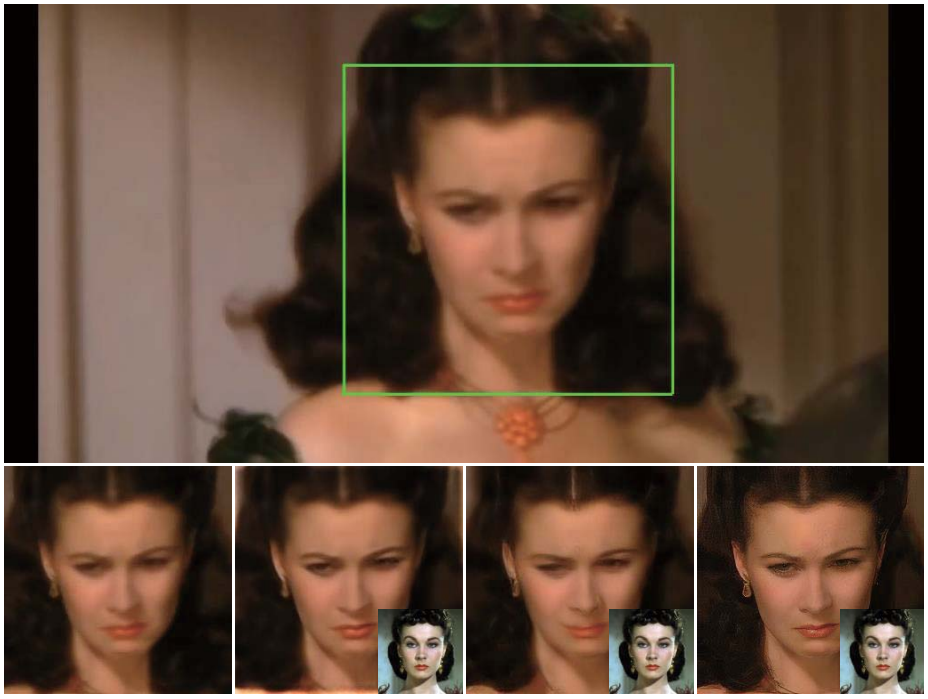
Enhanced Blind Face Restoration with Multi-Exemplar Images and Adaptive Spatial Feature Fusion

(Supplementary Material)

This supplementary file includes an animated figure of old film restoration result (Fig. A), network architecture details of ASFFNet (Section A), comparisons on random guidance (Section B), discussion on WLS for guidance selection (Section C), comparison of WarpNet and MLS (Section D), and more visual comparisons in ablation studies (Section E), $\times 4$ and $\times 8$ (Section F), and real LQ images (Section G).



Bicubic GFRNet [4] GWAINet [1] Ours

Figure A: Restoration results of frames from an old film. This is an animated figure. Please view it by zooming in Adobe Acrobat X Pro Reader or later versions.

A. Network Architecture of ASFFNet

Our ASFFNet consists of three feature extraction sub-nets, MLS, AdaIN, four ASFF blocks and reconstruction sub-net. Details of each module are presented in Tables C and D. We note that Conv. (Conv0.) (d, k, s) denotes a convolutional layer with (without) bias, where d, k and s are output dimension, kernel size and stride, respectively. BN is batch normalization, and LReLU (c) is leaky ReLU with negative slope c. Dilated ResBlock (k, s, r) is a composition of dilated convolutions, and is constructed as [dilated conv. (k, s, r), BN, LReLU (0.2), dilated conv. (k, s, r)], where k, s, r are kernel size, stride and dilation rate.

B. Comparisons on Random Guidance

In this section, we report the performance of exemplar-based methods on the same random guidance. We conduct these experiments on the test data of Ours (#1), in which the guidance was randomly selected from 10 candidates. We can see that GFRNet [4], *GFRNet and GWAINet [1] in Table A are inferior to their performance in Table 3, indicating the benefits of similar poses and expressions for exemplar-based methods. Moreover, our ASFFNet still outperforms than their performance in Table A, which can be attributed to the adaptive and progressive fusion of restored and guidance features for better reconstruction.

C. Optimal WLS for Guidance Selection

We hereby discuss the guidance selection by WLS with (*i.e.*, optimal WLS) and without (*i.e.*, WLS (w/o uw)) updating w in Eqn. (2). In WLS (w/o uw), all the landmark weights are fixed as $w = 1$. In general, the average quantitative metrics by

Type	×4			×8		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
GFRNet [4]	27.32	.907	.142	22.83	.844	.310
*GFRNet	27.61	.920	.123	23.80	.879	.265
GWAINet [1]	-	-	-	23.47	.869	.279
ASFFNet	27.99	0.925	0.107	24.19	0.873	0.252

Table A: Comparisons on random guidance.

