MetaF2N: Blind Image Super-Resolution by Learning Efficient Model Adaptation from Faces
(Supplementary Material)

Zhicun Yin[1] Ming Liu[1](✉) Xiaoming Li[1] Hui Yang[2]
Longan Xiao[2] Wangmeng Zuo[1,3]
[1] Harbin Institute of Technology [2] Shanghai Transsion Co, Ltd [3] Peng Cheng Laboratory
cszcyin@outlook.com, csmliu@outlook.com, csxmli@gmail.com, wmzuo@hit.edu.cn

The content of this supplementary material is organized as follows:

- Settings for synthesizing test images in Sec. s1.
- Inference efficiency comparison in Sec. s2.
- More qualitative comparison in Sec. s3.

s1. Details of Test Degradation Settings

During test, *Synthetic datasets* are built upon the in-the-wild version of FFHQ [3] and CelebA [7], where 1,000 high-quality images are extracted from each dataset, and faces occupy around 10% areas of the whole image on average. With the high-quality images, we first construct two test datasets whose degradations are independent and identically distributed as the training set, which are denoted by FFHQ$_{iid}$ and CelebA$_{iid}$, respectively. For evaluating the generalization ability on out-of-distribution degradations, we further construct FFHQ$_{ood}$ and CelebA$_{ood}$ by changing the parameters of the degradation model, *e.g.*, Gaussian blur $\rightarrow$ motion blur, Gaussian/Poisson noise $\rightarrow$ Speckle noise. Specifically, the motion blur kernels are from [1,4,6]. More details are shown in Tab. s1. Note that the IID degradation operates two rounds as shown in Tab. s1, and in each round, the operations follow the column order of Tab. s1, *i.e.*, blurring, downsampling, adding noise, and JPEG compression. On the contrary, the OOD degradation has only a single pass, and we use degradations different from the IID ones.

s2. Inference Efficiency

In order to evaluate the efficiency of our methods compared to ReDegNet [5], we evaluate the inference time (including the fine-tuning steps) on the RF200 dataset. The results are shown in Tab. s2, one can see that ReDegNet [5] fine-tunes the degradation extraction model for 100 steps and the super-resolution model for 1,000 steps, while our method requires only 1 fine-tuning step, which is much more efficient than ReDegNet [5].

s3. Qualitative Comparison

To show the effectiveness of the proposed MetaF2N, we compare with several state-of-the-art blind image SR methods, including ESRGAN [10], RealSR [2], Real-ESRGAN [9], BSRGAN [12], MM-RealSR [8], and ReDegNet [5]. Among them, ESRGAN [10] and RealSR [2] are trained with simple degradation models, which lead to unsatisfactory performance under more complex and realistic degradations. Therefore, they are not present in the qualitative comparison for a better view of comparison with other methods.

s3.1. More Qualitative Comparison

Limited by the space of the main manuscript, we put more qualitative comparison results here with real word data in Fig. s1 and synthesized data in Fig. s2. As the figures show, our results are clearer and contain more photo-realistic textures, which can be ascribed to the effectiveness of our image-specific fine-tuning scheme.

s3.2. Qualitative Comparison of Ablation Studies

MaskNet Configuration. In the supplementary material, the qualitative comparison results of our ablation studies are presented in Fig. s3. With the visual comparison of the ablation study on the training scheme regarding data configuration and the MaskNet, it indicates that the deployment of MaskNet brings considerable improvements in detail expression, and utilizing the low-quality face image as a reference is also useful for our task.

Fine-Tune Steps of Inner Loop. Besides, the visual comparison between different fine-tuning steps of our method is

also given in Fig. s4. With the fine-tuning steps increasing, the result is slightly improved in detail and tends to be stable. However, because of the color prediction error caused by GPEN [11], more fine-tuning steps may lead to color deviation phenomena.

Replacement of Generated Faces. Moreover, we also try to replace the face region of our restoration result with the faces recovered by GPEN [11] because 1) the faces have already been generated by GPEN for fine-tuning, and 2) the faces recovered by GPEN usually have higher quality. The quantitative comparison is shown in Tab. s3, while the qualitative comparison is shown in Fig. s5.

