Supplementary material for "Complete Face Recovery GAN: Unsupervised Joint Face Rotation and De-Occlusion from a Single-View Image"

Yeong-Joon Ju[1*] Gun-Hee Lee[2*] Jung-Ho Hong[1] Seong-Whan Lee[1†]
[1]Department of Artificial Intelligence, Korea University, Seoul, South Korea
[2]Department of Computer and Radio Communications Engineering, Korea University, Seoul, South Korea
{yj_ju, gunhlee, jungho-hong, sw.lee}@korea.ac.kr

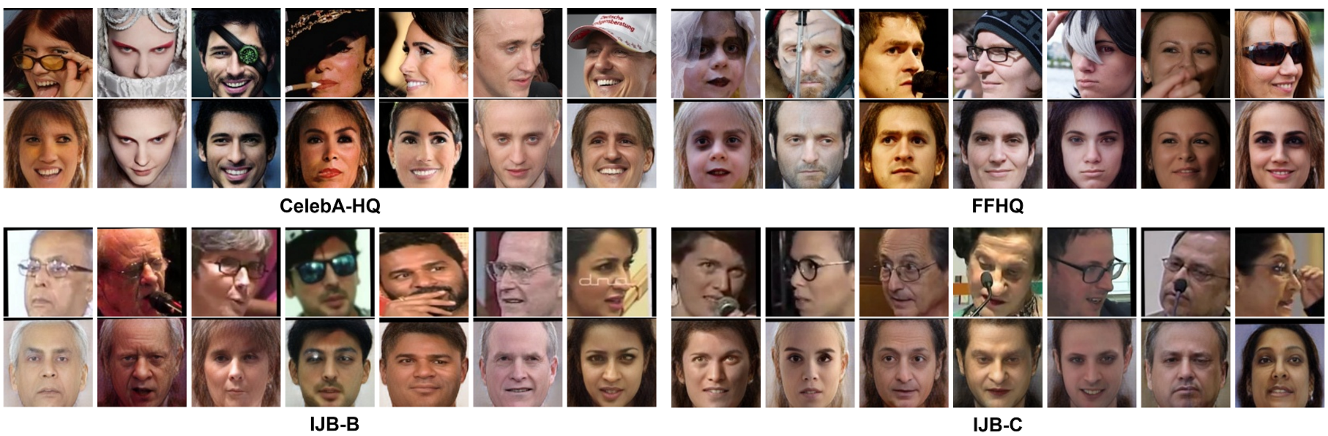

Figure 1: Additional frontalization and de-occlusion results. The first rows are the input images and the second rows are our results. We crop the results as alignments for input images of facial recognition networks.

1. Dataset for 3D Face Model

In first stage, we train using the occluded face images synthesized with hand-shape masks. The hand image is randomly transformed with rotation and scaling, and is located around a point of facial landmark points. To make the problem more difficult, its color is decided to a mean value of pixels around the location. We show examples and our training strategy in Fig 2.

2. Calculating Coarse Occlusion Mask

We describe how to calculate the coarse occlusion mask $\mathcal{M}$ to learn $\mathcal{O}$ path or to be used in Swap-R&R. To leverage texture differences, we compute L2 distance in CIE Lab [4] color space:

$$d_t = \sqrt{L^2 + a^2 + b^2}. \tag{1}$$

Structure differences are calculated with the product of contrast and structure in SSIM [5] as follows:

$$d_s = 1 - \frac{2\sigma_{\mathcal{R}_e\mathcal{R}_p} + C}{\sigma^2_{\mathcal{R}_e} + \sigma^2_{\mathcal{R}_p} + C}, \tag{2}$$

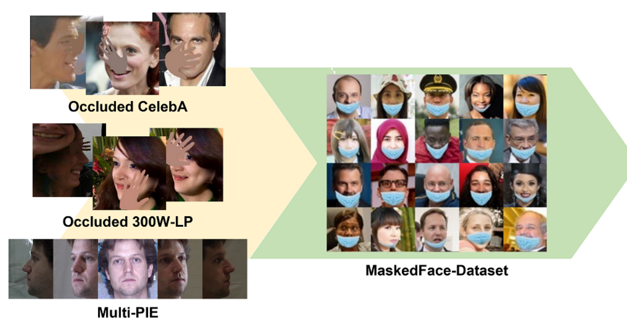

Figure 2: Training strategy for occlusion-robust 3D face reconstruction.

where σ_{ab} is the covariance of a and b. σ^2_a is the variance of a. To prevent division by zero, the constant C is added.

Then, d_t and d_s are normalized with their respective mean and standard deviation to ignore basic differences between estimated 3D face and real face images. A coarse oc-

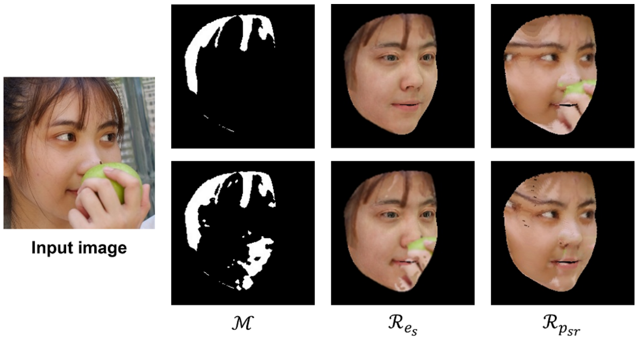

Figure 3: Comparison of results of Swap-R&R using a coarse occlusion mask calculated without or with our method leveraging structural and texture differences. The first row is resulted only used with maps parsed by BiseNet.

clusion mask $\mathcal{M}$ is calculated with normalized z_t and z_s:

$$\mathcal{M} = z_t z_s + z_t + \alpha z_s. \quad (3)$$

$\mathcal{M}$ is normalized and then noises are removed through median blur. We set values below 0 to 0 and values above 1 to 1, so values above 0 are considered as occlusion. Finally, we add the regions of eyeglasses and hairs and subtract regions of eyes, using maps parsed by BiseNet [7], which is the face parsing network. Our method can also detect hairs and eyeglasses, but it is difficult to find them in cases similar to the skin color of the face. We show the results of Swap-R&R for ablation study on our method in Fig 3. Our method detects occlusion types that face parsing networks cannot detect, making Swap-R&R work well.

3. Additional Results

We show more frontalization and de-occlusion results applied to several datasets, e.g., CelebA-HQ [1], FFHQ [2], IJB-B [6], and IJB-C [3] in Fig 1 to validate that our method performs well for in-the-wild images.

References

[1] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv:1710.10196*, 2017.

[2] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410, 2019.

[3] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165. IEEE, 2018.

[4] Alan R Robertson. The cie 1976 color-difference formulae. *Color Research & Application*, 2(1):7–11, 1977.

[5] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[6] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 90–98, 2017.

[7] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 325–341, 2018.