

Towards High-Fidelity, Identity-Preserving Real-Time Makeup Transfer: Decoupling Style Generation

Anonymous WACV Algorithms Track submission

Paper ID 2045

1. Technical Appendices and Supplementary Material

1.1. Network Structure Details

For generate, we adapt the architecture from Pix2Pix[2] which is based on U-Net 256. We only change the number of input channels and output channels to four, as the model takes concatenation of RGB makeup reference and the alpha map as input, and outputs the RGBA makeup mask. The architecture of discriminator and color regressor are outlined in 1.

Color Regressor	Discriminator
Conv2D(4,64,3,1,1)	Conv2D(7,64,3,1,1)
LeakyReLU(0.2)	LeakyReLU(0.2)
Conv2D(64,128,3,1,1)	Conv2D(64,128,3,1,1)
BatchNorm2D(128)	BatchNorm2D(128)
LeakyReLU(0.2)	LeakyReLU(0.2)
Conv2D(128,256,3,1,1)	Conv2D(128,256,3,1,1)
BatchNorm2D(256)	BatchNorm2D(256)
LeakyReLU(0.2)	LeakyReLU(0.2)
Conv2D(256,512,3,1,1)	Conv2D(256,512,3,1,1)
BatchNorm2D(512)	BatchNorm2D(512)
LeakyReLU(0.2)	LeakyReLU(0.2)
AdaptiveAvgPool2d((1,1))	Conv2D(512,1,3,1,0)
Linear(512,3)	Linear(36,18)
	LeakyReLU(0.2)
	Linear(18,1)

Table 1. Architecture for color regressor and discriminator. Conv2D parameters are: (input channels, output channels, kernel size, stride, padding); Linear layer parameters are: (input channels, output channels); AdaptiveAvgPool2d parameters are: ((image height, image weight)). Bias terms are disabled.

1.2. Broader Impacts

Previous makeup transfer methods primarily focus on preserving facial structure as the core aspect of identity, often overlooking skin tone as an important identity attribute. Consequently, when makeup is transferred between individuals with differing skin tones, the subject’s skin tone may be inadvertently altered. This not only affects identity preservation but also limits the applicability of makeup transfer technology across diverse demographic groups and raising fairness concerns. Our work addresses this limitation by explicitly disentangling makeup attributes from skin tone, ensuring that identity is preserved more holistically. By doing so, we aim to make makeup transfer more inclusive and accessible to a broader range of users, regardless of their skin tone.

Our method enables users to visualize realistic makeup effects before applying them in real life. However, excessive reliance on makeup transfer technology may reinforce unrealistic beauty standards and increase appearance-related anxiety.

1.3. Additional Experiment Result

More comprehensive qualitative comparison results on image data can be found in Figure 1 and Figure 2.

Additional video frame testing result can be found in Figure 3. In each row, we extract the makeup mask from a given reference image and apply the same makeup mask throughout the video.

Additional video results are attached in the supplementary materials.

1.4. Dataset Links

The following datasets are used in our experiments can be found in the following links:

MT dataset [5]: https://github.com/wtjiang98/BeautyGAN_pytorch

Wild-MT dataset [3]: <https://github.com/wtjiang98/PSGAN>

047 LADN dataset [1]: [https://github.com/](https://github.com/wanguanzhi/LADN)
048 [wanguanzhi/LADN](https://github.com/wanguanzhi/LADN)
049 FFHQ dataset [4]: [https://github.com/](https://github.com/NVlabs/ffhq-dataset)
050 [NVlabs/ffhq-dataset](https://github.com/NVlabs/ffhq-dataset)
051 CelebV-Text dataset [7]: [https://github.com/](https://github.com/celebv-text/CelebV-Text)
052 [celebv-text/CelebV-Text](https://github.com/celebv-text/CelebV-Text)

053 1.5. Metrics Code Links

054 The following implementation of metrics are used in our
055 quantitative analysis.

056 We utilize the PyTorch official implementation of
057 FID[6]: [https://github.com/mseitzer/](https://github.com/mseitzer/pytorch-fid)
058 [pytorch-fid](https://github.com/mseitzer/pytorch-fid)

059 We utilize the LPIPS calculation implemented
060 by [8]: [https://github.com/richzhang/](https://github.com/richzhang/PerceptualSimilarity)
061 [PerceptualSimilarity](https://github.com/richzhang/PerceptualSimilarity)

