SELECTIVE FREQUENCY NETWORK FOR IMAGE RESTORATION — APPENDIX

Yuning Cui[1], **Yi Tao**[2], **Zhenshan Bing**[1], **Wenqi Ren**[*3,5], **Xinwei Gao**[4], **Xiaochun Cao**[3,5], **Kai Huang**[3], **Alois Knoll**[1]
[1]Technical University of Munich [2]MIT Universal Village Program [3]Sun Yat-sen University
[4]Tencent [5]Chinese Academy of Sciences
{yuning.cui,bing,knoll}@in.tum.de,yitao@universal-village.org,
{renwq3,caoxiaochun,huangk36}@mail.sysu.edu.cn,vitogao@tencent.com

The appendix is organized as follows:

- Appendix A introduces the details of datasets and experimental settings for each task.
- Appendix B provides the theoretical proofs for the low-pass filter used in MDSF.
- Appendix C provides analyses for MCSF.
- Appendix D compares our method with other algorithms from a statistical perspective.
- Appendix E shows additional results for desnowing and defocus deblurring.
- Appendix F exhibits additional visual comparisons for five image restoration tasks.

A DATASETS AND EXPERIMENTAL DETAILS

We summarize the datasets in Tab. 1. We detail individual datasets and training settings as follows:

Image Motion Deblurring. Consistent with recent methods (Wang et al., 2022b; Zamir et al., 2021), we train SFNet using the GoPro dataset (Nah et al., 2017), which consists of 2103 blurry/sharp image pairs for training and 1111 pairs for evaluation. To validate the generalization capacity of our method, we directly apply the GoPro-trained model to HIDE dataset (Shen et al., 2020), which contains 2025 image pairs for evaluation. The images in GoPro and HIDE are both generated synthetically. To test the performance of our method on real-world images, we further evaluate on the newly proposed RSBlur (Rim et al., 2022). It has 8878 and 3360 image pairs for training and evaluation. On GoPro and RSBlur, we train out network for 3000 and 710 epochs, respectively.

Image Defocus Deblurring. We utilize the DPDD dataset (Abuolaim & Brown, 2020) to verify the effectiveness of our method following (Lee et al., 2021; Ruan et al., 2022). This dataset consists of images in 500 indoor/outdoor scenes, each with four images, labeled as right, left, center view and the all-in-focus ground truth. DPDD is split into training, validation and testing sets with 350, 74 and 76 scenes. For this task, there are two training patterns. One is to train using the center view image and corresponding ground truth, called single-image defocus deblurring task. The other one, dual-pixel defocus deblurring task, takes images from both left and right views as training samples. It is challenging to obtain all-in-focus and defocused image pairs with accurate correspondence in two shots (Ruan et al., 2022). To alleviate this issue, LFDOF is proposed by using refocusing techniques and light field synthetic aperture to get a large number of image pairs (Ruan et al., 2021).

For single-image defocus deblurring, we train SFNet following DRBNet (Ruan et al., 2022). Regarding the dual-pixel setting, SFNet is trained on DPDD for 200 epochs. Due to dual-pixel input images, we alter the convolution layers related to the input from 3-channel to 6-channel and use convolution to adjust the channel dimension for skip connections following (Zamir et al., 2022).

Image Deraining. Following (Jiang et al., 2020; Tu et al., 2022), we leverage a composite training dataset containing 13712 image pairs collected from various datasets. SFNet is evaluated on Rain100H (Yang et al., 2017), Rain100L (Yang et al., 2017), Test100 (Zhang et al., 2019a), Test1200 (Zhang & Patel, 2018a) and Test2800 (Fu et al., 2017b). The network is trained for 300 epochs.

*Corresponding author

Table 1: Details of the datasets for image restoration tasks.

Task	Dataset	Train Sample	Test Sample	Testset Rename
Motion Deblurring	GoPro (Nah et al., 2017)	2103	1111	
	HIDE (Shen et al., 2020)	0	2025	
	RSBlur (Rim et al., 2022)	8878	3360	
Defocus Deblurring	DPDD (Abuolaim & Brown, 2020)	350	76	
	LFDOF (Ruan et al., 2021)	11261	725	
Deraining	Rain14000 (Fu et al., 2017b)	11200	2800	Test2800
	Rain1800 (Yang et al., 2017)	1800	0	
	Rain800 (Zhang et al., 2019a)	700	100	Test100
	Rain100H (Yang et al., 2017)	0	100	Rain100H
	Rain100L (Yang et al., 2017)	0	100	Rain100L
	Rain1200 (Zhang & Patel, 2018a)	0	1200	Test1200
	Rain12 (Li et al., 2016)	12	0	
Dehazing	RESIDE/ITS (Li et al., 2018a)	13990	500	SOTS-Indoor
	RESIDE/OTS (Li et al., 2018a)	313950	500	SOTS-Outdoor
	Dense-Haze (Ancuti et al., 2019)	45	5	
Desnowing	CSD (Chen et al., 2021b)	8000	2000	CSD (2000)
	SRRS (Chen et al., 2020)	15005	15005	SRRS (2000)
	Snow100K (Liu et al., 2018)	50000	50000	Snow100K (2000)

Image Dehazing. We train our model on the commonly used dataset RESIDE (Li et al., 2018a) for dehazing. It consists of two training subsets, indoor training set (ITS) and outdoor training set (OTS), and a synthetic objective testing set (SOTS). We train our models on ITS and OTS separately, and test on the corresponding testsets. SFNet is trained for 300 epochs on ITS and 30 for OTS, and we change the batch size from 4 to 8 for OTS. Moreover, we also include the real-world Dense-Haze (Ancuti et al., 2019) as the experimental dataset, which contains 45 training images, 5 validation images and 5 testing images. Images are obtained in the homogeneous and dense hazy scenes. The network is trained for 2000 epochs with the patch size of 512×512.

