

Facial Makeup Transfer

Xintian Liu, Joseph Suh
University of California, Berkeley

xintian.liu@berkeley.edu, josephsuh@berkeley.edu

Abstract

This project develops a facial makeup transfer system that effectively transfers makeup styles as well as makeup removal between source and target images while preserving the identity and natural features of the face. Our solution effectively addresses critical challenges in identity preservation and accurate local detail transfer such as gradients in eye shadows. To achieve high-quality and natural-looking outputs, we introduce several loss functions including cycle consistency loss for reversibility, perceptual loss for feature preservation and color matching loss for accurate makeup style transfer. The model is trained on the Makeup Transfer Dataset, including both makeup and non-makeup faces to improve diversity. Experimental results shows that our model is able to produce natural and high-resolution outputs, addressing the limitations of the traditional pixel-based methods. This system will have various applications in digital beauty, virtual try-on systems, and personal identity protection. The code is publicly available at https://github.com/JosephJeesungSuh/Makeup_GAN.

1. Introduction

Facial makeup transfer refers to the process of transferring one or several makeup styles (e.g. lipsticks, eyeshadows, eyebrows, concealers) from the source image to the target image while preserving the identity of the target face.

With the current rapid expansion of social networks filled with photos, live photos and videos, people are becoming more aware of their appearance. This makeup transfer technology enables users to apply different looks without physical application, which is time and cost consuming. It will have broad applications in terms of digital beauty tools in cosmetic industry which allows customers to virtually try on facial products without the additional cost of product return. In the meantime, it is beneficial for content creation, especially for influencers, photographers, and designers, to make more realistic edits on their work.

The key challenges of facial makeup transfer lie in three

aspects: (1) protecting identity: sometimes heavy makeup modifies the eye and lip regions such as bold eyeshadow and eyeliner. When applied, these alterations make individual's key facial features difficult to recognize; (2) handling pose variations: many facial images are not captures with faces directly facing the camera. Slight variations of posing might cause distortions in makeup regions. For example, uneven lighting or sunlight on individual's face causes gradients in eyeshadow; (3) accurate local details and textures: accurately capturing and transferring details such as blush covering cheeks or sharp lip edges without blurring or distorting textures is also required. Failure in doing so results in unnatural or inconsistent makeup transfer.

Our project addresses these challenges by building on a generative adversarial network [2, 4, 5, 9] to achieve high-quality, realistic and customization makeup transfer.

2. Related Work

Previous approaches to facial makeup transfer can be broadly categorized into pixel-level transformations and neural network-based techniques. Compared to general style transfer tasks, makeup transfer has its unique challenges originating from the fact that makeup is applied locally around facial keypoints (nose, eyes, eyebrows, etc).

2.1. Traditional Pixel-Based Methods

Early methods often relied on image blending [Project 3] or histogram matching to transfer makeup styles. These methods typically fail to handle the spatial variations in makeup styles, such as gradients in eye shadow or contours, resulting in unnatural or inconsistent results.

2.2. GAN-based Methods

A generative adversarial network (GAN) has two parts: the generator learns to generate plausible data and the discriminator learns to distinguish the generator's fake data from real data [2]. Existing GAN-based techniques for makeup transfer often fail to disentangle makeup styles from facial features, leading to unintended modifications in the identity of the target image. Furthermore, these methods struggle to retain subtle details, such as skin texture or

eye shapes, especially when transferring intricate makeup styles.

This project builds on the BeautyGAN framework [5] which leverages a dual-input GAN architecture to separate facial structure and makeup style effectively. Specifically, we focus on and the facial identity preservation, transfer accuracy, and reversibility.

In addition to an adversarial loss, a cycle consistency loss guides the model to generate a reversible modification, and a perceptual loss term preserves high-level semantic features and content structure of the input images. There is also an additional loss term specifically designed for makeup transfer task aimed at matching the texture and color distribution of key facial regions [1].

As illustrated in Fig. 1, like common GAN frameworks, we employ an architecture that consists of one generator G and two discriminators, one performing on the non-makeup image domain and another performing on the makeup image domain.

