Focal Network for Image Restoration
Supplementary Material

Yuning Cui[1]   Wenqi Ren[2*]   Xiaochun Cao[2]   Alois Knoll[1]

[1]Technical University of Munich   [2]Shenzhen Campus of Sun Yat-sen University

{yuning.cui,knoll}@in.tum.de   {renwq3,caoxiaochun}@mail.sysu.edu.cn

1. Dataset and Experimental Setup

In this section, we describe more details of datasets and experimental setups for three image restoration tasks. Unless specified here, the hyper-parameters mentioned in the main text are adopted.

Image Dehazing. We evaluate the proposed network on both daytime and nighttime datasets. The daytime datasets include both synthetic (RESIDE [10]) and real-world (NH-HAZE [3], Dense-Haze [2], O-HAZE [4]) datasets.

The RESIDE [10] dataset contains two training subsets, *i.e.,* indoor training set (ITS) and outdoor training set (OTS). ITS consists of 13990 hazy images generated from 1399 sharp images. OTS comprises 313950 hazy images produced from 8970 clean images. RESIDE [10] contains a testing subset, *i.e.,* synthetic objective testing set (SOTS), which is composed of indoor and outdoor scenes, each including 500 hazy images. We evaluate the ITS-trained and OTS-trained models on SOTS-Indoor and SOTS-Outdoor datasets, respectively. The model is trained for 30 epochs on RESIDE-Outdoor with initial learning rate as $1e^{-4}$ and batch size as 8. On RESIDE-Indoor, the model is trained for 1000 epochs.

NH-HAZE [3] and Dense-Haze [2] both consist of 55 paired images, while O-HAZE [4] contains only 45 image pairs. For these real-world datasets, models are trained for 5000 epochs following [8] on 800×1200 patches with initial learning rate as $2e^{-4}$ and batch size as 2.

The nighttime dataset NHR [27] contains 16146 and 1794 image pairs for training and evaluation, respectively. The model is trained for 300 epochs with initial learning rate as $1e^{-4}$ and batch size as 8.

Image Desnowing. Snow100K [14], SRRS [5], and CSD [6] contain 50000, 15005, and 8000 image pairs for training, and 50000, 15005, and 2000 images for evaluation, respectively. Models are trained for 2000 epochs.

Single-Image Defocus Deblurring. As in previous methods [9, 21, 25], we choose the DPDD [1] dataset to

---

*Corresponding Author

Method	PSNR	SSIM	Time/s	Params/M	Memory/G
MPRNet [26]	32.66	0.959	1.148	20.1	10.415
MIMOUNet++ [7]	32.68	0.959	1.277	16.1	10.395
Restormer [25]	32.92	0.961	1.218	26.13	12.333
Stripformer [23]	33.08	**0.962**	1.054	20.0	12.149
Ours	**33.10**	**0.962**	**0.270**	**15.85**	**2.417**

Table 1. Comparisons on the GoPro [18] test set for image motion deblurring. The inference time is measured by using *torch.cuda.synchronize*() on an NVIDIA Tesla V100 GPU.

Method	NLRN [12]	DeamNet [20]	DAGL [17]	SwinIR [11]	Restormer [25]	Ours
σ=15	31.88	31.91	31.93	**31.97**	31.96	**31.97**
σ=25	29.41	29.44	29.46	29.50	29.52	**29.53**
σ=50	26.47	26.54	26.51	26.58	26.62	**26.65**
GFLOPs	-	146	256	759	141	143

Table 2. Gaussian grayscale image denoising results on the BSD68 [16] dataset.

demonstrate the efficacy of our method. We adopt the training strategy in [21] to train our model.

2. More Experimental Results

2.1. Image Motion Deblurring

For this task, we train and evaluate our model on the widely used GoPro [18] dataset with the initial learning rate as $1e^{-4}$ and batch size as 4. The model is trained for 3000 epochs. We deploy 20 residual blocks in each ResBlock. The results are presented in Table 1. Compared with other competitors, our method achieves a better trade-off between computation overhead and performance.

2.2. Image Denoising

We present the grayscale image denoising results in Table 2. The model is trained for 100 epochs on the same dataset as Restormer [10] and tested on BSD68 [16]. The initial learning rate is set as $1e^{-4}$ and batch size as 16.

Method	SSM	SSM*2	FSM	FSM*2	SSM+FSM
PSNR	33.71	34.97	32.60	32.73	**35.60**

Table 3. Design choices of DSM. SSM*2 means we use two SSM successively.

