Supplementary Material for MSFSR: A Multi-Stage Face Super-Resolution with Accurate Facial Representation via Enhanced Facial Boundaries

Yunchen Zhang, Yi Wu, Liang Chen*
Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application,
Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education,
Fujian Provincial Key Laboratory of Photonics Technology,
Fujian Normal University

cydiachen@cydiachen.tech, wuyi@fjnu.edu.cn, cl_0827@126.com

Abstract

In this supplementary material, we first provide more details on generating enhanced facial boundaries from facial landmark points. We then display more comprehensive experimental results, including failure cases. Finally, the generalization performance of the proposed method is demonstrated.

1. Implementation Details

Boundary Type	Index
Facial Outer Contour	1 - 17
Left Eye (Closed loop)	37 - 42
Right Eye (Closed loop)	43 - 48
Left Eyebrow	18 - 22
Right Eyebrow	23 - 27
Nose Bridge	28 - 31
Nose Tip	32 - 36
Upper Boundary of Upper Lip	49 - 55
Lower Boundary of Upper Lip	61 - 65
Upper Boundary of Lower Lip	66 - 68
Lower Boundary of Lower Lip	56 - 60

Table 1. The instruction of generating enhanced facial boundaries from 68 landmark points.

We give implementation details of connecting landmark points to form enhanced facial boundaries. The corresponding relationship between 11 enhanced facial boundaries and 68 facial landmark points is shown in Table 1. The enhanced facial boundaries are connected according to their semantic meanings. We form closed loops to deal with contours around human eyes. Figure 1 demonstrates 11 enhanced facial boundaries derives from 68 facial landmark points.

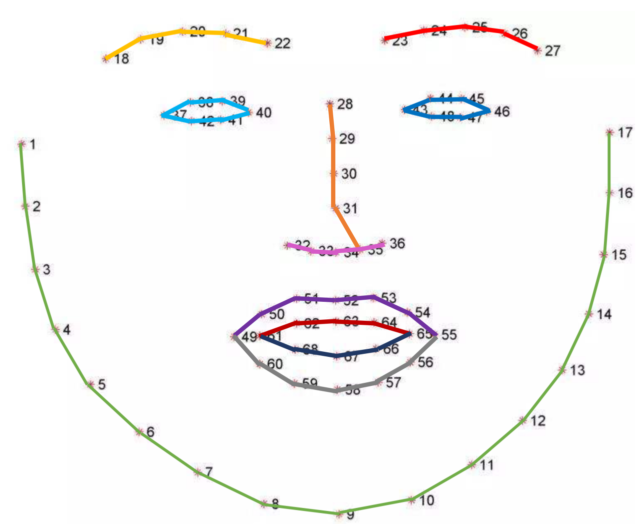

Figure 1. The demonstration of generating 11 enhanced facial boundaries from the 68 landmark points annotation.

2. More Qualitative Results

In this section, we present more visual results of different super-resolution methods. Figure 2 and Figure 3 show visual results of FSR with models trained on a mixed train-set. The mixed train-set is composed of CelebA-HQ [4], Helen [3] and WFLW [8] train-sets proposed in section 4.1. The reimplemented methods are trained with the same mixed dataset.

In Figure 3, we show some failure cases of the proposed

*Corresponding author.

