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 via 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 single-image defocus deblurring, image dehazing, image motion deblurring, image desnowing, and image deraining.

1 INTRODUCTION

Image restoration aims to recover a high-quality image by removing degradations, *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 address this problem based on various assumption (Zhang et al., 2022; Yang et al., 2020b) or hand-crafted features (Karaali & Jung, 2017), which are incapable of generating faithful results in real-world scenarios.

Recently, deep neural networks have witnessed the rapid development of image restoration and obtained favorable performance compared to conventional 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 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 frequency discrepancies between the sharp/degraded image pairs. To this end, a few works utilize the existing transformation tools, *e.g.,* Wavelet transform or Fourier transform, to decompose the feature into various frequency components and then treat the separate parts individually to reconstruct the corresponding feature (Selesnick et al., 2005; Yang & Fu, 2019; Zou et al., 2021; Mao et al., 2021). Nevertheless, Wavelet transform decouples 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. In addition, these methods need corresponding inverse Fourier/Wavelet transform, leading to the 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 filters 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 resulting frequency components. Our module has two key advantages. First, according to the input and task, the decoupling step dynamically generates the filter to decompose the feature map. Second, our module does not introduce 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 above dynamic module (MDSF) that processes features 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 the global and window-based average pooling to attain disparate frequency maps, and then use learnable parameters to modulate the resulting maps without resorting to any convolution layers. Compared to MDSF, besides the enlarged receptive fields, MCSF is lightweight enough to be embedded into the backbone in multiple positions. Our contributions are summarized as follows:

- We propose a multi-branch dynamic selective frequency module (MDSF) that is capable of decoupling the feature map into different frequency components dynamically via the theoretically proved filters, and then selecting the most informative components to recover.
- We design a multi-branch compact selective frequency module (MCSF) that performs frequency decoupling and recalibration based on multi-scale average pooling operations to pursue a large receptive field for the spatially-varying degradation kernels.
- Incorporating MDSF and MCSF into a U-shaped backbone, a selective frequency network (SFNet) is formed for image restoration. SFNet achieves state-of-the-art results on five image restoration tasks: defocus/motion deblurring, dehazing, deraining, and desnowing.

2 Related Work

Image Restoration. Prior to the deep learning era, a great number of methods have been proposed for image restoration problems based on various assumptions and hand-crafted features (Sezan & Tekalp, 1990; Kundur & Hatzinakos, 1996; Calvetti et al., 1999). In recent years, with the rapid development of deep learning, a flurry of approaches have been investigated utilizing convolutional neural networks for 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).

More recently, to capture the long-range dependencies, many works have borrowed the Transformer model (Vaswani et al., 2017) from natural language processing field into image restoration (Chen et al., 2021a; Liang et al., 2021; Zamir et al., 2022; Wang et al., 2022) and specific tasks such as 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 the more advanced backbone for image restoration, we pay more attention to the frequency selection mechanism based on CNN, which is more efficient than Transformer type network.

Frequency Based Image Restoration. Many algorithms have been developed to address various low-level vision problems from a frequency perspective. Specifically, Xu *et al.* (Chen et al., 2021c) propose a hierarchical desnowing network based on dual-tree complex wavelet representation (Selesnick et al., 2005). Yang *et al.* (Yang & Fu, 2019) develop the Wavelet based U-Net to replace up-sampling and down-sampling. Zou *et al.* (Zou et al., 2021) utilize Wavelet transform based module to help recover the texture details. Yang *et al.* (Yang et al., 2020a) devise a Wavelet structure similarity loss function for training. Mao *et al.* (Mao et al., 2021) use Fourier transform to integrate both high- and low-frequency residual information. Yoo *et al.* (Yoo et al., 2018) complete image restoration based on the estimation of the DCT coefficient distribution. In this work, we pursue a dynamic and efficient manner to select the useful frequency part at multiple receptive fields.

