FMRNet: Image Deraining via Frequency Mutual Revision

Kui Jiang[1], Junjun Jiang[1*], Xianming Liu[1], Xin Xu[2], Xianzheng Ma[3]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[2] School of Computer Science and Technology, Wuhan University of Science and Technology Wuhan, China
[3] Active Vision Lab, University of Oxford, United Kingdom
{jiangkui, jiangjunjun, csxm}@hit.edu.cn, xuxin@wust.edu.cn, maxianzheng@whu.edu.cn

Abstract

The wavelet transform has emerged as a powerful tool in deciphering structural information within images. And now, the latest research suggests that combining the prowess of wavelet transform with neural networks can lead to unparalleled image deraining results. By harnessing the strengths of both the spatial domain and frequency space, this innovative approach is poised to revolutionize the field of image processing. The fascinating challenge of developing a comprehensive framework that takes into account the intrinsic frequency property and the correlation between rain residue and background is yet to be fully explored. In this work, we propose to investigate the potential relationships among rain-free and residue components at the frequency domain, forming a frequency mutual revision network (FMRNet) for image deraining. Specifically, we explore the mutual representation of rain residue and background components at frequency domain, so as to better separate the rain layer from clean background while preserving structural textures of the degraded images. Meanwhile, the rain distribution prediction from the low-frequency coefficient, which can be seen as the degradation prior is used to refine the separation of rain residue and background components. Inversely, the updated rain residue is used to benefit the low-frequency rain distribution prediction, forming the multi-layer mutual learning. Extensive experiments demonstrate that our proposed FMRNet delivers significant performance gains for seven datasets on image deraining task, surpassing the state-of-the-art method ELFormer by 1.14 dB in PSNR on the Rain100L dataset, while with similar computation cost. Code and retrained models are available at https://github.com/kuijiang94/FMRNet.

1 Introduction

Images captured under adverse weather conditions, such as rain, snow, and fog, suffer from noticeable degradation of scene visibility and clarity. It is harmful to many outdoor computer vision systems, such as autonomous driving (Teichmann et al. 2018; Zhong et al. 2022) and video surveillance (Bae 2019; Huang et al. 2018).

Image deraining, aiming to produce a high-quality rain-free image from a given rain image, is a highly desirable component of intelligent decision-making in aforemen-
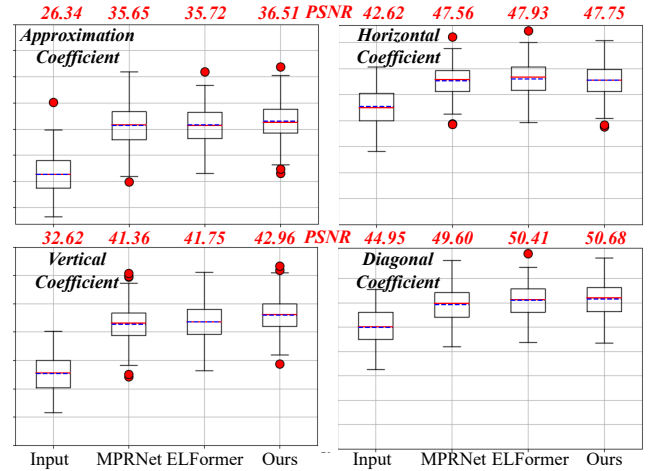
---

*Corresponding Author

Figure 1: Comparative results regarding PSNR plots on different wavelet coefficients on the R100L dataset. As expected, the wavelet-free based methods (MPRNet (Zamir et al. 2021) and ELFormer (Jiang et al. 2022)) achieve considerable performance on the horizontal coefficient, which suffers from the lightest degradation. By contrast, our proposed FMRNet still gains impressive scores on the vertical and approximation coefficients, where the structure and textural information are seriously destroyed by the rain perturbation due to similar directions to rain streaks.

tioned intelligence systems. Conventional methods (Barnum, Narasimhan, and Kanade 2010; Garg and Nayar 2005) provide available schemes, but behave poor generalization to the highly complex and varied rainy scenes because of the specific hand-crafted priors and assumptions.

To rectify this weakness, deep-learning based methods (Fu et al. 2017a; Li et al. 2018; Wang et al. 2020) promote further progress in innovative architectures and training practices for rain removal tasks, and show considerable superiority over conventional algorithms in visual quality improvement. However, due to aliasing effects between rain residue and background details, existing CNN-based methods (Ren et al. 2019; Deng et al. 2020) struggle with generating consistent distributions with missing details (Liu et al. 2020). One reason lies in that the rain residue and back-

ground intrinsically overlap, inferring the pixel residue value to eliminate the rain perturbation in an image inevitably destroys the contextual and structural information.

