

# Supplementary Material for “FRNet: Improving Face De-Occlusion via Feature Reconstruction”

## 1 Comparisons With State-of-the-Art Methods

### 1.1 Datasets

We train the FRNet using our synthetic dataset which includes 88,932 sets of training data, 11,120 sets of validation data, and 10,244 sets of testing data. In addition, we also use CelebA-HQ [4], FFHQ [5], MeGlass [1], Gender Occlusion Data [9] and RMFD [14] to test the model.

**CelebA-HQ** is a real-world portrait data with 30,000 images, and each image is annotated with 40 binary attributes. Using attribute labels, we isolate 1,470 occluded face images.

**FFHQ** is a high-quality face image dataset with 70,000 PNG images. Following [8], we use face parsing [10] to roughly split 11,778 occluded face images.

**MeGlass** is a dataset containing 1,710 different identities. Each identity has at least two face images with glasses and two face images without glasses.

**Gender Occlusion Data** is a masked face synthesis dataset based on FFHQ, including 12,941 female face images and 10,357 male face images.

**RMFD** is a large masked face dataset, including real-world masked face data.

We use face landmark points to align all images of these datasets to the size of  $256 \times 256$ .

### 1.2 Experiments Details

We compare our method with state-of-the-art image inpainting methods including CTSRG [2], MADF [18], DSNet [12], WaveFill [15], AOT-GAN [16], MISF [6], LGNet [11], image translation methods including CycleGAN [17], pix2pixHD [13], and glasses removal methods including ERGAN [3], HiSD [7].

For the training process, image inpainting methods and image translation methods all use the proposed synthetic dataset. ERGAN and HiSD directly use their released models. In the testing process, for the image inpainting methods, we use the occluded face images and the occlusion masks predicted by our network as input. For the image translation methods and glasses removal methods, we use occluded face images as input. All the models are not trained on the test data, to ensure fair conclusions.

### 1.3 More Results

In order to illustrate the robustness of our proposed FRNet, we have presented additional qualitative results.

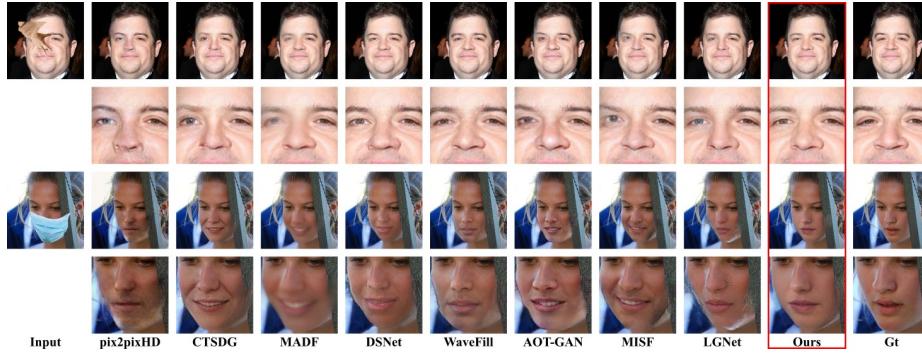
Fig. 1, Fig. 2, and Fig. 3 show the details of the generated images. By comparison, the face images generated by our method are more realistic, with richer details and clearer textures. As shown in the first group of Fig. 1, the eyes generated by our method exhibit consistent characteristics with known eyes. In the second group of images in Fig. 1, as well as in Fig. 3, when the occluded area is large, our method can effectively restore the global structure of the image while generating realistic content with fine-grained textures. From the images presented in Fig. 2, while the ERGAN method can remove occlusions, it struggles with effective inpainting of the occluded area. CycleGAN is unable to completely remove occlusions. Though pix2pixHD and HiSD can successfully remove occlusions, the resulting images still exhibit noticeable artifacts. The images generated by MISF display structural disharmony. LGNet’s generated images exhibit blurred edges in the restored areas. In contrast, our method effectively removes various types of eyeglasses and produces high-quality face images that are free of occlusions.

Fig. 4 and Fig. 5 show the results in our synthetic dataset. From these figures, we can observe that compared with the existing methods, our method can restore the global structures and details of the faces more effectively, while maintaining consistency within the features.

