BeautyBank: Encoding Facial Makeup in Latent Space (*Supplementary Material*)

In this supplemental material, we first provide additional training details in Section A. Then, we present more details about the dataset in Section B and C, as well as additional experiments.

A. Training Details

A.1. Bare-Face Encoding

Following the training method for g' in DualStyleGAN, after initially training the generator g with the FFHQ dataset, we performed finetuning on g_f using images from our BMS dataset. This approach enabled g_f to effectively generate images within the makeup domain. The outputs from g and g_f are illustrated in Fig. 1, which demonstrates the network's enhanced ability to generate various makeup colors and patterns.

Subsequently, in the bare-face code optimization (in Section 3.2.2), we fixed the parameters of g_f and used the 1,412 makeup images as label images. The initial latent code, z^+ ($z^+ = E_b(I_m) \in \mathbb{R}^{18\times512}$), was optimized using facial enhancement loss of the reconstructed facial images and the label images. Additionally, in the facial enhancement loss, a facial mask was derived by performing face segmentation on the 1,412 makeup images using a face-parsing method [6]. This mask exclusively contains the facial region of the images to improve the learning of identity features.

A.2. Makeup Encoding

In Makeup Encoding, the makeup encoding module, E_m, and the bare-face encoding module, E_b, share identical network architectures and parameters. The latent code output by E_m prepares for subsequent makeup style encoding.

We utilized FFHQ dataset for pre-training BeautyBank, and fine-tuned BeautyBank with 130 selected images from 1,412 images, as detailed in Section 3.3.1. Subsequently, in the fine-tuning of the makeup code (Section 3.3.2), stage 1 involved computing the objective function using facial makeup images reconstructed from the initial makeup code $E_m(I_m)$ and using 1,412 images as label images. The objective function includes L_{pm}, L_{em}, and L_{lm}, all applying the Hadamard product for perceptual loss. Eye and lip masks were obtained using a face-parsing method [6]. The eye mask includes areas corresponding to the bounding rectangles of both sets of eyes and eyebrows. To encompass the richly detailed makeup region beneath the eyes, we extended the bounding rectangles downward by 1.3 times their height and also excluded areas within the eye socket and any part of the rectangle extending beyond the face. The lip mask solely includes the areas of the upper and lower lips, excluding the interior of the mouth. In this stage, the first 7 rows of z_m^+ had a learning rate of 0.005, while the last 11 rows of z_m^+ had a learning rate of 0.1. In stage 2, we utilized a foreground mask including the face and neck areas and a background mask for other areas. Given the changes in features like facial shape during reconstruction, to ensure a smooth transition between the face and other parts, we applied Gaussian blurring with a kernel size of 11 to both the foreground and background masks. In this stage, we used label images from 1412 makeup images for L_{pf}, and label images from reconstructed images using bare-face codes for L_{pb}. During training of this stage, the first 7 rows of z_m^+ had a learning rate of 0.005 or 0.001, while the last 11 rows of z_m^+ used learning rates of 0.01 or 0.005. Additionally, in the makeup transfer task, the source image can be used instead of the reconstructed image as the label for L_{pb} to achieve better performance.

B. More Information on our BMS Dataset

We will publicly release the Bare-Makeup Synthesis Dataset. We analyzed 324,000 makeup images of 512x512 resolution from the BMS dataset, synthesized based on the FFHQ dataset, using an open-source gender-and-age detector [3]. As shown in Fig. 2, the proportion of male and female images is 63.71% and 36.29%, respectively. The images are distributed across the following age ranges: 0-20, 21-32, 33-53, and 54-100, with respective proportions of 44.48%, 36.91%, 16.62%, and 1.99%. It should be noted that our analysis includes only those images that were successfully detected by the gender-and-age detector, due to occasional failures in face detection. This demonstrates the diversity in gender and age of the facial images within the BMS dataset. Additionally, examples of paired bare-face and makeup images in our BMS dataset can be seen in

w/o fine-tuning w/ fine-tuning

Figure 1. Examples of generated images by the g and g_f networks before and after fine-tuning.