WLS (w/o uw) and optimal WLS are comparable. Fig. B shows several examples of selected guidance using optimal WLS and WLS(w/o uw), along with the final restoration results. One can see that the guidance images selected by optimal WLS tends to have more consistent pose and expression with degraded observations (*e.g.*, mouth open or close) in comparison with those selected by WLS (w/o uw). The final restoration results by WLS (w/o uw) are blurry or suffer from artifacts at inconsistent regions (*e.g.*, mouth), while the results by optimal WLS are more visually favorable, indicating the effectiveness of our proposed optimal WLS model for guidance selection.

D. Comparison of WarpNet and MLS

In GFRNet [4], a WarpNet is adopted for spatial alignment of degraded and guidance images, while moving least square (MLS) is simply employed in our ASFFNet. To validate the effectiveness of MLS, we retrain a variant of ASFFNet by substituing MLS with WarpNet. Denote these two ASFFNet models as Ours (w/ WarpNet) and Ours (w/ MLS), respectively. These two models are evaluated on VGGFace2, and the quantitative metrics are reported in Table B. Ours (w/ MLS) performs on par with Ours (w/ WarpNet), indicating that the effectiveness of simple MLS. Meanwhile, MLS is more efficient and has less parameters than WarpNet. Therefore, MLS is a good choice for spatial alignment in ASFFNet.

Type	×4			×8			Param	Time
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	(M)	(ms)
Ours (w/ WarpNet)	28.07	0.930	0.102	24.35	0.881	0.236	32.6	39.1
Ours (w/ MLS)	28.07	0.930	0.103	24.34	0.881	0.238	23.1	31.4

Table B: Comparisons of ASFFNet with different spacial alignment methods.

E. More Visual Results on Ablation Studies

In this section, we demonstrate more visual comparisons in ablation studies, including multiple-examplars, different feature fusion methods as well as MLS and AdaIN. (i) From Fig. C, one can see that a small number of examplars usually cannot provide sufficient good guidance with similar expression and pose, and thus the selected guidance cannot guarantee the satisfying restoration results. (ii) The proposed progressive and adaptive spatial feature fusion is more flexible and can generate richer texture on restoration results, as shown in Fig. D. (iii) In Fig. E, we present the comparison examples by removing MLS (*i.e.*, w/o MLS) and AdaIN (*i.e.*, w/o AdaIN) in ASFFNet. Without MLS, the guidance cannot be well aligned to the pose of degraded image, yielding blurry results along with visual artifacts. The restoration results without AdaIN are with fine details, but the inconsistency of color and illumination is still inherited from the guidance image.

F. More Visual Results on $\times 4$ and $\times 8$

First, we report restoration results of all the competing methods (*i.e.*, RCAN [8], ESRGAN [5], DeblurGANV2 [3], TDAE [7], WaveletSR [2], SCGAN [6], GWAINet [1], GFRNet [4]) on $\times 4$ and $\times 8$ in Figs. F and G . Then we select five methods (*i.e.*, RCAN [8], ESRGAN [5], WaveletSR [2], GFRNet [4], GWAINet [1]) with top quantitative performance to give more comparison examples in Figs H and I. Our ASFFNet obviously outperforms all the other competing methods in generating fine and visually photo-realistic details.

G. More Visual Results on Real LQ Images

In this section, we first evaluate the proposed ASSFNet on real-world LQ images, and compare it with GFRNet. These real-world LQ images are collected from *Google Image*, and the corresponding HQ guidance images are searched by restricting person ID. The resolution of LQ image is lower than 80×80. As shown in Fig. J, our ASFFNet can generate visually realistic results, even though the degradation is unknown. Moreover, it may be difficult to collect multi-exemplar guidance images for some cases, and thus the selected HQ guidance is with different poses and expressions from the LQ image. It is inspiring to see that our ASFFNet performs well in these cases, and is more robust for different poses (right part in Fig. J).