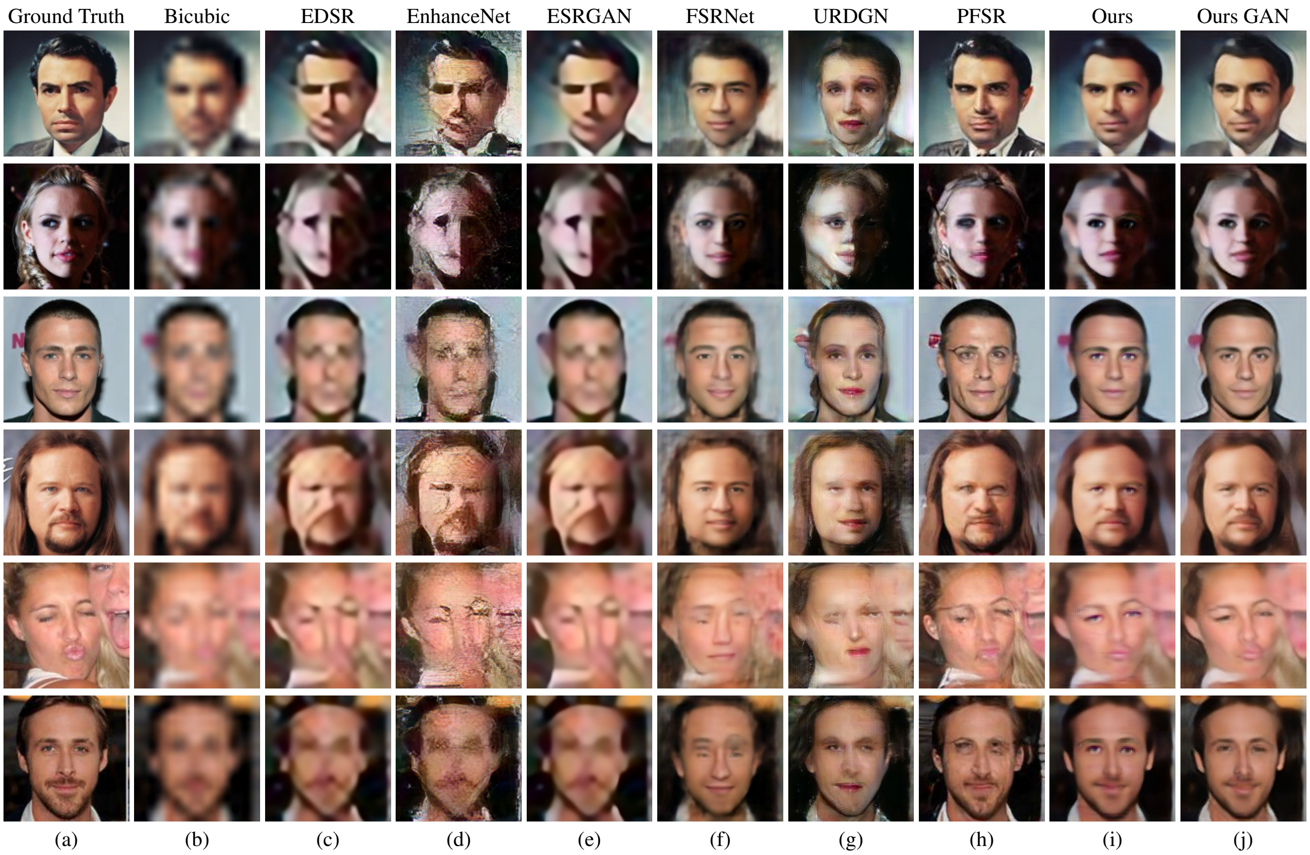

Figure 2. Visual results of different super-resolution methods on upscale factor of 8×. (a) Ground Truth. (b) Bicubic. (c) EDSR [5]. (d) EnhanceNet [6]. (e) ESRGAN [7]. (f) FSRNet [1]. (g) URDGN [10]. (h) PFSR [2]. (i) Ours. (j) Ours with GAN.