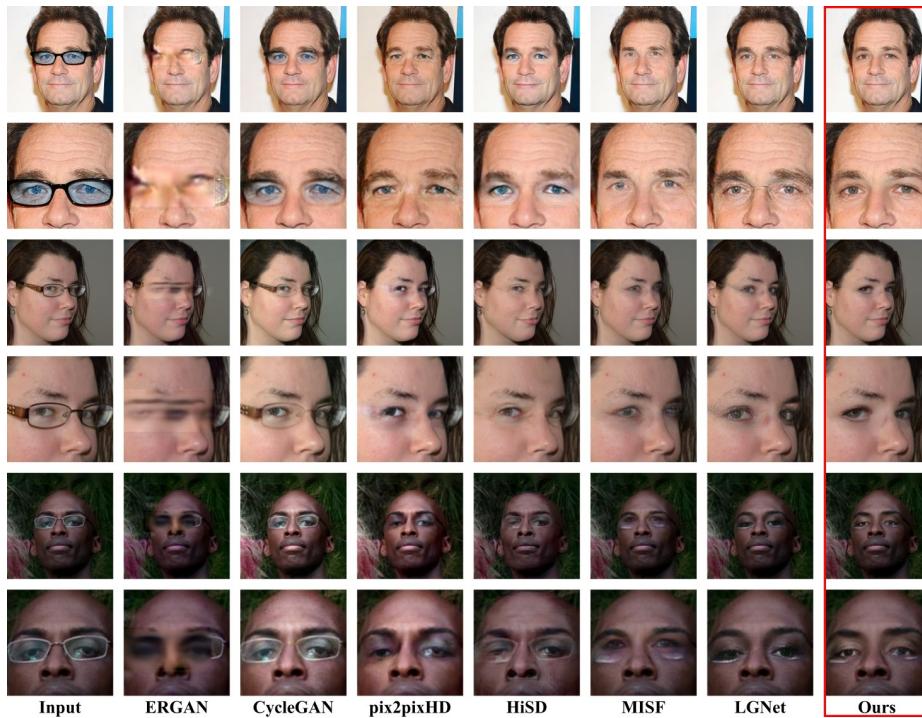
Fig. 6 and Fig. 7 show the results in Gender Occlusion Data. In Fig. 6, the 4th and 10th rows, and in Fig. 7, the 6th, 7th, 10th, 11th, and 12th rows, demonstrate that our model can effectively detect and remove multiple occlusions simultaneously on different datasets, producing visually realistic results.

Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12 and Fig. 13 show the results of removing glasses on CelebA-HQ, FFHQ and MeGlass. Our method also performs well on images with extreme head poses and eyeglasses with colored lenses.

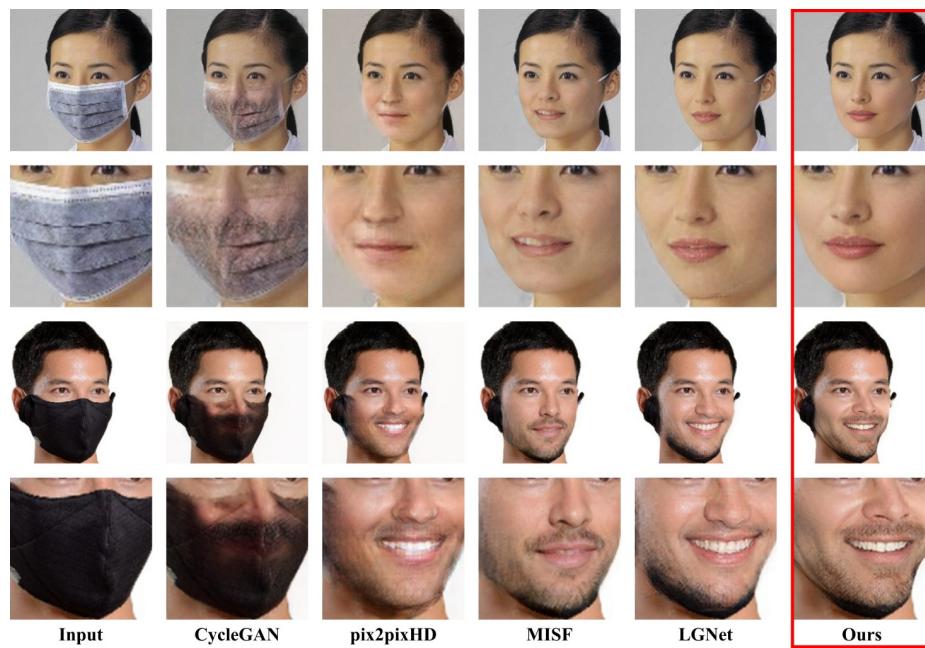
Fig. 14 shows the results of removing masks on RMFD. Our method can naturally extend to real-life masked face images and achieve superior results compared to existing methods.