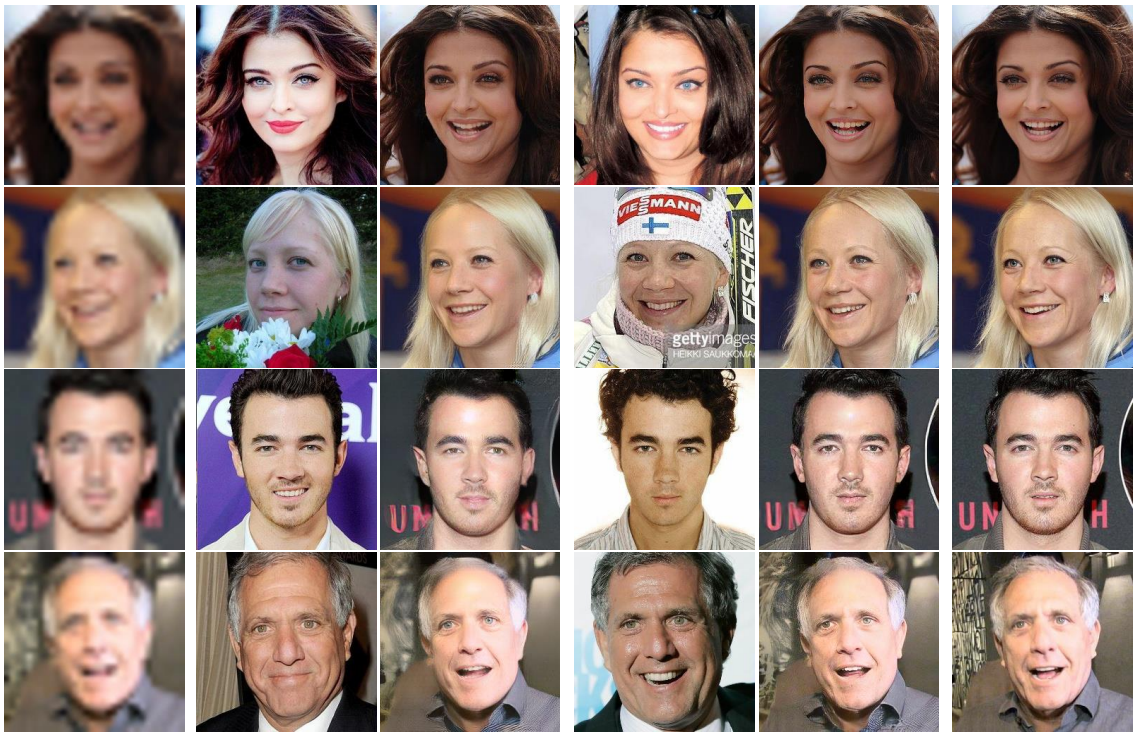
Finally, we apply ASFFNet to handle a real old photo with many famous scientists as shown in Fig. K, which was taken in 1927. Since it is hard to collect HQ guidance images for some scientists, we only process these faces having at least one HQ guidance. One can see that ASFFNet can generalize well to these LQ faces and generate plausible restoration results.

Input	Degraded Image I^d $(3 \times 256 \times 256)$	Guidance I^g_{k*} $(3 \times 256 \times 256)$	L^d Binary Image $(1 \times 256 \times 256)$
Feature Extraction	Conv. (64,3,1), BN, LReLU (0.2) Dilated ResBlock (3,1,7) Dilated ResBlock (3,1,5) Conv. (128,3,2), BN, LReLU (0.2) Dilated ResBlock (3,1,5) Dilated ResBlock (3,1,3) Conv. (128,3,2), BN, LReLU (0.2) Dilated ResBlock (3,1,3) Dilated ResBlock (3,1,1) Conv. (128,3,1), LReLU (0.2)	Conv. (64,3,1), BN, LReLU (0.2) Dilated ResBlock (3,1,7) Dilated ResBlock (3,1,5) Conv. (128,3,2), BN, LReLU (0.2) Dilated ResBlock (3,1,5) Dilated ResBlock (3,1,3) Conv. (128,3,2), BN, LReLU (0.2) Dilated ResBlock (3,1,3) Dilated ResBlock (3,1,1) Conv. (128,3,1), LReLU (0.2)	Conv0. (64,9,2), LReLU (0.2) Conv0. (64,3,1), LReLU (0.2) Conv0. (64,7,1), LReLU (0.2) Conv0. (128,3,1), LReLU (0.2) Conv0. (128,5,2), LReLU (0.2) Conv0. (128,3,1), LReLU (0.2) Conv0. (128,3,1), LReLU (0.2) Conv0. (128,3,1), LReLU (0.2) Conv0. (128,3,1), LReLU (0.2) Conv0. (128,3,1), LReLU (0.2)
		MLS AdaIN	
Output	F^d	$F^{g,w,a}$	F^l
Feature Fusion	ASFF Bock 1 ASFF Bock 2 ASFF Bock 3 ASFF Bock 4		
Output	F^c		
Reconstruction	Conv. (256,3,1) Dilated ResBlock (3,1,1) Dilated ResBlock (3,1,1) PixelShuffle(2) Conv. (128,3,1) Dilated ResBlock (3,1,1) Dilated ResBlock (3,1,1) PixelShuffle(2) Conv. (32,3,1) Dilated ResBlock (3,1,1) Dilated ResBlock (3,1,1) Tanh()		
Output	Restoration Results $\hat{I}^h$ $(3 \times 256 \times 256)$		

Table C: Architecture of ASFFNet.