Image Desnowing. We use CSD (Chen et al., 2021b), SRRS (Chen et al., 2020) and Snow100K (Liu et al., 2018) datasets for desnowing. The dataset settings follow previous works (Chen et al., 2020; 2022c), where we randomly sample 2500 image pairs from the training set for training, and 2000 images from testing set for evaluation. The model is trained for 800 epochs on each dataset.

B Proof for Low-pass Filter

Our proof is established on the knowledge that if we iteratively apply a low-pass filter to an image infinitely, only the extreme low frequency can be remained. Here we consider the low frequency as the lowest frequency. We first provide the formal computational process of *low*- and *high*-frequency signals, and then prove that the resulting filter of Eq. 1 in Sec. 3.2 is a low-pass filter.

Since the computational method of the lowest frequency component in a spectrum is dimension-agnostic, without loss of generality, we consider 1D Fourier transform for simplicity. For an input signal $X \in \mathbb{R}^n$, the matrix for discrete Fourier transform is given by,

$$F_n = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix} \tag{1}$$

where n is the length of the input signal, and $\omega = e^{-j\frac{2\pi}{n}}$. Accordingly, the matrix for inverse discrete Fourier transform can be described as $F_n^{-1} = \frac{1}{n}\overline{F_n}$. Then we can apply F_n and F_n^{-1} successively

to obtain the *low*-frequency operator by,

$$\begin{aligned}\mathcal{L}[X] &= F_n^{-1} diag(1,0,0,...0) F_n X \\ &= \frac{1}{n} MX\end{aligned} \tag{2}$$

in which, M is the matrix with all values as 1, *diag*$(\cdot)$ denotes the diagonal matrix. Then the *high*-frequency operator can be described as,

$$\begin{aligned}\mathcal{H}[X] &= F_n^{-1} diag(0,1,1,...1) F_n X \\ &= F_n^{-1}(E - diag(1,0,0,...0)) F_n X \\ &= (E - \frac{1}{n} M)X\end{aligned} \tag{3}$$

where E is the unit matrix.

Before providing the proof of the filter, we first model the filtering operation. Recall that in MDSF we utilize the convolution operation to apply the filter to the input feature map within a region of size $k \times k$, where the value of the central pixel is the sum of all re-weighted pixel values in the region. Then this operation is used iteratively across the spatial dimension in a sliding manner with a specific stride. Given a region of size $k \times k$, some pixels contributing to the edge values are outside this region. To make the problem feasible, we only consider a scope with the size of $k \times k$, and each pixel is computed restrictedly by the k^2 pixels in this scope with its individual filter. Next, based on this simplification and above operators, we prove the validity of the low-pass filter (Wang et al., 2022a).

Theorem 1. Given W, $D \in \mathbb{R}^{n \times n}$, $W_i = softmax(D_i)$, $i = 0, 1, ..., n-1$. Then W is a low-pass filter. For all $m \in \mathbb{R}^n$,

$$\lim_{t \to \infty} \frac{\|\mathcal{H}[W^t m]\|_2}{\|W^t m\|_2} = 0.$$

In our case, n is equal to k^2.

Proof. Note that after Softmax operation, W is a positive matrix, and each line adds up to 1. Thus $We = e$ implies that it has an eigenvalue of 1, and the corresponding eigenvector is $e \in \mathbb{R}^n$ with all values being 1. Let scalars λ_1, λ_2,..., λ_s denote the eigenvalues of W, in which $\lambda_1 = 1$. According to Perron-Frobenius Theorem (Meyer, 2000), λ_1 is the spectral radius of the matrix W, and absolute values of other eigenvalues are less than λ_1.

To compute the high-order equation of W, the Jordan canonical form of W is given as,

$$W = PJP^{-1} = [a_1 \ \ a_2 \ \ ... \ \ a_n] \begin{bmatrix} \lambda_1 & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{bmatrix} \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_n^T \end{bmatrix} \tag{4}$$

where P is the transition matrix, a, $b \in \mathbb{R}^n$, $a_1 = e$, $J(\lambda)$ is the Jordan block. Then $W^t m$ can be rewritten as,

$$W^t m = PJ^t P^{-1} m = P \begin{bmatrix} \lambda_1^t & & & \\ & J_2(\lambda_2)^t & & \\ & & \ddots & \\ & & & J_s(\lambda_s)^t \end{bmatrix} P^{-1} m \tag{5}$$

For a Jordan block $J_i(\lambda_i) \in \mathbb{R}^{z_i \times z_i}$, $i \in [2, s]$, $J_i(\lambda_i)^t$ can be computed by,

$$J_i(\lambda_i) = \begin{bmatrix} \lambda_i^t & (\lambda_i^t)' & \frac{(\lambda_i^t)''}{2!} & \cdots & \frac{(\lambda_i^t)^{(z_i-1)}}{(z_i-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{(\lambda_i^t)''}{2!} \\ & & & \ddots & (\lambda_i^t)' \\ & & & & \lambda_i^t \end{bmatrix}_{z_i} \tag{6}$$

For arbitrary $1 \leq q \leq z_i$ and $|\lambda_i| < 1$, the limit of the values except the main diagonal is,

$$\lim_{t\to\infty} \frac{t(t-1)...(t-q+1)}{q!}\lambda_i^{t-q} = \lim_{t\to\infty} \frac{t(t-1)...(t-q+1)}{q!\lambda_i^q(\frac{1}{\lambda_i})^t} = 0 \tag{7}$$

The above limit stands due to the the larger growing rate of exponential function (denominator) than that of power function (numerator). It is easy to obtain that the limit of the main diagonal is equal to zero by $\lim_{t\to\infty} \lambda_i^t = 0$. As a consequence, for Jordan blocks $J_i(\lambda_i)^t$, $i \in [2, s]$, the limits of their elements are zeros. Therefore, based on Eq. 5, the limit of $W^t m$ is,

$$\begin{aligned} \lim_{t\to\infty} W^t m &= Pdiag(\lambda_1^t, 0, ..., 0)P^{-1}m \\ &= \lambda_1^t a_1 b_1^T m \\ &= e b_1^T m \end{aligned} \tag{8}$$