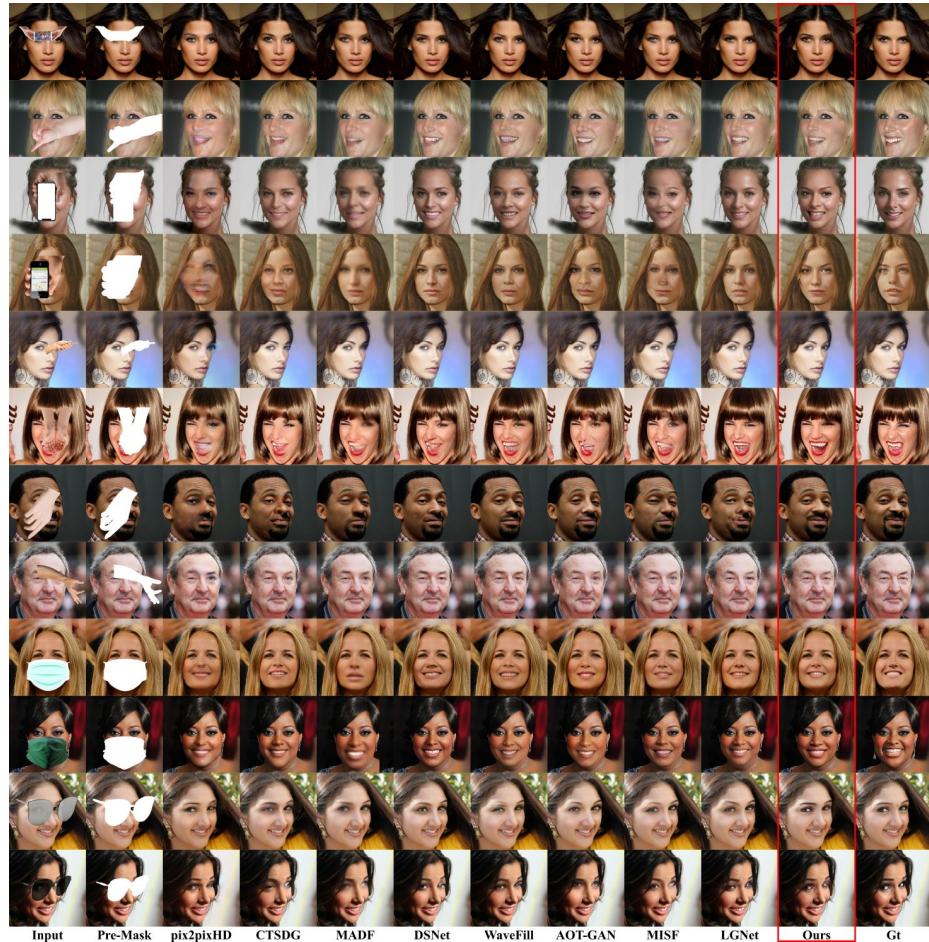
**Fig. 1.** Qualitative results on our synthetic dataset (top) and Gender Occlusion Data (bottom).



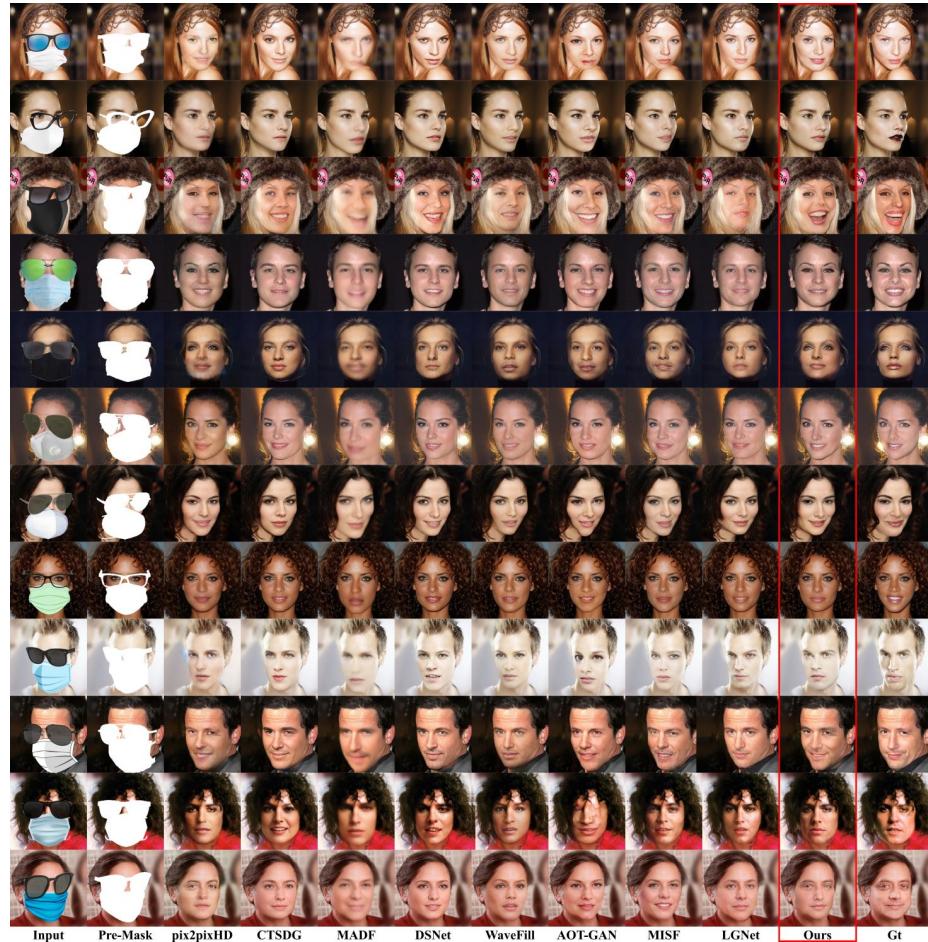
**Fig. 2.** Qualitative results on CelebA-HQ (top), FFHQ (middle), and MeGlass (bottom).



