SELECTIVE FREQUENCY NETWORK FOR IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Image restoration aims to reconstruct the latent sharp image from the corrupted observation. Besides dealing with this long-standing task in the spatial domain, a few approaches seek solutions in the frequency domain based on the large discrepancy between spectra of sharp/degraded image pairs. However, these works utilize the existing transformation tools, *e.g.*, Wavelet Transform, to split the feature into several parts, which is not flexible enough to select the most informative frequency component to recover. In this paper, we exploit a multi-branch and content-aware module to decompose the feature into separate frequency subbands dynamically and locally, and then accentuate the useful ones via the channel-wise attention mechanism. In addition, aiming to cope with the large-scale degradation kernel, we propose an extremely simple decoupling and modulation module to enlarge the receptive field based on global and window-based average pooling. Integrating two developed modules into a U-Net backbone, the proposed **S**elective **F**requency **Net**work (SFNet) performs favorably against state-of-the-art algorithms on five image restoration tasks, including image dehazing, image motion/defocus deblurring, image desnowing, and image deraining.

1 INTRODUCTION

Image restoration aims to recover a high-quality image by removing degradation, *e.g.*, noise, blur, and snowflake. In view of its important role in surveillance, self-driving techniques, and remote sensing, image restoration has gathered considerable attention from industry and academic communities. However, due to its ill-posed property, many conventional approaches deal with this challenging problem based on the assumption (Zhang et al., 2022a; Yang et al., 2020b) or hand-crafted features (Karaali & Jung, 2017), which are incapable of generating faithful results in complicated and real-world scenarios.

In recent years, deep neural networks (DNN) have witnessed the rapid development of image restoration and obtained favorable performance compared with traditional restoration methods. A flurry of convolutional neural networks (CNN) based methods have been developed for diverse image restoration tasks by inventing or borrowing advanced modules, including dilated convolution (Luo et al., 2022; Zou et al., 2021), U-Net (Ronneberger et al., 2015), residual learning (Zhang et al., 2017), multi-stage pipeline (Zhang et al., 2019b), and attention mechanisms (Liu et al., 2019). However, with convolution units, these methods have the nature of limited receptive fields, and thus, they are not capable of capturing long-range dependencies. This requirement is essential for restoration tasks, since a single pixel needs information from its surrounding region to be recovered. More recently, many frameworks tailor Transformer (Vaswani et al., 2017) for a variety of image restoration tasks, including motion deblurring (Tsai et al., 2022a), dehazing (Guo et al., 2022; Song et al., 2022) and desnowing (Chen et al., 2022b;c).

Nonetheless, the above-mentioned methods conduct restoration in the spatial domain, which do not sufficiently leverage the discrepancies between sharp/degraded image pairs in the frequency domain. Thus, many researchers attempt to solve restoration tasks from the frequency perspective. To this end, a few works utilize the Wavelet transform or its more advanced variant, dual-tree complex wavelet representation (Selesnick et al., 2005), to decompose the feature into various frequency components and then treat the separate part individually to reconstruct the corresponding feature (Selesnick et al., 2005; Yang & Fu, 2019; Zou et al., 2021). Nevertheless, Wavelet transform de-

couples the feature map into different subbands in a fixed manner, and thus it is not capable of distinguishing the most informative or useless frequency components to enhance or suppress. Some other works use convolution layers (Pang et al., 2020) or Fourier transform (Mao et al., 2021) to achieve frequency map separation. These methods need inverse transform, *e.g.*, inverse Fourier transform or up-sampling, which causes additional computation overhead.

To overcome the above drawbacks and select the most informative frequency component to reconstruct, we propose a novel decoupling and recalibration module for image restoration tasks, named Multi-branch Dynamic Selective Frequency module (MDSF). Specifically, we utilize the multi-branch learnable filter to generate high- and low-frequency maps dynamically and locally, and then leverage the channel-wise attention mechanism, modified from (Li et al., 2019), to emphasize or suppress the decomposed frequency components. Our module has two key advantages. First, according to the input and task, the decoupling step dynamically generates the filter to separate the feature map. Second, our module does not need any extra inverse transformation.

Receptive field is another critical factor for image restoration tasks due to the various sizes of degradation kernels (Suin et al., 2020; Son et al., 2021). To complement the dynamic module that processes feature locally, we further propose a simple yet effective module, dubbed Multi-branch Compact Selective Frequency module (MCSF), to enhance the helpful frequency signals based on the multiple and relatively global receptive fields. Specifically, based on the analyses, we utilize global and window-based average pooling to attain disparate frequency maps, and then use learnable parameters to modulate the resultant maps without introducing any convolution layers. Compared with MDSF, besides the enlarged receptive fields, MCSF is lightweight enough to be embedded into the backbone in multiple positions.

Our main contributions are as follows:

- We propose a multi-branch dynamic selective frequency module (MDSF) that is capable of decoupling the feature map into different frequency parts in a flexible manner and then selecting the most informative components to recover.
- We design a multi-branch compact selective frequency module (MCSF) to conduct frequency decoupling and recalibration based on global receptive field to deal with the large-size degradation kernel.
- Incorporating MDSF and MCSF into a U-shaped backbone, a selective frequency network (SFNet) is formed to address image restoration. Our network achieves state-of-the-art results on five image restoration tasks.

2 Related Work

Image Restoration. Image restoration tasks aim to recover the high-quality image from its degraded counterpart. Prior to the deep learning era, a great number of methods have been developed to address this ill-posed problem (Sezan & Tekalp, 1990; Kundur & Hatzinakos, 1996; Calvetti et al., 1999). With the development of deep learning, a flurry of approaches have been proposed utilizing convolutional neural networks to deal with diverse image restoration tasks, including image motion deblurring (Zamir et al., 2021; Yuan et al., 2020; Zhang et al., 2019b; Purohit et al., 2021), defocus deblurring (Ruan et al., 2022; Abuolaim & Brown, 2020; Son et al., 2021), desnowing (Chen et al., 2021c; 2020a), dehazing (Dong et al., 2020; Liu et al., 2019; Ren et al., 2016; Zhang et al., 2018), and deraining (Yang et al., 2020b; Wang et al., 2019).

Recently, aiming to attain the long-range dependency, many works borrow the Transformer model (Vaswani et al., 2017) from natural language tasks to address image restoration such as (Chen et al., 2021a; Liang et al., 2021; Zamir et al., 2022; Wang et al., 2022). Researches also tailor Transformer model for individual restoration tasks, *e.g.*, image motion deblurring (Tsai et al., 2022a), dehazing (Song et al., 2022) and desnowing (Chen et al., 2022b). Additionally, a few works attempt to deal with image restoration using multi-layer perceptron (Tu et al., 2022) and adder neural networks (Song et al., 2020). In this paper, instead of exploiting advanced backbone for image restoration, we focus on the frequency selection mechanism based on CNN, which is more efficient than Transformer type network.

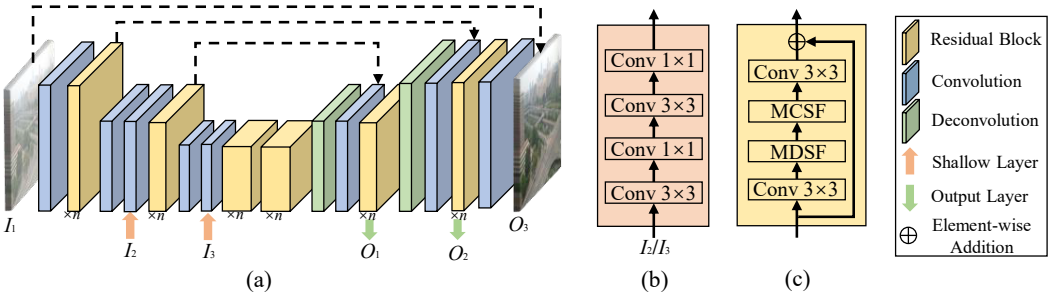

Figure 1: **(a)** The overview of the proposed selective frequency network. I_i, $i \in \{1, 2, 3\}$ are the multiple input images with different resolutions, which are involved in the mainstream via concatenation after shallow feature extraction. O_i, $i \in \{1, 2, 3\}$ are the multiple output images to supervise training. The output layer is a convolution with 3×3 kernel size. The down-sampling and up-sampling are accomplished by stride convolution and deconvolution, respectively. **(b)** The shallow layer to extract the feature for down-sampled input image. **(c)** The modified residual block accommodating MDSF and MCSF.

Frequency based image restoration. Many algorithms have been investigated to complete various computer vision tasks from a frequency perspective (Gao et al., 2021; Cai et al., 2020; Tao et al., 2021). In the context of image restoration, Xu *et al.* propose a novel hierarchical desnowing network based on dual-tree complex wavelet representation (Selesnick et al., 2005) to handle the complicated snow scenario (Chen et al., 2021c). Yang *et al.* develop the Wavelet based U-Net to replace up-sampling and down-sampling to extract edge features (Yang & Fu, 2019). Zou *et al.* design a module to help the network recover the texture details using Wavelet transform and convolution layers (Zou et al., 2021). Yang *et al.* devise a Wavelet structure similarity loss function for training (Yang et al., 2020a). Mao *et al.* use Fourier transform to integrate both high- and low-frequency residual information. Zhang *et al.* borrow OctConv (Chen et al., 2020b) to separate frequency (Zhang et al., 2022b). In this work, we pursue a dynamic and efficient manner to select the useful frequency part at multiple receptive fields.

3 METHODOLOGY

We first detail the two proposed modules, MDSF and MCSF. Then we move to the manner for integrating two modules with a U-shaped backbone. The loss function follows in the final part.

3.1 MULTI-BRANCH DYNAMIC SELECTIVE FREQUENCY MODULE (MDSF)

To select the informative frequency to reconstruct, as shown in Fig. 2a, MDSF mainly consists of two components: frequency decoupler and recalibrator. Decoupler (Fig. 2b) is capable of decomposing the feature map into separate frequency components dynamically, and recalibrator (Fig. 2c) utilizes channel attention to accentuate the helpful frequency. Besides, MDSF contains multiple branches, individually operating two steps, to provide various local receptive fields. For convenience, we only show two branches in Fig. 2a. More branches are experimented in Appendix G.