The limit in the theorem 1 can be computed as,

$$\begin{aligned} \lim_{t\to\infty} \frac{\|\mathcal{H}[W^t m]\|_2}{\|W^t m\|_2} &= \lim_{t\to\infty} \frac{\|(E - \frac{1}{n}M)W^t m\|_2}{\|W^t m\|_2} \\ &= \lim_{t\to\infty} \frac{\|(E - \frac{1}{n}M)eb_1^T m\|_2}{\|W^t m\|_2} \\ &= \lim_{t\to\infty} \frac{\|Eeb_1^T m - \frac{1}{n}Meb_1^T m\|_2}{\|W^t m\|_2} \\ &= \lim_{t\to\infty} \frac{\|Eeb_1^T m - eb_1^T m\|_2}{\|W^t m\|_2} \\ &= 0 \end{aligned} \tag{9}$$

□

C Analysis for MCSF

In this part, we analyze the validity of our MCSF, where we use average pooling to decompose the feature into different frequency parts. It is known that degraded/sharp image pairs have similar low-frequency components (Liu et al., 2020), while high-frequency signals are more informative for high-quality image. A natural question is that which threshold in the frequency domain should we select for frequency separation. We consider that the qualified threshold should satisfy the requirement: after being split with this threshold, there should a big difference between the image pair's *high*-frequency discrepancies and *low*-frequency discrepancies. In our case, we choose the lowest frequency and its complementary counterpart as the *low* and *high* frequency, respectively. Next, we provide the analysis to demonstrate that our choice meets the above requirement in the context of image restoration tasks, instead of obtaining this threshold via blindly testing in the frequency domain. Without loss of generality, we carry out analyses based on motion deblurring task.

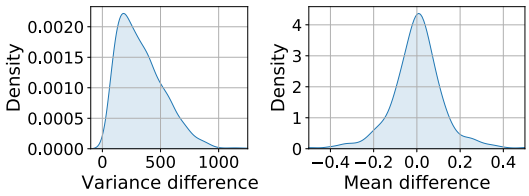

Figure 1: The probability density curve of the difference between blurry/sharp image pairs in the GoPro (Nah et al., 2017) dataset.

As we all know, the lowest frequency component of Fourier transform is equal to the mean value in the spatial domain. We firstly compare the low-frequency differences between sharp/blurry image pairs by computing the mean difference in Fig. 1 (Right). From the figure we can see that, the mean difference between image pair is very small. Next, we explore the discrepancies between high frequency, the opposite of lowest frequency. Based on the low-frequency feature, given any input $x_{h,w}$, where h, w are the spatial coordinates and we omit the channel dimension for simplicity, we can generate the relative high-frequency map via $x^H = x - f_0$, in which f_0 is the mean value or lowest frequency component in the frequency domain. Since bigger variance value means more high-frequency signals, we then resort to the tool of variance to demonstrate that the so-called high-frequency map of the sharp image indeed has more high-frequency signals than that of the blurry

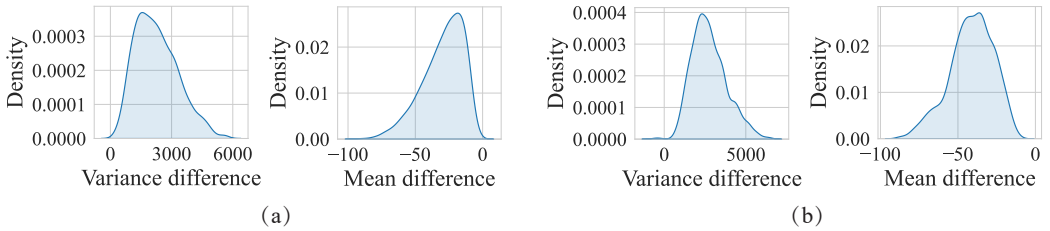

Figure 2: The probability density curves of the differences between degraded/sharp image pairs in the **(a)** RESIDE-Outdoor (Li et al., 2018a), **(b)** CSD (Chen et al., 2021b) datasets.

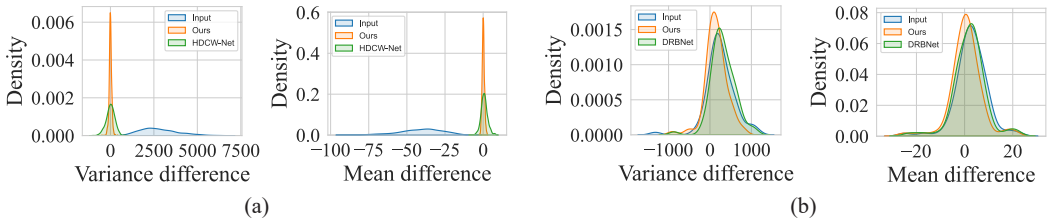

Figure 3: The differences between ground truth and input/results of several methods. **(a)** Comparisons between HDCW-Net (Chen et al., 2021a) and SFNet on CSD (Chen et al., 2021b). **(b)** Comparisons between DRBNet (Ruan et al., 2022) and SFNet on DPDD (Abuolaim & Brown, 2020).

image. The variance differences between image pairs are given in Fig. 1 (Left) and we can obtain the variance relationships for the most image pairs,

$$\frac{\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}(x_{h,w}^{SH})^2}{HW} > \frac{\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}(x_{h,w}^{BH})^2}{HW} \tag{10}$$

where $x_{h,w}^{SH}$ and $x_{h,w}^{BH}$ indicate so-called high-frequency maps of the sharp and blurry images, respectively. As $x_{h,w}^{SH}$ and $x_{h,w}^{BH}$ are computed by subtracting f_0, and thus they become zero-centered. Eq. 10 actually shows the variance relationship between the high-frequency feature of image pair.