2.3. Additional Loss for Makeup Task

The primary purpose of makeup is to modify the color information (lipsticks, shadows) while secondary effects such as texture and edges are also important consideration point for cosmetics consumers. By breaking up the makeup into several individual regions-lips, eyebrows, skin tones, *etc.*-and comparing the source image and generated image at those regions in terms of color distribution and texture, we can systematically establish the makeup loss.

2.4. Implementation Details

We implement the architecture that includes a StyleGAN-based generator for higher resolution and realism, with the addition of two discriminators to enforce both global and local consistency. The training is performed on face datasets that come with binary makeup / non-makeup classification label and possibly augmented with additional high-resolution facial images for diversity. Exploratory hyperparameter tuning are performed on the weightings of loss metrics as well as a set of training parameters.

2.5. Architecture Detail

The facial makeup transfer system builds on a modified BeautyGAN framework and consists of a single generator (G) and two discriminators (D_A, D_B).

2.5.1 Generator

A dual-input generator starts with two downsampling convolution blocks, one encoding non-makeup images and another encoding makeup images, followed by a concatenation of encodings along the channel dimension. Concatenated encodings go through repeated residual blocks, up-sampled, and passed to two convolution blocks which decode images. The generator is designed to learn the makeup style transfer between two images. Separate encoding / decoding blocks are employed since two images are on the different domain, one on the makeup and one on the non-makeup. The total number of trainable parameters in the generator is 9.2M.

2.5.2 Discriminator

Two discriminators are employed, each specialized at discriminating the makeup image and non-makeup image, respectively. A discriminator is consisted of sequential convolution layers which downsample the input by a factor of 2 and doubles the number of channels. A spectral normalization layer is inserted between convolution layers for training stabilization. The final convolution layer maps the input to a single channel, thereby transforming a 3-channel input image of 256×256 to a patch of 30×30 [3]. The fake / real classification is not done in a binary fashion but based on a number of patches in 1 (real) or 0 (fake).

2.6. Loss Detail

At each training step, discriminator weights are updated followed by a generator update. Discriminator loss is calculated from four images, two original images and two fake images (four fan-ins to D_A and D_B in Fig. 1). Generator loss is a linear combination of the following losses: *consistency loss*. When the fake images are passed again to the generator, the output should be mapped to the original image domain, ideally mapped back to the original images [9]. *Perceptual loss*. Even after makeup transfer, the generated face should preserve general facial structure (basic shape of facial keypoints and the overall structure). Therefore we employ a pretrained VGG-16 and use the intermediate feature (activation of layer 17, in particular) and compare the feature of original vs. generated images (two Perceptual Loss in Fig. 1). *Color matching loss*. This loss ensures the transferred makeup aligns with the source by matching the color distribution of facial regions (e.g., lips, eyes, skin). Histogram matching is used to align color intensity and tone, preserving gradients and textures for realistic results. These loss components are summed with weights λ_{cycle} , λ_{percep} , λ_{skin} , λ_{lip} , λ_{eye} for the best training result.

3. Experiments

3.1. Dataset

The Makeup Transfer dataset of our experiment is from [5], containing over 2500 makeup images and 1200 non-makeup images with varying resolutions. We resized all images to 256×256 and train our model on the fixed resolu-