To tackle this issue, the authors in (Jiang et al. 2021b) propose to learn the joint representation of rain residue and rain streak, as well as their blending relations, and employ the relations to refine the joint features in a coupled representation manner. Although these strategies have demonstrated its effectiveness in eliminating rain perturbation while partly recovering impaired background details, the image deraining remains to be a non-trivial problem to reconstruct pixel values as the output of network in image domain (refer to Figure 1). As a result, learning the separation of rain residue and background in image space fails to handle with aliasing artifacts and preserve structure well.

By contrast, wavelet transform (Mallat 1996) can depict the contextual and textural information of an image at different levels and is reversible, which shows impressive performance in deep networks for various computer vision tasks (Yu et al. 2021), including image super-resolution, dehazing, deraining (Yang, Yang, and Wang 2020; Huang et al. 2021) and so on. Although wavelet-based methods are professional in capturing more structural information, they mainly focus on the representation of individual coefficients, yet seldom consider the mutual relationship between them. Apart from the independent learning, these methods still ignore the aliasing effects among perturbation residue and background in the frequency domain, showing unsatisfied robustness to the complex degradation on image content by perturbation. Usually, it makes deraining results visually vulnerable and inconsistent with real contents on the contrast. That in turn naturally raises a question.

Whether the coupling and association learning between rain residue and background components can be extended to the frequency domain to explore the mutual relation among wavelet coefficients for better rain perturbation removal and background restoration?

To find a reasonable solution, unlike (Jiang et al. 2021b; Kui et al. 2022) characterising the coupling relation between rain residue and background in image domain, we propose to learn their mutual representation and refinement in the frequency space. In particular, we devise a multi-level mutuality learning mechanism between the low-frequency rain residue, complete rain residue and background to promote the model representation and compactness. The multi-level mutual relations are featured as: i) the predicted low-frequency rain residue can help the estimation of the complete rain distribution; ii) the rain residue provides the degradation prior (location and intensity) to promote background recovery; iii) the rain residue and background components are encoded jointly in a coupled learning manner, where the complementary and redundant components are adaptively extracted from each other for refinement.

To this end, we integrate the wavelet transform and mutual learning into a unified framework, and construct a frequency mutual revision network (FMRNet) for single image deraining. The philosophies behind FMRNet involve i) characterizing the rain residue and background distribution from the predicted low-frequency degradation via associa-

tion representation; ii) refining the separation between rain residue and background components via their mutual relation. Specifically, we devise a multi-level mutuality fusion module (MMFM) to achieve the low-frequency and complete rain residue estimation as well as background recovery via the mutual representation. More details regarding the design of MMFM are elaborated in Section 3.2. Similar to (Jiang et al. 2021b), the divide-and-conquer approach is introduced by dividing the separation task into multiple stages via a cascaded framework to progressively achieve the separation and refinement of rain distribution and background. Finally, a reconstruction module (RM) is designed to generate the predicted rain-free image and residual rain image, while compositing the rainy image, approaching to the original rainy input to form the closed loop self-supervision.

Overall, the main contributions are summarized as

- We propose a frequency mutual revision network (FMRNet) to eliminate rain perturbation while preserving background textures in frequency space. To the best of our knowledge, this is the first attempt to investigate the multi-level mutual representation of low-frequency rain residue, rain residue and background in frequency space.

- A novel multi-level mutuality fusion module (MMFM) is devised to characterize rain distribution while refining background components via the mutual learning. In addition, the predicted low-frequency rain distribution provides prior knowledge to complete the rain residue while guiding background recovery with degradation prior (location and intensity).

- Extensive experiments on synthetic and real datasets demonstrate that our FMRNet approach outperforms state-of-the-art methods quantitatively and qualitatively while enjoying considerable computational efficiency.

2 Related Work

In this section, we briefly review the advances in image deraining and wavelet transform.