Method	PSNR	Params/M
Share	35.26	1.46
Ours	35.60	1.47

Table 4. Design choices of MResBlock. *Share* means the convolution parameters are shared among two branches.

We employ 24 residual block in each ResBlock to have the comparable complexity with other algorithms [25, 20]. As shown in the table, our CNN-based method outperforms the strong Transformer model Restormer [25] for all noise levels.

3. More Ablation Studies

The experimental settings remain identical with that of the ablation study section in the main text.

Design Choices of DSM. We conduct experiments by using different combinations of SSM and FSM. As shown in Table 3, deploying two identical modules successively, the models obtain better performance than that of using single one. Our choice, employing dual-domain selection mechanism, produces the best performance.

In addition, we provide visualization comparisons between different designs to demonstrate the superiority of our method. As illustrate in Figure 1, using two SSM, the model focuses on degradation regions. The version with two FSM pays attention to edge signals. Our method only emphasizes signals that are difficult to recover, which facilitates high efficiency of our model.

Design Choices of MResBlock. In MResBlock, we split features and treat each component individually. During this process, convolution layers can be shared among two branches or not. As shown in Table 4, sharing parameters degrades performance by 0.34 dB. Therefore, we use separate convolutions in two branches for better performance.

More Visual Examples of DSM. Figure 2 and Figure 3 illustrate the effects of our DSM for defocus deblurring and desnowing, respectively. As shown in Figure 2, our spatial selection module (SSM) suppresses the simple regions, such as the trunk in the top image and the left stone in the bottom image. Our frequency selection module (FSM) highlights the edge information as marked by red arrows. The same conclusion can also be drawn for desnowing in Figure 3. Our modules accentuate the complicated buildings while attenuating the simple regions such as the sky.

4. More Visual Results

In this section, we provide more visual comparisons for three tasks: defocus deblurring (Figure 4), dehazing (Figure 5 and 6), and desnowing (Figure 7). The images produced by our models are visually closer to the ground-truth than those of other approches.

References

[1] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *European Conference on Computer Vision*, pages 111–126, 2020.

[2] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *IEEE International Conference on Image Processing*, pages 1014–1018, 2019.

[3] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020.

[4] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: A dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.

[5] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *European Conference on Computer Vision*, pages 754–770, 2020.

[6] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4196–4205, 2021.

[7] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4641–4650, 2021.

[8] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5812–5820, 2022.

[9] Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. Iterative filter adaptive network for single image defocus deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2034–2042, 2021.

[10] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.
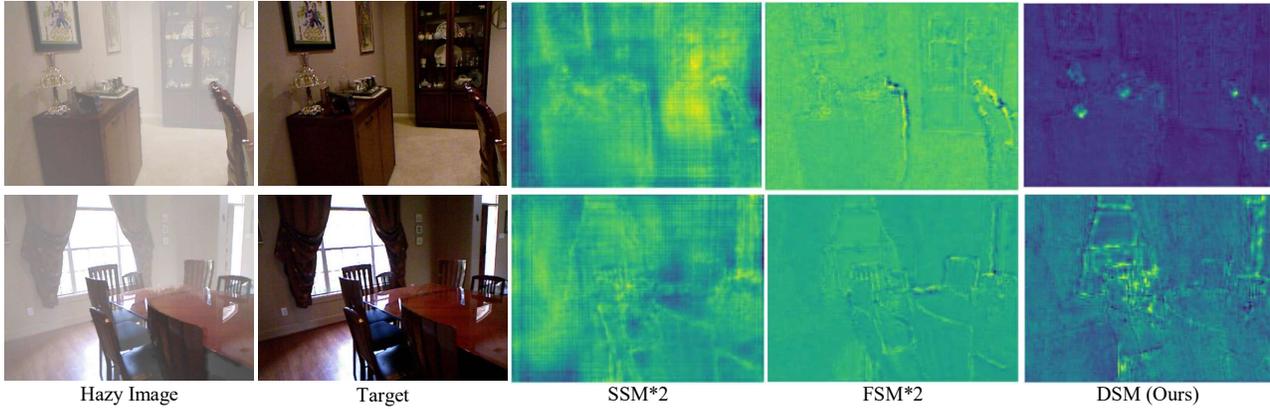
Figure 1. Comparisons between different designs of DSM. Images are obtained from SOTS-Indoor [10] dataset.
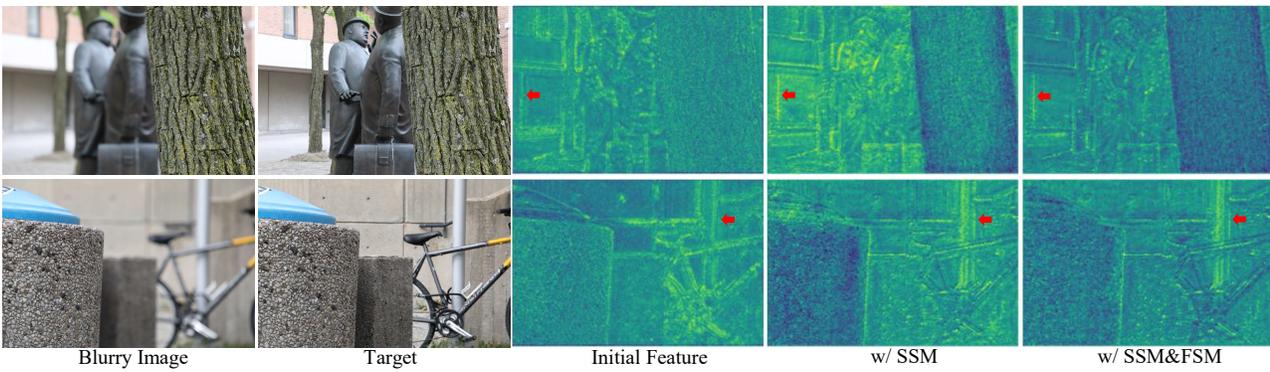
Hazy Image   Target   SSM*2   FSM*2   DSM (Ours)