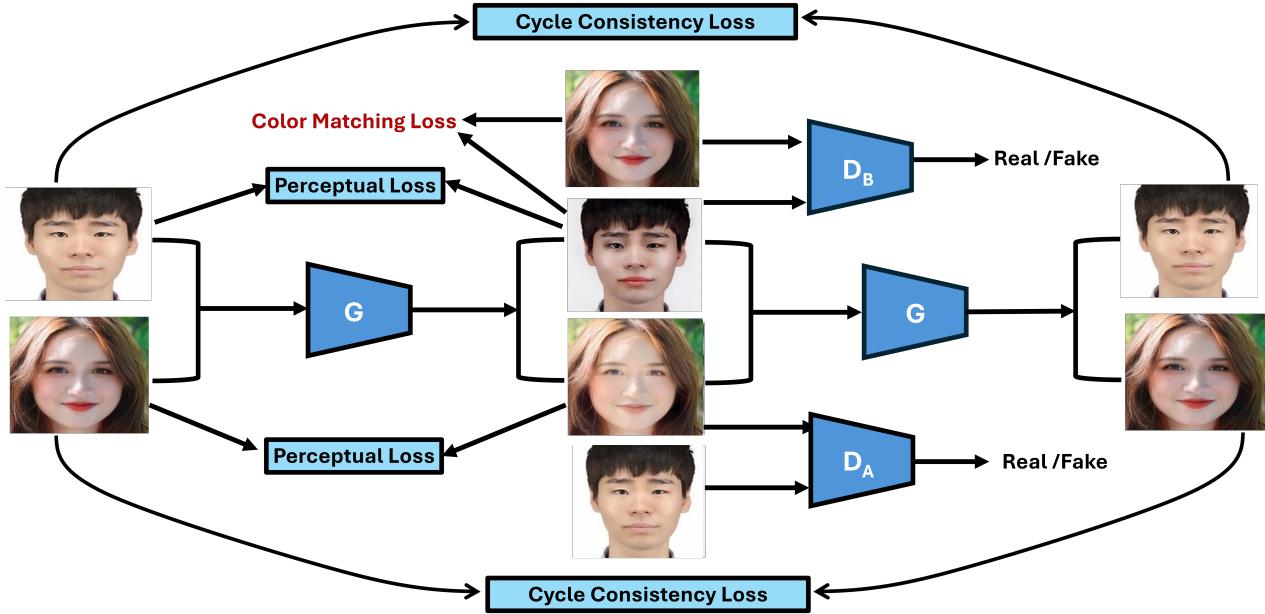


Figure 1. The architecture of the modified BeautyGAN model, which contains three modules in all, a generator G and two discriminators D_A and D_B .

tion. Each image comes with facial keypoint segmentation data such as eyebrows, ears, noses and eyes. This allows us to readily implement the one-to-one color matching loss for each of the facial keypoint.

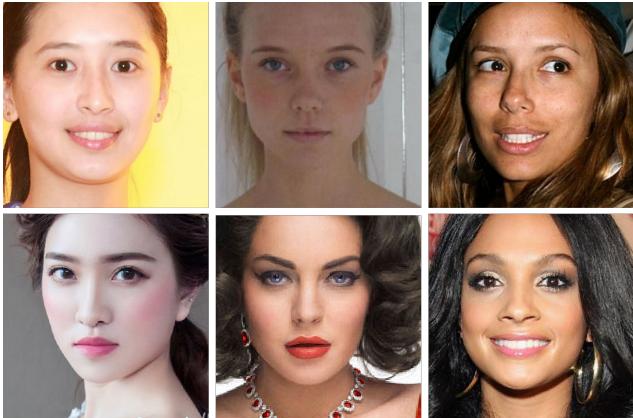


Figure 2. Example images of the Makeup Transfer dataset. Top row: no-makeup images, bottom row: makeup images. We aim to expand the dataset to encompass diverse demographics (race, age, gender, etc.) with data augmentation methods.

Fig. 2 shows unpaired makeup and non-makeup examples from the dataset. Each training sample consists of one makeup image and one randomly selected non-makeup image, sampled without replacement. Since the training image pair (image pair) is randomly generated for every epoch, we didn't spare a predetermined portion of the dataset for validation or test and all images for training.

Instead of using histogram loss over the entire face, in this project, each face is splitted into different cosmetic regions: lipsticks, eyes, faces. This is similar to [6]. Fig. 3 shows the examples of selected segments in eyes, lips, eyebrows regions in a non-makeup face.

3.2. Training

GAN is well known for its sensitivity to training hyperparameters [7]. Due to the limitation of compute, we didn't extensively search the optimal hyperparameters but rather adopted them from literature with similar architecture design or followed common GAN training practice.

