BeautyREC: Robust, Efficient, and Content-preserving Makeup Transfer Supplementary Material

Qixin Yan[1] Chunle Guo[2] Jixin Zhao[3] Yuekun Dai[3] Chen Change Loy[3] Chongyi Li[3*]

[1] WeChat, Tencent [2] TMCC, CS, Nankai University [3] S-Lab, Nanyang Technological University

qixinyan@tencent.com guochunle@nankai.edu.cn jixinzhao0101@gmail.com

YDAI005@ntu.edu.sg ccloy@ntu.edu.sg chongyi.li@ntu.edu.sg

https://li-chongyi.github.io/BeautyREC_files/

In this supplementary material, we present detailed network structure and parameters of our method and additional experimental results as follows:

- Details of our network structure and parameters.
- Discussion about the order of feature transfer.
- Additional experimental results of our method.
- More examples of our BeautyFace dataset.

We also discuss the broader impacts of our proposed method and dataset. Besides, we present a video demo in a separate file. We provide the code of our network in a separate file.

1. Details of Network Structure

In what follows, we detail our network structure and parameters, including content encoder, style encoder, component-specific correspondence, long-range dependencies, and image reconstruction.

The content encoder is composed of three convolutional layers with kernels $\{7, 4, 4\}$ and strides $\{1, 2, 2\}$, and three resblocks. Each convolutional layer is followed by an instance norm layer and a ReLU activation function. The number of feature channels is set to 48.

The global style encoder has the same architectures and parameters as the component style encoder. The style encoder has four convolutional layers with kernels $\{7, 4, 4, 1\}$ and strides $\{1, 2, 2, 1\}$. The number of feature channels is set to 48. We did not adopt the normalization operation after the convolutional layers for style encoders. All layers use the ReLU as the activation function.

In the component-specific correspondence module, we use a three-layer fully-connected network with an architecture $C \rightarrow \frac{C}{r} \rightarrow C (C = 48, r = 16)$ to learn channel-wise weights from statistic information of part-specific styles. For spatial attention, we separately apply mean and maximum pooling operations to the feature map that is re-weighted by the channel attention module along the channel direction. The resulting two maps by mean and maximum pooling operations are concatenated to learn a spatial attention mask using a 1×1 convolutional layer with the Sigmoid activation function.

In the long-range dependencies module, a transformer-based structure is employed to exploit the long-range dependencies between the source image and the reference image. The query of our model represents the global style features extracted by the global style encoder. The key and value of our model are the content features. We use sinusoidal positional encodings for the query and key features. The number of heads in the multi-head attention module is set to 8. A two-layer MLP with a residual connection between its input and output is placed at the end of the transformer unit.

In the image reconstruction stage, three residual blocks are used to progressively refine features. We exploit two upsample operations and convolutional layers to recover the image resolution. The component-specific transferred features and the global transferred features are concatenated with the corresponding decoder features. For the last convolutional layer, we use the Tanh activation function without normalization operation to produce the final result.

*Chongyi Li (chongyi.li@ntu.edu.sg) is the corresponding author.

2. The Order of Feature Transfer

Our method is insensitive to the order of feature transfer. We show a set of results of the model with different component-specific transfer orders below. As shown in Fig. 1, we achieve almost the same transferred results.

Figure 1. The results of the different orders of feature transfer.

3. Additional Experimental Results

In this part, we show more experimental results of our BeautyREC as follows:

- **Diversity of the results**: Our method can generalize well across age, race, expression, etc. To show the generalization capability, we download two images of baby girls from the internet and show the makeup transfer results below. As shown, our method successfully transfers the makeup from the face of an adult female to the two baby faces. Besides, the visual results shown in Fig. 5. also suggest the generalization ability of our method for diverse expressions, different makeup styles, and large misalignment.