Figure 2. Effectiveness of our DSM on defocus deblurring. Images are obtained from DPPD [1] dataset.

Blurry Image   Target   Initial Feature   w/ SSM   w/ SSM&FSM
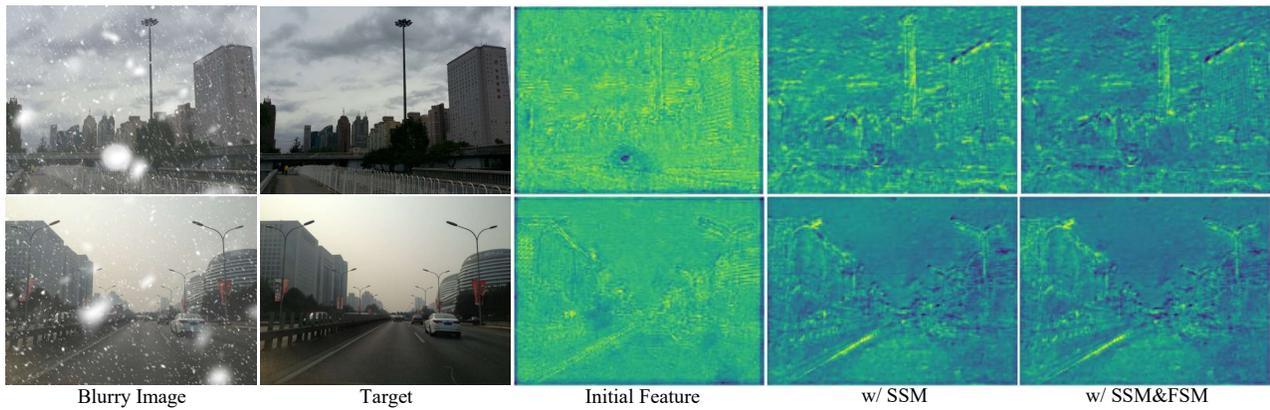
Figure 3. Effectiveness of our DSM on desnowing. Images are obtained from CSD [6] dataset.

Blurry Image   Target   Initial Feature   w/ SSM   w/ SSM&FSM

[11] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1833–1844, 2021.

[12] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. volume 31, 2018.

