Reference Guided Image Inpainting using Facial Attributes

-Supplementary material-

Dongsik Yoon[1]
kevinds1106@korea.ac.kr
Jeonggi Kwak[1]
kjk8557@korea.ac.kr
Yuanming Li[1]
lym7499500@korea.ac.kr
David Han[2]
dkh42@drexel.edu
Youngsaeng Jin[1]
youngsjin@korea.ac.kr
Hanseok Ko*[1]
hsko@korea.ac.kr

[1] Korea University
Seoul, South Korea
[2] Drexel University
Philadelphia, USA

6 Appendix

This section provides supplementary material of our reference guided image inpainting. It mainly contains the following:

- The illustration of our inpainting model architecture used in the main paper.
- More diverse and detailed implementation of the qualitative comparisons.
- To validate the proposed framework's loss function with ablation study.

6.1 Model Architecture

We supply several omitted architecture details from the main paper. Figure 1 shows the proposed generator. Our encoder and decoder adopt Learnable Bidirectional Attention Maps (LBAM)[3] in each layer to allow inpainting of any irregular mask.

6.2 Qualitative Comparisons

We compare an image inpainting quality against three state-of-the-art methods. Figure 2 shows images generated by the proposed method with those generated by other methods. Images generated by LBAM [3] have failed to maintain a facial structure such as a nose.

MLGN [2] has preserved important structures but has shown blurry results. Compared to existing methods, our method accomplished more remarkable and stable performance. Figure 4 shows an overview of the proposed facial image inpainting.

6.3 Ablation Study

To validate the effectiveness of the proposed loss function, we conduct qualitative and quantitative comparisons. Figure 3 shows the inpainting results with and without specific loss terms. Particularly, we demonstrate that MS-SSIM loss is an important component of our framework to maintain overall structure. Table 1 is quantitative comparisons with and without specific loss terms.