2.1 Single Image Deraining

Prior to the deep learning, conventional methods (Kang, Lin, and Fu 2012; Liu et al. 2013; Luo, Xu, and Ji 2015; Li et al. 2016) introduce hand-crafted priors into deraining task, and provide available schemes. However, since the priors and assumptions are designed for specific scenarios, these methods (Kang, Lin, and Fu 2012; Chen and Hsu 2013) struggle with poor generalization to highly complex and varied rainy scenes. Recently, deep learning technologies are introduced for rain removal tasks (Jiang et al. 2021a; Li, Cheong, and Tan 2017; Zhang and Patel 2017), which promote further progress in innovative architectures, optimization strategies and training practices, and show significant superiority over the conventional algorithms in visual quality improvement. For example, some researchers (Jiang et al. 2020; Zamir et al. 2021) view the complete rain distribution prediction as the combination of multiple sub-spaces and further estimate and aggregate the multi-scale features for textural reconstruction. Zou *et al.* (Zou et al. 2022) observe
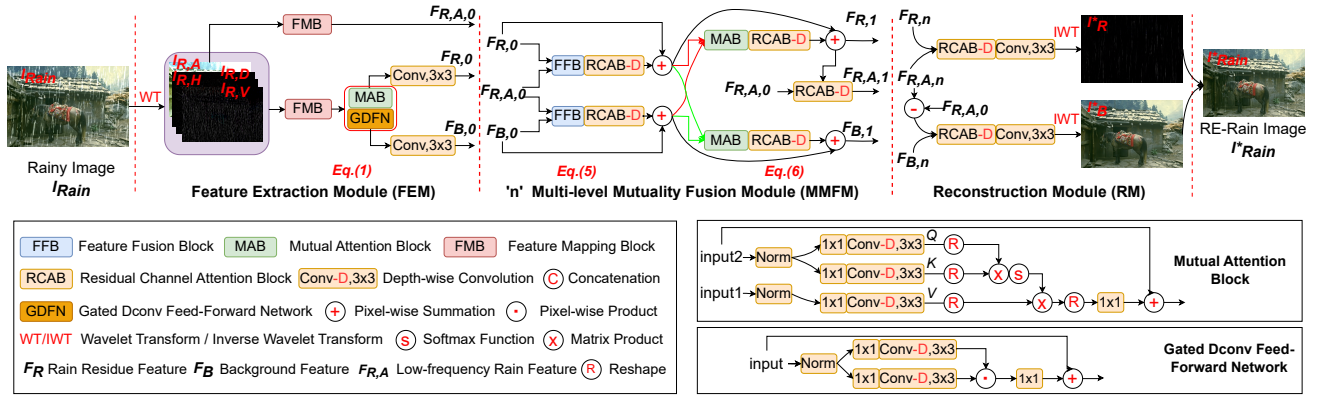
Figure 2: The architecture of our proposed frequency mutual revision network (FMRNet). It consists of a feature extraction module (FEM), several cascaded multi-level mutuality fusion modules (MMFMs), and a reconstruction module (RM). FEM learns the initial representation of rain residue $F_{R,0}$, background $F_{B,0}$ and low-frequency rain residue $F_{R,A,0}$. MMFM takes $F_{R,0}$, $F_{B,0}$ and $F_{R,A,0}$ as inputs, where the predicted low-frequency rain distribution ($F_{R,A,0}$) first provides the prior (local and degree) to complete the rain residue estimation and background recovery. Then the rain residue and background are encoded jointly to explore their mutual relations for further refinement. After that the refined rain residue is used to guide the representation of low-frequency rain distribution prediction, forming the multi-level mutual revision. Followed by a reconstruction module, the predicted rain residue ($F_{R,n}$) and background components ($F_{B,n}$) are transformed into the image domain to generate rain-free image (I_B^*) and predicted rainy image (I_{Rain}^*), composing the closed loop self-supervision.

that deep degradation representations can be clustered by degradation characteristics (types of rain) while independent of image content, and "dream" diverse in-distribution degraded images using a deep inversion paradigm, thus leveraging them to distill the pruned model. To further promote background recovery, some researchers employ the associated learning (Jiang et al. 2021b, 2022) and semantic context (Nanba, Miyata, and Han 2022) to encode the joint representation of rain residue and background, where the predicted rain distribution serves as an extra prior to guide texture recovery. However, these methods still struggle with generating results with visually pleasing contents since the aliasing effects between rain residue and background are unable to be completely eliminated via the decomposition in image domain, consequently destroying contextual and textural information.

2.2 Wavelet Transform

wavelet transform (Mallat 1989) are widely used for signal processing tasks (Szu, Telfer, and Kadambe 1992) due to its reversible property and preeminent ability in depicting the contextual and textural information of an image at different levels. Recent studies tend to harmonize the merits to boost image deraining performance (Yang et al. 2019; Huang et al. 2021). However, besides overlooking the latent interaction between rain residue and background, these methods (Yang et al. 2019; Yang, Yang, and Wang 2020) adopt the same framework or inference strategy for each wavelet coefficient, ignoring the intrinsic relations and heterogenous representation among different coefficients. Unlike them, we incorporate the mutual learning into wavelet transform, and explore the multi-level mutual revision among rain residue, background and low-frequency rain residue. Compared to the existing technologies (Yang et al. 2019; Jiang et al. 2021b),

our proposed frequency mutual revision network (FMRNet) can take advantage of the predicted low-frequency rain priors and mutual relations to guide the rain residue prediction and background recovery. It is more flexible and practical regarding the learning process.

3 Method

This section first offers the overview of our frequency mutual revision network (FMRNet) and then details the architecture of our proposed multi-level mutuality fusion module (MMFM) and its essential components.

3.1 Architecture and Model Optimization

Architecture Overview. Figure 2 outlines the framework of FMRNet, which involves wavelet transform, initial rain ($F_{R,0}$)/background ($F_{B,0}$) feature extraction, mutual-learning representation via multi-level mutuality fusion module (MMFM), and wavelet reconstruction.