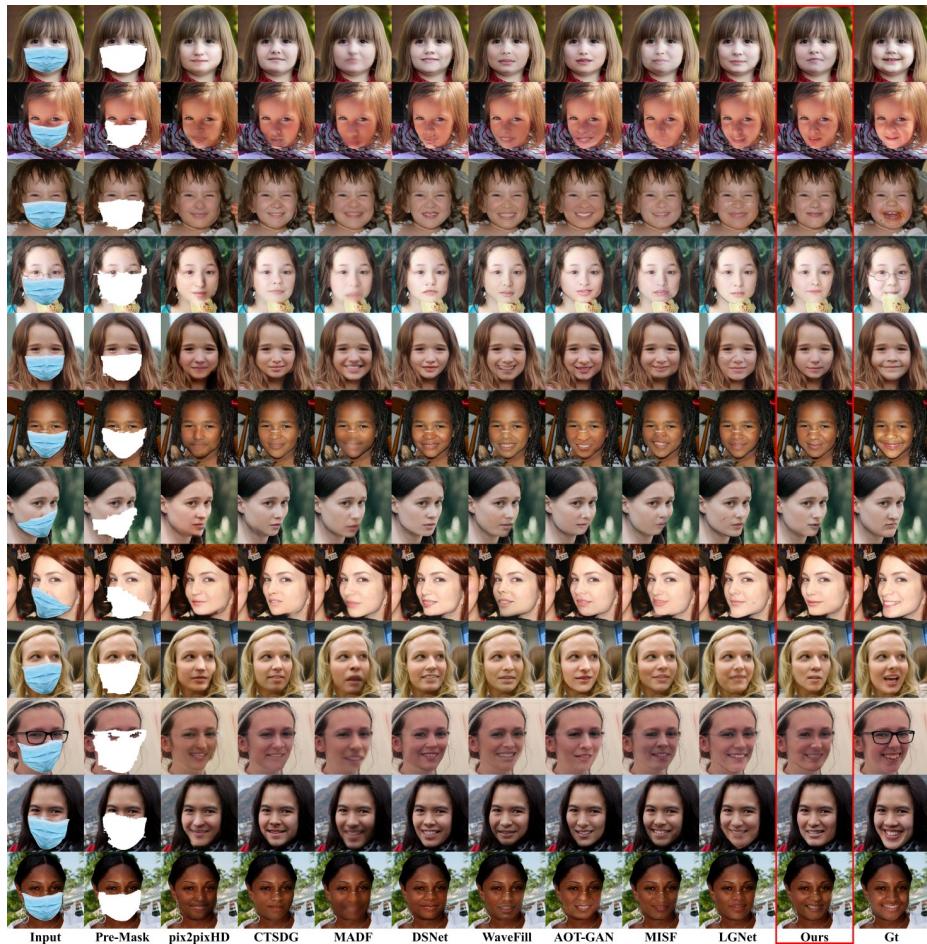
**Fig. 3.** Qualitative results on RMFD.



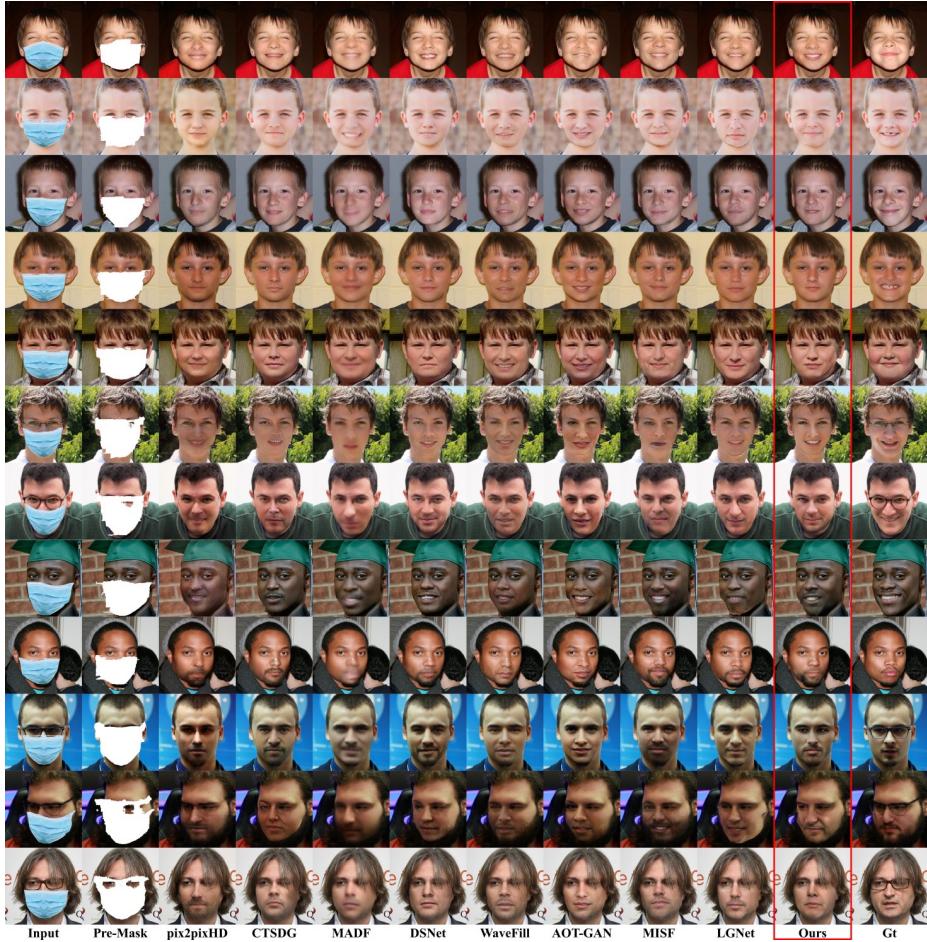
**Fig. 4.** Qualitative results of removing single occlusions on our synthetic dataset.