[13] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7314–7323, 2019.
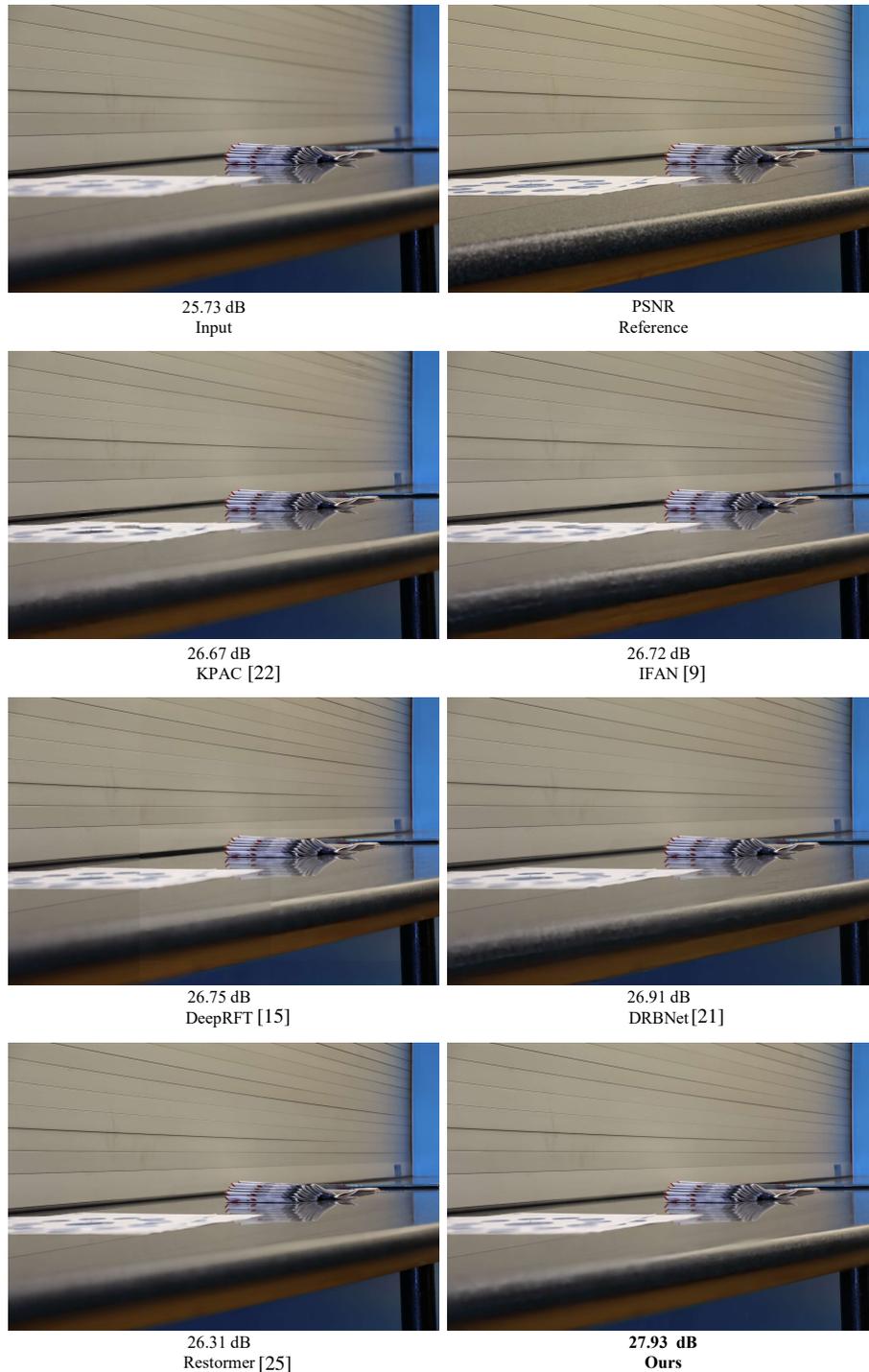
25.73 dB
Input

PSNR
Reference

26.67 dB
KPAC [22]

26.72 dB
IFAN [9]

26.75 dB
DeepRFT [15]

26.91 dB
DRBNet[21]

26.31 dB
Restormer [25]

27.93 dB
Ours

Figure 4. Single-image defocus deblurring comparisons on the DPDD [1] dataset.

[14] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073, 2018.

[15] Xintian Mao, Yiming Liu, Wei Shen, Qingli Li, and Yan Wang. Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745*, 2021.