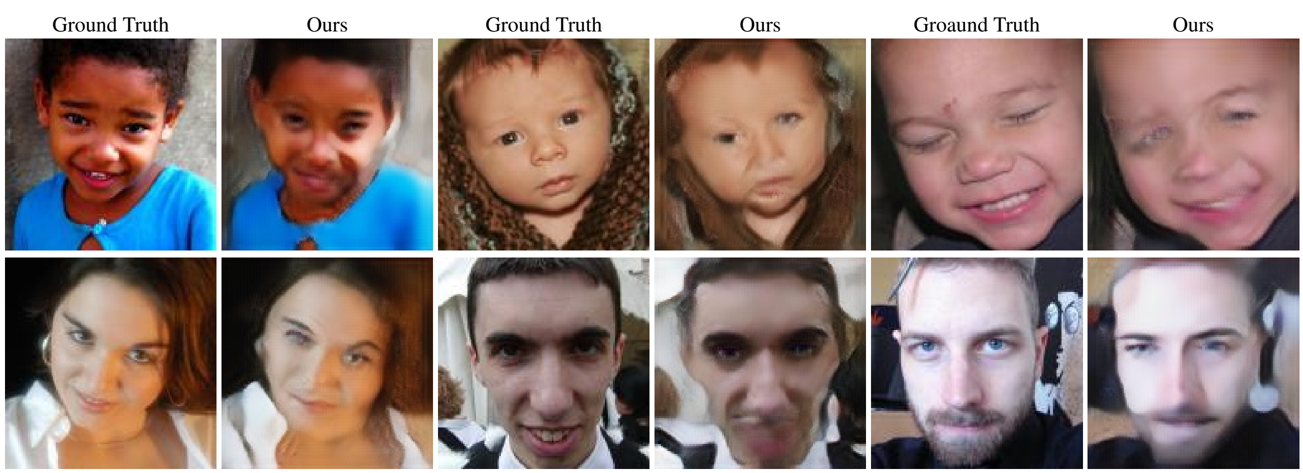

Figure 3. Visual results of failure cases on upscale factor of 8×. We can observe that our network generates unsatisfying facial images when dealing with facial images with large poses.

method. The failure cases can be divided into two categories. As shown in the first line of Figure 3, the proposed method generates facial images with blurry facial components. The deficiency could attributes to the distribution of the existing face dataset, with fewer infant and teenage faces. As demonstrated in the second line of Figure 3, these face images are captured from different angles, which results in severe shape distortions of facial components. These faces have their unique geometric structures so that the proposed method fails to extract accurate facial representation from these images.

3. The demonstration of generalization on different datasets

The generalization ability of the proposed network is important when dealing with different datasets. We display the generalizability of the proposed network to test cross-dataset generalization in Figure 4. The model is trained on CelebA-HQ [4] dataset, the test-set comes from WebFace [9] dataset.

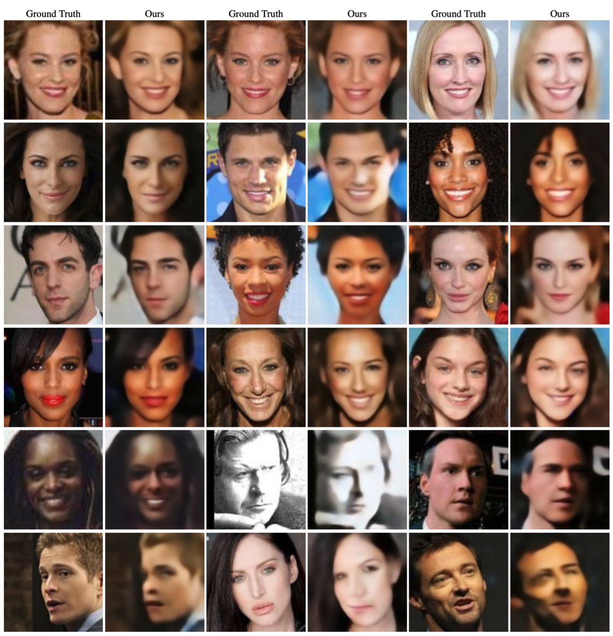

Figure 4. Visual results of the model trained on CelebA-HQ [4] dataset to deal with images from the test-set of WebFace [9] datasets on upscale factor of 8×.

We see that the proposed method is capable of dealing with LR face images from a different face dataset and hallucinating HR faces with realistic textures. The proposed method still suffers from performance degradation when dealing with face images captured from different angles. You can refer to Figure 5 for observing details on reconstructed faces.

References

[1] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2492–2501, 2018. 2

[2] Deokyun Kim, Minseon Kim, Gihyun Kwon, and Dae-Shik Kim. Progressive face super-resolution via attention to facial landmark. *arXiv preprint arXiv:1908.08239*, 2019. 2

[3] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang. Interactive facial feature localization. In *European conference on computer vision*, pages 679–692. Springer, 2012. 1

[4] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: towards diverse and interactive facial image manipulation. *arXiv preprint arXiv:1907.11922*, 2019. 1, 3, 4

[5] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 136–144, 2017. 2

[6] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4491–4500, 2017. 2

[7] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision (ECCV) Workshops*, 2018. 2

[8] Wayne Wu, Chen Qian, Shuo Yang, Quan Wang, Yici Cai, and Qiang Zhou. Look at boundary: A boundary-aware face alignment algorithm. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2129–2138, 2018. 1