Fig. 3.

C. Encoded Makeup Codes

We carefully selected 1412 makeup data from our BMS dataset and BeautyFace [5] for encoding. As shown in Fig. 4, these encoded makeup data are rich and diverse in color, texture, and pattern. We aligned all the makeup data based on facial landmarks following the FFHQ [2]. In future work, we plan to select more high-quality makeup images for encoding to expand the application of our method across various makeup scenarios.

D. More Makeup Transfer Results

We provide additional results that highlight the robustness and superiority of our BeautyBank in the task of makeup transfer. As shown in Fig. 6, BeautyBank successfully generates makeup images that preserve the identity features of the source image while faithfully transferring the makeup attributes from the reference image, including its colors, textures, and detailed patterns. These generated images demonstrate the effectiveness of our approach.

E. Ablation Study of Weights

Our method enables editing of generated images by adjusting 18 weights between the makeup code and bare-face code, each ranging from 0 to 1. We randomly selected three encoded makeups and conducted makeup transfer on bare-faced photos by setting these 18 weights to 0.2, 0.4, 0.6, 0.8, and 1, respectively. As demonstrated in Fig. 7, we can progressively increase the weights to generate makeup results that more closely match the reference makeup in terms of color, texture, and pattern.

F. Comparative analysis with different masks

Our method enables the editing and control of makeup by modifying the masks for facial areas, M_{fore}, and for non-facial areas, M_{back}, as mentioned in Section 3.3.2. To prevent the influence of makeup style on the iris during makeup transfer, we utilized the foreground mask (a) and background mask (b) shown in Fig. 8 for obtaining the transferred images. As illustrated in Fig. 8, the (d) makeup transfer results (d) using masks (a) and (b) exhibit an iris color that is much closer to the iris color in the source image compared to the image (c) without using the masks. This experiment demonstrates that our method supports flexible control over makeup transfer results through the editing of masks, thereby ensuring more natural and precise makeup application in targeted regions.

G. Comparison with DualStyleGAN

We conducted comparative analyses between BeautyBank and DualStyleGAN using a self-reconstruction approach. Fig. 5 shows the makeup transfer results ((a) and (c)) using both source and reference images, and makeup removal results ((b) and (d)) after bare-face encoding of (a) and (c), for both BeautyBank and DualStyleGAN. These results demonstrate that our method preserves identity more effectively and retains finer makeup details, while successfully disentangling information irrelevant to the makeup, such as the background.

Additionally, our method exhibits superior maintenance of facial identity throughout the makeup transfer and the subsequent removal process. We evaluated the transferred images from both BeautyBank and DualStyleGAN against the source images using ArcFace [1] and BlendFace [4] cosine similarity metrics, with the results shown in the '*Transfer*' column of Table 1. Similarly, images resulting from the self-reconstruction with BeautyBank and DualStyleGAN were also evaluated against the source images, as shown in the '*Removal*' column. The results indicate that our method more effectively maintains facial identity features during both makeup transfer and self-reconstruction phases. It should be noted that since our method primarily aims to reconstruct bare facial images during the bare-face encoding stage, areas not associated with the face do not require high fidelity reconstruction in our tasks.

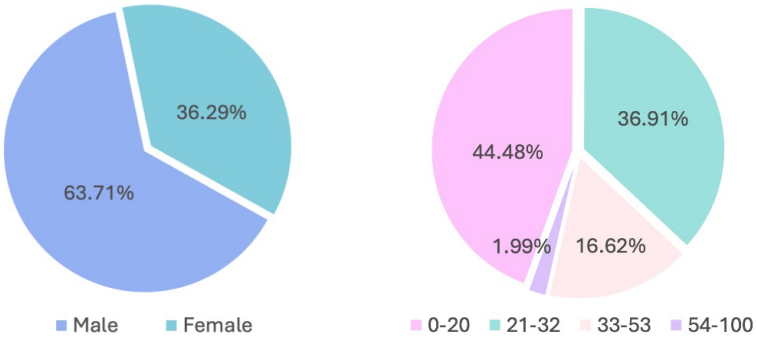

Figure 2. Gender and age distribution of our BMS dataset.