Based on Eq. 10 and the amplitude differences between x-axes of mean and variance (Fig. 1), we can draw two conclusions: (1) The discrepancies between high-frequency maps of image pairs are larger than that of low-frequency maps. (2) The so-called high-frequency part of sharp image really has more high-frequency signals than that of blurry image, obtained by the property of variance.

Till now, we have shown that our choice meets the presupposed requirement to serve the threshold for low- and high-frequency separation in the context of motion deblurring. In Fig. 2, we provide statistic differences between image pairs for more tasks[1]. We can see that there is a large gap between the data distribution of variance (high-frequency) and mean (low-frequency) difference in terms of sign, scope and amplitude, demonstrating that our MCSF is reasonable.

D Statistic Analysis of Produced Images

In this section, we verify the effectiveness of our method by comparing the data distributions of degraded images and outcomes of several methods. The variance and mean differences are computed between degraded/output images and ground-truth images. For desnowing (Fig. 3a), our method obtains the resulting images closer to the ground truth of CSD (Chen et al., 2021b) testset in terms of variance and mean than wavelet transform based method HDCW-Net (Chen et al., 2021a). In Fig. 3b, we compare SFNet with DRBNet (Ruan et al., 2022) on the single-image defocus deblurring task. DRBNet narrows down the distribution difference of variance compared to the degraded input, while its center line shifts to the right. By contrast, our result is closer to the ground truth.

[1]We randomly select 2000 images from the training set to produce the results.

Table 2: Desnowing results on SRRS (Chen et al., 2020) and Snow100K (Liu et al., 2018) datasets.

	SRRS (2000)		Snow100K (2000)	
Method	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DesnowNet (Liu et al., 2018)	20.38	0.84	30.50	0.94
CycleGAN (Engin et al., 2018)	20.21	0.74	26.81	0.89
All in One (Li et al., 2020)	24.98	0.88	26.07	0.88
JSTASR (Chen et al., 2020)	25.82	0.89	23.12	0.86
HDCW-Net (Chen et al., 2021a)	27.78	0.92	31.54	0.95
TransWeather (Valanarasu et al., 2022)	28.29	0.92	31.82	0.95
MSP-Former (Chen et al., 2022c)	30.76	0.95	33.43	**0.96**
NAFNet (Chen et al., 2022a)	29.72	0.94	32.41	0.95
SFNet	**32.40**	**0.98**	**33.79**	0.95

Table 3: Dual-pixel defocus deblurring results on DPDD (Abuolaim & Brown, 2020).

Method	PSNR↑	SSIM↑	MAE↓	LPIPS↓
DPDNet (Abuolaim & Brown, 2020)	25.13	0.786	0.041	0.223
RDPD (Abuolaim et al., 2021)	25.39	0.772	0.040	0.255
Uformer (Wang et al., 2022b)	25.65	0.795	0.039	0.243
IFAN (Lee et al., 2021)	25.99	0.804	**0.037**	0.207
DRBNet (Ruan et al., 2022)	26.33	0.811	-	**0.154**
SFNet	**26.34**	**0.817**	**0.037**	0.207

E MORE EXPERIMENTS

In Tab. 2, we provide more desnowing results on two datasets, SRRS (Chen et al., 2020) and Snow100K (Liu et al., 2018). Our method shows superiority over other algorithms. Specifically, compared to the recent algorithm MSP-Former (Chen et al., 2022c), SFNet achieves a significant gain of 1.64 dB PSNR on SRRS. Furthermore, we provide experimental results on DPDD (Abuolaim & Brown, 2020) for dual-pixel defocus deblurring. As reported in Tab. 3, our method outperforms DRBNet (Ruan et al., 2022), obtaining 0.01 dB higher PSNR, although DRBNet needs additional central view images for training.

F ADDITIONAL VISUAL RESULTS

We provide images recovered by various methods for different image restoration tasks, organised as,

- Image motion deblurring: Fig. 4, 5, 6.
- Image deraining: Fig. 7, 8, 9, 10.
- Image defocus deblurring: Fig. 11.
- Image dehazing: Fig. 12.
- Image desnowing: Fig. 13.

REFERENCES

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

Abdullah Abuolaim, Mauricio Delbracio, Damien Kelly, Michael S. Brown, and Peyman Milanfar. Learning to reduce defocus blur by realistically modeling dual-pixel data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2289–2298, October 2021.

C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In *arXiv*, 2019.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, 2022a.

Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2022b.

Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Yi Ye, and Erkang Chen. Msp-former: Multi-scale projection transformer for single image desnowing. *arXiv preprint arXiv:2207.05621*, 2022c.

W. T. Chen, H. Y. Fang, C. L. Hsieh, C. C. Tsai, I. H. Chen, J. J. Ding, and S. Y. Kuo. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *International Conference on Computer Vision*, 2021a.

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*, pp. 754–770. Springer, 2020.

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/CVF International Conference on Computer Vision*, pp. 4196–4205, 2021b.

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/CVF International Conference on Computer Vision (ICCV)*, pp. 4641–4650, October 2021.

Deniz Engin, Anil Genc, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, PP(99), 2017a.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017b.

Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8346–8355, 2020.

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/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2034–2042, 2021.

B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single image dehazing and beyond. *IEEE Transactions on Image Processing*, PP:1–1, 2018a.

Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *European Conference on Computer Vision*, pp. 262–277. Springer, 2018b.

Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2736–2744, 2016.

Keng-Hao Liu, Chia-Hung Yeh, Juh-Wei Chung, and Chuan-Yu Chang. A motion deblur method based on multi-scale high frequency residual image learning. *IEEE Access*, 8:66025–66036, 2020.

Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7314–7323, 2019.

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. doi: 10.1109/TIP.2018.2806202.

Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.

S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Computer Vision and Pattern Recognition*, 2017.

Kuldeep Purohit and AN Rajagopalan. Region-adaptive dense network for efficient motion deblurring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11882–11889, 2020.

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 Conference on Artificial Intelligence*, volume 34, pp. 11908–11915, 2020.

