GCFSR: a Generative and Controllable Face Super Resolution Method Without Facial and GAN Priors
Supplementary File

Jingwen He[1] Wu Shi[2,3] Kai Chen[1] Lean Fu[1] Chao Dong[2,4,*]
[1]ByteDance Inc, [2]ShenZhen Key Lab of Computer Vision and Pattern Recognition,
SIAT-SenseTime Joint Lab, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences,
[3]Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems,
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China,
[4]Shanghai AI Laboratory, Shanghai, China.

Abstract

In this supplementary file, we first give the detailed descriptions of our GCFSR in Section 1. Secondly, we provide the quantitative results to demonstrate the effectiveness of our proposed style modulation module (Section 2) and feature modulation module (Section 3). Then, we show more qualitative comparison with state-of-the-art methods (Section 4) and qualitative results of modulation on generative strength (Section 5). In Section 6, more visualization on feature modulation for different levels are illustrated. In the end, we present the qualitative results of GCFSR$_{adv}$ in Section 7.

1. More detailed descriptions on GCFSR.

As for the architecture of GCFSR, the encoder contains six 3×3 convolutional layers with stride 2, and each convolutional layer is followed by a leakyrelu activation layer. These strided convolutional layers gradually downsample the feature maps from resolution 1024^2 to resolution 16^2. Besides, the encoder uses another two strided convolutional layers and one fully connected layer to generate the latent codes $\boldsymbol{w}$. Then, the generator takes the topmost encoded 16×16 feature maps as well as the latent codes to generate realistic facial details by style modulation. The number of parameters in GCFSR is 66.69M.

2. The effectiveness of style modulation module in quantitative evaluation.

In this section, we provide the quantitative comparison of GCFSR with and w/o style modulation module for $16\times$, $32\times$ and $64\times$ SR. In general, the style modulation module improves the overall performance in most metrics.

Table 1. Quantitative comparison of GCFSR with and w/o style modulation module on CelebA-HQ for $16\times$, $32\times$ and $64\times$ SR. **Bolded** texts represent the best performance.

	style modulation	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	similarity ↑
16×	w	**27.17**	**0.7100**	**0.2604**	**30.48**	**0.9631**
	w/o	27.16	0.7081	0.2713	33.18	0.9602
32×	w	**24.95**	**0.6748**	**0.3061**	**43.34**	**0.7911**
	w/o	24.94	0.6735	0.3207	44.97	0.7887
64×	w	22.39	0.6315	**0.3663**	**57.15**	**0.6620**
	w/o	**22.42**	**0.6336**	0.3750	60.03	0.6463

* Corresponding author (e-mail: chao.dong@siat.ac.cn)

3. The effectiveness of feature modulation in quantitative evaluation.

In this section, we aim to demonstrate the effectiveness of our proposed feature modulation. Specifically, we remove the feature modulation from GCFSR and use the remaining network architecture (namely as GFSR) to train three single models for three upscaling factors ($16\times$, $32\times$, $64\times$), separately. The quantitative results are presented in Table 2 (the results of GLEAN are also provided as reference). As can be seen, in contrast with GCFSR, the single model GFSR achieves slightly better performance on $16\times$ SR, performs comparably on $32\times$ SR, but obtains worse results on $64\times$ SR. Nevertheless, the difference between GCFSR and GFSR in performance is minor (see Figure 2), indicating the superiority of our proposed Feature Modulation. On the other hand, our GFSR outperforms GLEAN for all upscaling factors, which again demonstrates the effectiveness of our proposed network architecture and the end-to-end training strategy.

Table 2. Comparison of GCFSR, GFSR (our single version), and GLEAN on CelebA-HQ for $16\times$, $32\times$, and $64\times$ SR. GFSR and GLEAN both use three models for three different SR tasks. **Red** and blue indicate the best and the second best performance. Similarity. represents Cosine similarity of ArcFace Embeddings. The absolute difference between GCFSR and GFSR is given.

		PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	similarity ↑
$16\times$	GLEAN [1]	26.88	0.6953	0.2693	29.99	0.9682
	GCFSR	**27.17**	**0.7100**	0.2604	30.48	0.9631
	GFSR	27.16	0.7085	**0.2560**	**29.57**	**0.9747**
	diff(abs)	0.01	0.0015	0.0044	0.91	0.0116
$32\times$	GLEAN	24.34	0.6534	0.3257	46.57	0.7750
	GCFSR	**24.95**	0.6748	0.3061	**43.34**	0.7911
	GFSR	24.89	**0.6761**	**0.3059**	45.08	**0.7979**
	diff(abs)	0.06	0.0013	0.0002	1.74	0.0068
$64\times$	GLEAN	21.38	0.6016	0.4109	62.93	0.6118
	GCFSR	**22.39**	0.6315	**0.3663**	**57.15**	**0.6620**
	GFSR	22.21	**0.6377**	0.3689	59.79	0.6261
	diff(abs)	0.18	0.0062	0.0026	2.64	0.0359