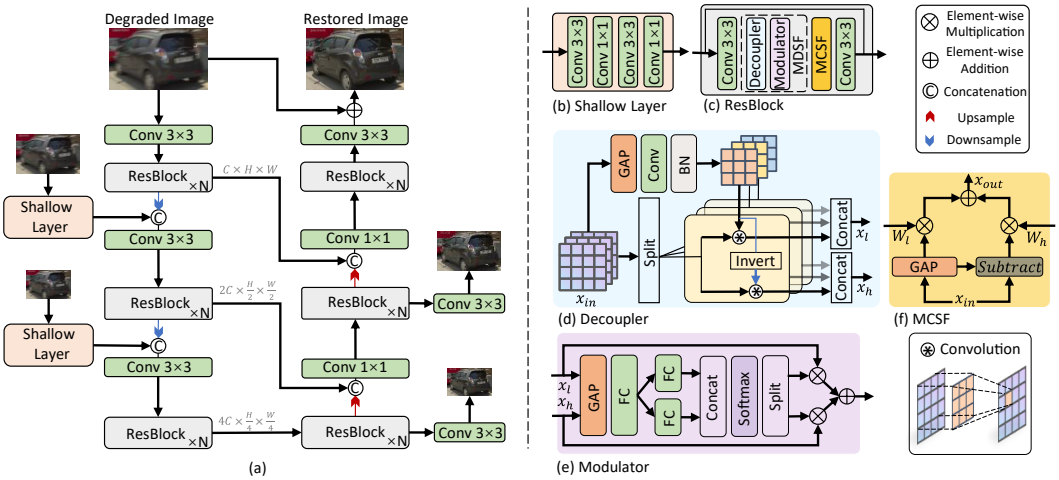


Figure 1: (a) Overall framework of the proposed SFNet. (b) Shallow layer extracting the shallow feature for the low-resolution images. (c) ResBlock containing our proposed modules: MDSF (Decoupler (d) and Modulator (e)) and MCSF (f). Two proposed modules are shown in the one-branch case for clarity. *Invert* depicts the operation of subtracting the low-pass filter from the identity filter.

3 METHODOLOGY

We first describe the overall pipeline of our SFNet (Fig. 1 (a)). Then we present the proposed modules, MDSF (Fig. 1 (d,e)) and MCSF (Fig. 1 (f)). The loss function follows in the final part.

3.1 OVERALL PIPELINE

Our network applies the popular encoder-decoder architecture to learn hierarchical representations. Specifically, SFNet consists of three decoders and three encoders, each containing multiple ResBlocks (Fig. 1 (c)). Our modules are inserted into the ResBlock successively where MDSF is only deployed in the last ResBlock of each en/decoder while MCSF exists in all blocks. Following (Cho et al., 2021; Mao et al., 2021; Tu et al., 2022), multi-input and multi-output mechanisms are used to ease the difficulty of training. To be specific, the images with reduced sizes are merged into the main path via the shallow layer (Fig. 1 (b)) and the predicted images are produced by 3×3 convolutional layers. Besides, we adopt the feature-level and image-level skip connections to assist training. In Fig. 1 we only show the top-level image skip connection for clarity. The up-sampling and down-sampling layers are implemented by transpose and stride convolutions, respectively.

3.2 MULTI-BRANCH DYNAMIC SELECTIVE FREQUENCY MODULE (MDSF)

To select the informative frequency component to reconstruct, MDSF mainly consists of two elements: frequency decoupler (Fig. 1d) and modulator (Fig. 1e). Decoupler is capable of decomposing the feature map into separate frequency parts dynamically based on the learned filters, and modulator utilizes channel-wise attention to accentuate the helpful frequency. Besides, to provide various local receptive fields, MDSF contains multiple branches by splitting channel dimension, each with a different filter kernel size. For convenience, we only show one-branch case in Fig. 1 (d).

To dynamically decompose feature maps, 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. The learned filters are shared across the group dimension to strike a balance between complexity and feature diversity. Specifically, given 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 produce the low-pass filter for each group of the input, formulated as

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

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, respectively. The group-based operation has fewer parameters and complexities than generating filters for each pixel. The number of groups is discussed in Sec 4.3.

To attain the high-pass filter, we subtract the resulting low-pass filter from the identity kernel with central value as one and everywhere else zero. Next, for each group feature map $X_i \in \mathbb{R}^{C_i \times H \times W}$, where i is the group index and $C_i = \frac{C}{g}$, its low- and high-frequency components can be obtained by using the corresponding reshaped filter F^L and F^H ($\in \mathbb{R}^{g \times k \times k}$), which is expressed as:

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

where c is the index of a channel; h and w denote the spatial coordinates; and $p, q \in \{-1, 0, 1\}$.