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

Lingyan Ruan, Bin Chen, Jizhou Li, and Miu-Ling Lam. Aifnet: All-in-focus image restoration network using a light field-based dataset. *IEEE Transactions on Computational Imaging*, 7:675–688, 2021. doi: 10.1109/TCI.2021.3092891.

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/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16304–16313, 2022.

Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao. Human-aware motion deblurring. In *International Conference on Computer Vision*, 2020.

Hyeongseok Son, Junyong Lee, Sunghyun Cho, and Seungyong Lee. Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2642–2650, 2021.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *arXiv preprint arXiv:2204.03883*, 2022.

Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3606–3615, 2020.

Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *ECCV*, 2022.

Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. *CVPR*, 2022.

Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2353–2363, 2022.

Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. In *International Conference on Learning Representations (ICML)*, 2022a.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17683–17693, 2022b.

Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

Rajeev Yasarla and Vishal M. Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

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/CVF conference on computer vision and pattern recognition*, pp. 14821–14831, 2021.

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/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018a.

He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, 2018b.

He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11): 3943–3956, 2019a.

Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5978–5986, 2019b.

Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2737–2746, 2020.

Figure 4: **Image motion deblurring** comparisons on GoPro (Nah et al., 2017) among DBGAN (Zhang et al., 2020), DMPHN (Zhang et al., 2019b), MIMO-UNet++ (Cho et al., 2021), MPRNet (Zamir et al., 2021), Restormer (Zamir et al., 2022), Stripformer (Tsai et al., 2022), and SFNet.

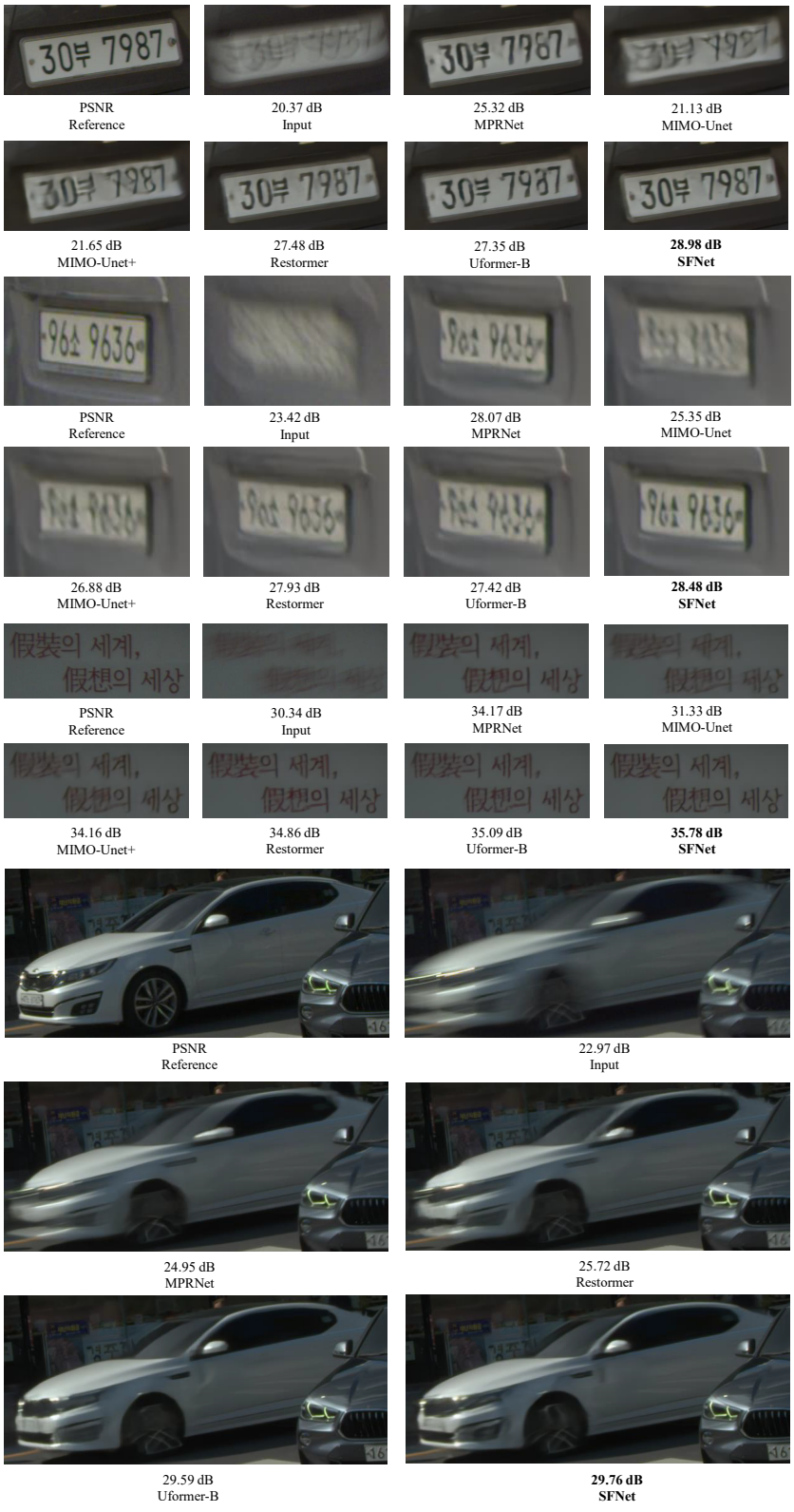

Figure 5: **Image motion deblurring** comparisons on the RSBlur dataset (Rim et al., 2022)among MPRNet (Zamir et al., 2021), MIMO-UNet (Cho et al., 2021), MIMO-UNet+ (Cho et al., 2021), Restormer (Zamir et al., 2022), Uformer-B (Wang et al., 2022b), and SFNet.

