

Delving into High-Quality Synthetic Face Occlusion Segmentation Datasets

Supplementary Material

Kenny T. R. Voo Liming Jiang Chen Change Loy
S-Lab, Nanyang Technological University
{kvoo001, liming002, ccloy}@ntu.edu.sg

A. Additional Quantitative Results

More examples of the generated datasets and quantitative results are presented in this section.

Non-occluded Face Dataset. Table 1 shows the test results on the CelebAMask-HQ-WO (Test), which is a face dataset without any occlusion. This test aims to verify that our generated dataset will not affect the models' performance in segmenting non-occluded faces. The models trained with the synthetic dataset generated by NatOcc and RandOcc are improved compared to the models trained on C-original and C-WO.

Cropped and Aligned COFW Dataset. Additional test has been carried out on cropped and aligned COFW dataset [1]. 400 face images was successfully obtained using the same method as FFHQ [4]. Table 1 shows that the overall performance is better compared to the COFW dataset [1] without cropped and aligned.

B. Additional Qualitative Results

The additional qualitative results in Figure 1 and Figure 2 show that models trained with both NatOcc and RandOcc datasets perform as well as or better than models trained on the real-world dataset (C-WO). In some examples, they are even better than real-world datasets, showing the effectiveness of our data generation methods.

C. More Examples of NatOcc and RandOcc Datasets

Additional examples of the NatOcc dataset are shown in Figure 3, Figure 4, and Figure 5. Figure 3 shows the hand-occluded faces without color transfer while Figure 4 shows the hand-occluded faces with color transfer. Figure 5 shows the COCO [5] objects-occluded faces after image harmonization. Moreover, some examples of the RandOcc dataset are shown in Figure 6.

D. More Examples of Validation Sets

Figure 7 shows additional examples of the RealOcc while Figure 8 shows some examples of occluded faces in the wild. The images are resized to $1,024 \times 1,024$ to ease the visualization.

References

- [1] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1513–1520, 2013. 1
- [2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision*, 2018. 2
- [3] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 7
- [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] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1, 6
- [6] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34, 2021. 2
- [7] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2

Table 1. **Additional quantitative test results:** PSPNet [7], DeepLabv3+ [2], and SegFormer [6] with different combination of datasets. The best results for each validation set are marked in bold. The metrics are mIoU (higher is better).

	Quantity	CelebAMask-HQ-WO (Test) (mIoU)			COFW (Train) (cropped and aligned) (mIoU)		
		PSPNet	DeepLabv3+	SegFormer	PSPNet	DeepLabv3+	SegFormer
C-Original	29,200	97.71	97.23	97.18	93.21	92.37	92.87
C-CM	29,200	97.78	97.79	97.88	95.34	95.32	95.62
C-WO	24,602	97.66	97.70	97.84	92.88	92.74	93.54
C-WO + C-WO-NatOcc	24,602 + 49,204	97.77	97.76	97.86	94.45	94.46	94.87
C-WO + C-WO-NatOcc-SOT	24,602 + 49,204	97.71	97.77	97.87	94.61	94.47	94.63
C-WO + C-WO-RandOcc	24,602 + 49,204	97.68	97.76	97.83	93.97	93.83	94.19
C-WO + C-WO-Mix	24,602 + 49,204	97.70	97.76	97.76	93.92	94.55	94.67
C-CM + C-WO-NatOcc	29,200 + 49,204	97.74	97.79	97.87	95.38	95.32	95.53
C-CM + C-WO-NatOcc-SOT	29,200 + 49,204	97.78	97.76	97.85	95.26	95.23	95.46