062 See the metrics code in the supplementary material.

063 References

- 064 [1] Qiao Gu, Guanzhi Wang, Mang Tik Chiu, Yu-Wing Tai, and
065 Chi-Keung Tang. Lادن: Local adversarial disentangling net-
066 work for facial makeup and de-makeup. In *Proceedings of*
067 *the IEEE/CVF International conference on computer vision*,
068 pages 10481–10490, 2019. 2, 4
- 069 [2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros.
070 Image-to-image translation with conditional adversarial net-
071 works. In *Proceedings of the IEEE conference on computer*
072 *vision and pattern recognition*, pages 1125–1134, 2017. 1
- 073 [3] Wentao Jiang, Si Liu, Chen Gao, Jie Cao, Ran He, Jiashi
074 Feng, and Shuicheng Yan. Psgan: Pose and expression ro-
075 bust spatial-aware gan for customizable makeup transfer. In
076 *Proceedings of the IEEE/CVF Conference on Computer Vi-*
077 *sion and Pattern Recognition*, pages 5194–5202, 2020. 1, 3,
078 5
- 079 [4] Tero Karras, Samuli Laine, and Timo Aila. A style-based gen-
080 erator architecture for generative adversarial networks. In *Pro-*
081 *ceedings of the IEEE/CVF conference on computer vision and*
082 *pattern recognition*, pages 4401–4410, 2019. 2
- 083 [5] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan,
084 Wenwu Zhu, and Liang Lin. Beautygan: Instance-level facial
085 makeup transfer with deep generative adversarial network.
086 In *Proceedings of the 26th ACM international conference on*
087 *Multimedia*, pages 645–653, 2018. 1, 3
- 088 [6] Maximilian Seitzer. pytorch-fid: FID Score for PyTorch.
089 <https://github.com/mseitzer/pytorch-fid>,
090 2020. Version 0.3.0. 2
- 091 [7] Jianhui Yu, Hao Zhu, Liming Jiang, Chen Change Loy, Wei-
092 dong Cai, and Wayne Wu. CelebV-Text: A large-scale facial
093 text-video dataset. In *CVPR*, 2023. 2, 5
- 094 [8] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman,
095 and Oliver Wang. The unreasonable effectiveness of deep
096 features as a perceptual metric. In *Proceedings of the IEEE*
097 *conference on computer vision and pattern recognition*, pages
098 586–595, 2018. 2