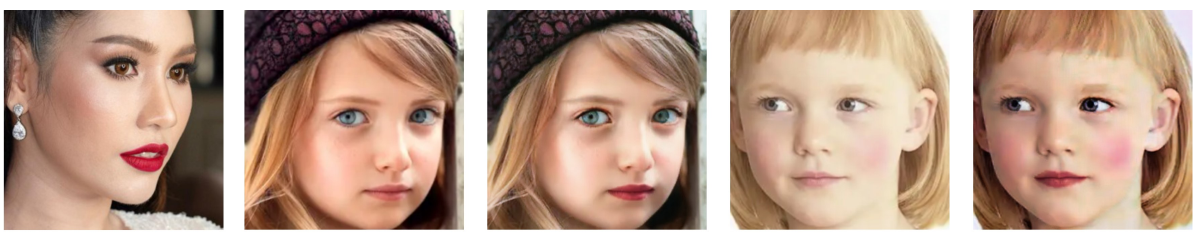

Figure 2. The results of the different ages and expressions.

- **Types of makeup:** Our method does not simply bond the color but learns the component-specific styles guided by the reference image. We show two more complex cases in Fig. 3. Our method can successfully transfer more complex cases, e.g., different lips' color and eye shadows' style.

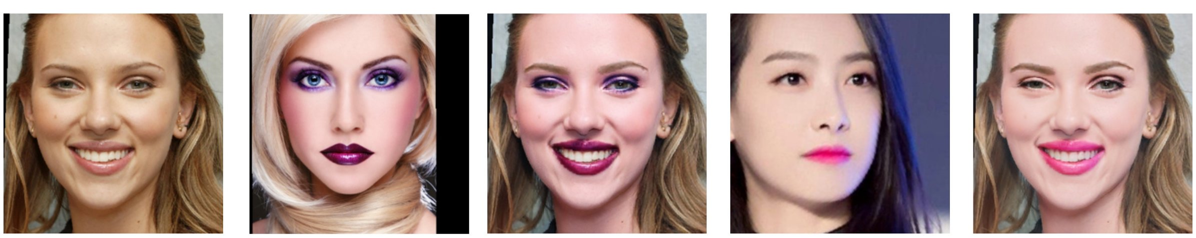

Figure 3. The transferred results of the different types of makeup.

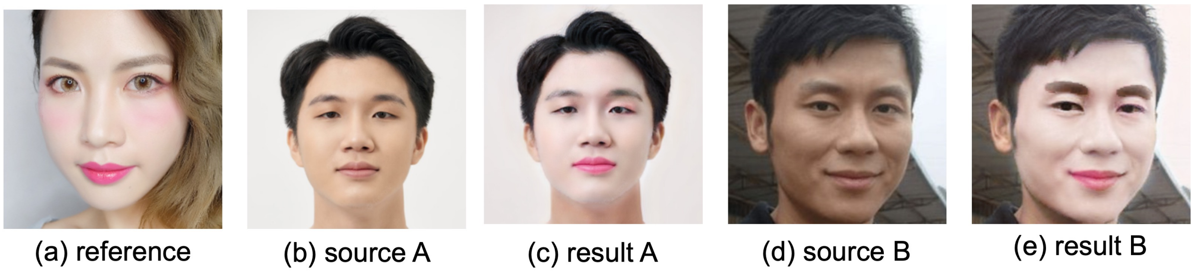

Figure 4. The transferred results of male artists as source images.

- **Experiments for male artists:** We show two cases in Fig. 4 for male artists as source images. Our method can successfully transfer the makeup, which verifies the robustness of our method.
- **More makeup transfer results:** We present more makeup transfer results for diverse makeup styles, different backgrounds, and large misalignment in Fig. 5.

4. More Examples of BeautyFace Dataset

We provide more examples of the proposed BeautyFace dataset in Fig. 6. As Fig. 6 shows, our BeautyFace dataset contains high-quality makeup images with diverse makeup styles, various facial expressions and poses, and large misalignment.