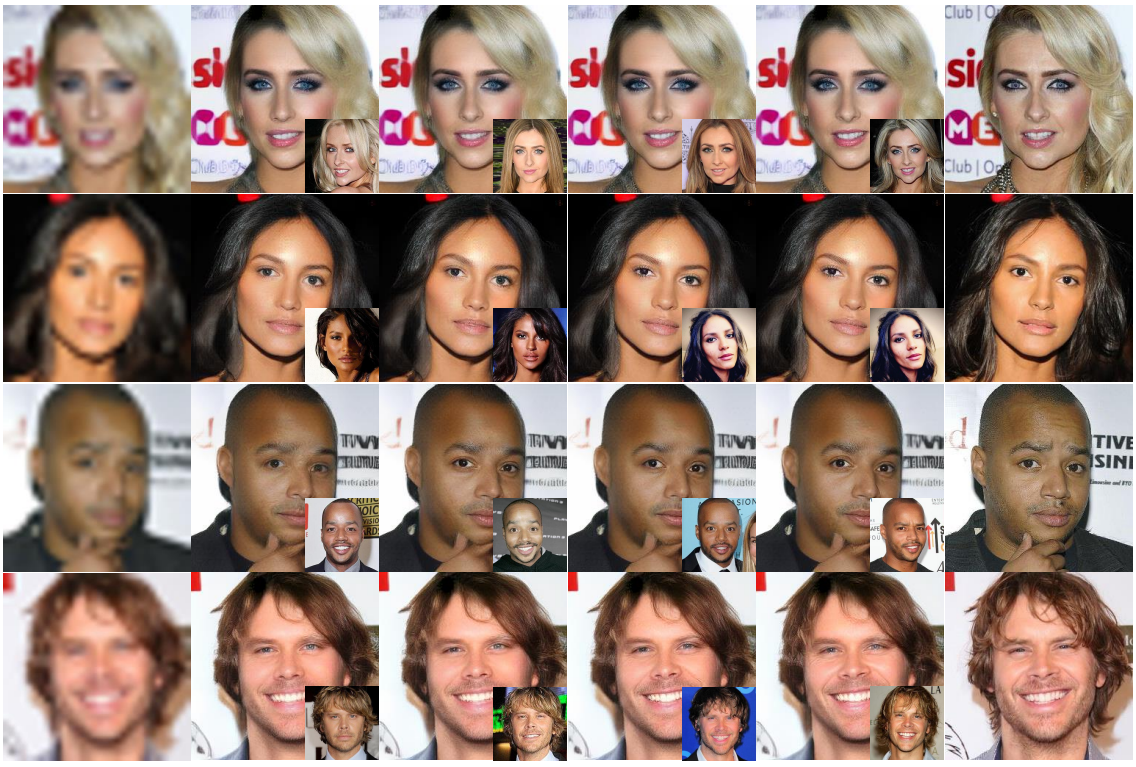
Input	F^d	$F^{g,w,a}$	F^l	F^d	$F^{g,w,a}$
ASFF Block	Conv. (64,1,1)	Conv. (64,1,1)	Conv. (64,1,1)	Conv. (128,3,1) BN, LReLU (0.2) Conv. (128,1,1)	Conv. (128,3,1) BN, LReLU (0.2) Conv. (128,1,1)
	Concat				
	Conv0. (128,3,1), BN, LReLU (0.2) Conv0. (128,3,1), BN, LReLU (0.2)			$\mathcal{F}_d(F^d)$	$\mathcal{F}_g(F^{g,w,a})$
	Attentiom Mask F^m			Element-wise Subtract	
	Element-wise Product				
	Element-wise Addition with $\mathcal{F}_d(F^d)$				
Output	F^d				

Table D: Details of ASFF block.