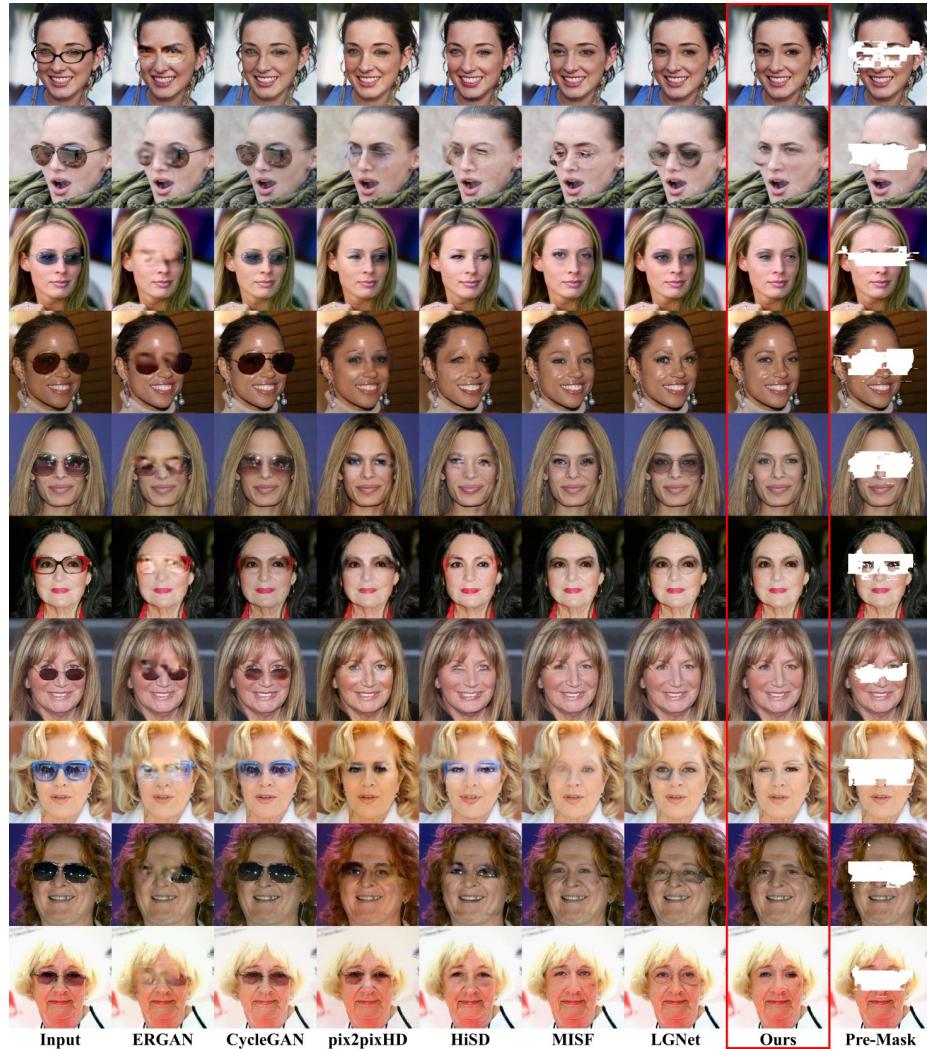
**Fig. 5.** Qualitative results of removing multiple occlusions on our synthetic dataset.