To dynamically decompose feature maps into frequency subbands, we utilize the learnable and theoretically proven low-pass filter (refer to Appendix B for the proof) and the corresponding high-pass filter to generate low- and high-frequency maps, respectively. We filter the input feature in groups to enhance the feature diversity. Specifically, for any feature map $X \in \mathbb{R}^{C \times H \times W}$, where C is the number of channels and $H \times W$ denotes the spatial dimension, we firstly leverage the filter-generating layer to generate the low-pass filter kernel for each group of the input feature,

$$F^l = \mathcal{B}(W(\mathrm{GAP}(X))) \tag{1}$$

where $F^l \in \mathbb{R}^{k^2 g \times 1 \times 1}$, k^2 is the kernel size of low-pass filter, g denotes the number of groups, $\mathcal{B}, W$, GAP are Batch Normalization (Ioffe & Szegedy, 2015), the parameters of convolution and global average pooling operation, respectively. The group based operation has fewer parameters and complexities than generating filters for each channel. The number of groups is discussed in Sec 4.8.

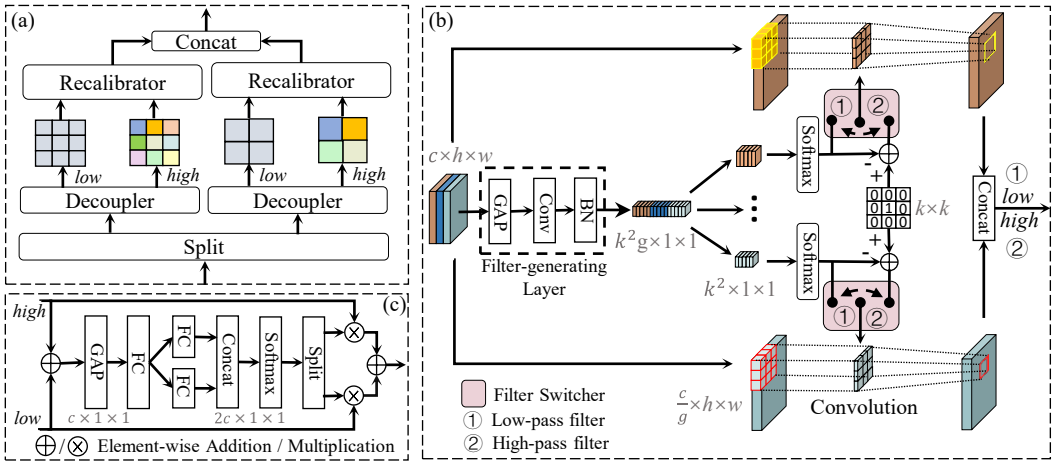

Figure 2: **(a)** The thumbnail of the proposed frequency selection strategy for MDSF. The split input features go through frequency decoupler and recalibrator to select the informative frequency to recover. Two branches have different receptive fields, which are controlled by the kernel size in the decoupler. We only depict two branches here for simplicity. *high* and *low* denote high- and low-frequency map, respectively. **(b)** The frequency decoupler for MDSF generates high- and low-frequency signals under the control of the filter switcher. **(c)** The frequency recalibrator network for MDSF emphasizes the useful components in the decomposed frequency.

Based on the generated low-pass filter, as shown in Fig. 2b, we utilize the filter switcher to get the high-pass filter via subtracting low-pass filter from the identity kernel. Next, for each group feature $X_i \in \mathbb{R}^{C_i \times H \times W}$, where i is the group index and $C_i = \frac{C}{g}$, the low- and high-frequency features can be obtained with the corresponding reshaped filter F^L and $F^H \in \mathbb{R}^{g \times k \times k}$,

$$X^l_{i,c,h,w} = \sum_{p,q} F^L_{i,p,q} \otimes X_{i,c,h+p,w+q}; \quad X^h_{i,c,h,w} = \sum_{p,q} F^H_{i,p,q} \otimes X_{i,c,h+p,w+q} \tag{2}$$

where $c \in \mathbb{R}^{C_i}$, h and w denote the spatial coordinates, $p, q \in \{-1, 0, 1\}$. Note that F_i is shared among the spatial dimension in a group, which leads to the fewer parameters and complexities.

After decoupling the feature map into different frequency components, we move to the second step, using the frequency recalibrator to emphasize the most needed part for reconstruction. The recalibrator works among the channel dimension based on the modified SKNet (Li et al., 2020). Formally, given two frequency maps (Fig. 2c), X^l and X^h, we first generate the fused feature following the setting in (Li et al., 2020),

$$Z = W_{fc}(\mathrm{GAP}(X^l + X^h)) \tag{3}$$

where W_{fc} is the parameters of fully connected layer. To attain the channel-wise weights, we use two other fully connected layers followed by concatenation and Softmax function,

$$[W^l, W^h]_c = \frac{e^{[W_l(Z), W_h(Z)]_c}}{\sum_j^{2C} e^{[W_l(Z), W_h(Z)]_j}} \tag{4}$$

where W^l and W^h are channel weights for two frequency parts, W_l and W_h are parameters of fully connected layers, $[\cdot, \cdot]$ denotes concatenation, C is the number of channels, $c \in \mathbb{R}^{2C}$. Compared with SKNet (Li et al., 2020), which performs Softmax function on each channel as $W^l_c = \frac{e^{W_l(Z)_c}}{e^{W_l(Z)_c} + e^{W_h(Z)_c}}$, we consider all channels to facilitate interactions between different channels of two input maps , showing superior performance (Appendix G). Then, the final weights can be obtained by split operation.

In summary, given two branches in Fig. 2a, MDSF works at two local receptive fields as,

$$Y = [\mathrm{Rec}(\mathrm{Dec}(X_1), \mathrm{Rec}(\mathrm{Dec}(X_2)] \tag{5}$$

where Rec and Dec denote recalibrator and decoupler, and X_i is the equally split map among the channel dimension.

3.2 MULTI-BRANCH COMPACT SELECTIVE FREQUENCY MODULE (MCSF)

Since the receptive field plays a critical role in computer vision tasks (Luo et al., 2017), especially for image restoration, where degradation kernels are different in size (Son et al., 2021; Mao et al., 2021), we propose MCSF to enlarge the receptive field of SFNet. This module shares the same pipeline with MDSF in Fig. 2a, but implements two steps, frequency decoupler and recalibrator, in a much more efficient and cheap fashion. MCSF also has multiple branches with different receptive fields. Considering these branches share a similar paradigm, we only describe the details of the window-based branch (right half part of Fig. 3), which is inspired from the idea of window partition in window-based attention (Liu et al., 2021).

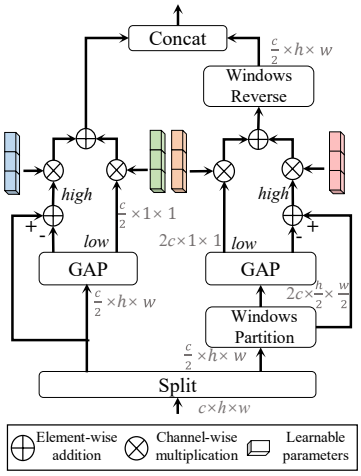

Figure 3: Diagram of MCSF.

Specifically, given the split feature $X \in \mathbb{R}^{\frac{C}{2}\times H\times W}$, it is partitioned into four windows, each with the size of $2c \times \frac{H}{2} \times \frac{W}{2}$. To get the low-frequency part, a mean filter is applied to the resultant windows (refer to Appendix C for analyses of this option). The corresponding high-frequency part can be obtained by subtracting the low-frequency map from the window feature. To select the useful frequency subbands, we rescale these two maps by learnable weights, which are directly optimized by back propagation. Finally, the updated frequency maps are reversed to the same size as the original input. The left branch in Fig. 3 has a similar operation, yet with a global receptive field.

Compared with MDSF, besides the enlarged receptive field, MCSF does not accomplish frequency decoupling and modulating with the aid of convolution layers, resulting in fewer parameters and complexities. Hence, MCSF can be embedded in multiple positions. Note that MDSF operates in a convolution style, therefore, cooperating with MDSF, MCSF does not need to exchange information between adjacent windows.

3.2.1 INTEGRATION MANNER AND LOSS FUNCTION

We utilize a multi-input multi-output (Cho et al., 2021) U-shaped backbone to ease the difficulty of training. The network has 6 residual groups, as shown in Fig. 1a. Each group has 16 residual blocks. MCSF and MDSF are integrated into the residual block successively (Fig. 1c). We put MCSF in all residual blocks and MDSF only in the last residual block of each group, and in remaining blocks MDSF is replaced by identity operation.

We adopt L1 loss in the spatial and frequency domain for training,

$$L_{spatial} = \sum_{r=1}^{3} \frac{1}{S_r} \|O_r - y_r\|_1; \quad L_{frequency} = \sum_{r=1}^{3} \frac{1}{S_r} \|\mathcal{F}(O_r) - \mathcal{F}(y_r)\|_1 \tag{6}$$

where $\mathcal{F}$ denotes Fourier transform, S_r is the number of total elements for normalization, and y is the ground truth with the corresponding resolution. The final loss function is given by $L = L_{spatial} + \lambda L_{frequency}$, where λ is set as 0.1.

4 EXPERIMENTAL RESULTS

4.1 SETTINGS

We evaluate the proposed SFNet on five image restoration tasks: image motion/defocus deblurring, image deraining, image dehazing and image desnowing. More details of all used datasets and specific training setting for each task are provided in Appendix A. In tables, the best and second-best scores are **bold** and underlined. FLOPs is tested on the size of 256×256.