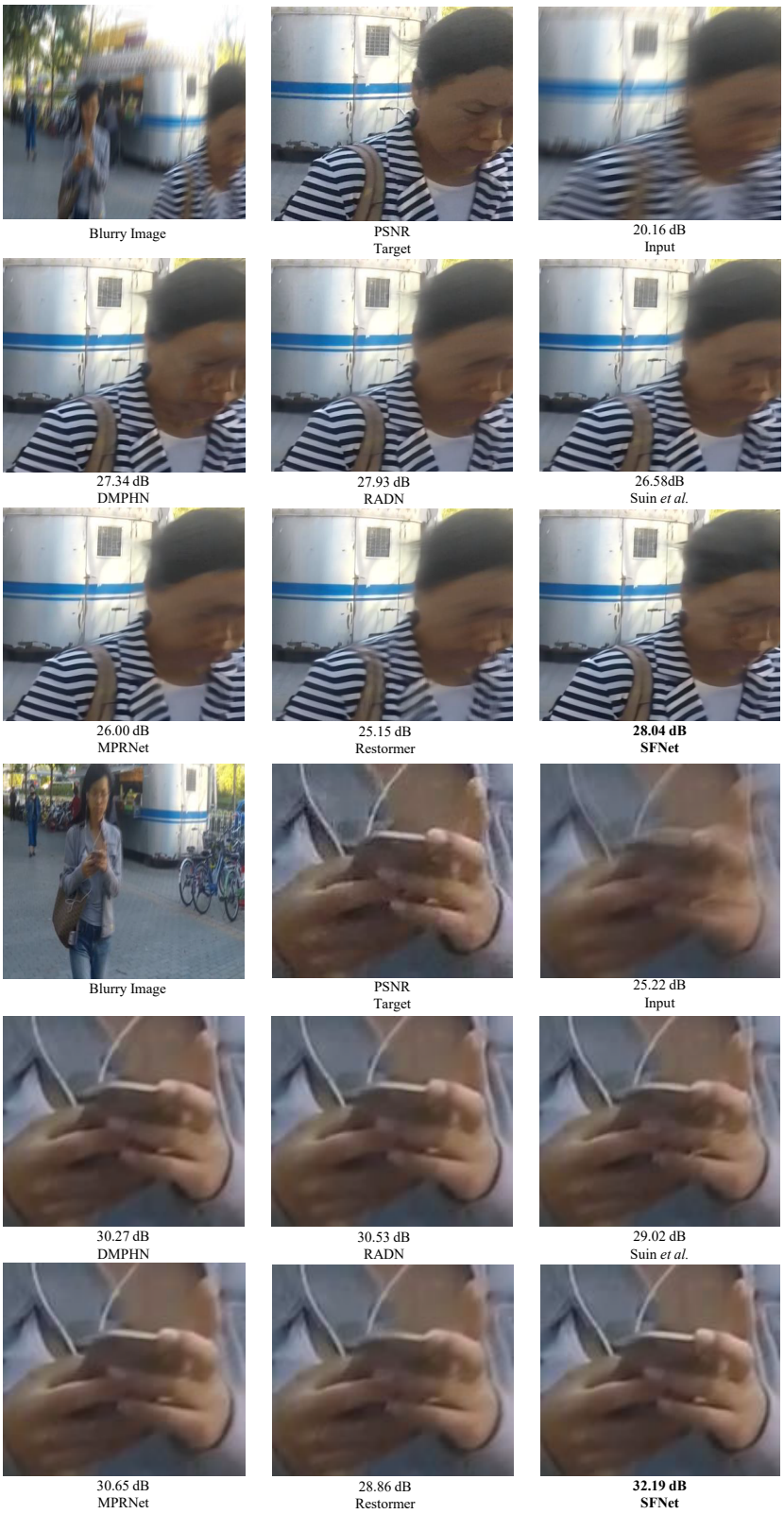

Figure 6: **Image motion deblurring** comparisons on the HIDE dataset (Shen et al., 2020) among DMPHN (Zhang et al., 2019b), RADN (Purohit & Rajagopalan, 2020), Suin *et al.* (Suin et al., 2020), MPRNet (Zamir et al., 2021), Restormer (Zamir et al., 2022), and SFNet.

Figure 7: **Image deraining** comparisons on the Rain100H dataset (Yang et al., 2017) among DerainNet (Fu et al., 2017a), PreNet (Ren et al., 2019), RESCAN (Li et al., 2018b), SEMI (Wei et al., 2019), UMRL (Yasarla & Patel, 2019), MIXIM-2S (Tu et al., 2022), and SFNet.

Figure 8: **Image deraining** comparisons on the Rain100L dataset (Yang et al., 2017) among PreNet (Ren et al., 2019), RESCAN (Li et al., 2018b), SEMI (Wei et al., 2019), UMRL (Yasarla & Patel, 2019), MPRNet (Zamir et al., 2021), MIXIM-2S (Tu et al., 2022), and SFNet.

Figure 9: **Image deraining** comparisons on the Test100 dataset (Zhang et al., 2019a) among PreNet (Ren et al., 2019), RESCAN (Li et al., 2018b), SEMI (Wei et al., 2019), UMRL (Yasarla & Patel, 2019), DerainNet (Fu et al., 2017a), MPRNet (Zamir et al., 2021), and SFNet.