Given a rainy image $I_{Rain} \in \mathbb{R}^{H \times W \times 3}$ and its clean version $I_B \in \mathbb{B}^{H \times W \times 3}$, where H and W denote the spatial height and width, we first use the Haar wavelet to generate its corresponding frequency components with the same size of $W/2 \times H/2 \times C$, including the approximation coefficient map $I_{R,A}$, vertical coefficient map $I_{R,V}$, horizontal coefficient map $I_{R,H}$, and diagonal coefficient map $I_{R,D}$. As illustrated before, the core concept of FMRNet is to learn the mutual representation between rain residue and background. We thus devise a feature extraction module (FEM) to encode the initial distribution. Specifically, a feature mapping block (FMB) takes these wavelet coefficients as inputs to generate the initial features. And then a mutual attention block (MAB) and gated deconv feed-forward network (GDFN) are used to aggregate the global response with the positive

importance-weights, followed by two convolutions to generate the initial representation of rain residue $F_{R,0}$ and background $F_{B,0}$. Based on the previous observation in (Jiang et al. 2022), the rain distribution in low-frequency wavelet coefficient gets the similar statistical distribution with that of the original rainy image space. We extract the initial feature $F_{R,A,0}$ of rain distribution in the low-frequency wavelet coefficient ($I_{R,A}$) to help the estimation of rain distribution and background recovery. The aforementioned procedures are expressed as

$$F_{fea} = H_{GDFN,MAB}(H_{FMB}(I_{R,A}, I_{R,V}, I_{R,H}, I_{R,D})),$$
$$F_{R,0} = H_{conv,R}(F_{fea}),$$
$$F_{B,0} = H_{conv,B}(F_{fea}),$$
$$F_{R,A,0} = H_{FMB,R,A}(I_{R,A}),$$
(1)

where $H_{FMB}(\cdot)$ and $H_{FMB,R,A}(\cdot)$ denote feature mapping functions in FMB, involving an initial convolution and residual attention block. $H_{GDFN,MAB}(\cdot)$ is the self-attention calculation to aggregate global response, followed by two convolutions to generate the initial representation.

Then we pack the initial representation of low-frequency rain distribution, rain residue and background into multiple multi-level mutuality fusion modules (MMFMs) to achieve the progressive separation between rain and background components. Specifically, the predicted rain distribution of low-frequency coefficient help the complete rainy estimation. Meanwhile the rain distribution provides the degradation prior (location and intensity) to guide the recovery of background contents. More design details are shown in Figure 2, and the formulaic representation is depicted as

$$F_{R,A,n}, F_{R,n}, F_{B,n} = \mathcal{G}_{MMFM,n}(F_{R,A,0}, F_{R,0}, F_{B,0}).$$
(2)

Benefiting from the progressively multi-level mutual representation, the network holds a perfect way to eliminate rain perturbation while preserving the background contents. Finally, the inverse wavelet transform is performed on both rain and background coefficients to produce the predicted rain and derained results (Huang et al. 2017).

Model Optimization. Similar to existing studies (Wang et al. 2023; Jiang et al. 2021b), the Charbonnier penalty loss (Lai et al. 2017) is used to mediates between the predicted rain-free image (I_B^*) and its ground-truth (I_B). Meanwhile, the predicted rainy image I_{Rain}^* is also to approach the original rainy image I_{Rain} to form the closed loop self-supervision. These constraints are formulated as

$$\mathcal{L}_{RGB} = \sqrt{(I_B^* - I_B)^2 + \varepsilon^2} + \alpha \cdot \sqrt{(I_{Rain}^* - I_{Rain})^2 + \varepsilon^2},$$
(3)

where the penalty coefficient ε is set to 10^{-3} with the α set to 0.2 to balance the loss components. To encourage the fidelity of both the structural information and texture restoration, the structural similarity (SSIM) (Wang et al. 2004) loss is introduced, depicted as

$$\mathcal{L}_Y = \text{SSIM}(I_{B,Y}^*, I_{B,Y}) + \alpha \cdot \text{SSIM}(I_{Rain}^*, I_{Rain}). \quad (4)$$

The final loss function is defined as $\mathcal{L} = \mathcal{L}_{RGB} + \lambda \cdot \mathcal{L}_Y$, where λ is used to balance the loss components, and experimentally set as -0.15.

3.2 Multi-level Mutuality Fusion Module

For a better separation of rain residue and background, in this paper, we propose the multi-level mutuality fusion module (MMFM) to learn the multi-level mutual representation, involving the low-frequency rain residue, complete rain residue and background contents. As shown in Figure 2, MMFM contains three progressive operations: i) exploring the association learning between the low-frequency rain residue and complete rain residue and background where the low-frequency degradation prior is used to guide the refined representation of both complete rain residue and background; ii) exploiting the mutual representation of rain residue and background for a better perturbation elimination and background restoration; ii) using the refined rain residue to re-feed the low-frequency rain distribution prediction.