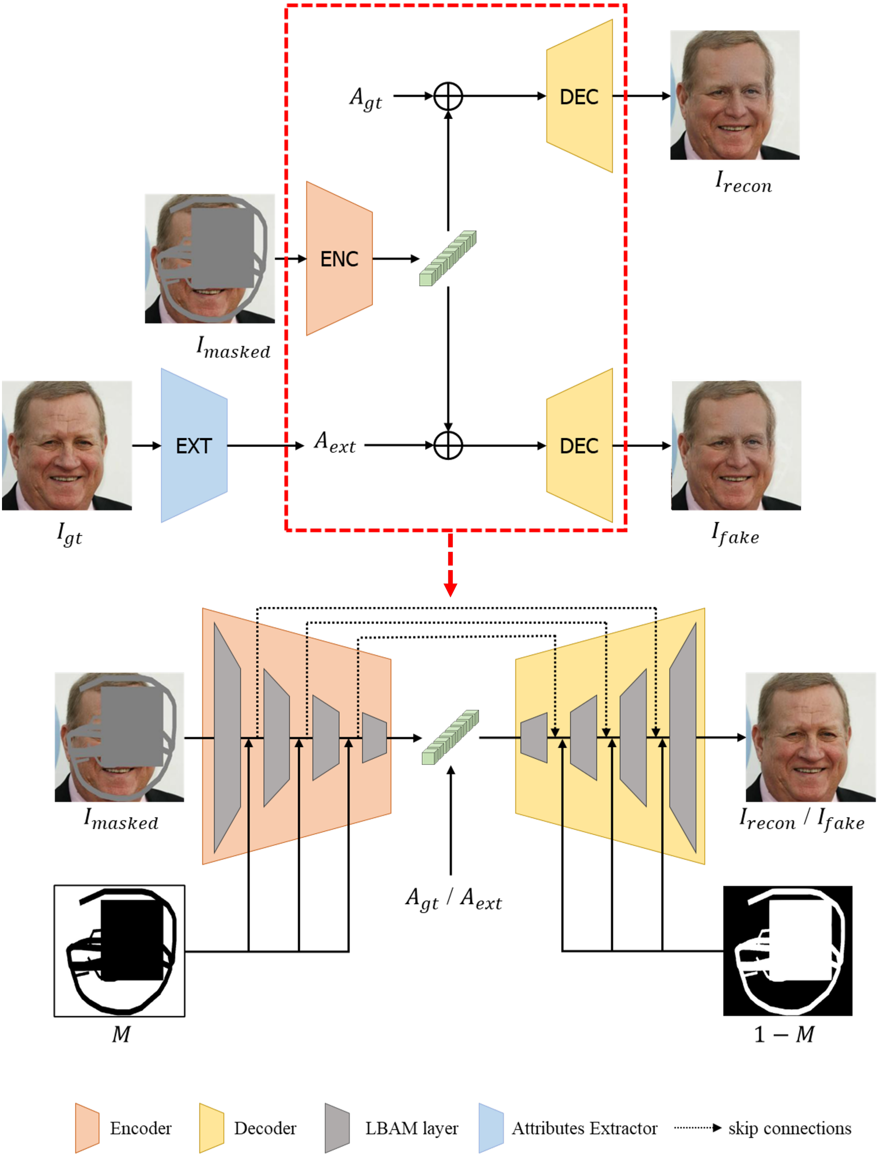

Figure 1: Details of our encoder and decoder.

Figure 2: Qualitative comparisons with LBAM [3], MLGN [2], MEDFE[1] and the proposed method.

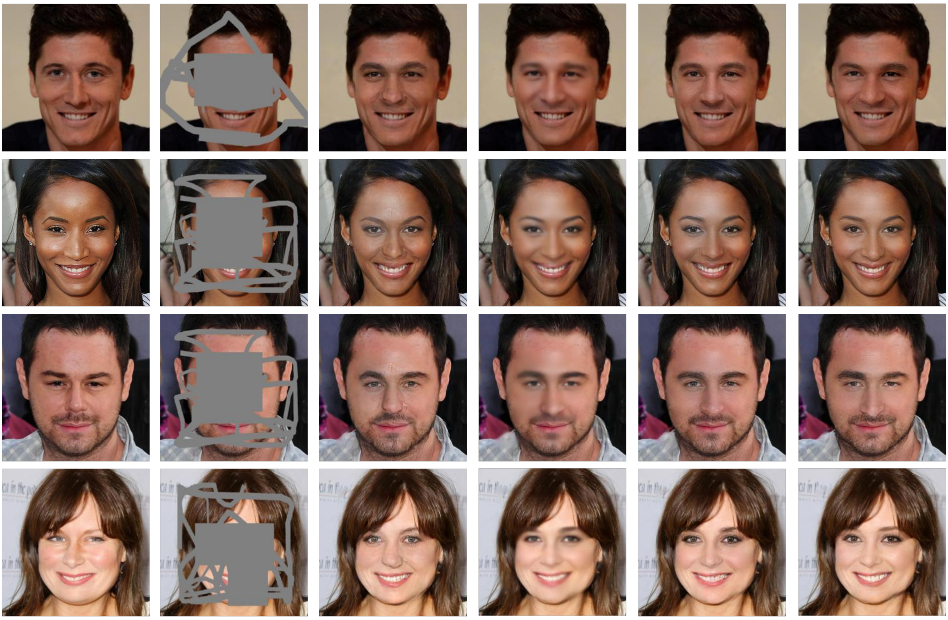

Figure 3: Qualitative comparisons with and without specific loss terms.

	Mask	Ours	W/O MS-SSIM	W/O VGG style	W/O perceptual
SSIM	Quickdraw	0.833	0.824	**0.840**	0.831
	10-20%	0.811	0.808	**0.813**	0.808
	20-30%	0.740	0.736	**0.741**	0.734
	30-40%	**0.660**	0.656	0.659	0.652
	40-50%	**0.571**	0.570	0.568	0.563
LPIPS	Quickdraw	**0.042**	0.046	0.062	0.045
	10-20%	0.068	**0.063**	0.088	0.070
	20-30%	0.103	**0.096**	0.137	0.106
	30-40%	0.146	**0.137**	0.196	0.150
	40-50%	0.198	**0.187**	0.266	0.202
FID	Quickdraw	**24.91**	25.60	27.92	25.41
	10-20%	**28.53**	29.74	32.80	28.72
	20-30%	**35.26**	39.79	43.81	36.52
	30-40%	**43.29**	49.60	58.50	45.58
	40-50%	**56.67**	67.57	80.41	62.25

Table 1: Quantitative comparison on CelebA-HQ. The best results of each row is boldfaced.