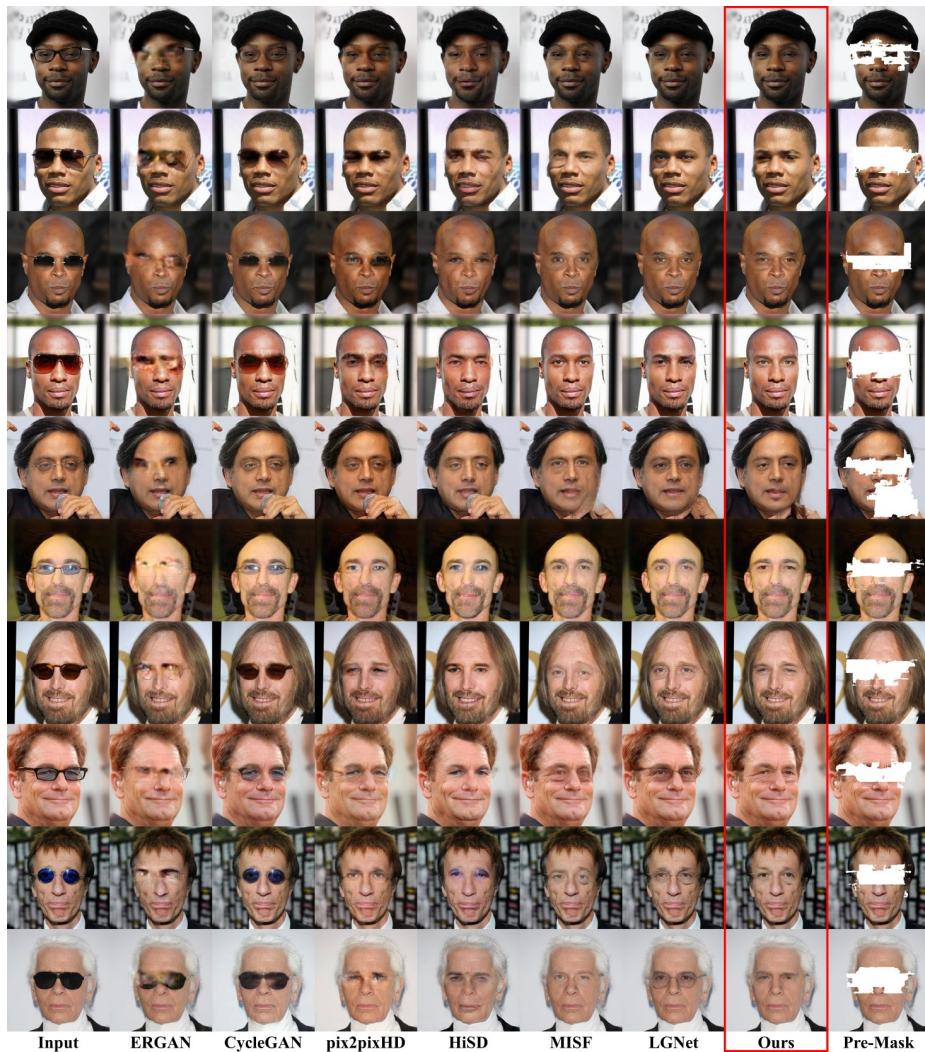
**Fig. 6.** Qualitative results of removing occlusions from female images in Gender Occlusion Data.



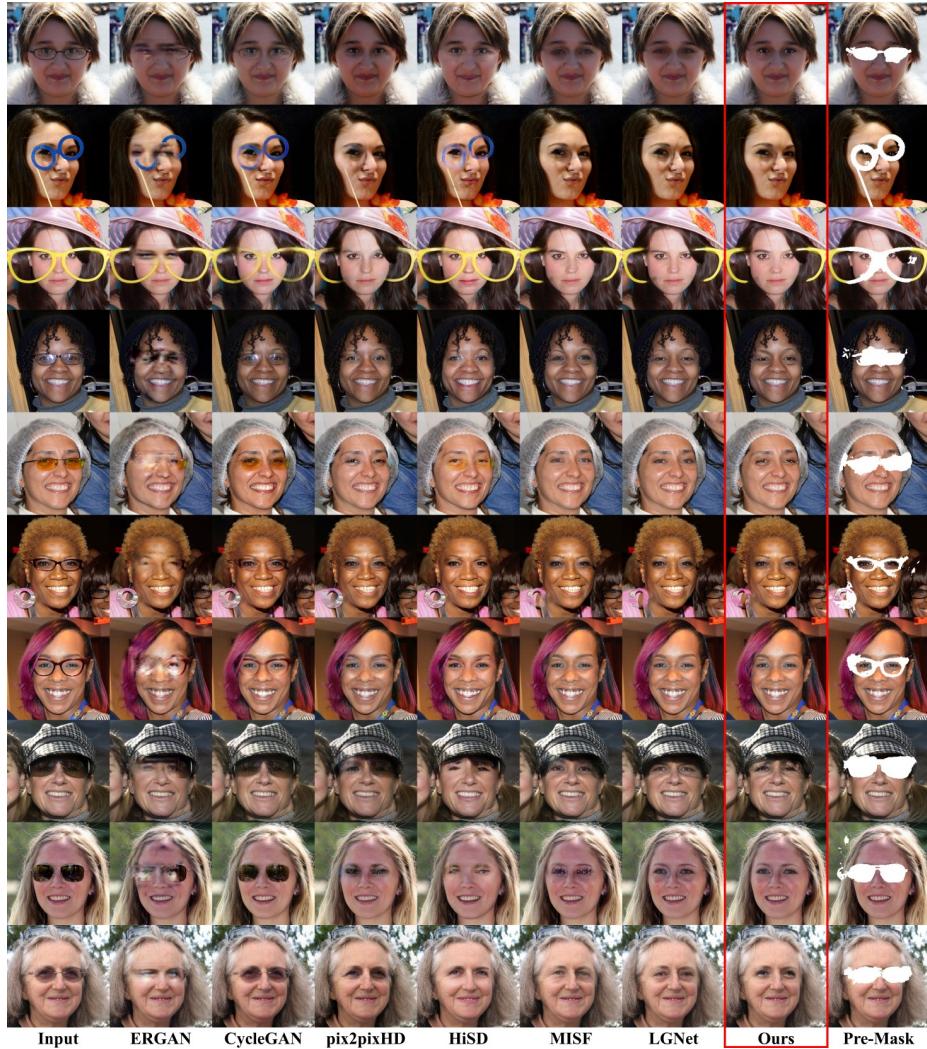
**Fig. 7.** Qualitative results of removing occlusions from male images in Gender Occlusion Data.



