Supplementary Material for
BlendFace: Re-designing Identity Encoders for Face-Swapping

Kaede Shiohara[1] Xingchao Yang[2] Takafumi Taketomi[2]
[1]The University of Tokyo [2]CyberAgent AI Lab
shiohara@cvm.t.u-tokyo.ac.jp {you_koutyo,taketomi_takafumi}@cyberagent.co.jp

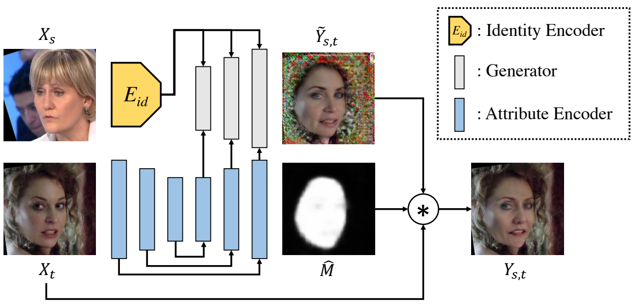


Figure 1: **Architecture of our model.** We simply add a convolution and sigmoid layer to the last layer of the original attribute encoder from AEI-Net [7].

A. Architecture

Here, we detail the architecture of our face-swapping model. As described in the main paper, our architecture is based on AEI-Net [7] but we incorporate a mask predictor into the model inspired by previous studies [12, 13, 15]. We illustrate the architecute of our model in Fig. 1. We simply add a convolution layer and sigmoid layer to the last layer of the attribute encoder to predict blending masks $\hat{M}$. We blend the foreground face image $\tilde{Y}_{s,t}$ and target image X_t using the predicted $\hat{M}$ as follows:

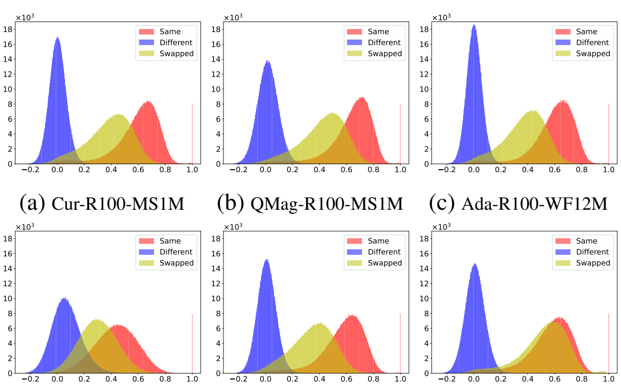
$$Y_{s,t} = \tilde{Y}_{s,t} \odot \hat{M} + X_t \odot (1 - \hat{M}). \tag{1}$$

In our model, we assume the blending masks for the same target images should be the same independently of source images. Note that losses are computed on the blended result $Y_{s,t}$; therefore, the intermediate generated face $\tilde{Y}_{s,t}$ is noisy outside of the face.

B. Comparison with Additional Baselines

We compare our model with two additional baselines in Fig. 2: 1) We input masked source images into a pretrained face-swapping model that is the same model as Arc-Arc in Fig. 6, denoted as *Masked Infer.*. To only include the face area below the eyebrow and between ears, we generate the masks by computing the convex hull of 68 facial landmarks by FAN [1]. 2) We train Arc-Arc model from scratch with masked source images, denoted as *Masked Train.*. As can be seen, the models still suffer from the attribute inconsistency. Therefore, we can conclude our method is a unique solution for the attribute leakage problem.

Figure 2: **Comparison with additional baselines.**



(a) Cur-R100-MS1M (b) QMag-R100-MS1M (c) Ada-R100-WF12M

(d) Arc-R50-Syn (e) Arc-ViT-MS1M (f) **Blend**-ViT-MS1M

Figure 3: **Different models and dataset.**

C. Attribute Biases on Different Face Recognition Models

We visualize the similarity distributions of different face recognition models in Fig. 3: (a) CurricularFace (Cur) [5], (b) QMagFace (QMag) [11], (c) AdaFace (Ada) [6] on a lager dataset WebFace12M (WF12M) [17], (d) ArcFace (Arc) on GAN-generated face images (Syn) [9], and (e) ArcFace with VisionTransformer (ViT) [3]. We can see that the attribute leakage problem still exists in these various face recognition models. Additionally, we train ViT backbone using our BlendFace pretraining. As shown in Fig. 3 (f), our training strategy works well even on ViT, which proves the generality of our method.

D. More Comparisons on FF++

We show more qualitative comparisons on FaceForensics++ [10] in Fig. 4. Our model performs consistent face-swapping for a range of source and target images compared to previous methods [2, 4, 8, 12, 14, 16].

References

[1] Adrian Bulat and Georgios Tzimiropoulos. How Far Are We from Solving The 2D & 3D Face Alignment Problem? (and a dataset of 230,000 3D facial landmarks). In *ICCV*, 2017. 1

[2] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. SimSwap: An Efficient Framework For High Fidelity Face Swapping. In *ACMMM*, 2020. 2

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020. 2

[4] Gege Gao, Huaibo Huang, Chaoyou Fu, Zhaoyang Li, and Ran He. Information bottleneck disentanglement for identity swapping. In *CVPR*, 2021. 2

[5] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, and Feiyue Huang Jilin Li. Curricularface: Adaptive curriculum learning loss for deep face recognition. In *CVPR*, 2020. 1

[6] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *CVPR*, 2022. 1

[7] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. Advancing high fidelity identity swapping for forgery detection. In *CVPR*, 2020. 1

[8] Yuval Nirkin, Yosi Keller, and Tal Hassner. FSGAN: Subject agnostic face swapping and reenactment. In *ICCV*, 2019. 2

[9] Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao. Synface: Face recognition with synthetic data. In *ICCV*, 2021. 2

[10] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. Faceforensics++: Learning to detect manipulated facial images. In *ICCV*, 2019. 2

[11] Philipp Terhörst, Malte Ihlefeld, Marco Huber, Naser Damer, Florian Kirchbuchner, Kiran Raja, and Arjan Kuijper. Qmagface: Simple and accurate quality-aware face recognition. In *WACV*, 2023. 1

[12] Yuhan Wang, Xu Chen, Junwei Zhu, Wenqing Chu, Ying Tai, Chengjie Wang, Jilin Li, Yongjian Wu, Feiyue Huang, and Rongrong Ji. Hififace: 3d shape and semantic prior guided high fidelity face swapping. In *IJCAI*, 2021. 1, 2

[13] Chao Xu, Jiangning Zhang, Miao Hua, Qian He, Zili Yi, and Yong Liu. Region-aware face swapping. In *CVPR*, 2022. 1

[14] Yangyang Xu, Bailin Deng, Junle Wang, Yanqing Jing, Jia Pan, and Shengfeng He. High-resolution face swapping via latent semantics disentanglement. In *CVPR*, 2022. 2

[15] Zhiliang Xu, Hang Zhou, Zhibin Hong, Ziwei Liu, Jiaming Liu, Zhizhi Guo, Junyu Han, Jingtuo Liu, Errui Ding, and Jingdong Wang. Styleswap: Style-based generator empowers robust face swapping. In *ECCV*, 2022. 1

[16] Yuhao Zhu, Qi Li, Jian Wang, Chengzhong Xu, and Zhenan Sun. One shot face swapping on megapixels. In *CVPR*, June 2021. 2

[17] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, et al. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *CVPR*, 2021. 2

Figure 4: **More comparisons on FF++.**