We train separate models for different tasks. Unless mentioned otherwise, the following parameters are adopted. The batch size is set as 4 with patch size of 256×256. Each patch is randomly flipped

Table 1: Quantitative comparisons with previous leading single-image defocus deblurring methods on DPDD testset (Abuolaim & Brown, 2020), which contains 39 outdoor and 37 indoor scenes.

	Indoor Scenes				Outdoor Scenes				Combined			
Method	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓
EBDB	25.77	0.772	0.040	0.297	21.25	0.599	0.058	0.373	23.45	0.683	0.049	0.336
DMENet	25.50	0.788	0.038	0.298	21.43	0.644	0.063	0.397	23.41	0.714	0.051	0.349
JNB	26.73	0.828	0.031	0.273	21.10	0.608	0.064	0.355	23.84	0.715	0.048	0.315
DPDNet	26.54	0.816	0.031	0.239	22.25	0.682	0.056	0.313	24.34	0.747	0.044	0.277
KPAC	27.97	0.852	0.026	0.182	22.62	0.701	0.053	0.269	25.22	0.774	0.040	0.227
IFAN	28.11	0.861	0.026	0.179	22.76	0.720	0.052	0.254	25.37	0.789	0.039	0.217
DRBNet	-								25.72	0.791	-	0.183
Restormer	28.87	**0.882**	0.025	**0.145**	23.24	0.743	0.050	**0.209**	25.98	**0.811**	0.038	**0.178**
SFNet	**29.16**	0.878	**0.023**	0.168	**23.45**	**0.747**	**0.049**	0.244	**26.23**	**0.811**	**0.037**	0.207

Table 2: Image dehazing comparisons on the synthetic dehazing datasets: SOTS-Ourdoor and SOTS-Indoor (Li et al., 2018a).

	SOTS-Indoor		SOTS-Outdoor	
Method	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCP	16.62	0.818	19.13	0.815
DehazeNet	19.82	0.821	24.75	0.927
AOD-Net	20.51	0.816	24.14	0.920
GFN	22.30	0.880	21.55	0.844
GridDehazeNet	32.16	0.984	30.86	0.982
MSBDN	33.67	0.985	33.48	0.982
UHD	21.75	0.879	26.48	0.942
FFA-Net	36.39	0.989	33.57	0.984
ACER-Net	37.17	0.990	-	
DeHamer	36.63	0.988	35.18	0.986
MAXIM-2S	38.11	0.991	34.19	0.985
SHA-Net	38.41	-	-	
DehazeFormer-L	40.05	**0.996**	-	
SFNet (Ours)	**41.24**	**0.996**	**40.05**	**0.997**

Table 3: Image motion deblurring results on GoPro (Nah et al., 2017) and HIDE (Shen et al., 2020) datasets.

	GoPro		HIDE	
Method	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DeblurGAN	28.70	0.858	24.51	0.871
DeblurGAN-v2	29.55	0.934	26.61	0.875
SRN	30.26	0.934	28.36	0.915
DBGAN	31.10	0.942	28.94	0.915
MT-RNN	31.15	0.945	29.15	0.918
DMPHN	31.20	0.940	29.09	0.924
SPAIR	32.06	0.953	30.29	0.931
MIMO-UNet+	32.45	0.957	29.99	0.930
IPT	32.52	-	-	
MPRNet	32.66	0.959	30.96	0.939
HINet	32.71	0.959	30.32	0.932
Restormer	32.92	0.961	**31.22**	**0.942**
Stripformer	33.08	0.962	31.03	0.940
SFNet (Ours)	**33.27**	**0.963**	31.10	0.941

horizontally with a probability of 0.5. The initial learning rate is $1e^{-4}$ and gradually reduced to $1e^{-6}$ with the cosine annealing (Loshchilov & Hutter, 2016). Adam serves the optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$). We use PyTorch (Paszke et al., 2017) to implement our models on NVIDIA Tesla V100 GPU and Intel Xeon Platinum 8255C CPU.

4.2 Single-image defocus deblurring results

Tab. 1 shows comparisons of defocus deblurring results with previous leading methods on DPDD dataset (Abuolaim & Brown, 2020). Our SFNet surpasses state-of-the-art on the most metrics. Particularly on the combined scene item, SFNet obtains 0.25 dB PSNR improvement compared with the strong Transformer method Restormer (Zamir et al., 2022). In addition, our method provides a significant gain of 0.51 dB PSNR accuracy over the pre-trained network DRBNet (Ruan et al., 2022). The visual results in the first row of Fig. 5 show that our method recovers more details than other algorithms. Due to space limitations, results of dual-pixel setting is provided in Appendix E.

4.3 Image dehazing results

We perform dehazing experiments on the synthetic benchmark RESIDE (Li et al., 2018a) and real hazy dataset Dense-Haze (Ancuti et al., 2019). For RESIDE, we train our model for the indoor and outdoor scenarios separately, and then evaluate on the corresponding SOTS-Indoor and SOTS-outdoor testsets. The quantitative results are shown in Tab. 2. As presented, our method obtains the highest scores on all metrics. Particularly on the outdoor scene, our network generates a sub-

Table 4: Image dehazing results on Dense-Haze dataset (Ancuti et al., 2019).

Method	PSNR↑	SSIM↑
DehazeNet	9.48	0.438
AOD-Net	12.82	0.468
GridDehazeNet	14.96	0.533
FFA-Net	12.22	0.444
MSBDN	15.13	0.555
DeHamer	16.62	0.560
SFNet (Ours)	**17.46**	**0.578**

Table 5: Image motion deblurring results on RSBlur dataset (Jaesung Rim).

Method	PSNR↑	SSIM↑
SRN-DeblurNet	32.53	0.849
MIMO-UNet	32.73	0.846
MIMO-UNet+	33.37	0.856
MPRNet	33.61	0.861
Restormer	33.69	0.863
Uformer-B	33.98	0.866
SFNet (Ours)	**34.20**	**0.897**

Table 6: Image desnowing results on CSD (2000) dataset (Chen et al., 2021b).

Method	PSNR↑	SSIM↑
DesnowNet	20.13	0.81
JSTASR	27.96	0.88
HDCW-Net	29.06	0.91
TransWeather	31.76	0.93
MSP-Former	33.75	0.96
SnowFormer	38.26	0.98
SFNet (Ours)	**38.41**	**0.99**

Table 7: Deraining comparisons with previous methods on five deraining datasets: 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).

	Test100		Rain100H		Rain100L		Test2800		Test1200		Average	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	23.38	0.835	22.48	0.796
SEMI	22.35	0.788	16.56	0.486	25.03	0.842	24.43	0.782	26.05	0.822	22.88	0.744
DIDMDN	22.56	0.818	17.35	0.524	25.23	0.741	28.13	0.867	29.65	0.901	24.58	0.770
UMRL	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	30.55	0.910	28.02	0.880
RESCAN	25.00	0.835	26.36	0.786	29.80	0.881	31.29	0.904	30.51	0.882	28.59	0.857
PreNet	24.81	0.851	26.77	0.858	32.44	0.950	31.75	0.916	31.36	0.911	29.42	0.897
MSPFN	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903
MAXIM-2S	31.17	**0.922**	30.81	0.903	38.06	**0.977**	**33.80**	**0.943**	32.37	**0.922**	33.24	**0.933**
SFNet (Ours)	**31.47**	0.919	**31.90**	**0.908**	**38.21**	0.974	33.69	0.937	**32.55**	0.911	**33.56**	0.929

stantial gain of 4.87 dB PSNR over DeHamer (Guo et al., 2022). Compared with recent works, DehazeFormer-L (Song et al., 2022) and SHA-Net (Ye et al., 2021), our method receives 0.19 dB and 2.83 dB higher PSNR on SOTS-Indoor testset, respectively. Additionally, we validate the performance of our approach on the real hazy dataset Dense-Haze (Ancuti et al., 2019). The results shown in Tab. 4 demonstrate the superior ability of our method to deal with real-world dehazing problem. The dehazed results in the second line of Fig. 5 illustrate that SFNet is more effective in removing haze than other methods.

4.4 Image motion deblurring results

We report in Tab. 3 numerical comparisons on the GoPro (Nah et al., 2017) and HIDE (Shen et al., 2020) datasets. As we can see, on the GoPro dataset SFNet shows 0.35 dB PSNR performance improvement over Restormer with $\sim$3$\times$ faster speed, shown in Tab. 8. Moreover, our method receives a performance gain of 0.19 dB compared with the recent algorithm Stripformer (Tsai et al., 2022b) with fewer parameters (20 M vs. ours 13.27 M). Notably, our SFNet demonstrates stronger generalization capability to HIDE dataset than Stripformer on all metrics. Besides the synthetic datasets, we further evaluate the effectiveness of our network on the real-world dataset. Tab. 5 shows result comparisons on RSBlur (Jaesung Rim) dataset. SFNet sets new state-of-the-art on this dataset, providing a gain of 0.22 dB PSNR and 0.031 SSIM over previous best method Uformer-B (Wang et al., 2022). The third line in Fig. 5 shows that our method produces sharper results.

4.5 Image desnowing results

We compare our method on the CSD (Chen et al., 2021b) dataset with existing state-of-the-art methods (Chen et al., 2022c; Valanarasu et al., 2022; Chen et al., 2022a). As shown in Tab. 6, our framework yields 0.15 dB PSNR improvement over the concurrent Transformer model SnowFormer (Chen et al., 2022b). The visual results in the fourth row of Fig. 5 show that our method is more effective in removing snowflakes with various sizes. More outcomes on SRRS (Chen et al., 2020a) and Snow100K (Liu et al., 2018) datasets are provided in Appendix E.

4.6 IMAGE DERAINING RESULTS

Following recent works (Jiang et al., 2020; Purohit et al., 2021; Tu et al., 2022), we compare PSNR/SSIM scores on the Y channel in YCbCr color space. Tab. 7 shows that our method achieves the best performance on the average PSNR metrics compared with competing methods. Moreover, on the Test100 dataset (Zhang et al., 2019a), SFNet obtains an accuracy boost of 0.30 dB PSNR over MLP model MAXIM-2S (Tu et al., 2022) with fewer parameters (14.1 M vs. ours 13.27 M). Visual results shown in Fig. 5 (the last line) show the superiority of our method.