4. More qualitative comparison with state-of-the-art methods.

Figure 1. Qualitative comparisons of GLEAN [1], GPEN [7], GFPGAN [5], and GCFSR (ours) on CelebA-HQ [3] for $16\times$ SR. The GT image (Right) has a resolution of 1024^2. **Zoom in for best view.**

Figure 2. Qualitative comparisons of GLEAN [1], GPEN [7], GFPGAN [5], and GCFSR (ours) on CelebA-HQ [3] for 32× SR. The GT image (Right) has a resolution of 1024^2. **Zoom in for best view.**

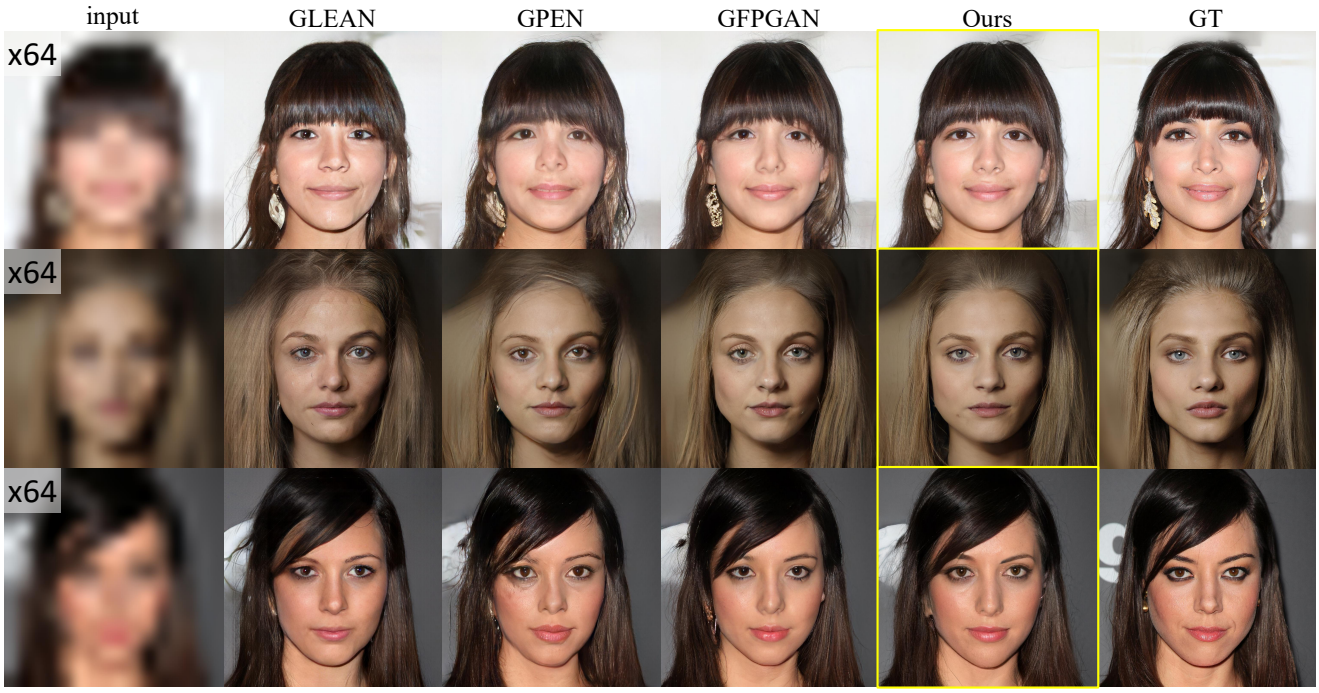

Figure 3. Qualitative comparisons of GLEAN [1], GPEN [7], GFPGAN [5], and GCFSR (ours) on CelebA-HQ [3] for 64× SR. The GT image (Right) has a resolution of 1024^2. **Zoom in for best view.**

5. More qualitative results of modulation on generative strength.

Figure 4. The results obtained by modulation on the generative strength. We change the conditional upscaling factor s from $s = 4$ to $s = 64$ continuously, and find satisfactory results (e.g., results denoted by yellow rectangles) between two ends. **Zoom in for best view.**