References

[1] Giacomo Boracchi and Alessandro Foi. Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing*, pages 3502–3517, 2012. 1

[2] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *IEEE International Conference on Computer Vision*, pages 3086–3095, 2019. 1

[3] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 1

[4] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971, 2009. 1

[5] Xiaoming Li, Chaofeng Chen, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. From face to natural image: Learning real degradation for blind image super-resolution. In *European Conference on Computer Vision*, pages 376–392, 2022. 1, 3, 4, 5

[6] 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, 2020. 1

[7] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision*, pages 3730–3738, 2015. 1

[8] Chong Mou, Yanze Wu, Xintao Wang, Chao Dong, Jian Zhang, and Ying Shan. Metric learning based interactive modulation for real-world super-resolution. In *European Conference on Computer Vision*, pages 723–740, 2022. 1, 4, 5

[9] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *IEEE International Conference on Computer Vision*, pages 1905–1914, 2021. 1, 4, 5

[10] 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 *European Conference on Computer Vision Workshops*, pages 63–79, 2018. 1

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

[12] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *IEEE International Conference on Computer Vision*, pages 4791–4800, 2021. 1, 4, 5

Table s1: More details about our degradation setting during test.

Phase	Blur				Downsample		Noise		JPEG
	Type	PDF	σ	Kernel Size	Type	Scale	Type	σ / Scale	Quality Factor
IID (first round)	Gaussian Blur	$\frac{1}{N}\exp(-\frac{1}{2}\mathbf{C}^\mathbf{T}\mathbf{\Sigma}\mathbf{C})$ $\frac{1}{N}\exp(-\frac{1}{2}\mathbf{C}^\mathbf{T}\mathbf{\Sigma}\mathbf{C})^\beta$ $\frac{1}{N}\frac{1}{1+(\mathbf{C}^\mathbf{T}\mathbf{\Sigma}\mathbf{C})^\beta}$	[0.2, 3]	[7, 21]	bicubic bilinear area	[0.67, 6.67]	Gaussian Poission	[1, 30] [0.05, 3]	[35, 95]
IID (second round)	Gaussian Blur	$\frac{1}{N}\exp(-\frac{1}{2}\mathbf{C}^\mathbf{T}\mathbf{\Sigma}\mathbf{C})$ $\frac{1}{N}\exp(-\frac{1}{2}\mathbf{C}^\mathbf{T}\mathbf{\Sigma}\mathbf{C})^\beta$ $\frac{1}{N}\frac{1}{1+(\mathbf{C}^\mathbf{T}\mathbf{\Sigma}\mathbf{C})^\beta}$	[0.2, 1.5]	[7, 21]	bicubic bilinear area	[0.83, 8.83]	Gaussian Poission	[1, 25] [0.05, 2.5]	[35, 95]
OOD	Motion Blur	-	-	[7, 35]	nearest	[0.5, 10]	Speckle	-	[30, 100]

Table s2: The efficiency comparison during inference between our MetaF2N and ReDegNet [5].

Methods	Fine-tuning Steps	Time
ReDegNet	100+1,000	~15min
Ours	1	1.80s
Ours	10	3.67s
Ours	20	5.86s

Table s3: Quantitative comparison on synthetic datasets with independent and identically distributed degradations as the training set, out-of-distribution degradations and the collected real-world dataset RF200. Note that the results are both produced by MetaF2N with one-step fine-tuning. The best results are highlighted by **bold**.