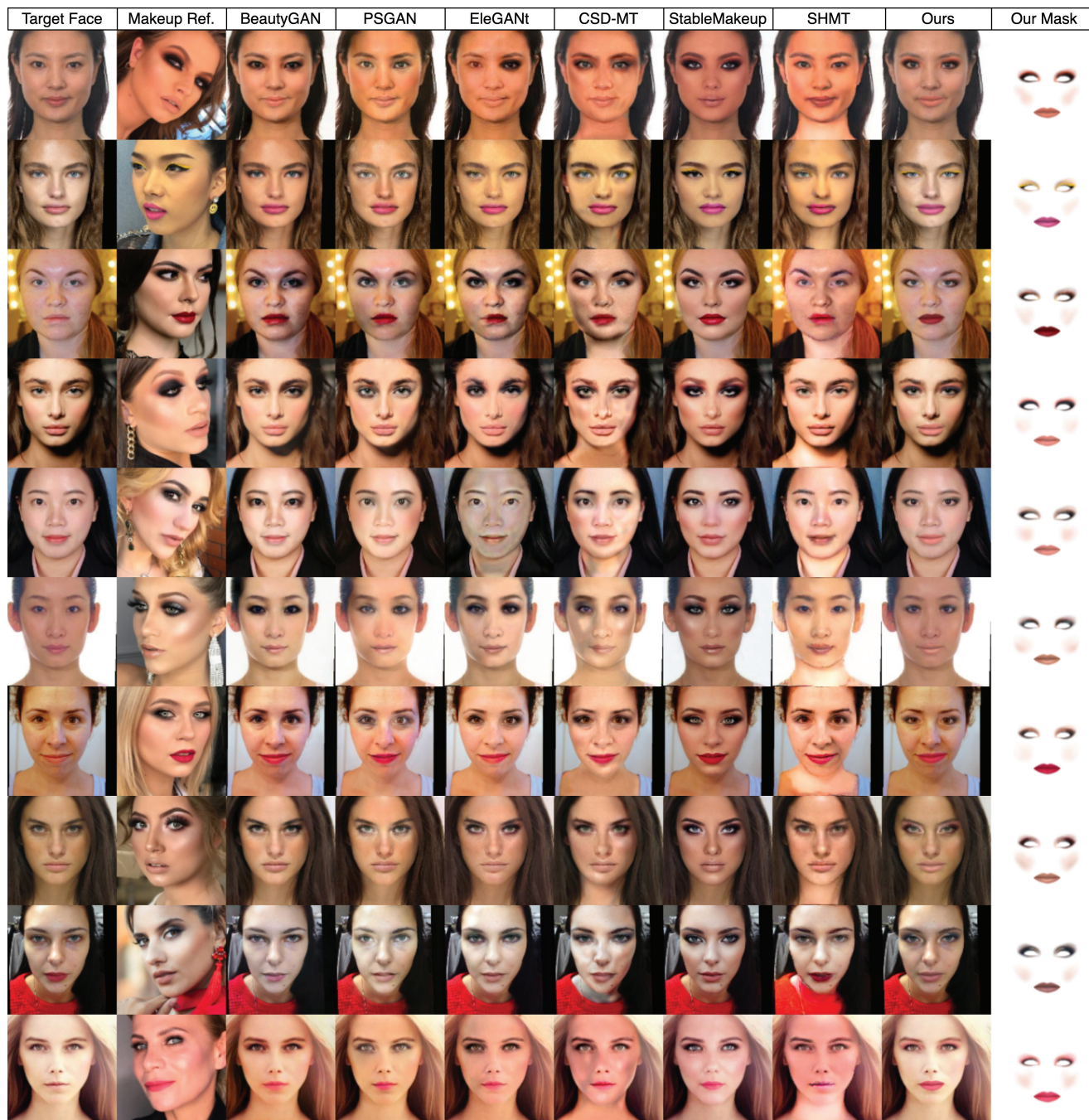


Figure 1. Additional qualitative comparison with state of the art methods. The makeup references are from Makeup-Wild dataset [3] and the natural faces are from makeup transfer dataset [5].

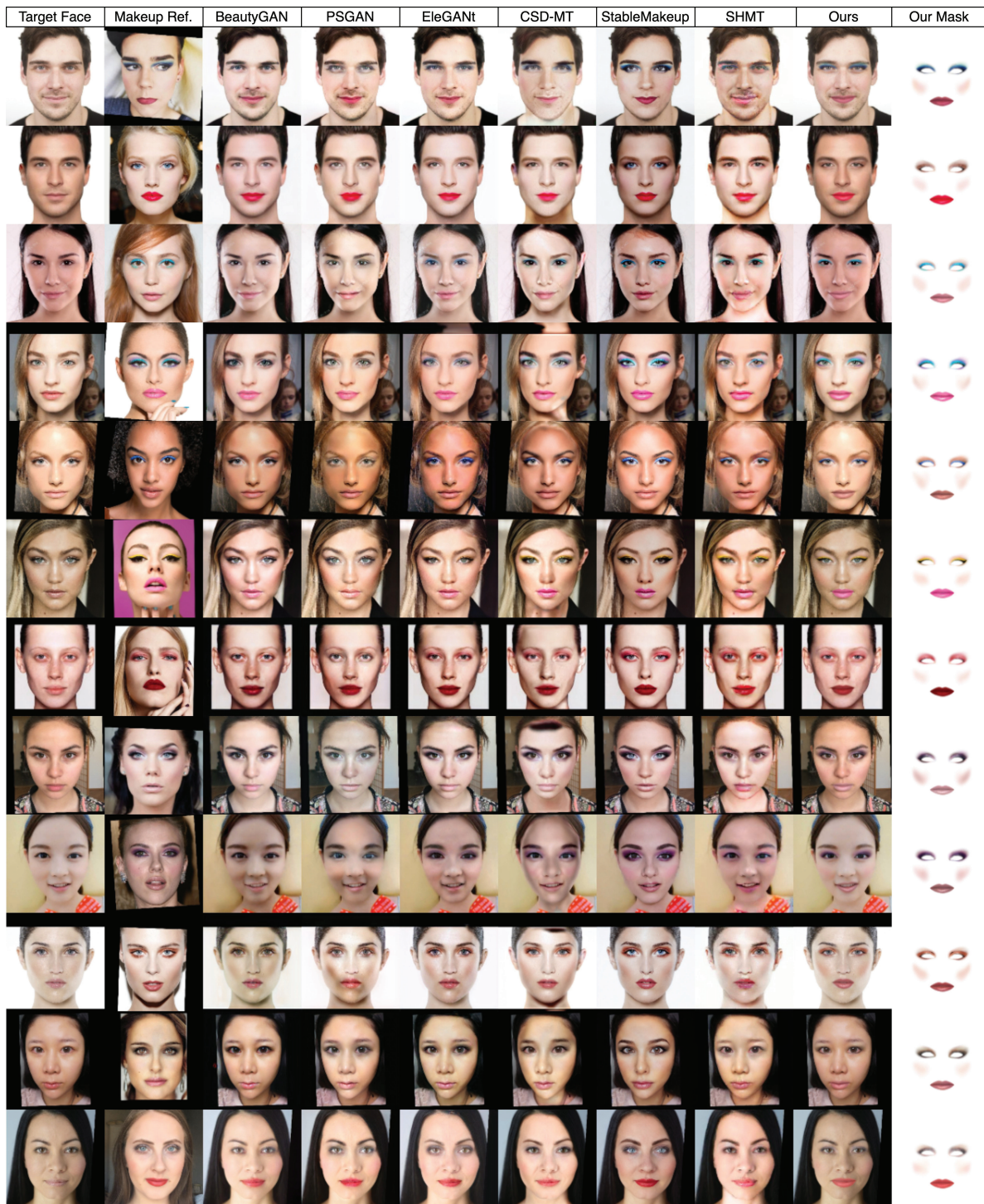


Figure 2. Additional qualitative comparison with state of the art methods. The makeup references and faces are both from LADN dataset [1].



Figure 3. Additional video testing result. The makeup reference images are from Makeup-Wild dataset [3] and videos are from CelebV-Text dataset [7]. More videos are in the Supplementary Material.