After decoupling the feature map into different frequency components, we leverage the frequency modulator to emphasize the genuinely helpful part for reconstruction, as shown in Fig. 1 (e). The modulator works among the channel dimension based on the modified SKNet (Li et al., 2020). Formally, given two frequency maps, X^l and X^h, we first generate the fused feature by,

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

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

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

where W^l and W^h are channel-wise attention weights for two frequency parts; W_l and W_h are parameters of fully connected layers; $[\cdot, \cdot]$ denotes concatenation; and c is the channel index of the concatenated features. Compared to SKNet (Li et al., 2020), which performs Softmax 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. Then, the final frequency weights can be obtained by split operation.

Based on the above one-branch case, the multiple branches with varied filter sizes can be express as:

$$\hat{X} = [\mathcal{M}_1(\mathcal{D}_1(X_1)), ..., \mathcal{M}_m(\mathcal{D}_m(X_m))] \tag{5}$$

where $\mathcal{D}$ and $\mathcal{M}$ denote decoupler and modulator, and X_m is the equally split feature over channels.

3.3 Multi-branch compact selective frequency module (MCSF)

Since the receptive field plays a critical role in image restoration, where degradation kernels always differ in size (Son et al., 2021; Mao et al., 2021), we develop MCSF to efficiently enlarge the receptive field of SFNet. MCSF has two branches with different receptive fields, the global branch and window-based branch. Considering these branches share a similar paradigm, we only detail the window-based one, which is inspired from the idea of window-based attention (Liu et al., 2021).

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 resulting 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 (W_h, W_l in Fig. 1 (f)), which are directly optimized by back propagation. Finally, the updated frequency maps are reversed to the original resolution. The global branch has a similar operation, yet with a global receptive field.

Compared to 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 (see Tab. 9 for details). Hence, MCSF can be embedded in multiple positions.

3.4 Loss Function

To facilitate the frequency selection process, we adopt L_1 loss in the spatial and frequency domains:

$$\mathcal{L}_{spatial} = \sum_{r=1}^{3} \frac{1}{S_r} \|\hat{X}_r - Y_r\|_1; \quad L_{frequency} = \sum_{r=1}^{3} \frac{1}{S_r} \|\mathcal{F}(\hat{X}_r) - \mathcal{F}(Y_r)\|_1 \tag{6}$$

where $r = 3$ means we use a degraded image plus two scaled-down ones for training; $\mathcal{F}$ denotes Fourier transform; S_r is the number of total elements for normalization; and $\hat{X}_r, Y_r$ are the output and target images. The loss function is given by $L = L_{spatial} + \lambda L_{frequency}$, where λ is set as 0.1.

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

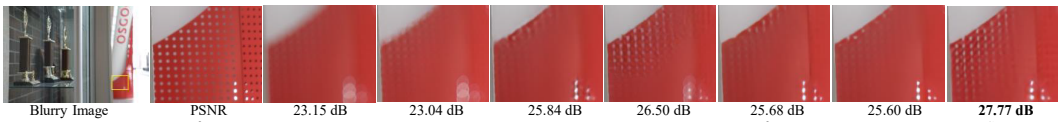


Figure 2: Single-image defocus deblurring results on the DPDD dataset (Abuolaim & Brown, 2020).

4 Experimental Results

4.1 Settings

We evaluate the proposed SFNet on five restoration tasks: image motion/defocus deblurring, image deraining, image dehazing, and image desnowing. More details of the used datasets and training settings for each task are provided in Appendix A. FLOPs is computed 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 horizontally for data augmentation. The initial learning rate is $1e^{-4}$ and gradually reduced to $1e^{-6}$ with the cosine annealing (Loshchilov & Hutter, 2016). Adam ($\beta_1 = 0.9, \beta_2 = 0.999$) is used for training . There are 16 ResBlocks in each encoder and decoder. MDSF has two branches with filter sizes of 3^2 and 5^2 respectively, and the number of groups is 8. We use PyTorch (Paszke et al., 2017) to implement our models on an NVIDIA Tesla V100 GPU and Intel Xeon Platinum 8255C CPU.