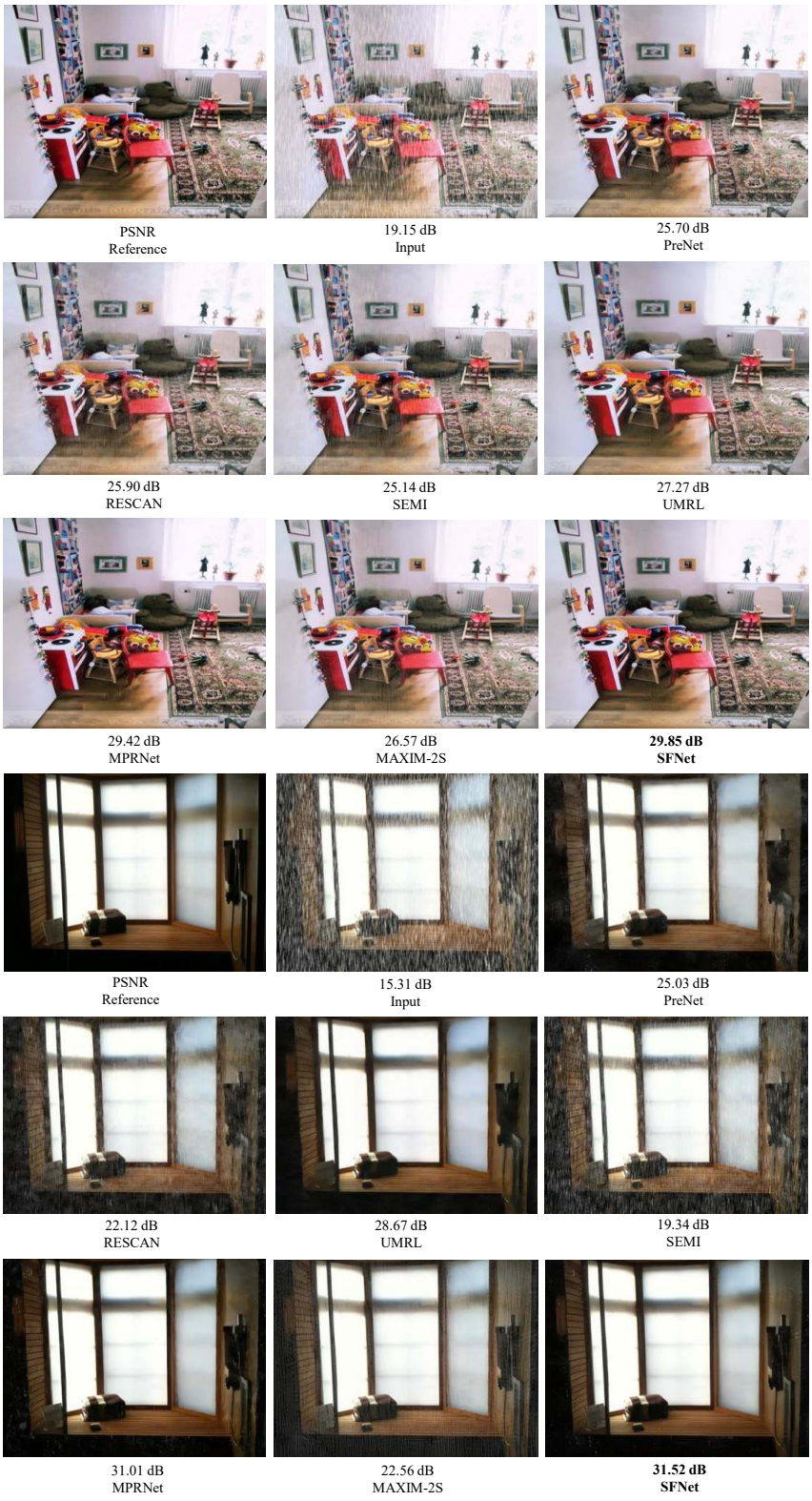

Figure 10: **Image deraining** comparisons on the Test1200 dataset (Zhang & Patel, 2018b) among PreNet (Ren et al., 2019), RESCAN (Li et al., 2018b), SEMI (Wei et al., 2019), UMRL (Yasarla & Patel, 2019), MPRNet (Zamir et al., 2021), MIXIM-2S (Tu et al., 2022), and SFNet.

Figure 11: **Image defocus deblurring** comparisons on DPDD (Abuolaim & Brown, 2020) among KPAC (Son et al., 2021), DRBNet (Ruan et al., 2022), Restormer(Zamir et al., 2022), and SFNet.