Methods		w/o GPEN Faces	w/ GPEN Faces
FFHQ_{iid}	PSNR↑	**26.13**	25.74
	LPIPS↓	0.281	**0.275**
	FID↓	45.22	**40.78**
	NIQE↓	3.81	**3.57**
CelebA_{iid}	PSNR↑	**25.63**	25.33
	LPIPS↓	0.289	**0.284**
	FID↓	45.72	**40.64**
	NIQE↓	3.97	**3.79**
FFHQ_{ood}	PSNR↑	**25.49**	25.17
	LPIPS↓	0.284	**0.277**
	FID↓	45.07	**41.22**
	NIQE↓	4.22	**3.83**
CelebA_{ood}	PSNR↑	**24.87**	24.63
	LPIPS↓	0.297	**0.292**
	FID↓	47.03	**41.76**
	NIQE↓	4.32	**4.05**
RF200	KID↓	21.2	**20.08**
	NIQE↓	3.03	**3.02**

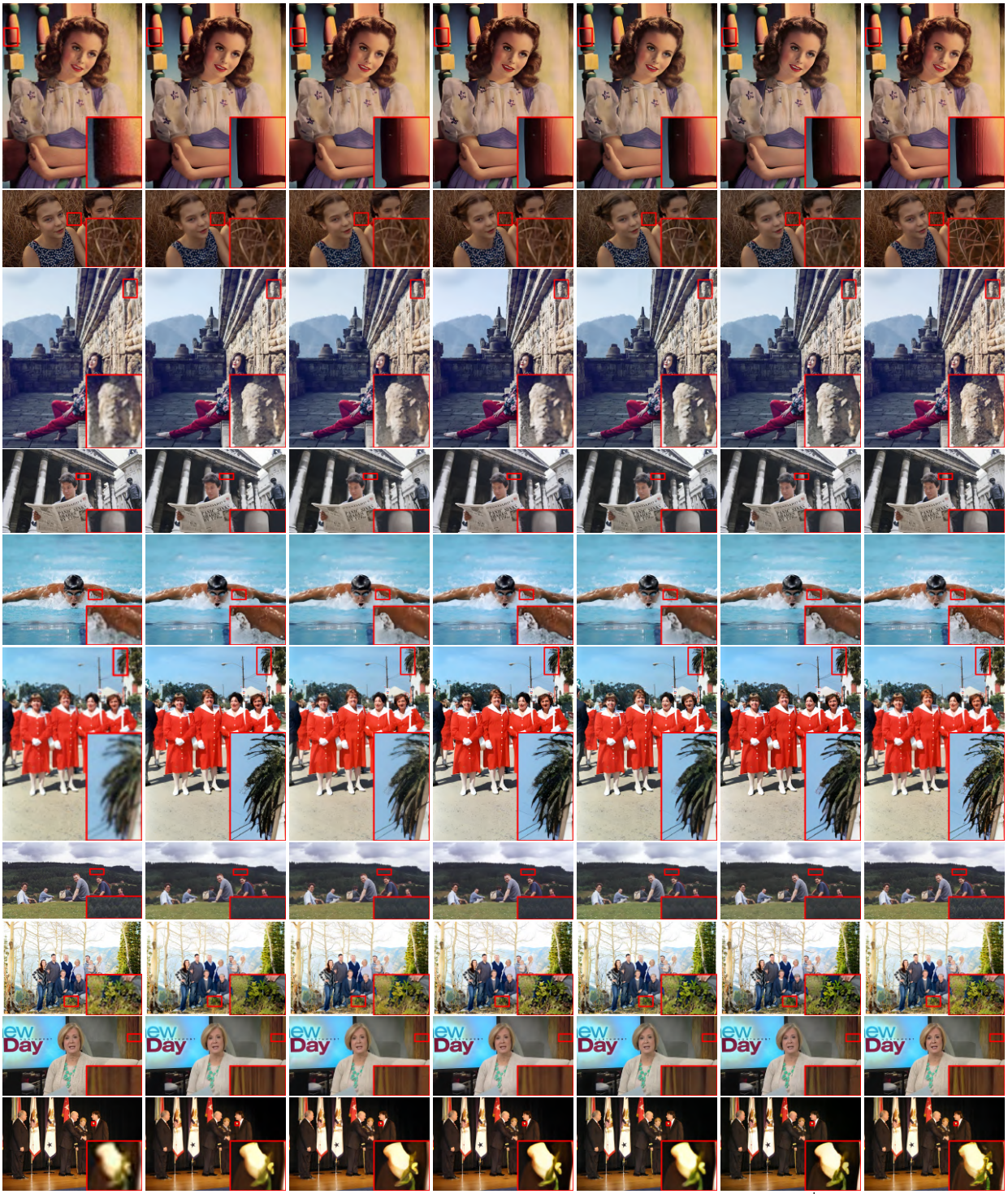


Figure s1: More visual comparison against state-of-the-art blind SR methods on real-world low-quality images in our RF200 dataset. ReDegNet† means that model is fine-tuned for each image following the official configurations of ReDegNet [5].