4.2 Main Results

Single-image defocus deblurring results. Tab. 1 shows comparisons of defocus deblurring methods on the DPDD (Abuolaim & Brown, 2020) dataset. Our SFNet surpasses state-of-the-art on the most metrics. Particularly on the combined scene, SFNet obtains 0.25 dB PSNR improvement compared to the strong Transformer method Restormer (Zamir et al., 2022). In addition, our method provides a significant gain of 0.51 dB PSNR over the pre-trained network DRBNet (Ruan et al., 2022). The visual results in Fig. 2 show that our method recovers more details than other algorithms. Due to space limitations, experimental results of dual-pixel setting are provided in Appendix E.

Figure 3: Image dehazing results on the SOTS-Indoor dataset (Li et al., 2018).

Image dehazing results. We perform dehazing experiments on the synthetic benchmark RESIDE (Li et al., 2018) and real hazy dataset Dense-Haze (Ancuti et al., 2019). For RESIDE, we train our models for the indoor and outdoor scenarios separately, and then evaluate on the corresponding

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

	SOTS-Indoor		SOTS-Outdoor	
Method	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DehazeNet	19.82	0.821	24.75	0.927
AOD-Net	20.51	0.816	24.14	0.920
GridDehazeNet	32.16	0.984	30.86	0.982
MSBDN	33.67	0.985	33.48	0.982
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
PMNet	38.41	0.990	34.74	0.985
DehazeFormer-L	40.05	**0.996**	-	
SFNet	**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-v2	29.55	0.934	26.61	0.875
DBGAN	31.10	0.942	28.94	0.915
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	**33.27**	**0.963**	31.10	0.941

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	**17.46**	**0.578**

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

Method	PSNR↑	SSIM↑
SRN-DeblurNet	32.53	0.840
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	**34.35**	**0.872**

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	**38.41**	**0.99**

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 substantial gain of 4.87 dB PSNR over DeHamer (Guo et al., 2022). Compared to the recent works, DehazeFormer-L (Song et al., 2022) and PMNet (Ye et al., 2022), 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 are shown in Tab. 4. As we can see, SFNet exhibits the superior ability of dealing with the real-world dehazing problem, receiving a gain of 0.84 dB over DeHamer (Guo et al., 2022). The dehazed results in Fig. 3 illustrate that SFNet is more effective in removing haze than other methods.

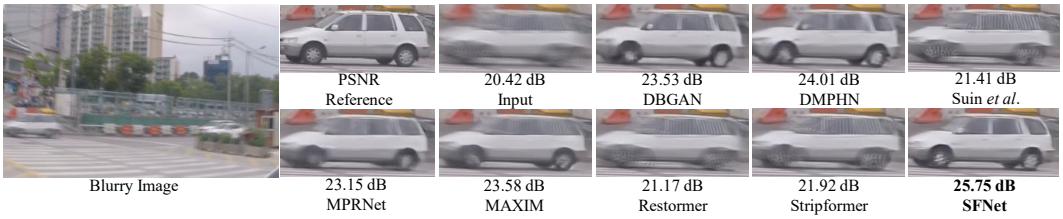
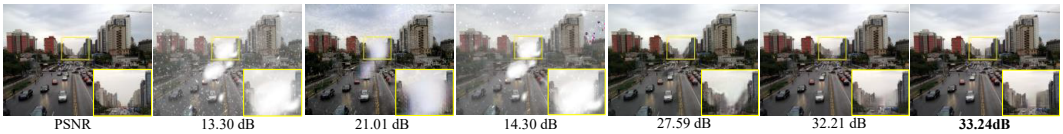


Figure 4: Image motion deblurring results on the GoPro dataset (Nah et al., 2017).

