pre-trained and fixed weights in our experiments. For StyleGAN, the generator G and the encoders E, T are inherited from StyleGAN-v2 [24] and HyperInverter [28], respectively. The FBP model is inherited from a AlexNet model pre-trained on the SCUT-FBP5500 dataset [31].

Implementation details. Our experiments were implemented using PyTorch 1.12. Before training, we perform preprocessing of each image, including aligning the face, cropping the face region, and resizing the image to 1024×1024. During training, we use the Adam [32] optimizer with a



(a) Faces with different races and genders.



(b) Faces with different ages.

Fig. 3. Results on different race, gender and age groups. The facial regions in the red and green rectangles are the most obviously-refined areas, which have been enlarged to show more details. Best viewed in colors.

	XT [33]	HYXJ [34]	TTPT [35]	QY [36]	Ours
BTP	0.2227	0.2025	0.1985	0.2091	**0.1865**

Table 1. Comparison with beautification applications.

learning rate of 0.01 to minimize the loss, with loss weights set to $\eta_{fbp} = 0.1, \eta_{rec} = 1.0, \eta_{per} = 0.6, \eta_{ver} = 1.0$. The target score is set to the highest beauty score ($s = 5$), and the number of iterations is set to 10 by default.

4.2. Experimental Results and Qualitative Evaluation

Results on different race, gender and age groups. We test the proposed method on different faces sampled from different age, race and gender groups. These faces differ greatly in appearance, which make our experiments even more challenging. The experimental results are shown in Fig. 3. It shows that our method achieves excellent beautification effects even for faces with different attributes.

Results on different time steps. We show the retouched faces generated at different time steps in Fig. 4, from which we can see that the beautification effect grows gradually with increasing time step, compared to the source image (see the difference map in Fig. 4). And the faces are visually more pleasingly beautified as the time step increases to a certain degree. However, too large time step will lead to a distorted effect, so that we set the optimal time step to 10.

Comparison with other methods. Our method is also compared with several commercial beautification applications, including XT [33], HYXJ [34], TTPT [35], and QY [36]. The comparison results are shown in Fig. 5. From this, we can see that our method can achieve comparable or even better beautification effect than related methods. And we also compare BTP among these methods in Table 1, where the result shows that our method achieves the closest beautification effect towards the target score.

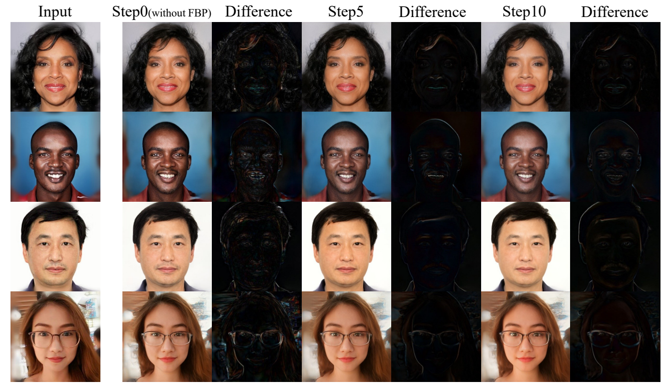


Fig. 4. Results on different time steps. The first difference map is the subtraction between the source image and the first-time generated image (w/o FBP), while the other difference maps are the subtraction between the beautified image and the first-time generated image.

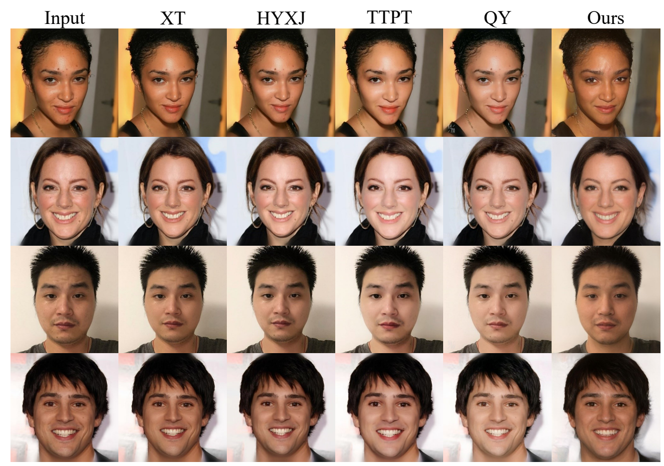


Fig. 5. Comparison with other beautification applications, including XT (xingtu) [33], HYXJ (Butter Camera) [34], TTPT (tiantianPtu) [35], QY (qingyan) [36].

5. CONCLUSION

In this paper, we define a new task for automatic face beautification and present a corresponding simple and innovative method called Human Esthetics Guided StyleGAN Inversion, which leverages a facial beauty prediction model to guide the manipulation of the embedding and achieve personalized results that align with user preferences. The experimental results demonstrate the efficacy of our approach for customized beautification and confirm the superiority of our method on this task. Future research could explore ways to compress the model by pruning less important layers and neurons to reduce computational resources.