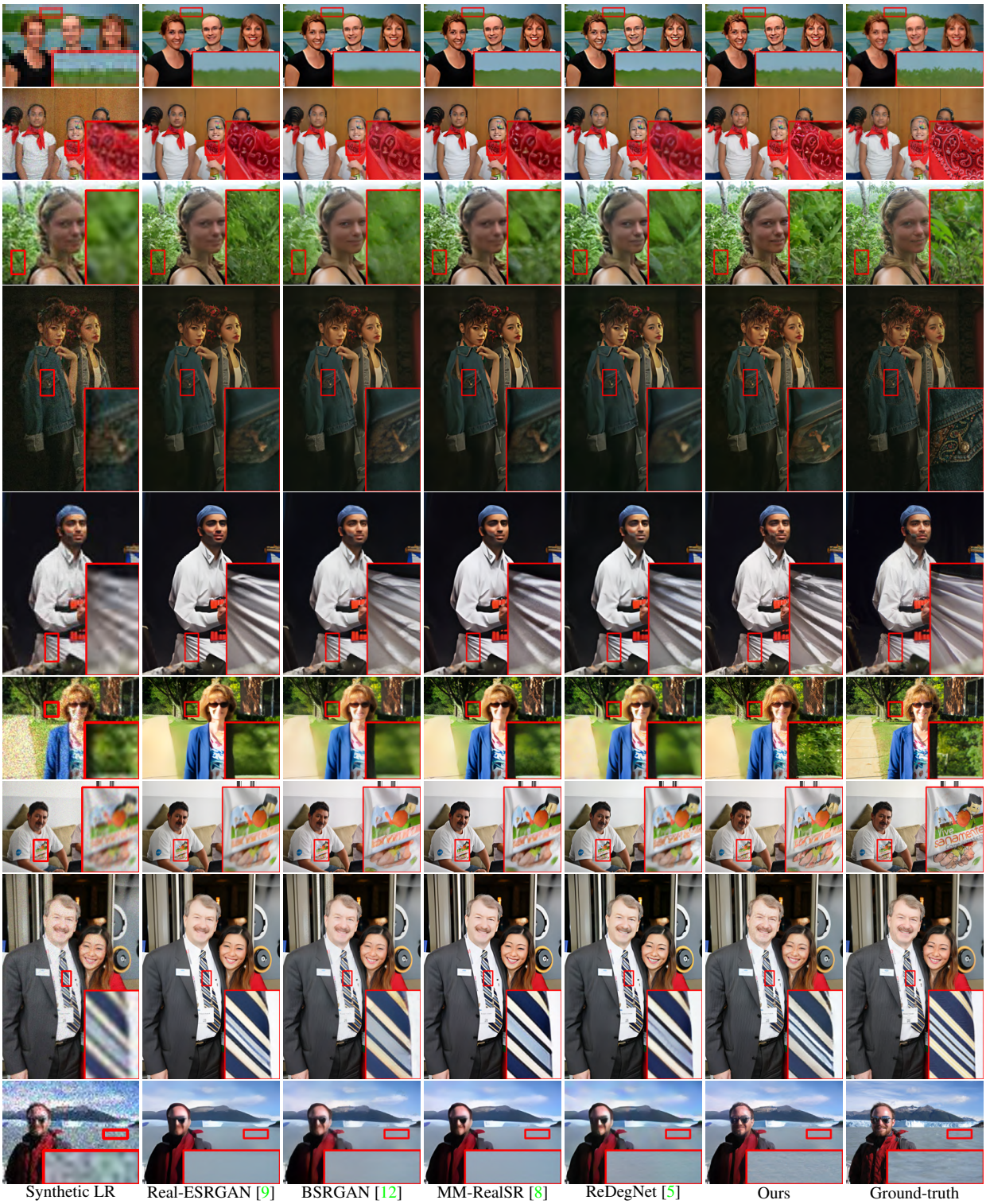
