# Portrait Eyeglasses and Shadow Removal by Leveraging 3D Synthetic Data Supplementary Material

Junfeng Lyu  Zhibo Wang  Feng Xu
School of Software and BNRist, Tsinghua University

## A. Ablation Study for Mask Operation

To further evaluate the effect of mask operation in the item removal stage, we additionally train an ablation that directly feeds the shadow-removed results $\hat{I}^g_{syn}$ (*w/o MO*) to the De-Glass Network. Qualitative results in Fig. 1 show that without mask operation, the network has worse results on real-world portraits with unusual eyeglasses textures. Quantitative results in Tab. 1 also show degradation of the realism metric FID and the identity preservation metrics without the mask operation.

| | FID↓ | TAR@FAR↑ $1e^{-2}$ | TAR@FAR↑ $1e^{-3}$ | Rank-1↑ |
|---|---|---|---|---|
| w/o MO | 27.57 | 0.6540 | 0.4095 | 0.4311 |
| w/ MO (ours) | **26.89** | **0.6702** | **0.4315** | **0.4621** |

Table 1. Quantitative comparison.

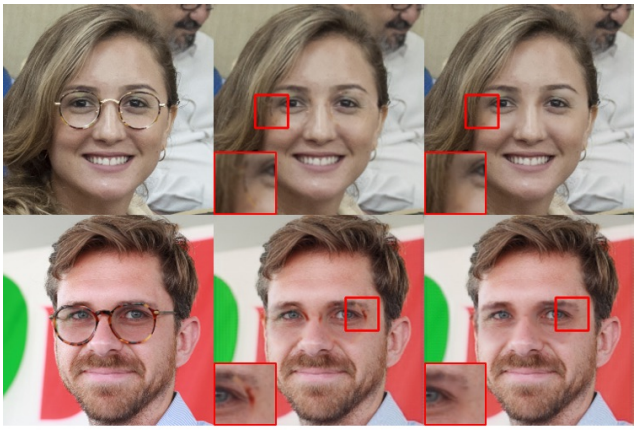


Figure 1. Evaluations on the mask operation.

## B. More Results

To demonstrate the robustness of our proposed method, we show more qualitative results for illuminations (Fig. 3), head poses (Fig. 4) and eyeglasses variation (Fig. 5).

## C. Qualitative Comparison on CelebA

Some methods in the previous comparisons, *e.g.*, ELEGANT [7], show limited performance in the task of eyeglasses removal. To validate the correct usage of their released code, we also show qualitative results on the testing set of CelebA [6] using the same trained models. Results in Fig. 6 show that most of the methods, *e.g.*, ELEGANT, have better performance compared to their results on FFHQ [4] or MeGlass [3].

## D. Eyeglasses Removal via StarGANv2

StarGANv2 [2] learns mixed style to transfer images of one domain to diverse images of a target domain. Although it inherits the assumption of StarGAN [1], the behaviour of the two networks is quite different. In practice, we adopt StarGANv2 to the task of eyeglasses removal via two ways: reference-guided and latent-guided. The results in Fig. 2 show that StarGANv2 prefers to change other facial attributes, *e.g.*, skin color and hairstyle, when removing the eyeglasses. Similar phenomena are also mentioned in [5]. Based on these experiments, it seems inappropriate to compare StarGANv2 on the task of eyeglasses removal. Therefore, we make comparisons with StarGAN as replace in our paper.

## E. Potential Negative Impact

Although our method hardly changes identity, it still requires people's consent to edit their portraits for public usage. The potential negative impact might include abuse of our method without any consent. Besides, applying the technique at sufficiently large scale may cause unintended societal effects. For example, it could worsen the already widespread stigmatization of corrective eyewear. This technique should be applied in a correct way, and removing eyeglasses or not should be considered as a personal choice.

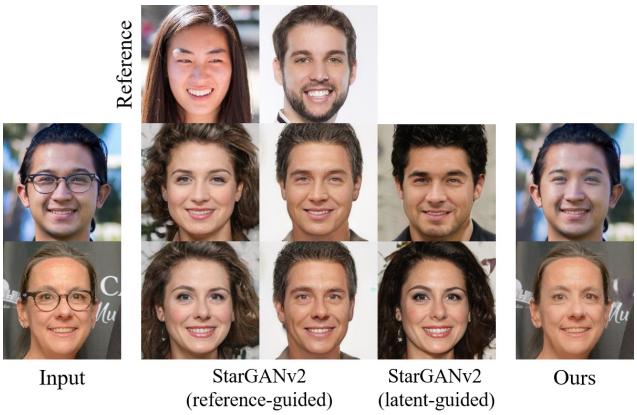


Figure 2. Comparison with StarGANv2 on the task of eyeglasses removal.