Figure 3. Paired bare-face and makeup images in our BMS dataset.

References

[1] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699. IEEE, 2019. 2, 5

[2] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410, 2019. 2

[3] Mahesh Sawant. Gender-and-age-detection. https://github.com/smahesh29/Gender-and-Age-Detection. 1

[4] Kaede Shiohara, Xingchao Yang, and Takafumi Taketomi. BlendFace: Re-designing identity encoders for face-swapping. In *ICCV 2023*, pages 7600–7610. IEEE, 2023. 2, 5

[5] Qixin Yan, Chunle Guo, Jixin Zhao, Yuekun Dai, Chen Change Loy, and Chongyi Li. BeautyREC: Robust, efficient, and component-specific makeup transfer. In *CVPRW*, pages 1102–1110, 2023. 2

[6] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. BiSeNet: Bilateral segmentation network for real-Time semantic segmentation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *ECCV*, volume 11217, pages 334–349. Springer, 2018. 1

Figure 4. Examples of selected 1412 makeup images.

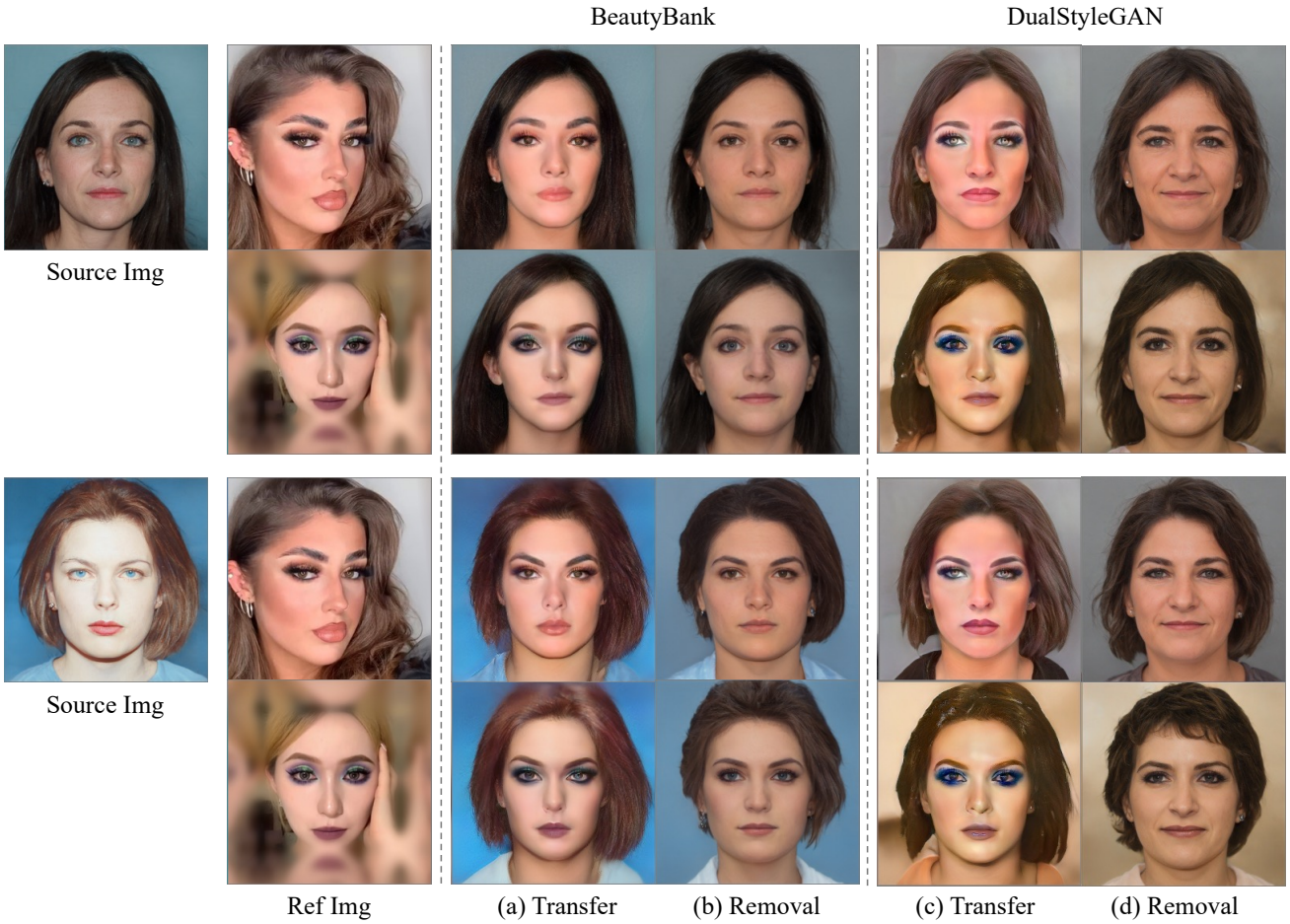