Input | Guidance of WLS (w/o uw) | Restoration Result | Guidance of Our WLS | Restoration Result | Ground-truth

Figure B: Comparison of guidance selected by optimal WLS and WLS (w/o uw), and the final restoration results.



Input | Ours (#1) | Ours (#3) | Ours (#5) | Ours (#10) | Ground-truth

Figure C: Visual comparison of our ASFFNet with different numbers of exemplars. Close-up in the bottom right is the selected guidance.

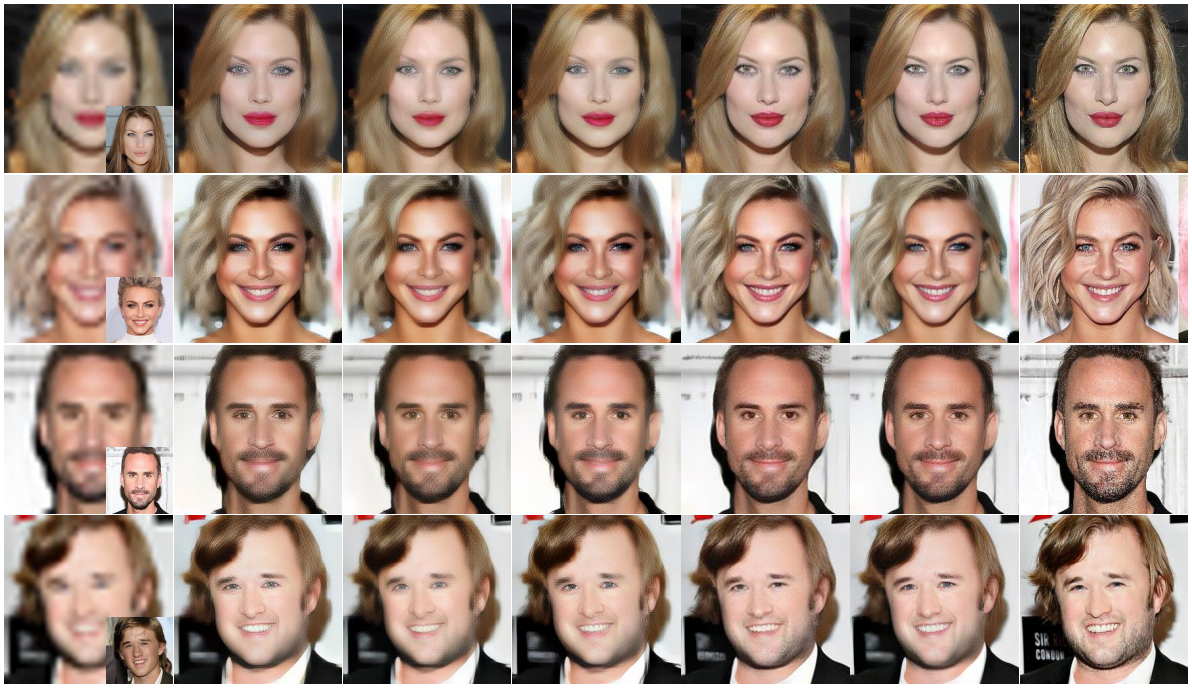


Figure D: More visual comparison of different feature fusion methods.