4.7 INFERENCE SPEED COMPARISONS

In Tab. 8, we evaluate the inference speed of five motion deblurring methods on GoPro testset (Nah et al., 2017). The inference time is tested on our equipment with the publicly released model and code. Our method achieves a better trade-off between performance and costs than other methods.

Table 8: Overall comparisons on GoPro testset.

Method	DBGAN	MPRNet	DeepRFT+	Restormer	SFNet
PSNR	31.10	32.66	33.23	32.92	**33.27**
Times (s)	1.447	1.148	0.806	1.218	**0.408**
Params (M)	**11.6**	20.1	23.0	26.13	13.27
FLOPs (B)	759.85	777.01	187.04	140.99	**125.43**

4.8 ABLATION STUDIES

We first show the effectiveness of the proposed modules, and then delve into each design. Unless otherwise mentioned, all models are trained on GoPro (Nah et al., 2017) dataset, and each residual group has 8 blocks in this section.

Influence of each module. The results reported in Tab. 9 show that MDSF and MCSF yield a PSNR gain of 0.29 and 0.26 dB with few increasing computing burdens. MDSF shows superiority than MCSF with the dynamic decoupling mechanism and advanced recalibrator. Furthermore, in Fig. 4, we plot the statistic differences between the ground truth and results of three methods on dehazing task. With the frequency selection mechanism, the statistics of our results are closer to ground truth. We provide statistic comparisons on more tasks in Appendix D.

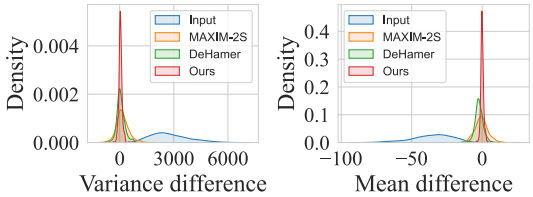

Figure 4: Variance (Left) and Mean (Right) difference between the ground truth and results of three methods on SOTS-Outdoor testset. Models are trained on OTS dataset (Li et al., 2018a).

Design choices for MCDF. We investigate the influence of the number of MCDF in Tab. 10. All models are trained for 1000 epochs. We increase the number of MCDF from back to front in each residual group, one MCDF for one residual block. As shown in the table, as we inject more MCDF, the performance raises up from 31.22 to 31.45 dB PSNR monotonously. Due to its few introduced parameters and complexities, we equip each residual block with our MCDF.

Design choices for MDSF. Towards understanding the impact of the number of groups in MDSF, we test various configurations in Tab. 11. All models are trained for 1900 epochs. Generally, the increasing number of groups leads to higher PSNR, demonstrating the effectiveness of the filter diversity. However, the accuracy saturates at group 16, which is caused by overfitting probably. We pick 8 groups for all tasks.

Table 9: Ablation studies for SFNet. Models are trained for 3000 epochs.

Method	PSNR	Params (M)	FLOPs (B)
Baseline	32.23	6.90	66.32
Baseline + MDSF	32.52	7.04	66.55
Baseline + MCSF	32.49	6.92	66.38
SFNet	32.68	7.05	66.61

Table 10: Ablation studies for MCSF.

Number	PSNR
2	31.22
4	31.33
6	31.42
8	31.45

Table 11: Ablation studies for MDSF.

Group	PSNR
2	32.09
4	32.10
8	32.15
16	32.13

Figure 5: From top to bottom are visual comparisons on DPDD defocus deblurring (Abuolaim & Brown, 2020),SOTS-Outdoor dehazing (Li et al., 2018a), GoPro motion deblurring (Nah et al., 2017), CSD desnowing (Chen et al., 2021c), and Test100 deraining (Zhang et al., 2019a) datasets.

Alternatives for decoupler. To examine the advantage of our design, we compare our decoupler with a few alternatives in Tab. 12. We borrow the idea from (Pang et al., 2020) to form *Conv* method (Tab. 12b), using stride convolution to generate different frequency parts. The *mean* item (Tab. 12e) utilizes average pooling with 3×3 kernel to obtain frequency parts without changing the resolution. More details of the architectures for *OctConv* (Tab. 12c) (Chen et al., 2020b) and *Wavelet* (Tab. 12d) are available in Appendix F. All methods are trained for 1500 epochs with the same setting. The results show that our MDSF obtains 0.3 dB higher PSNR than *OctConv* method. It is worth noting that mean filter, a local receptive filed version of MCSF, attains a higher score than the most other methods, demonstrating the validity of our MCSF. The common used method, Wavelet transform (Zou et al., 2021; Yang et al., 2020a), only gets a score of 31.50 dB PSNR, verifying our hypothesis of its rigid frequency separation nature.

Table 12: Comparisons between various frequency separation methods. Methods with † use our frequency recalibrator.

Method	PSNR	Params (M)	FLOPs (B)
(a) Gaussian†	31.51	7.02	66.81
(b) Conv†	31.58	8.65	79.35
(c) OctConv	31.46	6.99	66.86
(d) Wavelet†	31.50	7.06	66.33
(e) Mean†	31.52	7.02	66.75
(f) MDSF	**31.76**	7.04	66.55

5 Conclusion

We present an image restoration framework, SFNet, which is built on frequency selection mechanism. We develop two key modules, MDSF and MCSF, to conduct frequency decomposition and recalibration with different receptive fields. Specifically, our multi-branch dynamic selective frequency module (MDSF) builds a dynamic filter to decompose the feature map into various frequency parts and utilizes the modified channel attention to perform accentuation, thus effectively selecting the most informative frequency to recover. Furthermore, the proposed multi-branch compact selective frequency module (MCSF) introduces a simple yet effective manner to enlarge the receptive field and perform frequency selection. With two proposed modules, SFNet shows state-of-the-art results on five image restoration tasks, demonstrating the validity of our designs.

REFERENCES

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

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.

M. Cai, H. Zhang, H. Huang, Q. Geng, and G. Huang. Frequency domain image translation: More photo-realistic, better identity-preserving. 2020.

Daniela Calvetti, Lothar Reichel, and Qin Zhang. Iterative solution methods for large linear discrete ill-posed problems. In *Applied and computational control, signals, and circuits*, pp. 313–367. Springer, 1999.

H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao. Pre-trained image processing transformer. In *Computer Vision and Pattern Recognition*, 2021a.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 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*, 2021b.

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, 2020a.

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, 2021c.

Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2020b.

S. J. Cho, S. W. Ji, J. P. Hong, S. W. Jung, and S. J. Ko. Rethinking coarse-to-fine approach in single image deblurring. 2021.

Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2157–2167, 2020.

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.

Y. Gao, F. Wei, J. Bao, S. Gu, D. Chen, F. Wen, and Z. Lian. High-fidelity and arbitrary face editing. 2021.

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/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5812–5820, 2022.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *JMLR.org*, 2015.

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

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.

Ali Karaali and Claudio Rosito Jung. Edge-based defocus blur estimation with adaptive scale selection. *IEEE Transactions on Image Processing*, 27(3):1126–1137, 2017.

Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. *IEEE signal processing magazine*, 13(3):43–64, 1996.

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.

X. Li, W. Wang, X. Hu, and J. Yang. Selective kernel networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 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.

Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 510–519, 2019.

J. Liang, J. Cao, G. Sun, K. Zhang, and R. Timofte. Swinir: Image restoration using swin transformer. *IEEE*, 2021.

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.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

I. Loshchilov and F. Hutter. Sgdr: Stochastic gradient descent with warm restarts. 2016.

Pinjun Luo, Guoqiang Xiao, Xinbo Gao, and Song Wu. Lkd-net: Large kernel convolution network for single image dehazing. *arXiv preprint arXiv:2209.01788*, 2022.

W. Luo, Y. Li, R. Urtasun, and R. Zemel. Understanding the effective receptive field in deep convolutional neural networks. 2017.

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.

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.

Yingxue Pang, Xin Li, Xin Jin, Yaojun Wu, Jianzhao Liu, Sen Liu, and Zhibo Chen. Fan: Frequency aggregation network for real image super-resolution. In *European Conference on Computer Vision*, pp. 468–483. Springer, 2020.

A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. Devito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. 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.

Kuldeep Purohit, Maitreya Suin, AN Rajagopalan, and Vishnu Naresh Boddeti. Spatially-adaptive image restoration using distortion-guided networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2309–2319, 2021.

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.

Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

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.

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.

Ivan W Selesnick, Richard G Baraniuk, and Nick C Kingsbury. The dual-tree complex wavelet transform. *IEEE signal processing magazine*, 22(6):123–151, 2005.

M Ibrahim Sezan and A Murat Tekalp. Survey of recent developments in digital image restoration. *Optical Engineering*, 29(5):393–404, 1990.

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.

D. Song, Y. Wang, H. Chen, C. Xu, and D. C. Tao. Addersr: Towards energy efficient image super-resolution. 2020.

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.

G. Tao, X. Ji, W. Wang, S. Chen, C. Lin, Y. Cao, T. Lu, D. Luo, and Y. Tai. Spectrum-to-kernel translation for accurate blind image super-resolution. 2021.

F. J. Tsai, Y. T. Peng, Y. Y. Lin, C. C. Tsai, and C. W. Lin. Stripformer: Strip transformer for fast image deblurring. 2022a.

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*, 2022b.

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.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *arXiv*, 2017.

Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

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, 2022.

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.

Hao-Hsiang Yang and Yanwei Fu. Wavelet u-net and the chromatic adaptation transform for single image dehazing. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 2736–2740. IEEE, 2019.