[9] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014. 3, 4

[10] Xin Yu and Fatih Porikli. Ultra-resolving face images by discriminative generative networks. In *European conference on computer vision*, pages 318–333. Springer, 2016. 2

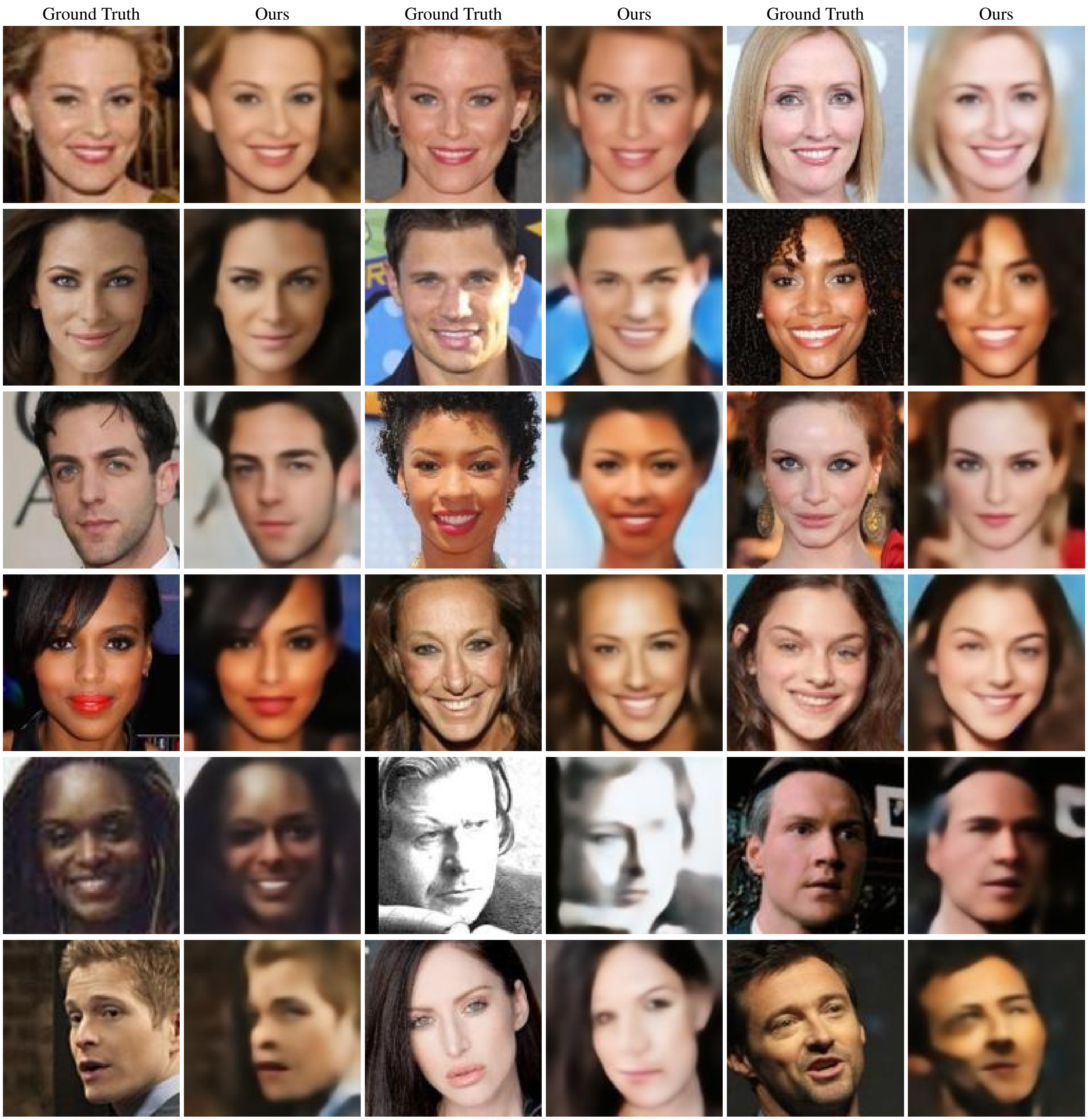

Figure 5. Visual results of the model trained on CelebA-HQ [4] dataset to deal with test images from WebFace [9] datasets on upscale factor of 8×.