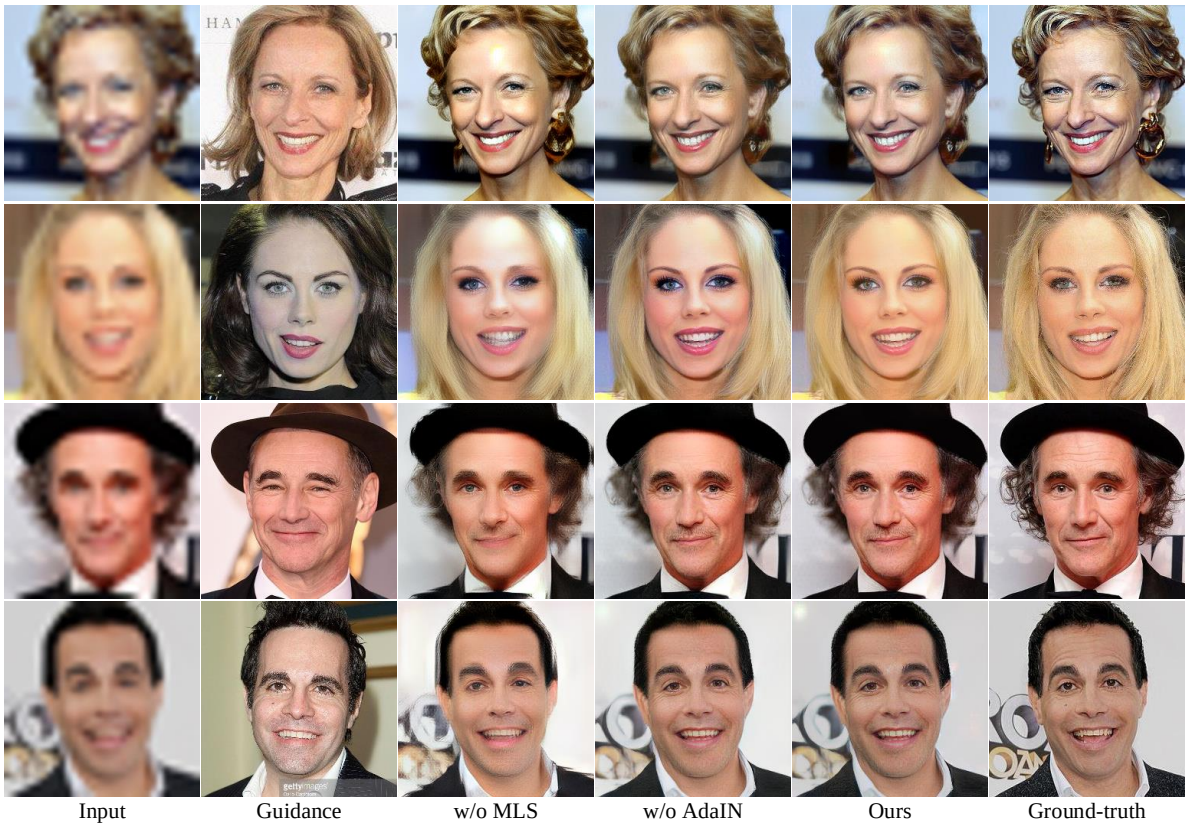


Figure E: Visual comparison of our ASFFNet by removing MLS and AdaIN.

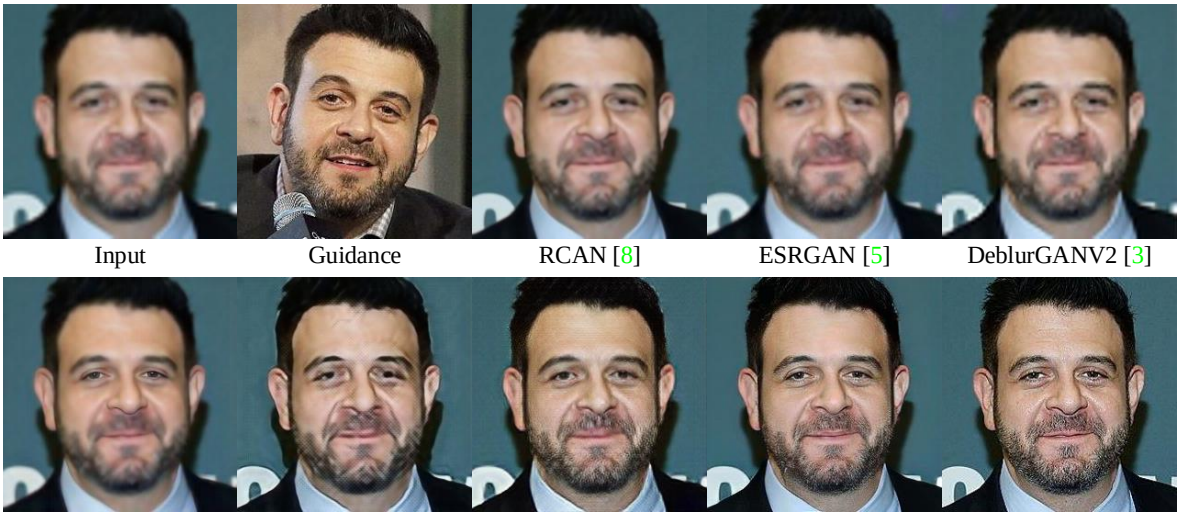


Figure F: The $4\times$ SR results compared with all the competing methods.