Image motion deblurring results. We evaluate our method on both the synthetic and real-world datasets. The numerical comparisons on the synthetic GoPro (Nah et al., 2017) and HIDE (Shen et al., 2020) datasets are reported in Tab. 3. As can be seen, on the GoPro dataset SFNet shows 0.35 dB PSNR performance improvement over Restormer (Zamir et al., 2022) with ∼3× faster inference speed, shown in Tab. 8. Moreover, our method receives a performance gain of 0.19 dB compared to the recent algorithm Stripformer (Tsai et al., 2022b) with 34% fewer parameters and 26% fewer FLOPs. Notably, our SFNet demonstrates stronger generalization capability to HIDE dataset than Stripformer on all metrics. In addition to the synthetic datasets, we further evaluate the effectiveness

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, 2018) and Test2800 (Fu et al., 2017).

	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	**31.47**	0.919	**31.90**	**0.908**	**38.21**	0.974	33.69	0.937	**32.55**	0.911	**33.56**	0.929



PSNR Reference | 13.30 dB Input | 21.01 dB JSTASR | 14.30 dB DesnowNet | 27.59 dB HDCW-Net | 32.21 dB SnowFormer | **33.24dB SFNet**

Figure 5: Image desnowing results on the CSD dataset (Chen et al., 2021c).

of our network on the real-world dataset. Tab. 5 shows quantitative comparisons on the newly proposed RSBlur (Jaesung Rim) dataset. SFNet sets new state-of-the-art on this dataset, providing a substantial gain of 0.37 dB PSNR over the previous best method Uformer-B (Wang et al., 2022). The qualitative comparisons on the GoPro (Nah et al., 2017) dataset are shown in Fig. 4. Our SFNet produces much sharper and visually pleasant result than competing algorithms.

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., 2022a; Valanarasu et al., 2022; Chen et al., 2022c). As shown in Tab. 6, our framework yields a 0.15 dB PSNR improvement over the concurrent Transformer model SnowFormer (Chen et al., 2022b). The visual results in Fig. 5 show that our method is more effective in removing spatially varying snowflakes than competitors. More results on SRRS (Chen et al., 2020a) and Snow100K (Liu et al., 2018) are provided in Appendix E.

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 category compared to competing approaches. Moreover, on the Test100 dataset (Zhang et al., 2019a), the proposed SFNet obtains a performance boost of 0.30 dB PSNR over MLP model MAXIM-2S (Tu et al., 2022). Visual results shown in Fig. 6 illustrate that our model recovers more fine details without artifacts.

Computational costs comparisons. In Tab. 8, we evaluate the computational costs of five motion deblurring methods on the GoPro testset (Nah et al., 2017). Evaluated on the full-resolution image, our method achieves the fastest speed compared to competitors. Furthermore, SFNet achieves the comparable performance with fewer parameters.

Table 8: Overall comparisons on the GoPro (Nah et al., 2017) testset. +0.57 means the performance improvement provided by the advanced patch-based testing strategy.

Method	NAFNet	MPRNet	DeepRFT+	Restormer	SFNet
PSNR	**33.08 (+0.57)**	32.66	33.23	32.92	33.27
Times/s	0.833	1.148	0.806	1.218	**0.408**
Params/M	67.8	20.1	23.0	26.13	**13.27**
FLOPs/B	**63.33**	777.01	187.04	140.99	125.43

PSNR Reference | 8.39 dB Input | 19.41 dB RESCAN | 21.33 dB PreNet | 20.79 dB UMRL | 24.29 dB MAXIM | **26.48 dB SFNet**

Figure 6: Image deraining results on the Rain100H dataset (Yang et al., 2017).

Table 9: Ablation studies for individual proposed modules.

Method	PSNR	Params (M)	FLOPs (B)
Baseline	31.20	6.90	66.32
MDSF	31.42	7.04	66.55
MCSF	31.45	6.92	66.38
Full	31.68	7.05	66.61

Table 10: Ablation study for the number for MCSF.

Number	PSNR	Params	FLOPs
2	31.22	6.91	66.34
4	31.33	6.91	66.35
6	31.42	6.92	66.37
8	31.45	6.92	66.38

Table 11: Ablation studies for MDSF.

Group	PSNR
2	31.40
4	31.40
8	31.42
16	31.30

4.3 Ablation Studies

In this section, we first show the effectiveness of the proposed modules, and then investigate the effects of different designs for each module. Finally, we delve into the mechanism of MDCF to demonstrate its validity. Following the recent method (Tu et al., 2022), all models are trained on the GoPro (Nah et al., 2017) dataset for 1000 epochs, and each decoder and encoder has 8 ResBlocks.