We grid searched primarily for learning rate $lr = \{1, 2, 4\}e - 4$ and batch size $\{1, 2, 4, 8, 16\}$ until the 10th training epoch and qualitatively found out that $lr = 2e - 4$ (both for generator and discriminator) and batch size of 8 works best. The remaining hyperparameters are: number of epochs = 250, $\beta_1 = 0.5$, $\beta_2 = 0.999$ (Adam optimizer), $\gamma = 0.99$ (decay per epoch). For loss hyperparameters, we have selected λ s in a way that each loss components have a comparable scale so that all losses are treated equally during backpropagation. This resulted in $\lambda_{cycle} = 10.0$, $\lambda_{percep} = 0.05$, $\lambda_{skin} = 0.1$, $\lambda_{lip} = 1.0$, $\lambda_{eye} = 1.0$.

For example, the weighting for skin color is much smaller than weights for lips / eyes because the number of pixels corresponding to skin is much larger, resulting in higher L1 loss. Training over 250 epochs took 2.5 days (15 min/epoch) on a single NVIDIA RTX A5000 GPU.

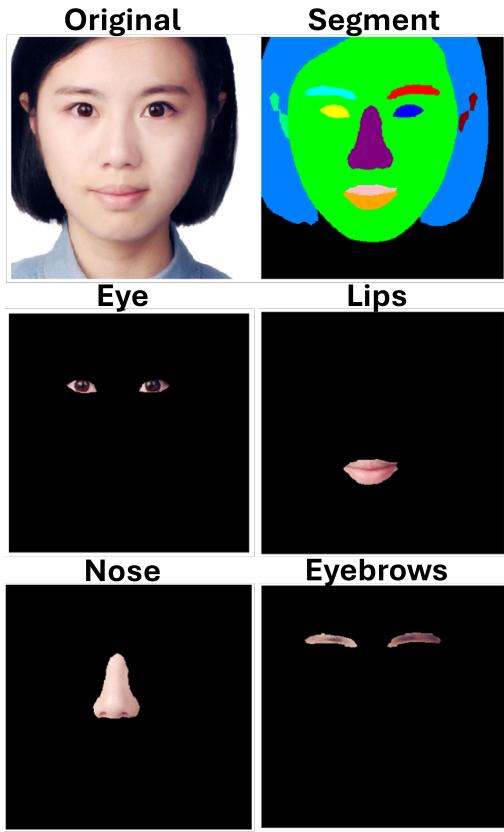


Figure 3. Examples of a non-makeup face with parsing mask and eyes, lip, nose and eyebrows regions.

3.3. Result

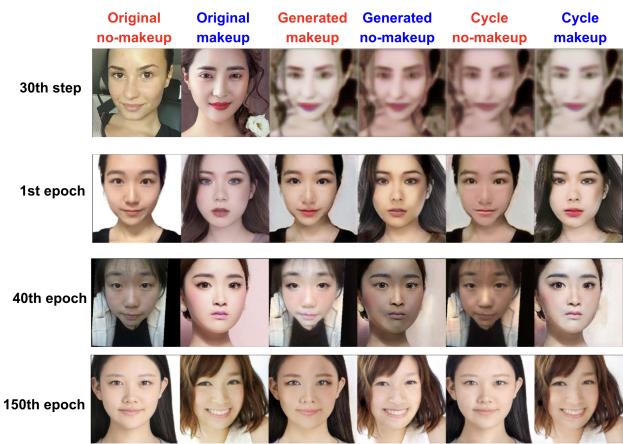


Figure 4. Intermediate results from training. Each rows correspond to training samples and their makeup transferred result after the 30th step, 1st epoch, and 40th epoch, respectively.

Fig. 4 shows intermediate results. At the very initial stage of training (before seeing less than 1,000 samples) the generator cannot produce sensible images, pretty much in-

terpolating between input images on the pixel space. At the end of the first epoch, the model can produce distinct images but still suffer from irregularities, such as a black hole (appearing above the left ear of makeup person in Fig. 4). After about 40 epochs of training, the model starts to produce makeup transferred images and stabilizes after on. Fig. 5 is an inference on the fully trained model with our own pictures.