Figure 5. Comparative analysis with BeautyBank and DualStyleGAN.

Table 1. Quantitative comparison of identity for the cycle self-reconstruction experiment. The first group compares the makeup transfer results with the source image, while the second group compares the results after makeup removal (following the transfer) with the source image, using ArcFace [1] and BlendFace [4] cosine similarity metrics.

Method	*Transfer*		*Removal*	
	ArcFace ↑	BlendFace ↑	ArcFace ↑	BlendFace ↑
DualStyleGAN	0.174	0.148	0.084	0.130
BeautyBank (Ours)	**0.177**	**0.164**	**0.206**	**0.188**

Figure 6. More results of our BeautyBank in the makeup transfer task.

Figure 7. Ablation study of makeup transfer results with different weights.

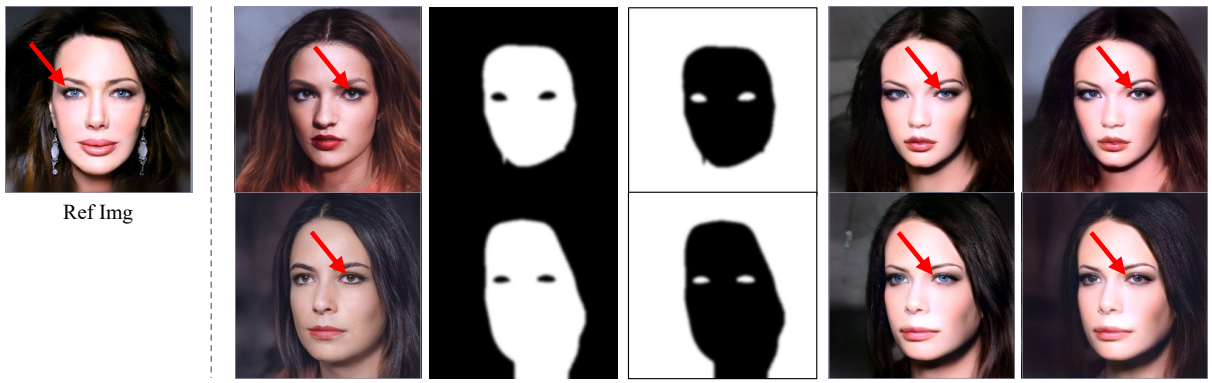

Figure 8. Comparative analysis of results using different masks to determine and control the color of iris. By utilizing the iris region mask, we can determine whether the iris color comes from the source or the reference image.