**Fig. 8.** Qualitative results of removing occlusions from female images in CelebA-HQ.



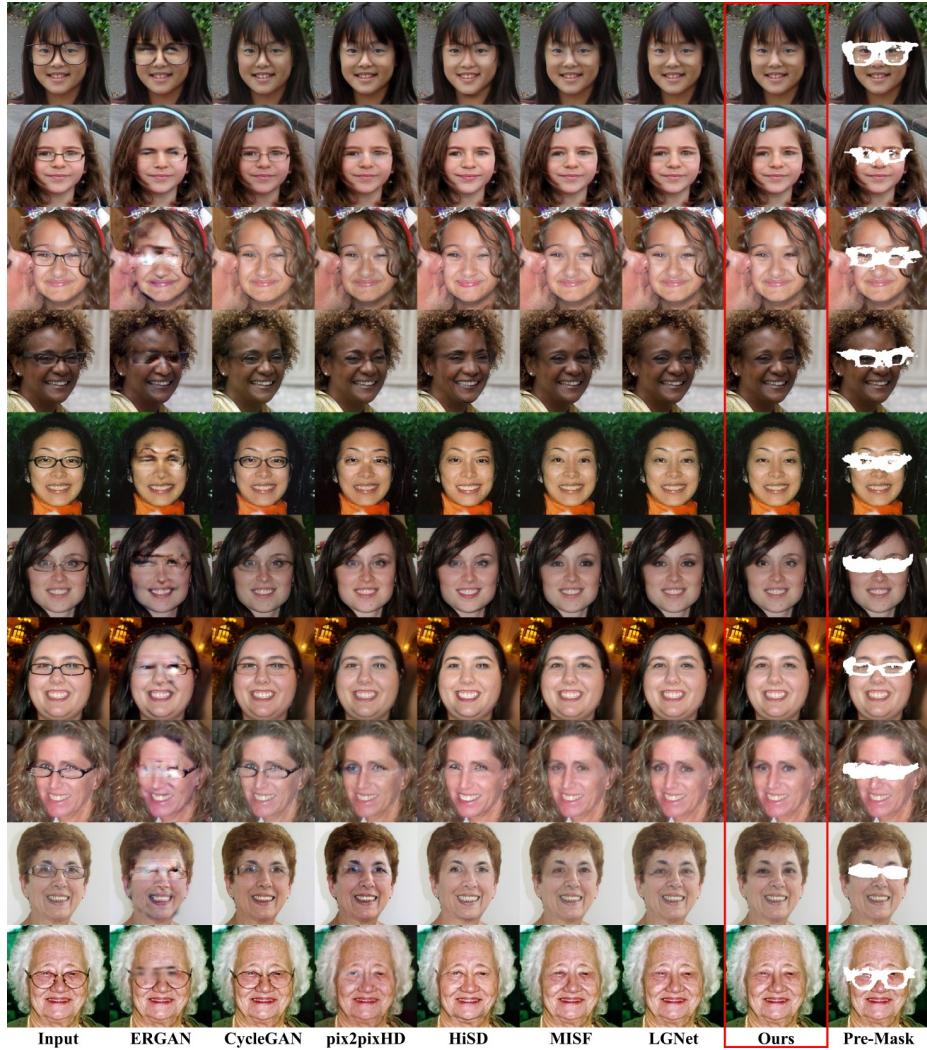
**Fig. 9.** Qualitative results of removing occlusions from male images in CelebA-HQ.