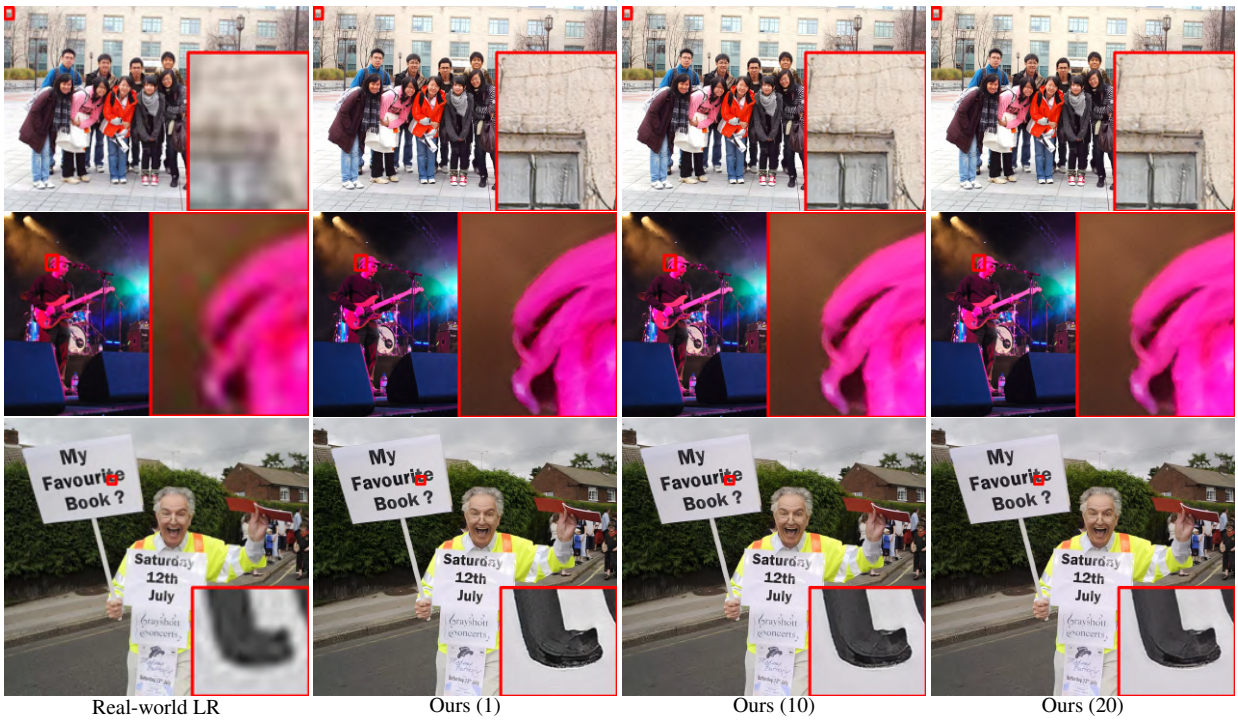