Hao-Hsiang Yang, Chao-Han Huck Yang, and Yi-Chang James Tsai. Y-net: Multi-scale feature aggregation network with wavelet structure similarity loss function for single image dehazing. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2628–2632. IEEE, 2020a.

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.

Wenhan Yang, Robby T Tan, Shiqi Wang, Yuming Fang, and Jiaying Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on pattern analysis and machine intelligence*, 43(11):4059–4077, 2020b.

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.

Tian Ye, Mingchao Jiang, Yunchen Zhang, Liang Chen, Erkang Chen, Pen Chen, and Zhiyong Lu. Perceiving and modeling density is all you need for image dehazing. *arXiv preprint arXiv:2111.09733*, 2021.

Yuan Yuan, Wei Su, and Dandan Ma. Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3555–3564, 2020.

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. Multi-scale single image dehazing using perceptual pyramid deep network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018.

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.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26 (7):3142–3155, 2017.

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.

Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li. Deep image deblurring: A survey. *International Journal of Computer Vision*, 130 (9):2103–2130, 2022a.

Yanni Zhang, Qiang Li, Miao Qi, Di Liu, Jun Kong, and jianzhong Wang. Multi-scale frequency separation network for image deblurring. *arXiv preprint arXiv:2206.00798*, 2022b.

Wenbin Zou, Mingchao Jiang, Yunchen Zhang, Liang Chen, Zhiyong Lu, and Yi Wu. Sdwnet: A straight dilated network with wavelet transformation for image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1895–1904, 2021.

The appendix is organized as follows:

- Appendix A introduces the details of the 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 tasks.
- Appendix F details the architecture of methods in Tab. 12.
- Appendix G provides additional ablation studies for MDSF.
- Appendix H exhibits visual comparisons for five image restoration tasks.

A Datasets and Experimental Details

We summarize the training and testing datasets for our experiments in Tab. 13. We describe the individual datasets and training details for each task as follows:

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

Defocus Deblurring. We utilize a newly developed 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, and for each scene it has four images, labeled as right, left, center view and the all-in-focus ground truth. For this task, there are two patterns for training. 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 the images from left and right views as the training samples. The DPDD dataset is split into training, validation and testing set with 350, 74 and 76 scenes. However, 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 using the advantages of 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 the recent work DRBNet (Ruan et al., 2022). Specifically, SFNet is initially trained on LFDOF dataset for 50 epochs and then fine-tuned on the DPDD dataset for 20 epochs. The initial learning rate is set as $1e^{-4}$ and $1e^{-5}$, respectively. Then we evaluate the fine-tuned network on DPDD dataset for comparisons with other methods. Regarding the configuration of dual-pixel deblurring task, we utilize DPDD dataset for training and testing. The training process lasts for 200 epochs. Due to the dual-pixel input images, we alter the convolution layers related to the input from 3-channel to 6-channel. In addition, we use convolution layer to adjust the channel dimension for skip connection 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, as described in Tab. 13. The trained network is directly 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.

Image Dehazing. Following (Song et al., 2022), we train SFNet on the widely used dataset RESIDE (Li et al., 2018a). It consists of two training subsets, indoor training set (ITS) and outdoor training set (OTS), and synthetic objective testing set (SOTS). We train our model on ITS and OTS separately, and validate on the corresponding testset. SFNet is trained for 300 epochs on ITS and 30 for

Table 13: Dataset details for diverse image restoration tasks.

Task	Dataset	Train Sample	Test Sample	Testset Rename
Motion Deblurring	GoPro	2103	1111	
	HIDE	0	2025	
	RSBlur	8878	3360	
Defocus Deblurring	DPDD	350	76	
	LFDOF	11261	725	
Deraining	Rain14000	11200	2800	Test2800
	Rain1800	1800	0	
	Rain800	700	100	Test100
	Rain100H	0	100	Rain100H
	Rain100L	0	100	Rain100L
	Rain1200	0	1200	Test1200
	Rain12	12	0	
Dehazing	RESIDE/ITS	13990	500	SOTS/Indoor
	RESIDE/OTS	313950	500	SOTS/Outdoor
	Dense-Haze	45	5	
Desnowing	CSD	8000	2000	CSD (2000)
	SRRS	15005	15005	SRRS (2000)
	Snow100K	50000	50000	Snow100K (2000)

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 patches of 512×512 size.

Image Desnowing. For this task, we use CSD (Chen et al., 2021c), SRRS (Chen et al., 2020a) and Snow100K (Liu et al., 2018) datasets to train and test our network. The training and testing dataset settings follow works (Chen et al., 2020a; 2022c), where we randomly sample 2500 image pairs from the whole training set for training, and choose 2000 images on each testing set for evaluation. The model is trained for 800 epochs on separate dataset.

B Proof of the low-pass filter

In this part, we firstly provide the formal computational process of *low*- and *high*-frequency signals in a feature map, and then theoretically prove that the resultant filter after softmax operation in Sec. 3.1 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 can be given by,

$$F_n = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix} \tag{7}$$

where n is the length of the input signal, $\omega = e^{-j\frac{2\pi}{n}}$ (the basis for this choice is provided in Appendix C). Accordingly, the matrix for inverse discrete Fourier transform can be described as $F_n^{-1} = \frac{1}{n}\overline{F_n}$. In Sec. 3.2, we consider the *low* frequency as the lowest frequency. Thus we can

apply F_n and F_n^{-1} successively to obtain the *low*-frequency operator by,

$$\begin{aligned}\mathcal{L}[X] &= F_n^{-1} diag(1,0,0,...0) F_n X \\ &= \frac{1}{n} M X\end{aligned} \tag{8}$$

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

$$\begin{aligned}\mathcal{H}[X] &= F_n^{-1} diag(0,1,1,...1) F_n X \\ &= F_n^{-1}(E - diag(1,0,0,...0)) F_n X \\ &= (E - \frac{1}{n} M) X\end{aligned} \tag{9}$$

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 re-weighted pixel values in the region. Then this operation is used iteratively on the whole spatial domain 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 operators for both low- and high-frequency subbands, we prove the validity of the low-pass filter.

Theorem 1. Given $W \in \mathbb{R}^{n \times n}$, $W_i = softmax(D_i)$, $i = 0, 1, ..., n-1$, where $D \in \mathbb{R}^{n \times n}$. 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 can be written as,

$$W = PJP^{-1} = [a_1 \ \ a_2 \ \ ... \ \ a_n] \begin{bmatrix} \lambda_1 & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{bmatrix} \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_n^T \end{bmatrix} \tag{10}$$

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,

$$W^t m = PJ^t P^{-1} m = P \begin{bmatrix} \lambda_1^t & & & \\ & J_2(\lambda_2)^t & & \\ & & \ddots & \\ & & & J_s(\lambda_s)^t \end{bmatrix} P^{-1} m \tag{11}$$

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,

$$J_i(\lambda_i) = \begin{bmatrix} \lambda_i^t & (\lambda_i^t)' & \frac{(\lambda_i^t)''}{2!} & \cdots & \frac{(\lambda_i^t)^{(z_i-1)}}{(z_i-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{(\lambda_i^t)''}{2!} \\ & & & \ddots & (\lambda_i^t)' \\ & & & & \lambda_i^t \end{bmatrix}_{z_i} \tag{12}$$

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

$$\lim_{t\to\infty} \frac{t(t-1)...(t-q+1)}{q!}\lambda_i^{t-q} = \lim_{t\to\infty} \frac{t(t-1)...(t-q+1)}{q!\lambda_i^q(\frac{1}{\lambda_i})^t} = 0 \tag{13}$$

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 get 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. 11, the limit of $W^t m$ is,

$$\begin{aligned} \lim_{t\to\infty} W^t m &= Pdiag(\lambda_1^t, 0, ..., 0)P^{-1}m \\ &= \lambda_1^t a_1 b_1^T m \\ &= e b_1^T m \end{aligned} \tag{14}$$

The limit in the theorem 1 can be computed as,

$$\begin{aligned} \lim_{t\to\infty} \frac{\|\mathcal{H}[W^t m]\|_2}{\|W^t m\|_2} &= \lim_{t\to\infty} \frac{\|(E - \frac{1}{n}M)W^t m\|_2}{\|W^t m\|_2} \\ &= \lim_{t\to\infty} \frac{\|(E - \frac{1}{n}M)eb_1^T m\|_2}{\|W^t m\|_2} \\ &= \lim_{t\to\infty} \frac{\|Eeb_1^T m - \frac{1}{n}Meb_1^T m)\|_2}{\|W^t m\|_2} \\ &= \lim_{t\to\infty} \frac{\|Eeb_1^T m - eb_1^T m)\|_2}{\|W^t m\|_2} \\ &= 0 \end{aligned} \tag{15}$$

□

C ANALYSIS FOR MCSF

In this part, we analyze the validity of our MCSF, where we use the average pooling in decoupler to decompose the feature into different frequency parts. It is known that the low-frequency components of a degraded image and its sharp counterpart are similar (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 to operate frequency separation. We consider that the qualified threshold should satisfy the requirement: 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 opposite counterpart as the *low* and *high* frequency, respectively. Next, we provide the analysis 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.

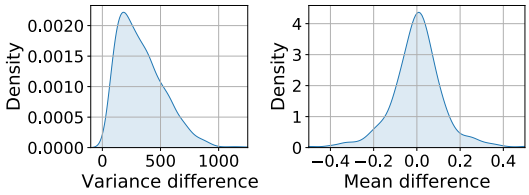

Figure 6: The probability density curve of the difference between blurry/sharp image pairs in the GoPro (Nah et al., 2017) training set.

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 discrepancies between sharp/blurry image pairs by computing the mean difference in Fig. 6 (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 image. The variance differences between image pairs are given in Fig. 6 (Left) and we can obtain the variance relationships for the most of image pairs,

$$\frac{\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}(x_{h,w}^{SH})^2}{HW} > \frac{\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}(x_{h,w}^{BH})^2}{HW} \tag{16}$$

where $x_{h,w}^{SH}$ and $x_{h,w}^{BH}$ indicate so-called high-frequency maps of sharp and blurry image, respectively. As $x_{h,w}^{SH}$ and $x_{h,w}^{BH}$ are computed by subtracting f_0, and thus they become zeros-centered. Eq. 16 actually shows the variance relationship between the high-frequency feature of image pair.