6. More visualization on feature modulation.

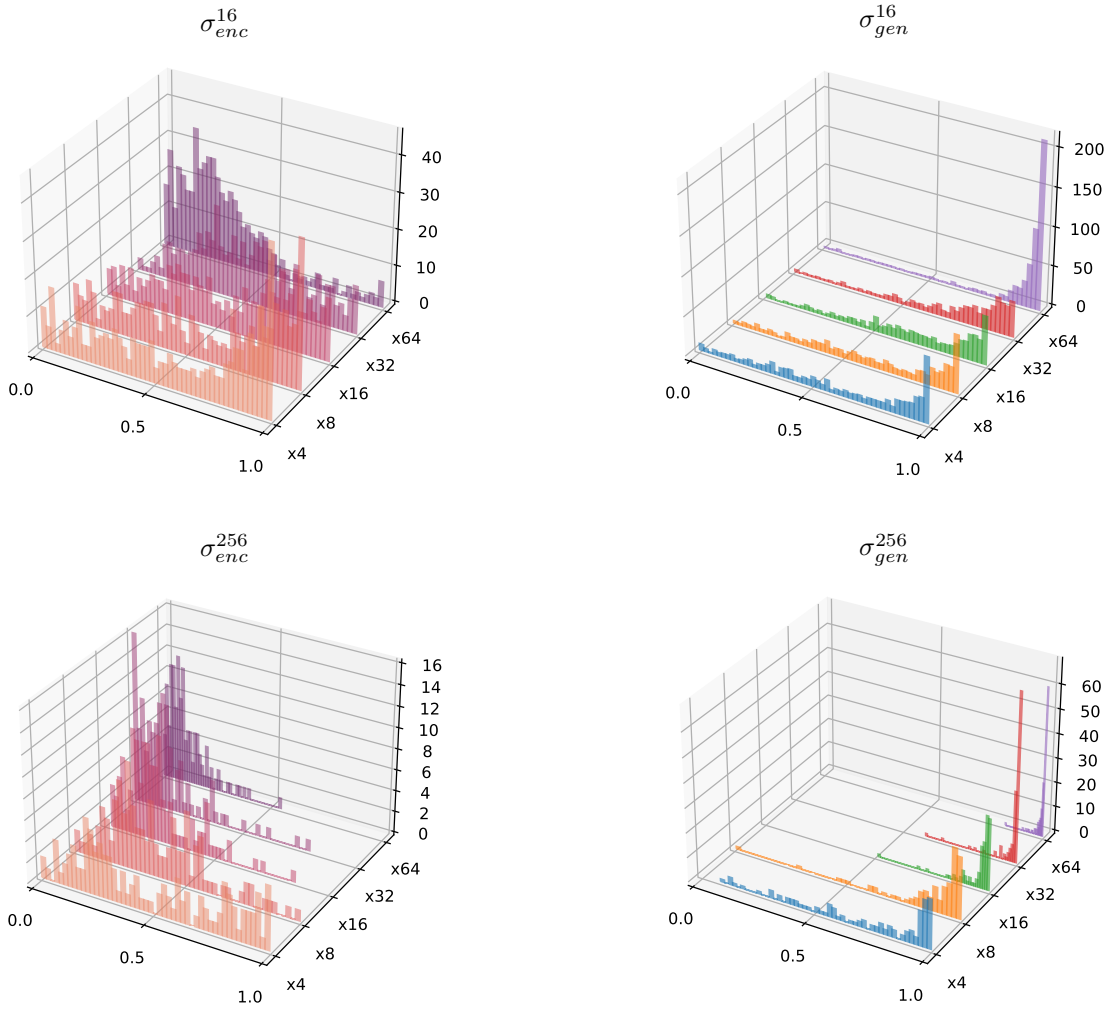

Figure 5. The visualization on feature modulation. The histograms of scaling vectors σ_{enc}^{16}, σ_{gen}^{16}, σ_{enc}^{256}, and σ_{gen}^{256} for different conditional upscaling factors are presented.

Here we provide the histograms of scaling vectors that correspond to level 16 and level 256: σ_{enc}^{16}, σ_{gen}^{16}, σ_{enc}^{256}, and σ_{gen}^{256}, which are illustrated in Figure 5. For σ_{enc}^{16} and σ_{enc}^{256}, their values are approaching 0 as the conditional upscaling factor s increases. Reversely, the values of σ_{gen}^{16} and σ_{gen}^{256} are approaching 1. This indicates that higher conditional upscaling factor corresponds to stronger generative effect, since the features from the encoder are weakened while the features from the decoder are strengthened.

7. Qualitative results of GCFSR$_{adv}$

Figure 6. Visual comparisons of GCFSR$_{adv}$ (trained with only one adversarial loss) and other blind face restoration methods (PSFRGAN [2], DFDNet [4], HiFaceGAN [6], GFPGAN [5], GPEN [7]) on $4\times$ and $8\times$ SR, evaluated on CelebA-HQ. The GT image has a resolution of 512×512. **Zoom in for best view.**

References

[1] Kelvin CK Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. Glean: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14245–14254, 2021. 2, 3

[2] Chaofeng Chen, Xiaoming Li, Lingbo Yang, Xianhui Lin, Lei Zhang, and Kwan-Yee K Wong. Progressive semantic-aware style transformation for blind face restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11896–11905, 2021. 6

[3] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. 2, 3

[4] Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. Blind face restoration via deep multi-scale component dictionaries. In *European Conference on Computer Vision*, pages 399–415. Springer, 2020. 6

[5] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9168–9178, 2021. 2, 3, 6

[6] Lingbo Yang, Shanshe Wang, Siwei Ma, Wen Gao, Chang Liu, Pan Wang, and Peiran Ren. Hifacegan: Face renovation via collaborative suppression and replenishment. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1551–1560, 2020. 6

[7] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Gan prior embedded network for blind face restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 672–681, 2021. 2, 3, 6