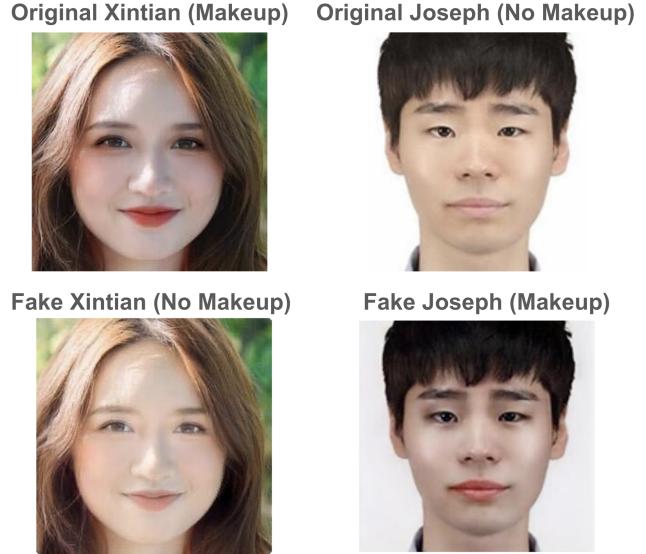


Figure 5. Final result with our face. Skin color, lip, and eye makeup styles are transferred from Xintian to Joseph without interfering with the key facial features of Joseph.

4. Conclusions

In this project, we proposed and implemented a robust modified BeautyGAN architecture for facial makeup transfer. With single generator presented, our model achieves both makeup application and removal on the target images given different source images with high efficiency. To ensure high-quality results, we introduced several loss functions including adversarial loss, cycle consistency loss, perceptual loss and color matching loss.

As shown in experiment results Fig. 4 and Fig. 5, our modified BeautyGAN produces natural-looking and high-resolution outputs. It addresses critical challenges in makeup transferring such as identity preservation and detail accuracy and can potentially provide a platform for practical applications including virtual try-on systems.

5. Future Work

Limitations. One limitation of our work is that the model does not account for variations in makeup perception caused by different lighting conditions. For example, in Fig. 5

Xintian’s photo was taken outdoors with pronounced shadows due to sunlight orientation, while Joseph’s photo was captured indoors. The model fails to consider these photographic conditions, leading to a generated image where Fake Joseph’s skin tone appears much darker than Xintian’s actual makeup style. Addressing this issue by incorporating lighting variations into image generation models, as explored in works like [8], could be a next for future projects.

Moving forward, we can explore the following improvement: (1) Dataset expansion: incorporating more diverse demographics such as age, gender to improve diversity and applicability to broad range of potential users; (2) Real-time processing: optimizing the architecture for video or real-time makeup transfer; (3) User interaction: enhancing customized control for makeup style editing, possibly with additional conditioning prompts.

References

- [1] Hung-Jen Chen, Ka-Ming Hui, Szu-Yu Wang, Li-Wu Tsao, Hong-Han Shuai, and Wen-Huang Cheng. Beautyglow: On-demand makeup transfer framework with reversible generative network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10042–10050, Long Beach, CA, USA, 2019. IEEE. [2](#)
- [2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27, pages 2672–2680. Curran Associates, Inc., 2014. [1](#)
- [3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. [2](#)
- [4] Tero Karras. A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv:1812.04948*, 2019. [1](#)
- [5] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of the 26th ACM International Conference on Multimedia*, pages 645–653, New York, NY, USA, 2018. ACM. [1, 2](#)
- [6] Si Liu, Xinyu Ou, Ruihe Qian, Wei Wang, and Xiaochun Cao. Makeup like a superstar: Deep localized makeup transfer network. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*, pages 2568–2575. AAAI Press, 2016. [3](#)
- [7] Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. How good is my gan? In *Proceedings of the European conference on computer vision (ECCV)*, pages 213–229, 2018. [3](#)
- [8] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. [5](#)
- [9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent

adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. [1, 2](#)