Taking the first MMFM as an example, the first operation in MMFM involves two feature fusion blocks (FFBs) ($H_{FFB,R}(\cdot)$ and $H_{FFB,B}(\cdot)$) and two improved residual channel attention blocks (RCAB-Ds) ($H_{RCABD,R}(\cdot)$ or $H_{RCABD,B}(\cdot)$). The former takes the initial rain residue feature ($F_{R,A,0}$) of low-frequency coefficient and rain residue feature ($F_{R,0}$) as inputs to learn the association representation with the guidance of $F_{R,A,0}$; the latter is used for a deep accurate learning of association representation and to produce the refined rain residue prediction. In RCAB-D, the standard convolution is replaced with depth-wise separable convolutions (Mehta et al. 2019), which are computationally more efficient while achieving similar or better performance. The same operation is performed on the low-frequency rain residue $I_{R,A,0}$ and background $I_{B,0}$. The procedure in the first operation can be expressed as

$$f_{R,1} = H_{RCABD,R}(H_{FFB,R}(F_{R,0}, F_{R,A,0})) + F_{R,0},$$
$$f_{B,1} = H_{RCABD,B}(H_{FFB,B}(F_{B,0}, F_{R,A,0})) + F_{B,0},$$
(5)

where $f_{R,1}$ and $f_{B,1}$ respectively denote the updated features of rain residue and background in the first operation.

The second operation takes $f_{R,1}$ and $f_{B,1}$ as inputs to learn the mutual representation of rain residue and background, involving two mutual attention blocks (MABs) ($H_{MAB,R}(\cdot)$ and $H_{MAB,B}(\cdot)$) and two RCAB-Ds ($H_{RCABD,R}(\cdot)$ and $H_{RCABD,B}(\cdot)$). Specifically, the rain residue and background components are encoded jointly in MAB, where the complementary and redundant components are adaptively extracted from each other for refinement. Similarly, RCAB-D is used to achieve the deep representation and generate the refinement prediction of rain residue and background. The aforementioned procedures are depicted as

$$F_{R,1} = H_{RCABD,R}(H_{MAB,R}(f_{R,1}, f_{B,1})) + f_{R,1},$$
$$F_{B,1} = H_{RCABD,B}(H_{MAB,B}(f_{B,1}, f_{R,1})) + f_{B,1},$$
(6)

where $F_{R,1}$ and $F_{B,1}$ refer to the refined representation of rain residue and background in the second operation. Following that, the refined $F_{R,1}$ is used to generate the updated prediction of the low-frequency rain distribution $F_{R,A,1}$.

Through wavelet transform and multi-level mutuality fusion learning, these operations allow the network to explore

intrinsic relations among spectra and the mutuality between rain residue and background, which achieves more impressive and effortless separation of rain residue and background contents. Our innovative scheme not only outperforms traditional separation methods in the image space in terms of accuracy, but also retains the intricate textures.

4 Experiments

To validate our proposed FMRNet, we conduct extensive experiments on synthetic and real rain-image datasets and compare it with typical image deraining methods. These methods include DAWN (Jiang et al. 2023), DANet (Kui et al. 2022), ELFormer (Jiang et al. 2022), MPRNet (Zamir et al. 2021), SWAL (Huang et al. 2021), DRDNet (Deng et al. 2020), MSPFN (Jiang et al. 2020), IADN (Jiang et al. 2020), and PreNet (Ren et al. 2019).

4.1 Implementation Details

Data Collection. Following (Jiang et al. 2020), we use $13,700$ clean/rain image pairs from (Zhang, Sindagi, and Patel 2020; Fu et al. 2017b) for training all compared methods to guarantee fairness since these methods are originally trained with different datasets. In particular, the compared methods are retrained with the publicly released codes by tuning the optimal settings. For testing, four synthetic (Test100 (Zhang, Sindagi, and Patel 2020), Test1200 (Zhang and Patel 2018), R100H, and R100L (Yang et al. 2017)) and three real-world datasets (Rain in Driving (RID), Rain in Surveillance (RIS) (Li et al. 2019) and Real127 (Zhang and Patel 2018)) are considered for evaluation.

Experimental Setup. In our baseline, the number of multi-level mutuality fusion module (MMFM) is empirically set to 10. To obtain training samples, the training images are coarsely cropped into small 256×256 patches. We use Adam optimizer with the learning rate (2×10^{-4} with the decay rate of 0.8 at every 80 epochs till 500 epochs) and batch size (8) to train FMRNet on a single NVIDIA 3090 GPU.