Based on Eq. 16 and the amplitude differences between x-axes of mean and variance (Fig. 6), 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. 7, 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 in each task, demonstrating that our MDCF is reasonable.

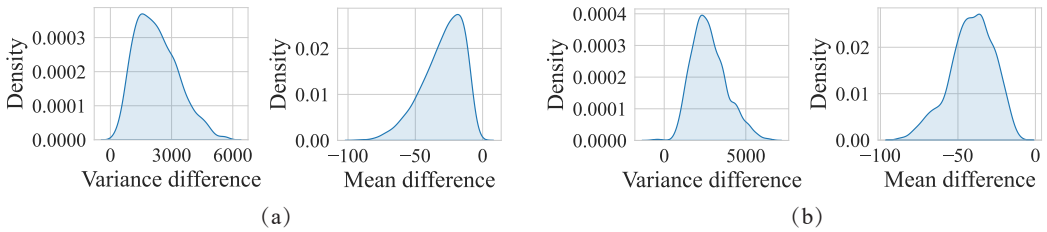

Figure 7: The probability density curves of the differences between degraded/sharp image pairs in the **(a)** RESIDE-Outdoor (Li et al., 2018a), **(b)** CSD (Chen et al., 2021c) datasets.

D STATISTIC ANALYSIS OF PRODUCED IMAGES

In this section, we verify the effectiveness of our method by comparing the data distributions of degraded input and outcomes of several methods. We compute the variance and mean difference in Fig. 8. For the desnowing task on CSD (Chen et al., 2021c) testset (Fig. 8a), our method obtains the resultant images closer to the ground truth in terms of variance and mean than Wavelet transform based method HDCW-Net (Chen et al., 2021b). In Fig. 8b, we compare SFNet with DRBNet (Ruan et al., 2022) on the single-image defocus deblurring task. DRBNet narrows down the distribution of variance difference compared with the degraded input, while its center line shifts to the right. By contrast, our result is closer to the ground truth (y-axis).

E ADDITIONAL EXPERIMENTS

In Tab. 14, we show more desnowing results on two other datasets, SRRS (Chen et al., 2020a) and Snow100K (Liu et al., 2018). Our method shows superiority over other methods. Specifically, compared with the recent algorithm MSP-Former (Chen et al., 2022c), SFNet achieves a significant gain of 1.64 dB PSNR on the SRRS dataset. Furthermore, we provide experimental results on DPDD (Abuolaim & Brown, 2020) dataset for dual-pixel defocus deblurring. As reported in Tab. 15, our method outperforms DRBNet (Ruan et al., 2022), obtaining 0.01 dB higher PSNR than DRBNet (Ruan et al., 2022), which needs additional central view dataset for training and testing.

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

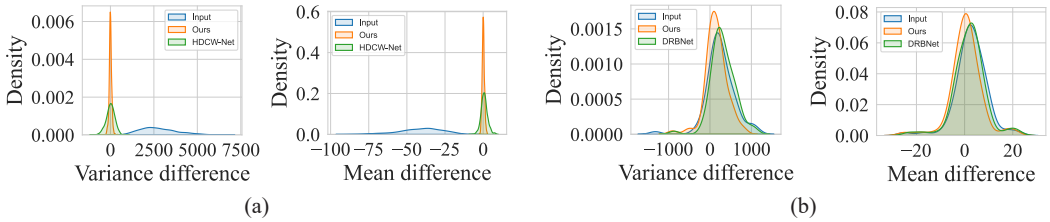

Figure 8: The statistic differences between the ground truth and results of several methods. **(a)** Comparisons between HDCW-Net (Chen et al., 2021b) and SFNet on CSD (Chen et al., 2021c) testset. **(b)** Comparisons between DRBNet (Ruan et al., 2022) and SFNet on DPDD (Abuolaim & Brown, 2020) testset.

Table 14: Desnowing results on SRRS (Chen et al., 2020a) and Snow100K (Liu et al., 2018) datasets.

	SRRS (2000)		Snow100K (2000)	
Method	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DesnowNet	20.38	0.84	30.50	0.94
CycleGAN	20.21	0.74	26.81	0.89
All in One	24.98	0.88	26.07	0.88
JSTASR	25.82	0.89	23.12	0.86
HDCW-Net	27.78	0.92	31.54	0.95
TransWeather	28.29	0.92	31.82	0.95
NAFNet	29.72	0.94	32.41	0.95
MSP-Former	30.76	0.95	33.43	**0.96**
SFNet	**32.40**	**0.98**	**33.79**	0.95

F MORE DETAILS ABOUT ALTERNATIVES FOR DECOUPLER

We detail the architectural designs of the OctConv (Chen et al., 2020b) and Wavelet transform based decoupler. In Fig. 9a, the left side shows the process of generating the high- and low-frequency components from a single feature map, where we set $\alpha_{in} = 0$ (refer to (Chen et al., 2020b) for more details of this parameter). The right side depicts the inverse operation, from high- and low-frequency parts to a single feature map by setting $\alpha_{out} = 0$. Specifically, given a single feature map, we use the average pooling to generate the low-frequency component in low resolution. After information updates via two convolution layers, the final feature map is obtained by adding the updated high-frequency map and the upsampled low-frequency map.

Regarding the implementation of the Wavelet transform based alternative (Fig. 9b), discrete wavelet transform is utilized to generate four subbands. Next, our recalibrator in MDSF is used for a comparison fairness. Finally, the enhanced feature map is attained by the inverse discrete Wavelet transform.

Table 15: Dual-pixel defocus deblurring results on DPDD testset (Abuolaim & Brown, 2020).

Method	PSNR↑	SSIM↑	MAE↓	LPIPS↓
DPDNet	25.13	0.786	0.041	0.223
RDPD	25.39	0.772	0.040	0.255
Uformer	25.65	0.795	0.039	0.243
IFAN	25.99	0.804	**0.037**	0.207
DRBNet	26.33	0.811	-	**0.154**
SFNet	**26.34**	**0.817**	**0.037**	0.207

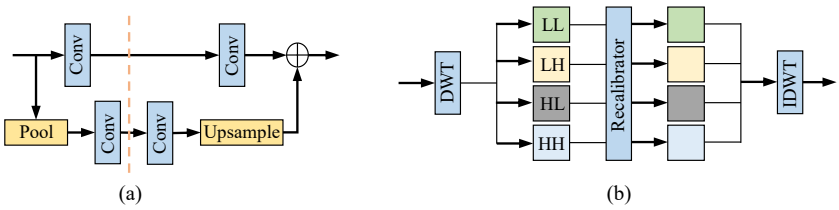

Figure 9: Diagram of **(a)** OctConv (Chen et al., 2020b) and **(b)** Wavelet transform based method.

G ADDITIONAL ABLATION STUDIES FOR MDSF

We investigate the effectiveness of multiple branches in MDSF, which provide the multiple local receptive fields. As shown in the first set of Tab. 16, increasing the number of branches leads to the better performance, demonstrating the utility of the enlarged receptive field. Next, we compare our channel-wise attention mechanism in recalibrator with that in SKNet (Li et al., 2019) as reported in Tab. 16a,b. Generating the attention weights from the global channel dimension, the network obtains a gain of 0.04 dB PSNR. We further inject the proposed channel-wise attention into MCSF to modulate the decomposed frequency components, replacing its simple recalibrator. The results (Tab. 16e,f) show that our method boosts the performance of MCSF from 32.22 to 32.39 dB.

Table 16: Different choices in MDSF. Variant b,c and d describe the network with different combinations of kernel size in decoupler. Variant f means we substitute the channel-wise attention in MDSF recalibrator for that of MCSF, and MCSF is used in the last two blocks in each residual block group.

Method	PSNR
a Branch-3-SKNet	33.42
b Branch-3	33.46
c Branch-3-5	33.52
d Branch-3-5-7	33.53
e MCSF-L2	32.22
f MCSF-L2 w/ MDCF-Recalibrator	32.39

H ADDITIONAL VISUAL RESULTS

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

- Image motion deblurring: Fig. 10, 11, 12.
- Image deraining: Fig. 13, 14, 15, 16.
- Image defocus deblurring: Fig. 17.
- Image dehazing: Fig. 18.
- Image desnowing: Fig. 19.

Figure 10: **Image motion deblurring** comparisons on the GoPro dataset (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., 2022a), and SFNet.