Influence of each module. Tab. 9 shows that MDSF and MCSF yield a PSNR gain of 0.22 and 0.25 dB over the baseline model with few introduced computing burdens. Deployed only in a single position in each en/decoder, MDSF produces the similar performance with MCSF, demonstrating the effectiveness of dynamic frequency selection mechanism. Furthermore, in Fig. 7, we plot the statistic differences between the ground truth and results of three methods on dehazing. With the frequency selection mechanism, the statistics of our results are closer to that of ground truth.

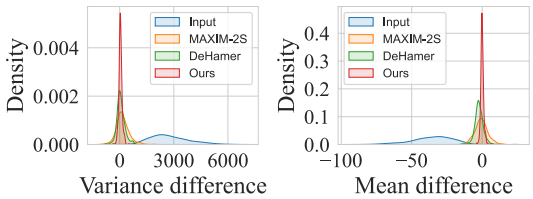


Figure 7: Variance (Left) and Mean (Right) difference between the ground truth and results of three methods on SOTS-Outdoor testset. Models are trained on OTS dataset (Li et al., 2018).

Design choices for MCDF. We study the influence of the number of MCDF in Tab. 10. We increase the number of MCDF from back to front in each residual group, where 2 MCSF means we employ MCSF in the last two ResBlocks in each en/decoder. As we can see from the table, the more MCSF leads to the consistently increasing performance from 31.22 to 31.45 dB PSNR. Due to its few introduced parameters and complexities, we insert MCSF in each ResBlock for frequency learning.

Design choices for MDSF. Towards understanding the impact of the number of groups in MDSF, we test various configurations in Tab. 11. 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 finally pick 8 groups for better performance.

Alternatives for MDSF. To examine the advantage of our design, we compare our decoupler with a few alternatives in Tab. 12. We first substitute the learning-based and fixed frequency separation methods for our decoupler. We borrow the idea from FAN (Pang et al., 2020) to form *Conv* method (Tab. 12a), using stride convolution to generate different frequency parts with reduced resolution. The *Octconv* (Tab. 12b) version (Chen et al., 2020b) shares the similar idea with *Conv*, which utilizes down-sampling to reduce network redundancy. These variants only introduce extra low-frequency signal to the network that is harmful for image restorations. We further utilize the fixed separation methods to replace the proposed decoupler. *Gaussian* (Tab. 12c) and *Wavelet* (Tab. 12d) produce the similar results, much lower than our MDSF. Additionally, *Wavelet* needs more parameters to deal with its four branches.

Table 12: Alternatives for MDSF.

Method	PSNR	Params (M)	FLOPs (B)
(a) Conv	31.11	8.65	79.35
(b) OctConv	31.07	6.99	66.86
(c) Gaussian	30.98	7.02	66.81
(d) Wavelet	30.97	7.06	66.33
(e) Window	31.22	6.98	66.73
(f) Local	31.23	7.00	66.34
(g) MDSF	**31.42**	7.04	66.55

Since our filter kernel is generated by learning, we further compare our MDSF with two attention approaches to verify the validity of the proposed selection mechanism. Specifically, we utilize the widely used window-based self-attention (Wang et al., 2022) Tab. (12e) and dynamic convolution (Han et al., 2021; Wu et al., 2019) (Tab. 12f) to conduct comparisons. As we can see from the table, our method has huge advantages over these methods, demonstrating the effectiveness of MDSF.

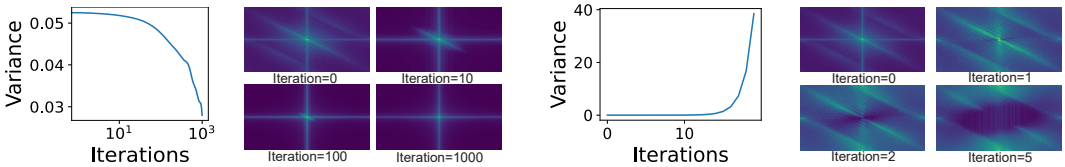


Figure 8: The variance and discrete Fourier transform of the resulting images as we iteratively impose the produced filters on the image. **Left:** The low-pass filter. **Right:** The high-pass filter.