## References


[1] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[2] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 1

[3] Jianzhu Guo, Xiangyu Zhu, Zhen Lei, and Stan Z Li. Face synthesis for eyeglass-robust face recognition. *arXiv preprint arXiv:1806.01196*, 2018. 1

[4] 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*, pages 4401–4410, 2019. 1

[5] Xinyang Li, Shengchuan Zhang, Jie Hu, Liujuan Cao, Xiaopeng Hong, Xudong Mao, Feiyue Huang, Yongjian Wu, and Rongrong Ji. Image-to-image translation via hierarchical style disentanglement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8639–8648, June 2021. 1

[6] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015. 1

[7] Taihong Xiao, Jiapeng Hong, and Jinwen Ma. Elegant: Exchanging latent encodings with gan for transferring multiple face attributes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 172–187, September 2018. 1


Figure 3. More results of different illuminations.


Figure 4. More results of different head poses.


Figure 5. More results of eyeglasses with different shapes and textures.

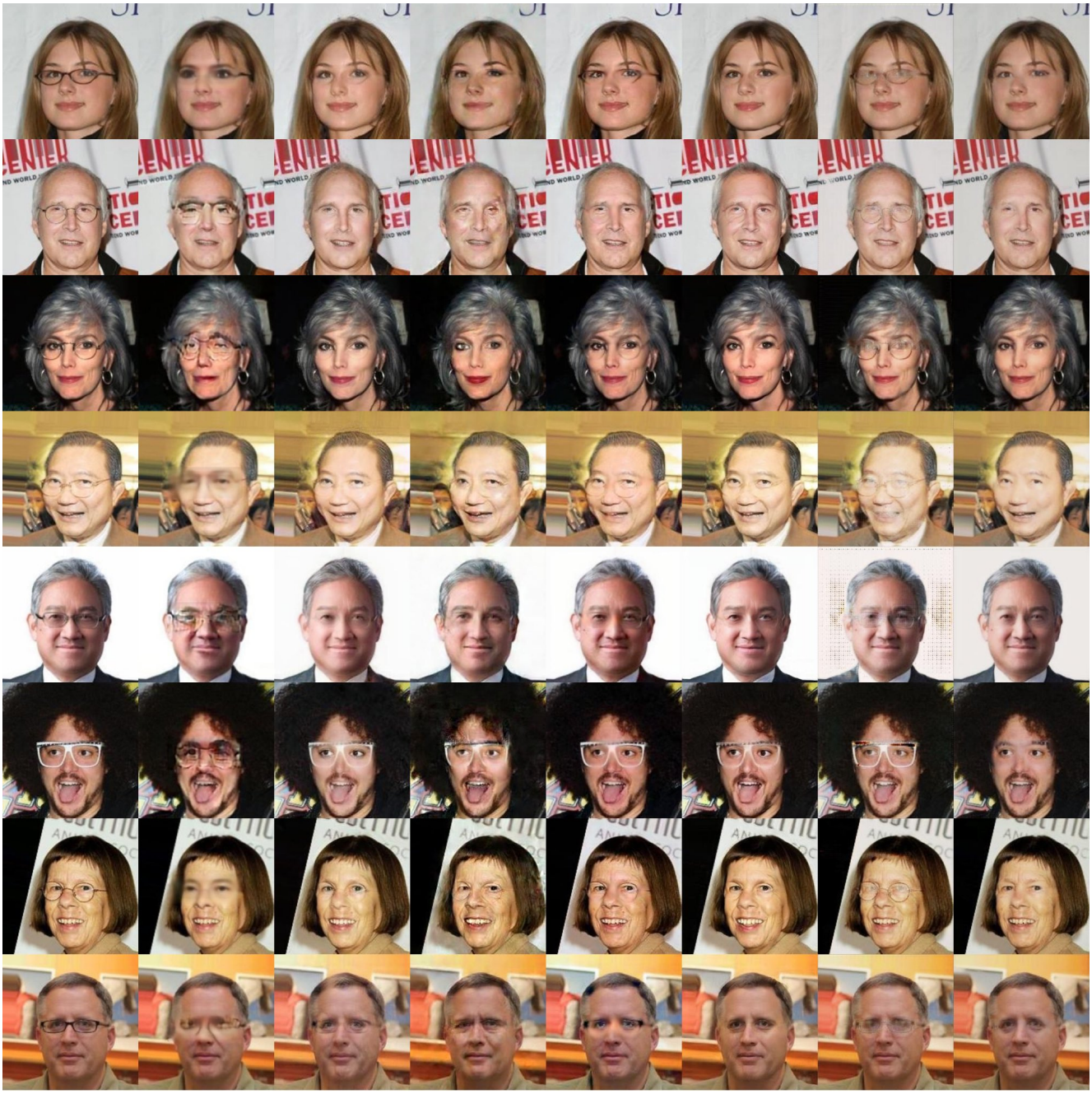


Figure 6. Qualitative comparisons on CelebA.