5. Broader Impacts

The proposed method enables efficient makeup transfer via a single-path structure, which is beneficial to the practical applications. The component-specific correspondence modeling and long-range dependencies capturing lead to accurate and robust makeup transfer. Besides, our high-quality BeautyFace dataset covers more recent makeup styles and more diverse face poses, backgrounds, expressions, races, illumination, etc. Thus, we believe our method and dataset can bring new sights and positive impacts on both academia and industry.

Figure 5. **Our BeautyREC for transferring diverse makeup styles.** The source and reference images contain diverse makeup styles, different facial expressions and poses, pure backgrounds, and large misalignment.

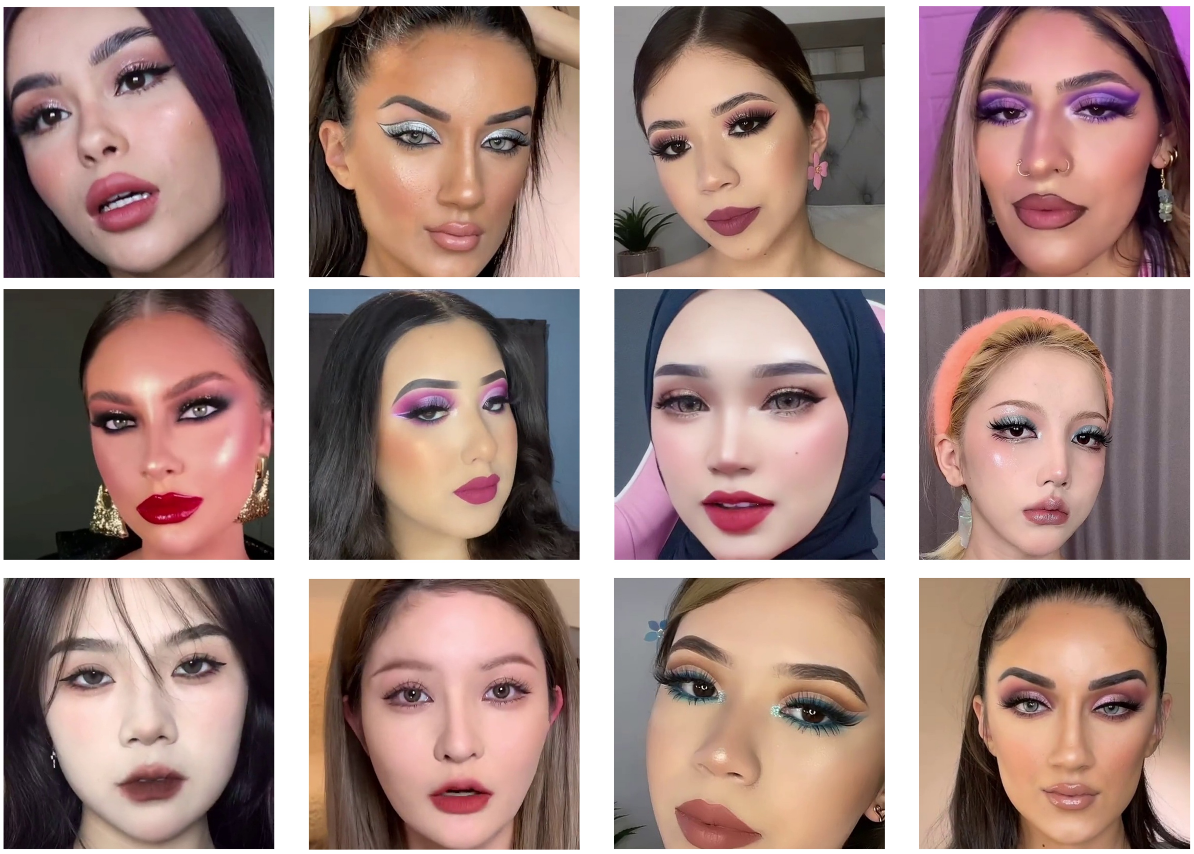

Figure 6. **Examples of BeautyFace dataset.** BeautyFace dataset contains high-quality makeup images with diverse makeup styles, various facial expressions and poses, and large misalignment.