Figure s2: Visual comparison against state-of-the-art blind SR methods on synthetic datasets. The first five low quality images are synthesized with degradation independent and identically distributed as the training set, while the other low quality images are synthesized with out of distribution degradation. Note that the results of Ours are produced by MetaF2N with one-step fine-tuning. Please zoom in for better observation.

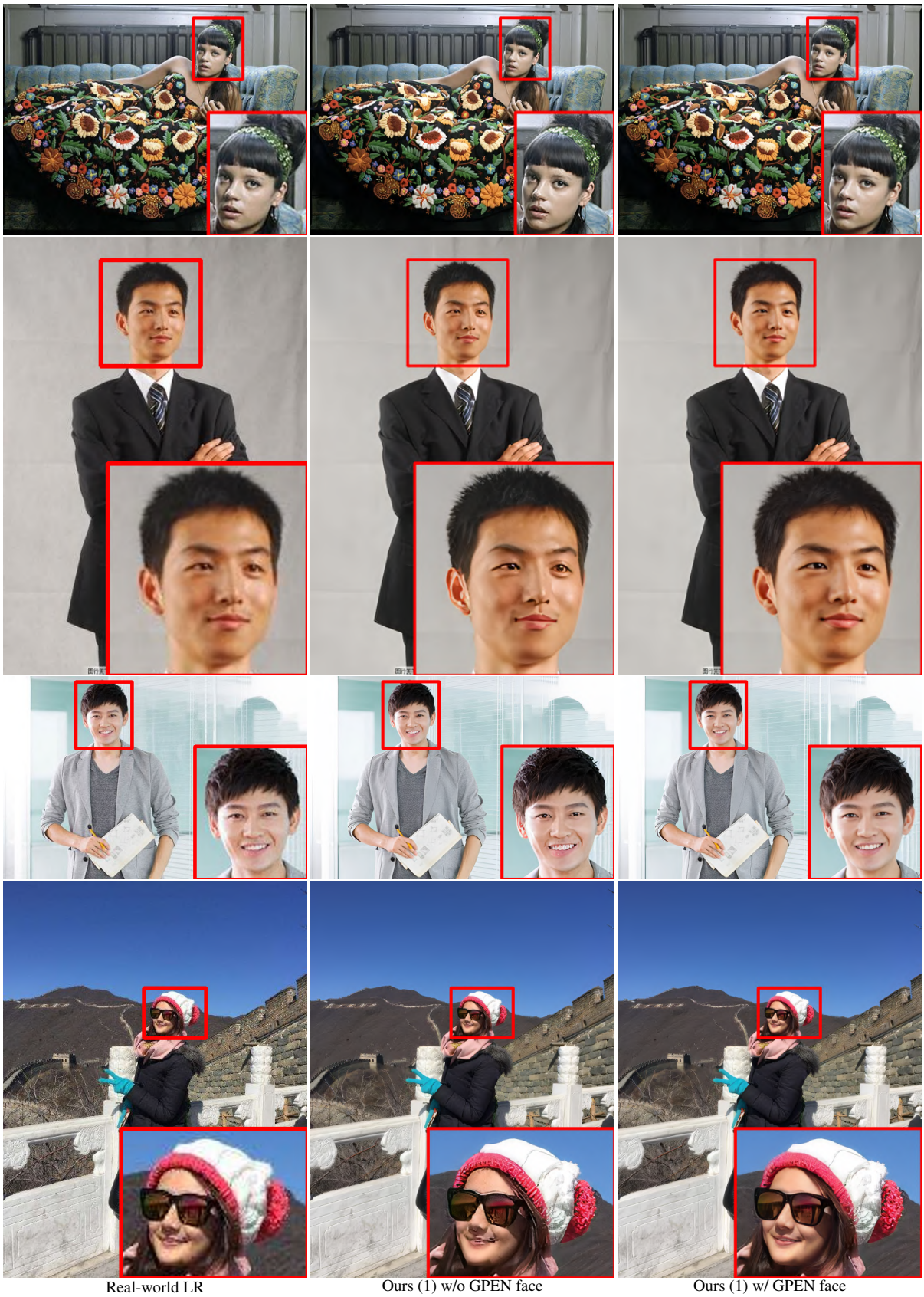
Synthetic LR | Trained with GT | Trained with GPEN [11] | MaskNet$_{NR}$ | Ours | Ground-truth

Figure s3: Visual comparison of ablation study on the training scheme regarding data configuration and the MaskNet with synthesized data.



Real-world LR | Ours (1) | Ours (10) | Ours (20)

Figure s4: Visual comparison of the results with different fine-tuning steps of our model.



Real-world LR | Ours (1) w/o GPEN face | Ours (1) w/ GPEN face

Figure s5: Visual comparison with the replacement of GPEN [11] faces.