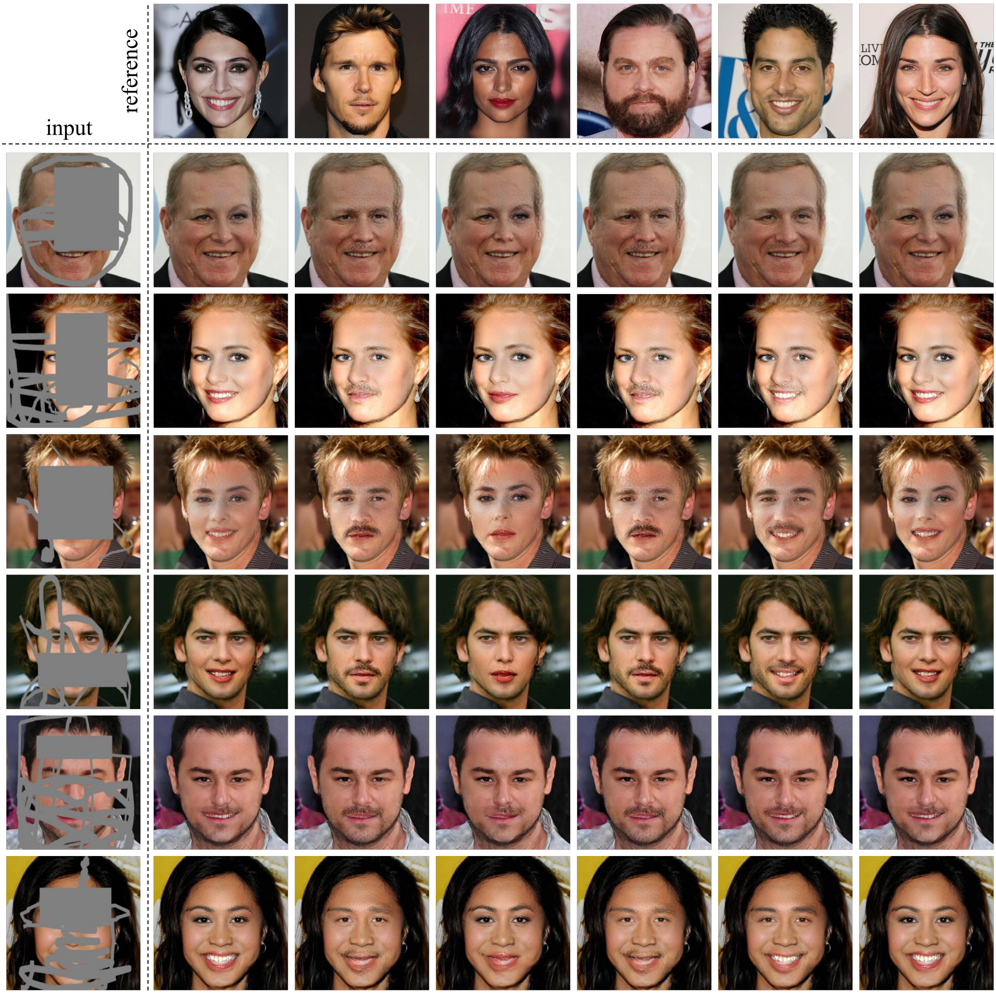

Figure 4: Overview of the proposed facial image inpainting.

References

[1] Hongyu Liu, Bin Jiang, Yibing Song, Wei Huang, and Chao Yang. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 725–741. 2020.

[2] Jie Liu and Cheolkon Jung. Facial image inpainting using multi-level generative network. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1168–1173. IEEE, 2019.

[3] Chaohao Xie, Shaohui Liu, Chao Li, Ming-Ming Cheng, Wangmeng Zuo, Xiao Liu, Shilei Wen, and Errui Ding. Image inpainting with learnable bidirectional attention maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8858–8867, 2019.