[16] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and

9.76 dB
Input

PSNR
Reference

16.67 dB
DeHamer [8]

16.73 dB
Ours

6.33 dB
Input

PSNR
Reference

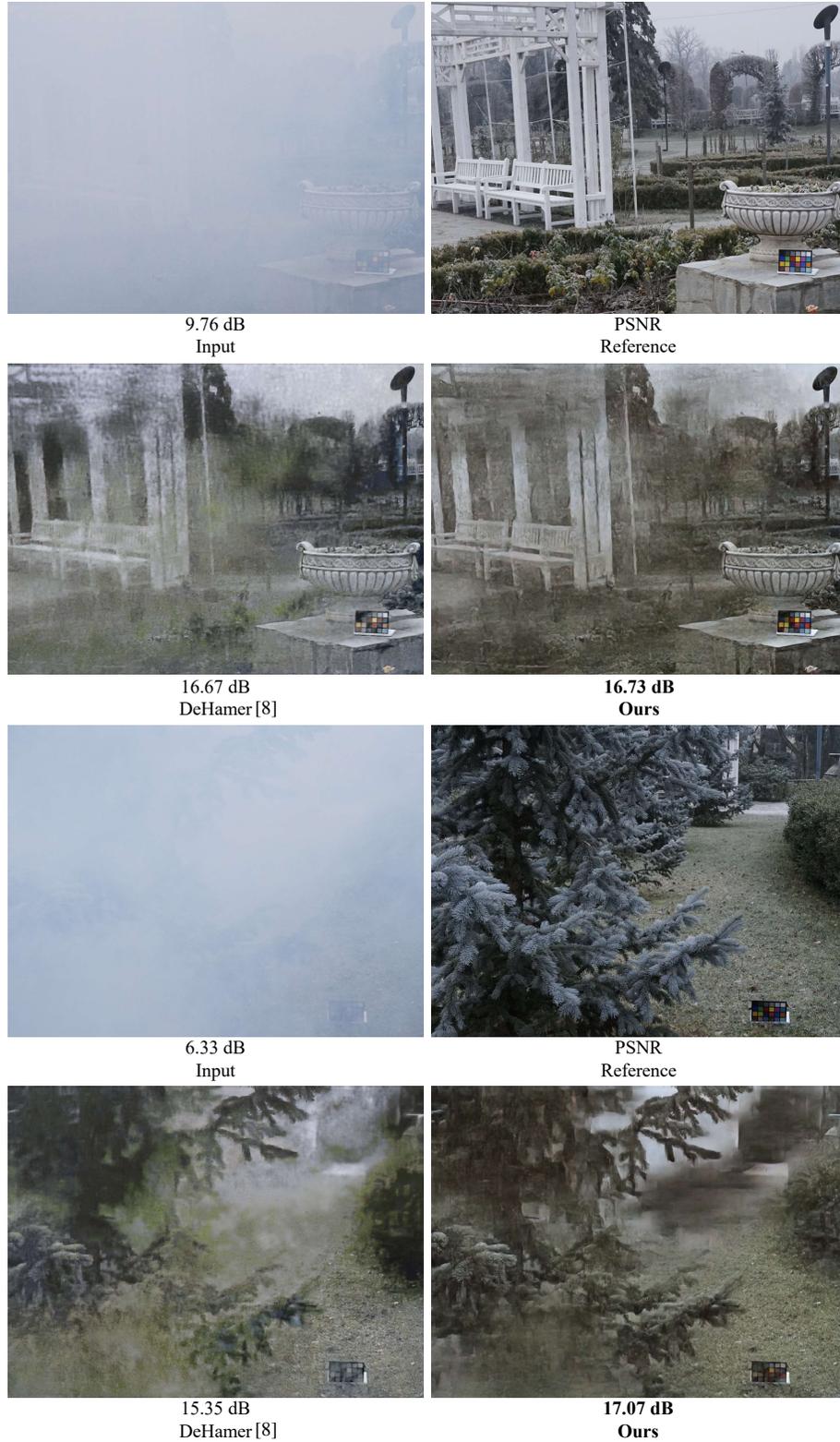15.35 dB
DeHamer [8]