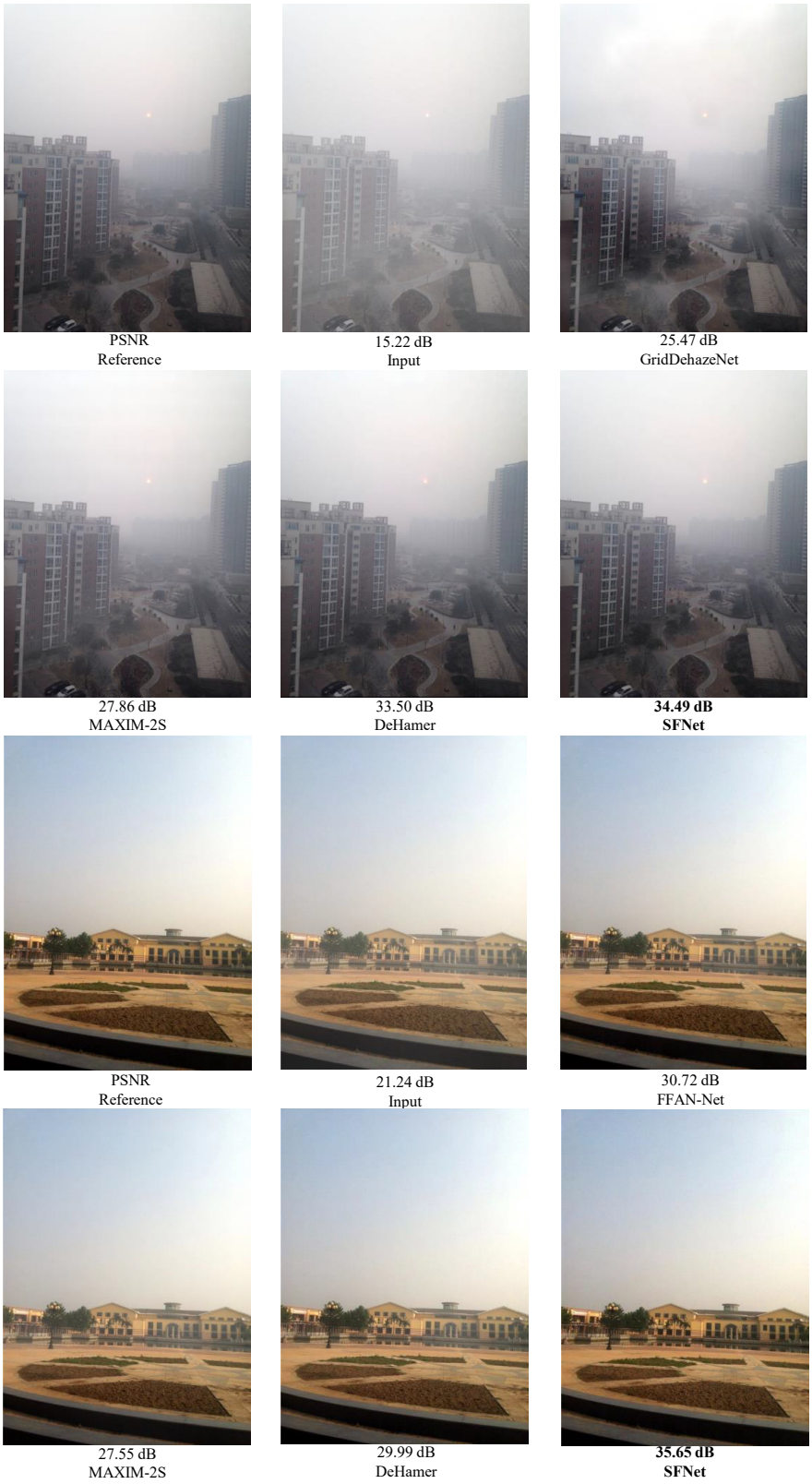

Figure 12: **Image dehazing** comparisons on the SOTS-Outdoor dataset (Li et al., 2018a) among GridDehazeNet (Liu et al., 2019), FFA-Net (Qin et al., 2020), MAXIM-2S (Tu et al., 2022), DeHamer (Song et al., 2022), and SFNet.

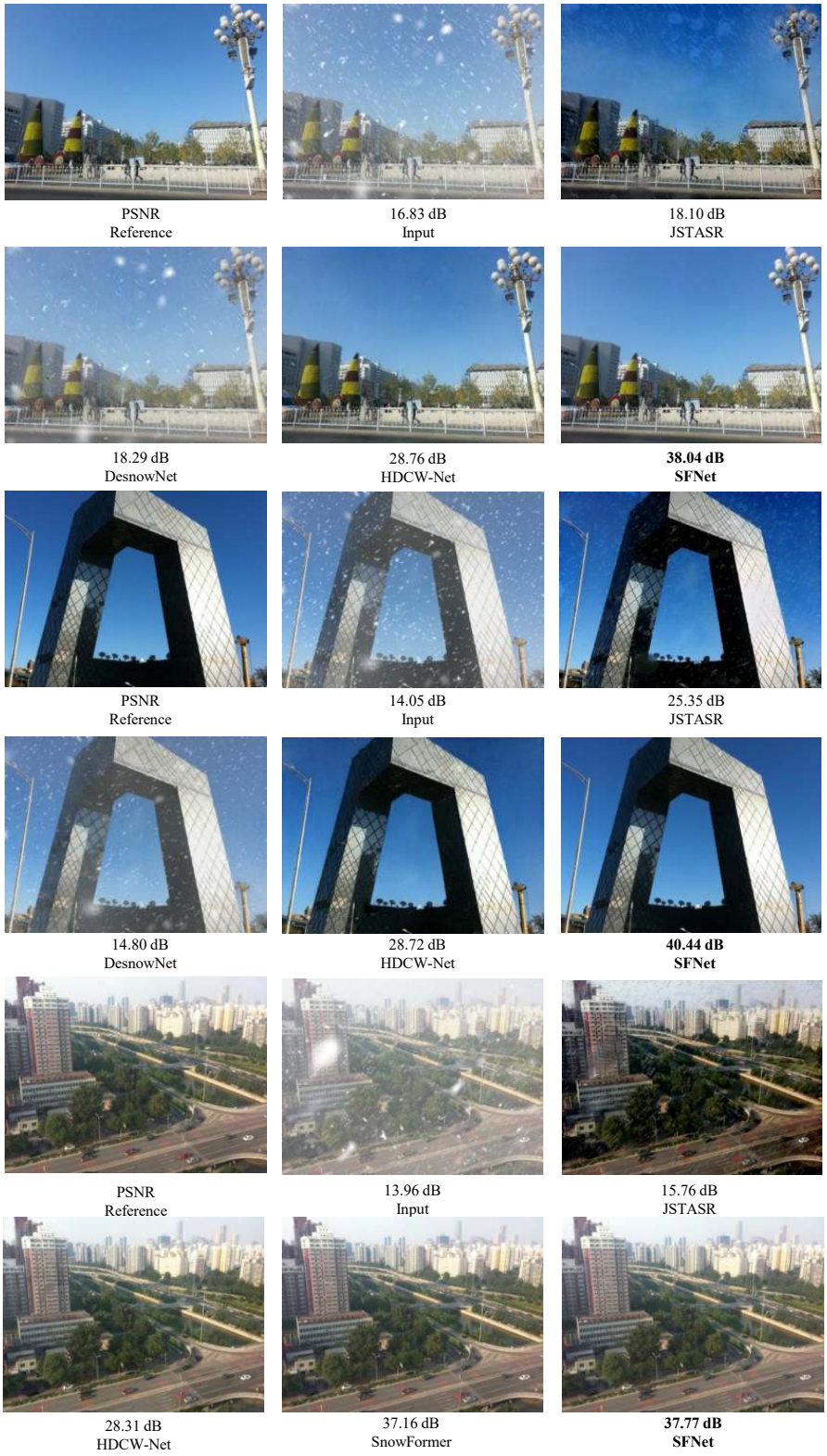

Figure 13: **Image desnowing** comparisons on the CSD dataset (Chen et al., 2021b) among JSTASR (Chen et al., 2020), HDCW-Net (Chen et al., 2021a), DesnowNet (Liu et al., 2018), SnowFormer (Chen et al., 2022b), and SFNet.