4.2 Ablation Studies

Validation on Basic Components. We conduct ablation studies to validate the contributions of individual components, including the self-attention (SA) in feature extraction module (FEM), low-frequency prior guidance (LPG), and mutual learning among rain residue and background in multi-level mutuality fusion module (MMFM), depth-wise separable convolutions (DSC), and "Y" channel ($\mathcal{L}_Y$) on the *Test1200* dataset, to the final deraining performance. For simplicity, we denote our final model as FMRNet, with the number of MMFM to 10. Then, we design a *w/o SA* model to evaluate the global representation via self-attention in FEM. We devise three models (*w/o LPG*, *w/o ML* and *w/o MMF*) to investigate the effect of multi-level mutuality fusion in MMFM, involving low-frequency guidance and mutual learning. In addition, the *w/o DSC* model is designed by replacing DSC in RCAB-D with standard convolutions to analyze its effect. Moreover, the Y-channel (structure) loss ($\mathcal{L}_Y$) is also validated. *It is worth noting that these models*

Model	SA	LPG	ML	DSC	$\mathcal{L}_Y$	PSNR	SSIM	Par.	Time	GFlops
Rain Image	–	–	–	–	–	22.16	0.732	–	–	–
w/o SA	×	✓	✓	✓	✓	32.98	0.919	1.548	0.136	86.57
w/o LPG	✓	×	✓	✓	✓	33.01	0.920	1.513	0.148	90.02
w/o ML	✓	✓	×	✓	✓	32.63	0.916	1.504	0.146	89.46
w/o MMF	✓	×	×	✓	✓	32.34	0.914	1.459	0.139	87.28
w/o DSC	✓	✓	✓	×	✓	32.96	0.919	1.576	0.155	92.48
w/o $\mathcal{L}_Y$	✓	✓	✓	✓	×	33.06	0.916	1.551	0.150	91.23
FMRNet	✓	✓	✓	✓	✓	33.21	0.924	1.551	0.150	91.23

Table 1: Ablation study on the self-attention (SA) (mutual attention block (MAB) and gated dconv feed-forward network (GDFN)) in feature extraction module (FEM), low-frequency prior guidance (LPG) and mutual learning (ML) among rain residue and background in multi-level mutuality fusion module (MMFM), depth-wise separable convolutions (DSC), and "Y" channel ($\mathcal{L}_y$) on the *Test1200* dataset. We obtain the model parameters (Million (M)), inference time (Second (S)), and computation complexity (GFlops (G)) of deraining on images with 512×512 pixels.

consume approximately the same number of parameters as that of FMRNet.

Quantitative results in terms of the deraining performance and efficiency on the Test1200 dataset are presented in Table 1, revealing that the complete deraining model FMRNet achieves significant improvements over its incomplete variants. The results show that combining the low-frequency guidance and mutual learning in MMFM exhibits considerable superiority (gaining 0.87 dB) in terms of restoration quality (referring to the results of FMRNet, *w/o MMF* models). We speculate that the predicted low-frequency rain distribution can provide valuable priors (location and intensity) to help the complete rain residue estimation and background restoration. Meanwhile, the mutual representation among rain residue and background allows the network to explore the complementary refinement to alleviate the aliasing effect, leading to more accurate separation. In addition, removing the self-attention in FEM may decline the capability of global fusion, leading to 0.23 dB performance drop (referring to the results of FMRNet and *w/o SA* models). Moreover, using the depth-wise separable convolutions allows increasing the model depth with approximately the same parameters, thus enhancing the representation power. Removing $\mathcal{L}_Y$ may greatly degrade the representation capability on the spatial structure, leading to an obvious performance drop (0.15 dB in PSNR) (referring to the results of FMRNet and *w/o $\mathcal{L}_Y$* models).

4.3 Comparison with State-of-the-arts

Synthesized Data. We compare the performance of FMRNet with that of 8 representative methods on four commonly used rain-image datasets. Quantitative results are provided in Table 2. Meanwhile, the complexities of inference time, computational cost and model parameters are also compared with image size of 512×512. Our proposed FMRNet model has demonstrated exceptional restoration performance, surpassing all other compared techniques. Specif-

Methods	IADN	MSPFN	DRDNet	MPRNet	SWAL	DANet	ELFormer	DAWN	FMRNet (Ours)
Test100/Test1200	26.71/32.29	27.50/32.39	28.06/26.73	30.27/32.91	28.47/30.40	29.90/33.10	30.45/**33.38**	29.86/32.76	**30.66**/33.21
	0.865/0.916	0.876/0.916	0.874/0.824	0.897/0.916	0.889/0.892	0.893/0.919	0.909/**0.925**	0.902/0.919	**0.915**/0.924
	0.924/0.958	0.928/0.960	0.925/0.920	0.939/0.960	0.936/0.950	0.938/0.962	0.945/0.964	0.941/0.960	**0.936/0.964**
R100H/R100L	27.86/32.53	28.66/32.40	21.21/29.24	30.41/36.40	29.30/34.60	29.96/35.85	30.48/36.67	29.89/35.97	**30.69/37.81**
	0.835/0.934	0.860/0.933	0.668/0.883	0.889/0.965	0.887/0.958	0.889/0.962	0.896/0.968	0.889/0.963	**0.900/0.974**
	0.875/0.942	0.890/0.943	0.797/0.903	0.910/0.969	0.908/0.963	0.911/0.967	0.915/0.972	0.911/0.969	**0.920/0.978**
Avg-PSNR	29.84	30.23	26.31	32.49	30.69	32.20	32.74	32.12	**33.09**
Par.(M)	**0.980**	13.35	5.230	3.637	9.792	2.943	1.532	1.489	1.551
Time (S)	0.132	0.507	1.426	0.207	0.116	0.109	0.125	**0.087**	0.150
FLOPs (G)	80.99	708.3	–	565.8	**39.00**	130.9	66.39	89.34	91.23