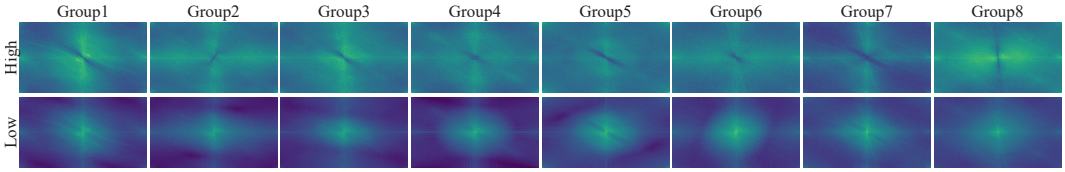


Figure 9: The discrete Fourier transform of the group-wise features generated by our MDSF decoupler. The results are sampled from the last decoder. **Top:** High-frequency. **Bottom:** Low-frequency.

4.4 Qualitative Analyses of MDSF

We provide qualitative analyses of MDSF based on discrete Fourier transform. Results are from the the branch of 3^2 filter size. The input image is sampled from the deblurring dataset GoPro (Nah et al., 2017), shown in Fig. 10. The features are obtained from the last ResBlock in the last decoder.

We first verify the properties of the alleged low-/high-pass filters in MDSF. To this end, we iteratively apply the produced filters to the image. The variance and corresponding spectral features of the intermediate images are provided in Fig. 8. Taking the low-pass filter as an example, along with the increasing of iteration times, the variance of the image decreases constantly, and the high-frequency signals in spectral features are reduced drastically. The high-pass filter exhibits the opposite properties, demonstrating the effectiveness of our filters. It is remarkable that the high-pass filter produces the large variance with few iterations, hence it is more effective than the low-pass filter. As a result, it is easy for MDSF to introduce more high-frequency signals into the network for reconstruction.

In MDSF, we generate different filters for each group to enhance the diversity of frequency features. To delve into this mechanism, we visualize the group-wise spectral features in Fig. 9. As expected, different groups focus on the learning of disparate low-/high-frequency signals, enriching the diversity of frequency representations for selection. We further compare the feature maps before and after our MDSF in Fig. 10. Using the attained filters, the decoupler of MDSF produces different frequency components. The high-frequency feature contains much edge information. The resulting feature after modulator recovers more details of the number plate that is blurry in the initial feature.

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 conduct frequency selection. With designed modules, SFNet shows state-of-the-art results on five image restoration tasks, demonstrating the validity of our frequency selection mechanism.

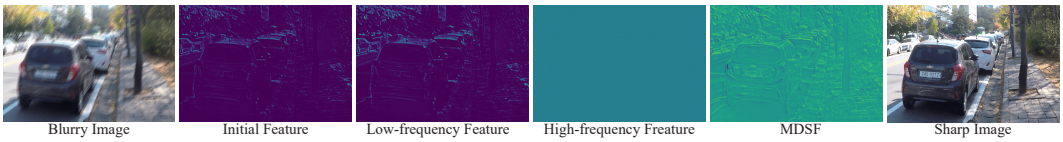


Figure 10: The internal features of MDSF. With our frequency selection mechanism, MDSF produces more fine details than the initial feature, *e.g.,* the number plate. Zoom in for the best view.

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.

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. 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*, 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.

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 2017.

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.

Qi Han, Zejia Fan, Qi Dai, Lei Sun, Ming-Ming Cheng, Jiaying Liu, and Jingdong Wang. On the connection between local attention and dynamic depth-wise convolution. In *International Conference on Learning Representations*, 2021.

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.

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, 2018.

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.

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.

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.

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.

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, 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.

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 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.

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.

Felix Wu, Angela Fan, Alexei Baevski, Yann N Dauphin, and Michael Auli. Pay less attention with lightweight and dynamic convolutions. *arXiv preprint arXiv:1901.10430*, 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.

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. In *European Conference on Computer Vision*, 2022.

Jaeyoung Yoo, Sang-ho Lee, and Nojun Kwak. Image restoration by estimating frequency distribution of local patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

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 2018.

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, 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, 2022.

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.