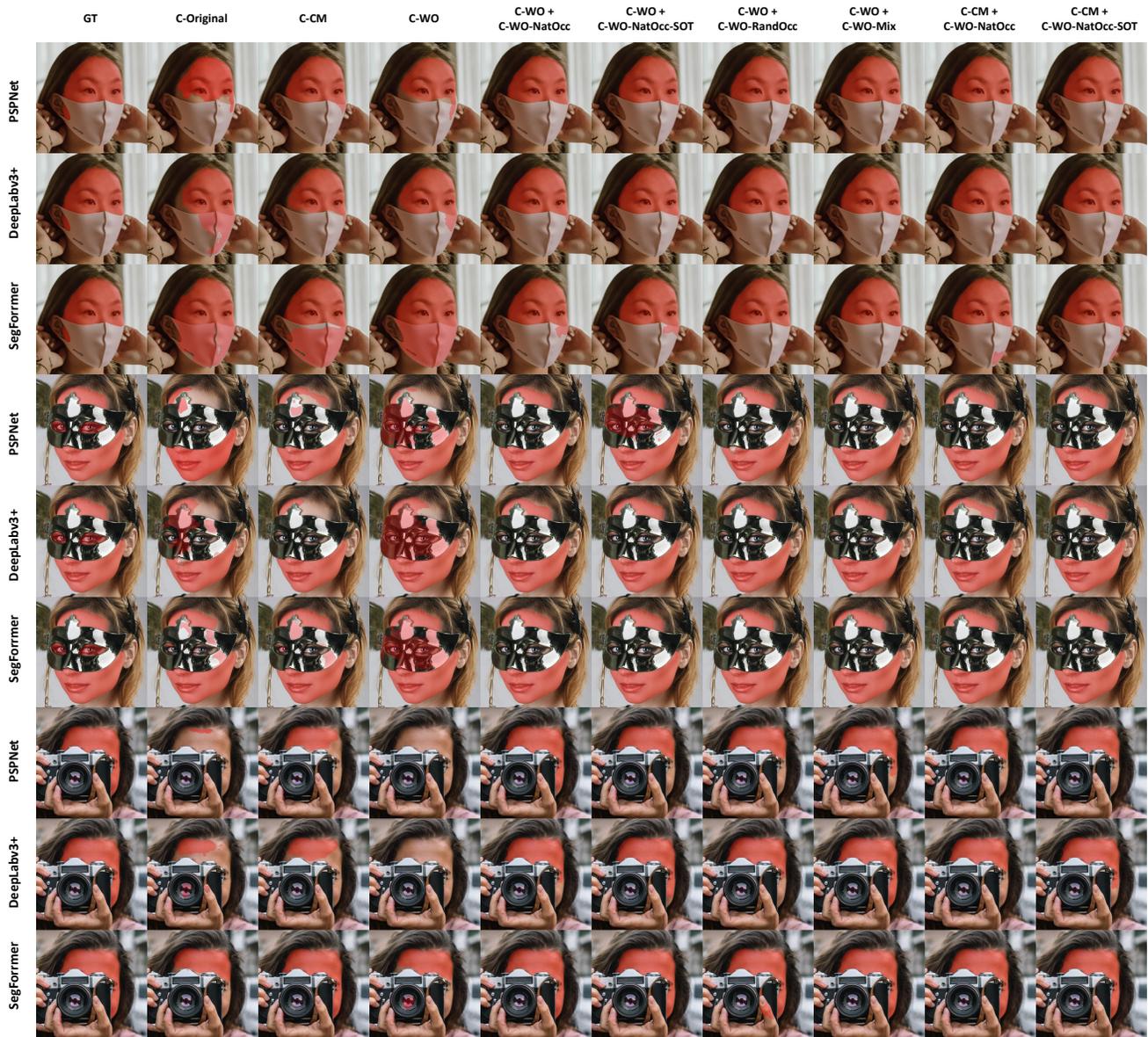


Figure 1. Examples of the inference results on hand-occluded faces. NatOcc and RandOcc are effective in simulating real-world occluded datasets.



Figure 2. Examples of the inference results on objects-occluded faces. NatOcc and RandOcc are effective in simulating real-world occluded datasets.



Figure 3. Examples of the hand-occluded faces generated by NatOcc without color transfer.



Figure 4. Examples of the hand-occluded faces generated by NatOcc with color transfer.



Figure 5. Examples of the COCO [5] objects occluded faces generated by NatOcc without color transfer.



Figure 6. Examples of the images generated by RandOcc by overlaying random shape with random transparency and texture from DTD [3].



Figure 7. Examples of the RealOcc, aligned and cropped real-world occluded faces.



Figure 8. Examples of the RealOcc-Wild, real-world occluded faces in the wild. The images were resized to $1,024 \times 1,024$ for easier visualization. The size and aspect ratio are not the same as the original images.