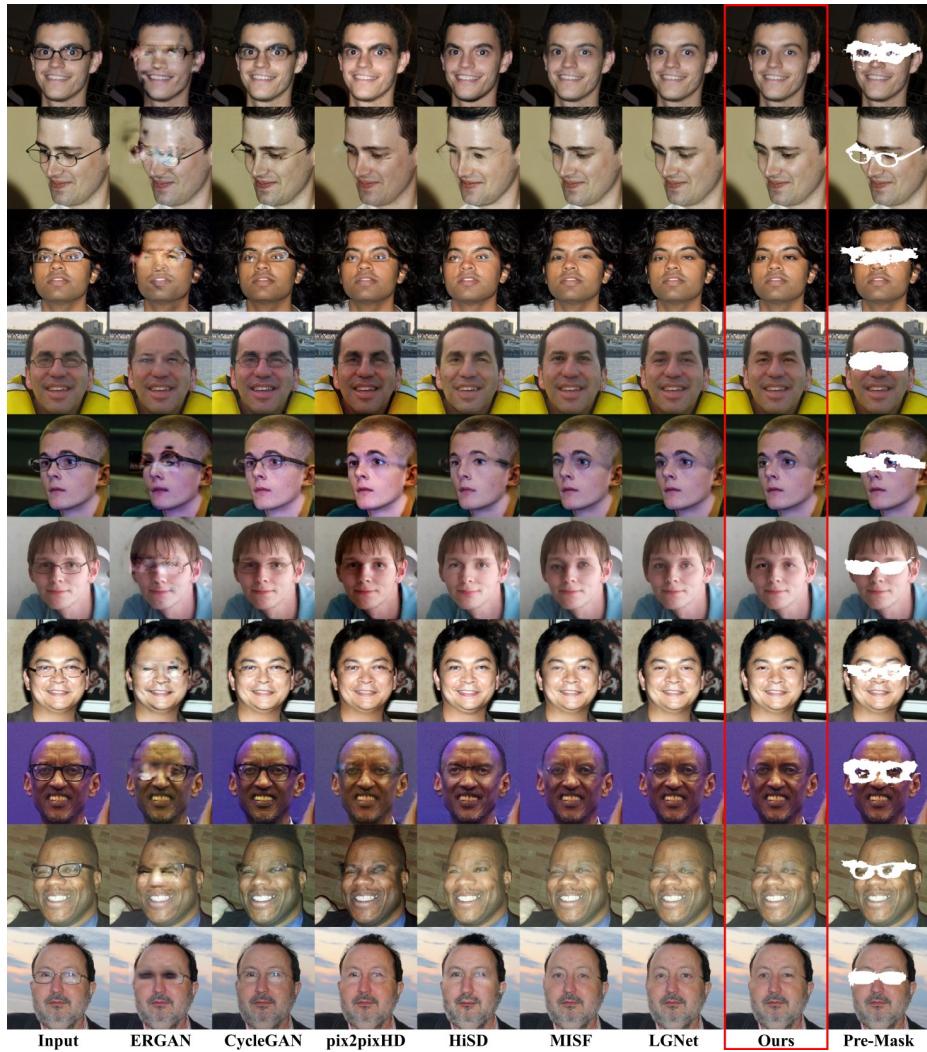
**Fig. 10.** Qualitative results of removing occlusions from female images in FFHQ.



**Fig. 11.** Qualitative results of removing occlusions from male images in FFHQ.



**Fig. 12.** Qualitative results of removing occlusions from female images in MeGlass.



**Fig. 13.** Qualitative results of removing occlusions from male images in MeGlass.

**Fig. 14.** Qualitative results on RMFD.

## References

1. Guo, J., et al.: Face synthesis for eyeglass-robust face recognition. In: CCBR. pp. 275–284. Springer (2018)
2. Guo, X., et al.: Image inpainting via conditional texture and structure dual generation. In: ICCV. pp. 14134–14143 (2021)
3. Hu, B., et al.: Unsupervised eyeglasses removal in the wild. TCYB **51**(9), 4373–4385 (2020)
4. Karras, T., et al.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)
5. Karras, T., et al.: A style-based generator architecture for generative adversarial networks. In: CVPR. pp. 4401–4410 (2019)
6. Li, X., et al.: Misf: Multi-level interactive siamese filtering for high-fidelity image inpainting. In: CVPR. pp. 1869–1878 (2022)
7. Li, X., et al.: Image-to-image translation via hierarchical style disentanglement. In: CVPR. pp. 8639–8648 (2021)
8. Lyu, J., et al.: Portrait eyeglasses and shadow removal by leveraging 3d synthetic data. In: CVPR. pp. 3429–3439 (2022)
9. Modak, G., et al.: A deep learning framework to reconstruct face under mask. In: CDMA. pp. 200–205. IEEE (2022)
10. Peng, J., et al.: Pp-liteseg: A superior real-time semantic segmentation model. arXiv preprint arXiv:2204.02681 (2022)
11. Quan, W., et al.: Image inpainting with local and global refinement. TIP **31**, 2405–2420 (2022)
12. Wang, N., et al.: Dynamic selection network for image inpainting. TIP **30**, 1784–1798 (2021)
13. Wang, T.C., et al.: High-resolution image synthesis and semantic manipulation with conditional gans. In: CVPR. pp. 8798–8807 (2018)
14. Wang, Z., et al.: Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093 (2020)
15. Yu, Y., et al.: Wavefill: A wavelet-based generation network for image inpainting. In: ICCV. pp. 14114–14123 (2021)
16. Zeng, Y., et al.: Aggregated contextual transformations for high-resolution image inpainting. TVCG (2022)
17. Zhu, J.Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV. pp. 2223–2232 (2017)
18. Zhu, M., et al.: Image inpainting by end-to-end cascaded refinement with mask awareness. TIP **30**, 4855–4866 (2021)