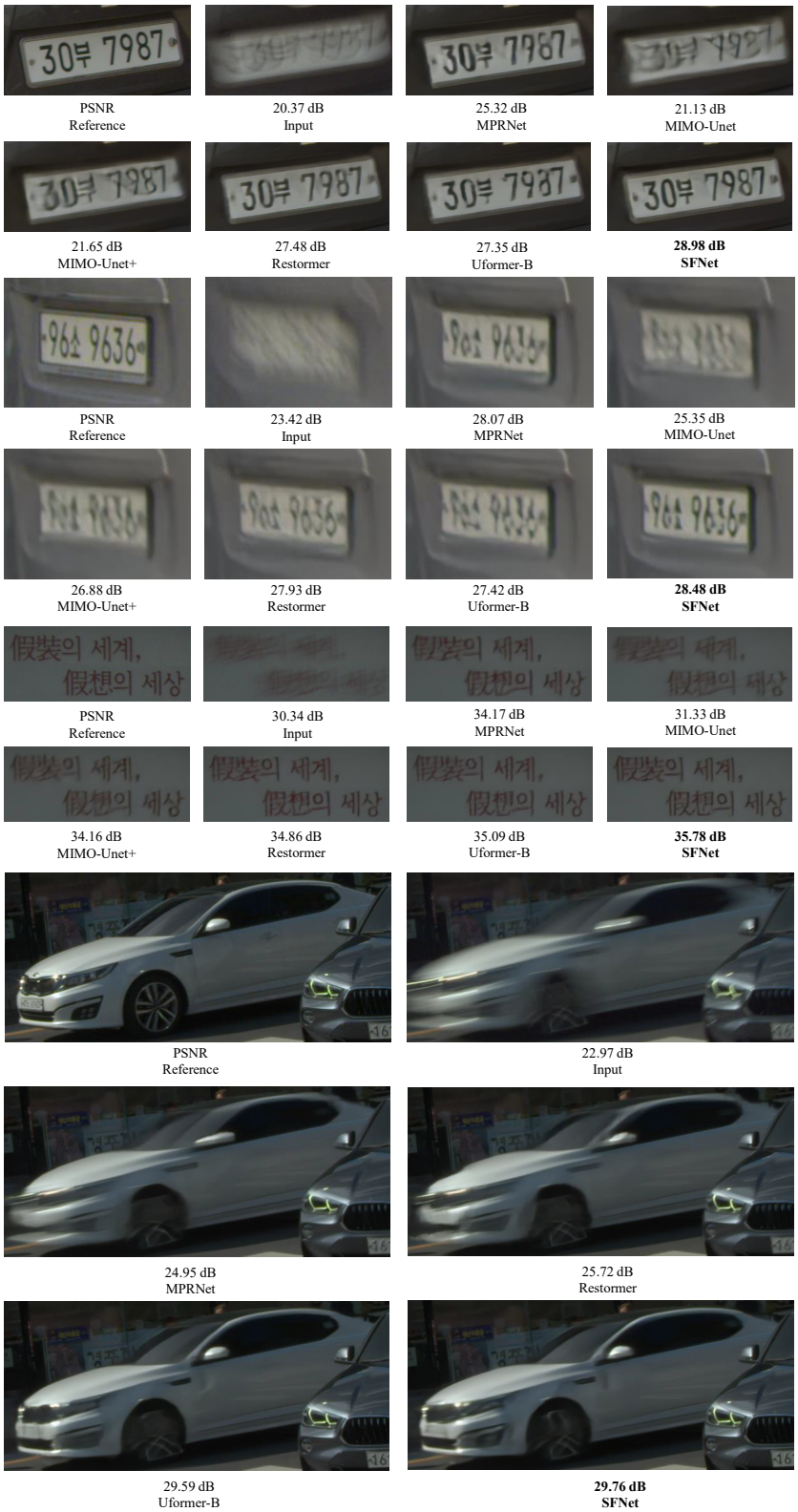

Figure 11: **Image motion deblurring** comparisons on the RSBlur dataset (Jaesung Rim)among MPRNet (Zamir et al., 2021), MIMO-UNet (Cho et al., 2021), MIMO-UNet+ (Cho et al., 2021), Restormer (Zamir et al., 2022), Uformer-B (Wang et al., 2022), and SFNet.

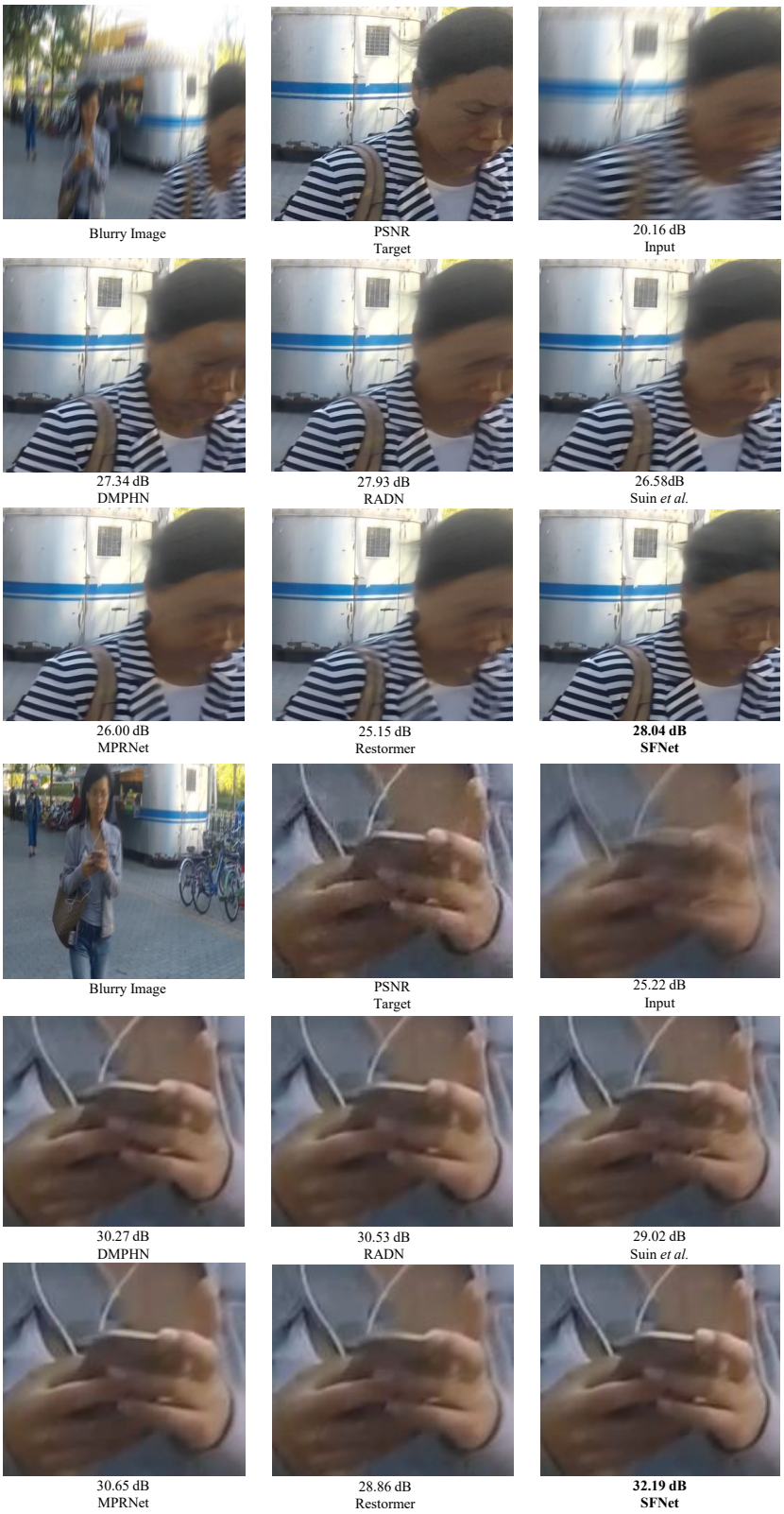

Figure 12: **Image motion deblurring** comparisons on the HIDE dataset (Shen et al., 2020) among DMPHN (Zhang et al., 2019b), RADN (Purohit & Rajagopalan, 2020), Suin *et al.* (Suin et al., 2020), MPRNet (Zamir et al., 2021), Restormer (Zamir et al., 2022), and SFNet.

Figure 13: **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 14: **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 15: **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.

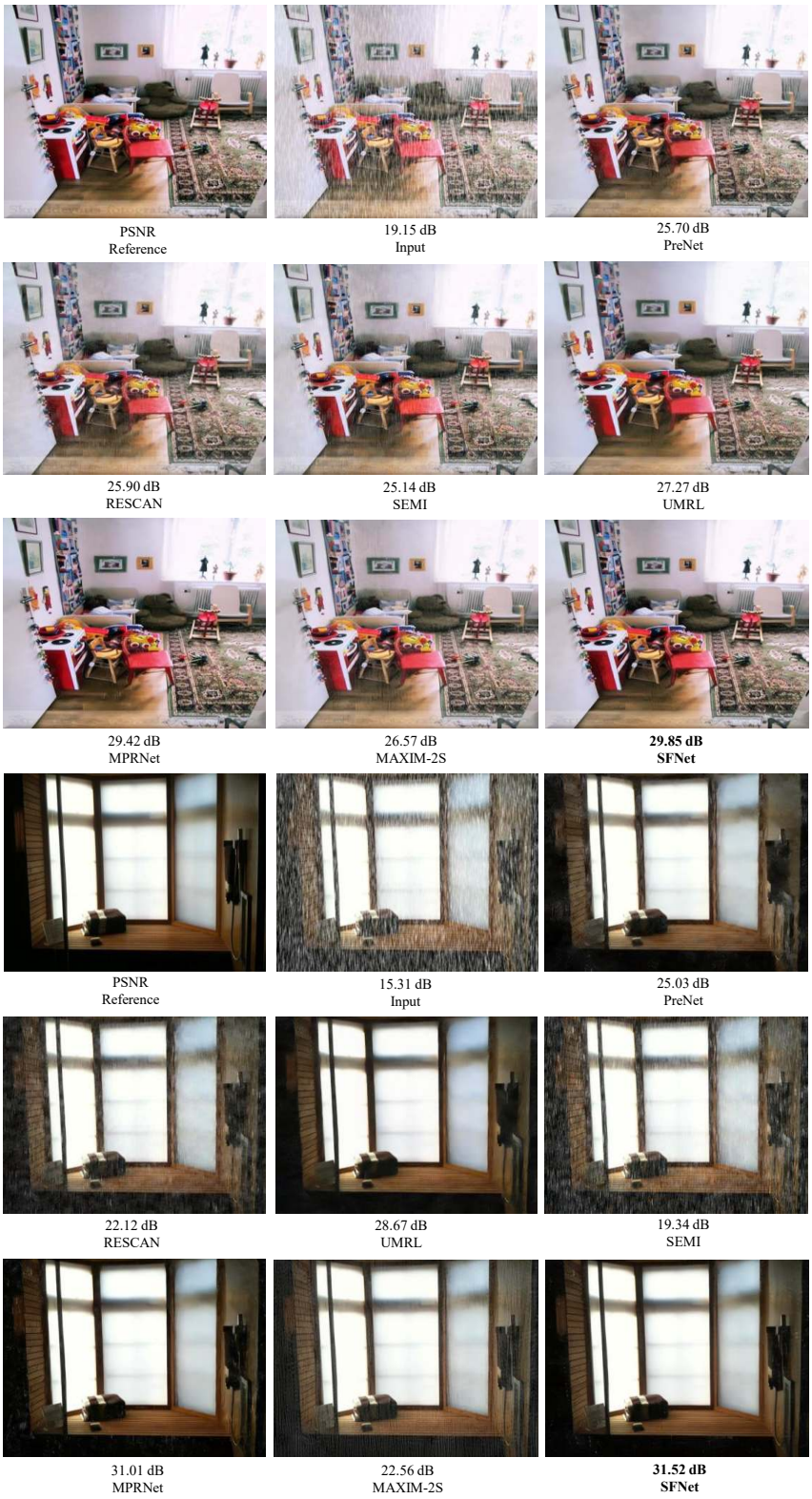

Figure 16: **Image deraining** comparisons on the Test1200 dataset (Zhang & Patel, 2018b) among PreNet (Ren et al., 2019), RESCAN (Li et al., 2018b), SEMI (Wei et al., 2019), UMRL (Yasarla & Patel, 2019), MPRNet (Zamir et al., 2021), MIXIM-2S (Tu et al., 2022), and SFNet.