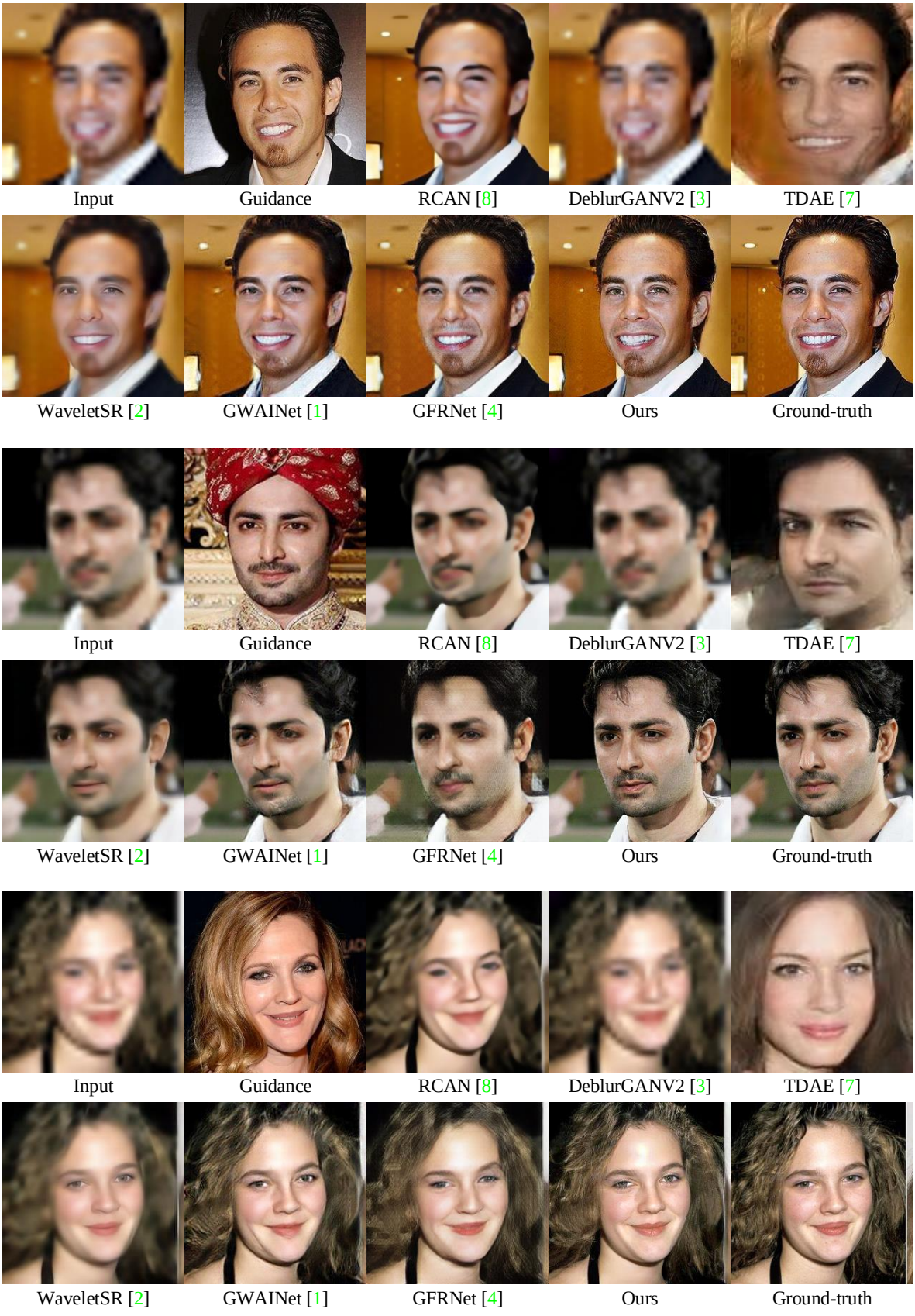
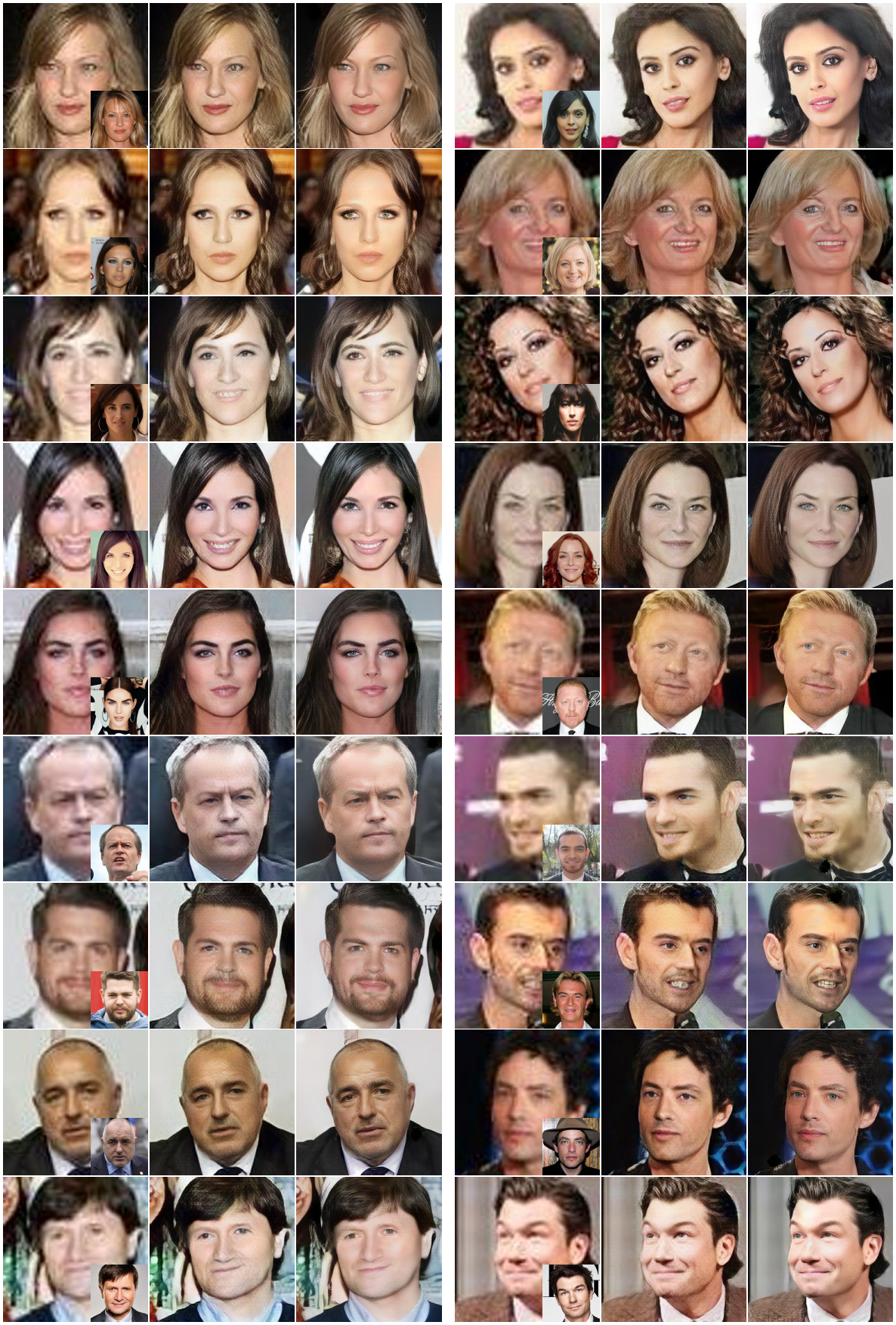