17.07 dB
Ours

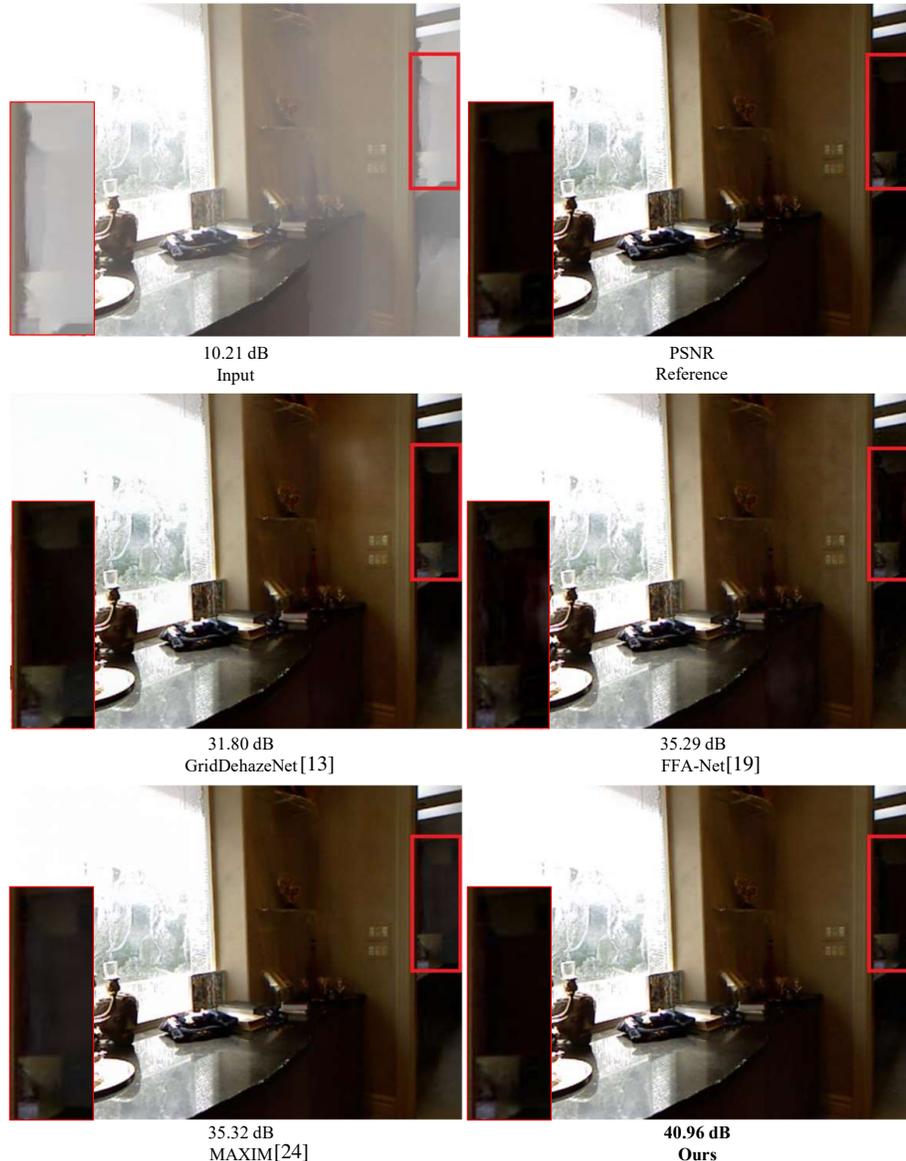Figure 5. Image dehazing comparisons on the Dense-Haze [2] dataset.

Figure 6. Image dehazing comparisons on the SOTS-Indoor [10] testset.

measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 416–423, 2001.

[17] Chong Mou, Jian Zhang, and Zhuoyuan Wu. Dynamic attentive graph learning for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4328–4337, 2021.

[18] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[19] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Confer-ence on Artificial Intelligence*, pages 11908–11915, 2020.

[20] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8596–8606, 2021.

[21] Lingyan Ruan, Bin Chen, Jizhou Li, and Miuling Lam. Learning to deblur using light field generated and real defocus images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 16304–16313, 2022.

[22] Hyeongseok Son, Junyong Lee, Sunghyun Cho, and Seungyong Lee. Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages

| 15.14 dB | PSNR |
| Input | Reference |

| 15.77 dB | 25.24 dB |
| DesnowNet [14] | JSTASR [5] |

| 28.66 dB | **40.28 dB** |
| HDCW-Net [6] | **Ours** |

Figure 7. Image desnowing comparisons on the CSD [6] dataset.

2642–2650, 2021.

[23] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *European Conference on Computer Vision*, pages 146–162, 2022.

[24] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5769–5780, 2022.

[25] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE Conference on Com-*

puter Vision and Pattern Recognition, pages 5728–5739, 2022.

[26] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[27] Jing Zhang, Yang Cao, Zheng-Jun Zha, and Dacheng Tao. Nighttime dehazing with a synthetic benchmark. In *Proceedings of the ACM International Conference on Multimedia*, pages 2355–2363, 2020.