Table 2: Comparison of average PSNR/SSIM/FSIM scores. We obtain model parameters (Million), average inference time (Second) and computational cost (Flops (G)) of deraining on images with the size of 512×512.
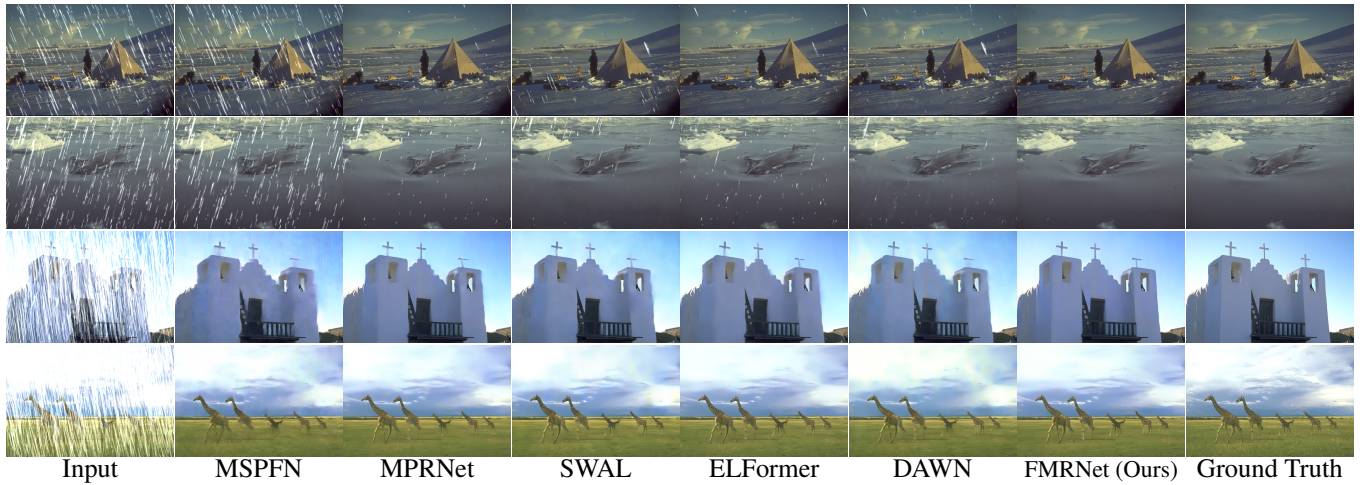

Figure 3: Visual comparison of derained images obtained by seven methods on R100H/R100L/Test100/Test1200 datasets.

ically, it achieves significant superiority over the current transformer-based (ELFormer) and wavelet-based (DAWN) SOTA approaches on four datasets, with an average improvement of 0.35 dB and 0.97 dB. Additionally, FMRNet has proven to be highly efficient, which consumes approximate parameters and computational cost to the SOTA, making it a highly effective and cost-efficient solution. Moreover, we have observed that most of the deraining models obtain impressive performance on light rain cases with high consistency. However, only our FMRNet and ELFormer still perform favorably well on heavy rain conditions, exhibiting great superiority over other competing methods in terms of PSNR and SSIM. On the Rain100L dataset, our proposed FMRNet method surpasses the wavelet-based DAWN (Jiang et al. 2023), transformer-based ELFormer (Jiang et al. 2022) and CNN-based MPRNet (Zamir et al. 2021) in PSNR by 1.84 dB, 1.14 dB and 1.41 dB, respectively.

For more convincing evidence, we also provide additional visual comparisons in Figure 3, involving light and heavy rain conditions. As expected, FMRNet has shown its exceptional ability to eliminate rain streaks and produce remarkable images in all kinds of rainy conditions, surpassing all of its competitors with ease. By contrast, spatial-domain-based methods, such as MPRNet (Zamir et al. 2021)

and ELFormer (Jiang et al. 2022) can partly eliminate rain perturbation in the image domain and thus bring an improvement in visibility. But they fail to generate visually appealing results due to **the aliasing and residual artifact in the frequency embedding space**. In particular, they tend to generate derained results with over-smooth contents when the structure and textures have the similar direction to rain streaks. Likewise, SWAL(Huang et al. 2021) and DAWN (Jiang et al. 2023) promote the spatial and frequency domains representation but shows unsatisfactory deraining performance due to the inter-frequency conflicts and compromise, leading to obvious color distortion. Besides recovering cleaner and more credible image textures, our FMRNet produces results with better contrast and less color distortion. **In particular for the heavy rain condition of the third and fourth scenarios in Figure 3 where the structure ("giraffe" and "church" scenarios) is destroyed due to the similar direction and distribution to rain streaks, only FMRNet can infer credible details while the other methods produce results with losing details and contrast distortion.** We speculate these visible improvements on restoration quality may benefit from elaborate deraining scheme – wavelet-based multi-level mutuality fusion and reconstruction. These strategies encourage the network simul-