Figure G: The $8\times$ SR results compared with all the competing methods.

Figure H: More examples on $4\times$ SR compared with selected top-5 competing methods.

Input & Guidance | RCAN [8] | WaveletSR [2] | GWAINet [1] | GFRNet [4] | Ours | Ground-truth

Figure I: More examples on 8× SR compared with selected top-5 competing methods.



Input & Guidance GFRNet [4] Ours Input & Guidance GFRNet [4] Ours

Figure J: Visual comparison on real-world LQ images. Best view it by zooming in the screen.

(a) A real old photo taken in 1927.

(b) Restoration result of ASFFNet. Best view it by zooming in screen.

(c) Restoration results of each image. Close-up in the right bottom is the guidance.

Figure K: Restoration results of an old photo.

References

[1] Berk Dogan, Shuhang Gu, and Radu Timofte. Exemplar guided face image super-resolution without facial landmarks. In *CVPRW*, 2019. 1, 2

[2] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In *ICCV*, pages 1689–1697, 2017. 2

[3] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, 2019. 2

[4] Xiaoming Li, Ming Liu, Yuting Ye, Wangmeng Zuo, Liang Lin, and Ruigang Yang. Learning warped guidance for blind face restoration. In *ECCV*, 2018. 1, 2

[5] 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 *ECCVW*, 2018. 2

[6] Xiangyu Xu, Deqing Sun, Jinshan Pan, Yujin Zhang, Hanspeter Pfister, and Ming-Hsuan Yang. Learning to super-resolve blurry face and text images. In *ICCV*, 2017. 2

[7] Xin Yu and Fatih Porikli. Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In *CVPR*, 2017. 2

[8] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 2