6. REFERENCES

[1] M. Bashour, "History and current concepts in the analysis of facial attractiveness," *Plastic and reconstructive surgery*, vol. 118, no. 3, pp. 741–756, 2006.

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

[3] L. Liang, D. Liu, and L. Jin, "Facemore: A face beautification platform on the cloud," in *ICSMC*, 2015, pp. 1798–1803.

[4] S. Liu, X. Ou, R. Qian, W. Wang, and X. Cao, "Makeup like a superstar: deep localized makeup transfer network," in *IJCAI*, 2016, pp. 2568–2575.

[5] T. V. Nguyen and L. Liu, "Smart mirror: Intelligent makeup recommendation and synthesis," in *MM*, 2017, pp. 1253–1254.

[6] W. Jiang, S. Liu, C. Gao, R. He, B. Li, and S. Yan, "Beautify as you like," in *MM*, 2020, pp. 4542–4544.

[7] T. S. Leite and J. M. De Martino, "Improving the attractiveness of faces in images," in *SIBGRAPI*, 2015, pp. 211–218.

[8] J. Li, C. Xiong, L. Liu, X. Shu, and S. Yan, "Deep face beautification," in *MM*, 2015, pp. 793–794.

[9] R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan: How to embed images into the stylegan latent space?" in *ICCV*, 2019, pp. 4432–4441.

[10] T. Matsui, K. Arakawa, and K. Nomoto, "A nonlinear filter system for beautifying face images with enhancement using interactive evolutionary computing," in *ISPACS*, 2005, pp. 534–537.

[11] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "Pairedcyclegan: Asymmetric style transfer for applying and removing makeup," in *CVPR*, 2018, pp. 40–48.

[12] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin, "Beautygan: Instance-level facial makeup transfer with deep generative adversarial network," in *MM*, 2018, pp. 645–653.

[13] H.-J. Chen, K.-M. Hui, S.-Y. Wang, L.-W. Tsao, H.-H. Shuai, and W.-H. Cheng, "Beautyglow: On-demand makeup transfer framework with reversible generative network," in *CVPR*, 2019, pp. 10 042–10 050.

[14] F. Chen, X. Xiao, and D. Zhang, "Data-driven facial beauty analysis: prediction, retrieval and manipulation," *IEEE Trans. on Affec. Compu.*, vol. 9, no. 2.

[15] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, "Data-driven enhancement of facial attractiveness," in *ACM SIGGRAPH*, 2008, pp. 1–9.

[16] C. Kong, B. Chen, W. Yang, H. Li, P. Chen, and S. Wang, "Appearance matters, so does audio: Revealing the hidden face via cross-modality transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 1, pp. 423–436, 2021.

[17] M. R. Cunningham, A. R. Roberts, A. P. Barbee, P. B. Druen, and C.-H. Wu, ""'their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness." *Journ. of perso. and socia. psych.*, vol. 68, no. 2, p. 261, 1995.

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

[19] H. Yan, "Cost-sensitive ordinal regression for fully automatic facial beauty assessment," *Neurocomputing*, vol. 129, pp. 334–342, 2014.

[20] L. Lin, L. Liang, and L. Jin, "R 2-resnext: A resnext-based regression model with relative ranking for facial beauty prediction," in *ICPR*, 2018, pp. 85–90.

[21] ——, "Regression guided by relative ranking using convolutional neural network (r3cnn) for facial beauty prediction," *IEEE Trans. on Affec. Compu.*, 2019.

[22] L. Lin, L. Liang, L. Jin, and W. Chen, "Attribute-aware convolutional neural networks for facial beauty prediction." in *IJCAI*, 2019, pp. 847–853.

[23] A. Creswell and A. A. Bharath, "Inverting the generator of a generative adversarial network," *TNNLS*, vol. 30, no. 7, pp. 1967–1974, 2018.

[24] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *CVPR*, 2020, pp. 8110–8119.

[25] S. Pidhorskyi, D. A. Adjeroh, and G. Doretto, "Adversarial latent autoencoders," in *CVPR*, 2020, pp. 14 104–14 113.

[26] J. Zhu, Y. Shen, D. Zhao, and B. Zhou, "In-domain gan inversion for real image editing," in *ECCV*, 2020, pp. 592–608.

[27] D. Roich, R. Mokady, A. H. Bermano, and D. Cohen-Or, "Pivotal tuning for latent-based editing of real images," *ACM TOG*, vol. 42, no. 1, pp. 1–13, 2022.

[28] T. M. Dinh, A. T. Tran, R. Nguyen, and B.-S. Hua, "Hyperinverter: Improving stylegan inversion via hypernetwork," in *CVPR*, 2022, pp. 11 389–11 398.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commu. of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[30] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019, pp. 4690–4699.

[31] L. Liang, L. Lin, L. Jin, D. Xie, and M. Li, "Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction," in *ICPR*, 2018, pp. 1598–1603.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.

[33] https://apps.apple.com/cn/app/xingtu-xiuchugaojimei/id1500526240.

[34] http://www.bybutter.com/.

[35] https://tu.qq.com/.

[36] https://m.ulikecam.com/.