Figure 17: **Image defocus deblurring** comparisons on the DPDD dataset (Abuolaim & Brown, 2020) among KPAC (Son et al., 2021), DRBNet (Ruan et al., 2022), Restormer(Zamir et al., 2022), and SFNet.

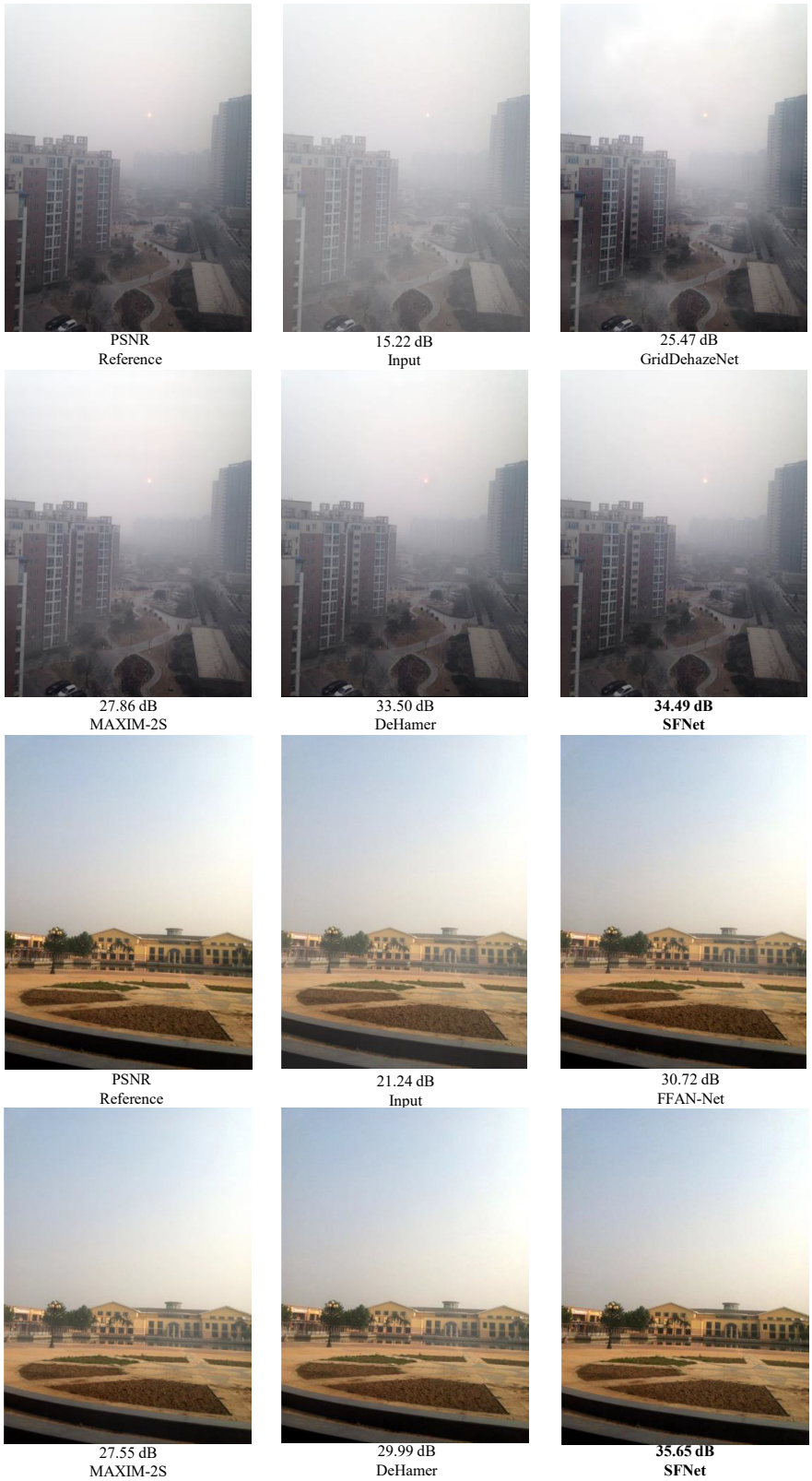

Figure 18: **Image dehazing** comparisons on the SOTS-Outdoor dataset (Li et al., 2018a) among GridDehazeNet (Liu et al., 2019), FFA-Net (Qin et al., 2020), MAXIM-2S (Tu et al., 2022), DeHamer (Song et al., 2022), and SFNet.

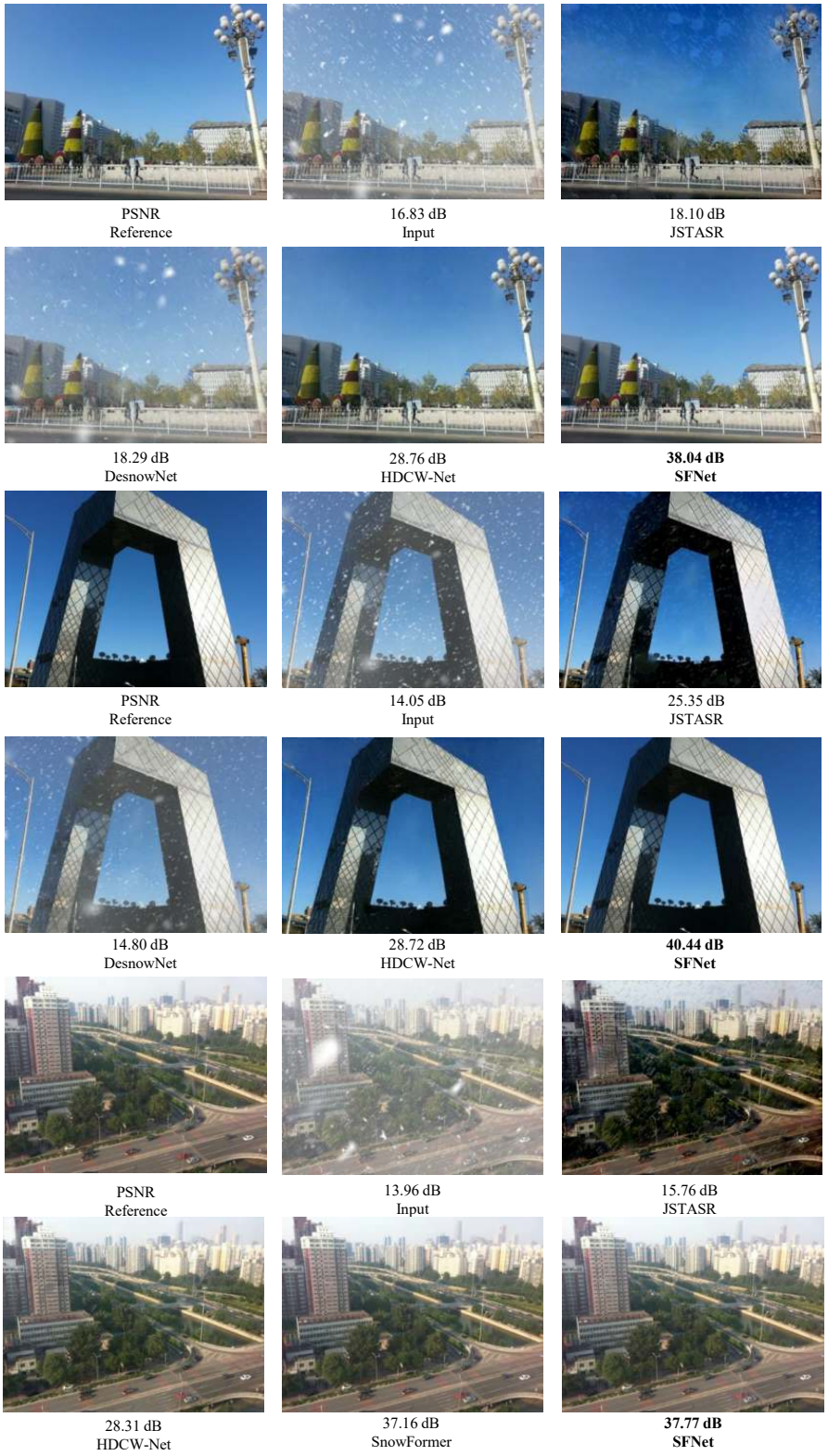

Figure 19: **Image desnowing** comparisons on the CSD dataset (Chen et al., 2021c) among JSTASR (Chen et al., 2020a), HDCW-Net (Chen et al., 2021b), DesnowNet (Liu et al., 2018), SnowFormer (Chen et al., 2022b), and SFNet.