Datasets	IADN	MSPFN	DRDNet	MPRNet	ELFormer	DAWN	FMRNet (Ours)
Real127	3.769/29.12	3.816/**29.05**	4.208/30.34	3.965/30.05	<u>3.735</u>/29.16	3.762/<u>29.19</u>	**3.714**/29.27
RID (2495)	6.035/40.72	6.518/40.47	5.715/39.98	6.452/40.16	**4.318**/37.89	4.785/<u>37.49</u>	<u>4.431</u>/**37.25**
RIS (2348)	5.909/42.95	6.135/43.47	6.269/45.34	6.610/48.78	5.835/**42.16**	**5.536**/43.16	<u>5.628</u>/42.29

Table 3: Comparison of average NIQE/SSEQ scores by seven deraining methods on three real-world datasets. The Bold and underline present the best and second performances, respectively.
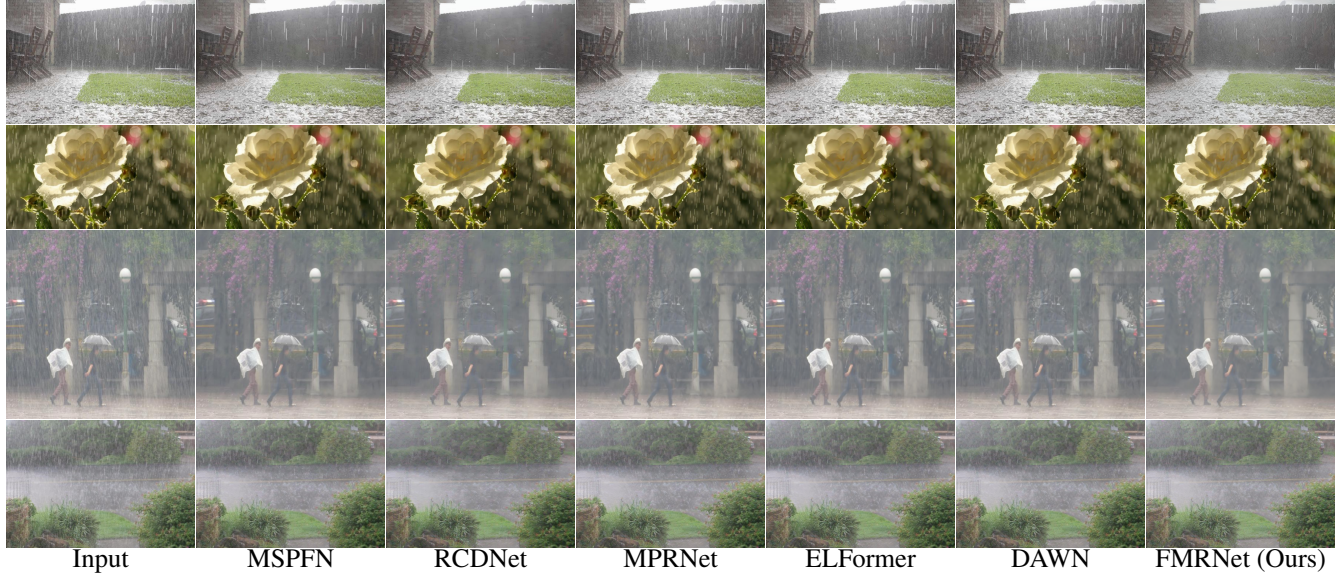

Figure 4: Visual comparison of derained images obtained by six methods on four real-world scenarios, covering heavy rain (1st), light rain (2st) and rain veiling effect (3st-4st). Please zoom in for a close up comparison.

taneously focus on rain perturbation on pixel image and frequency embedding spaces, while the mutual exploration facilitates the separation of rain residue and background.

Real-world Data. We further conduct experiments on real-world datasets, including the Real127 (Zhang and Patel 2018), Rain in Driving (RID), and Rain in Surveillance (RIS) (Li et al. 2019). The samples in RID and RIS are collected from car-mounted cameras and networked traffic surveillance cameras in rainy day, respectively involving 2,495 and 2,348 real-world rain samples. These images differ in rain types, image quality, object size, angle, etc., and represent real application scenarios where deraining may be desirable. Table 3 provides the comparison in terms of the quantitative results of NIQE (Mittal, Soundararajan, and Bovik 2012) and SSEQ (Liu et al. 2014), where smaller NIQE and SSEQ scores indicate better perceptual quality and clearer contents. Our FMRNet achieves competitive performance, gaining the lowest average values of NIQE on the Real127 and SSEQ on the RID datasets, and the second best average scores on the RIS dataset. The visual comparisons are shown in Figure 4. Spatial-domain-based methods, such as MPRNet (Zamir et al. 2021) and ELFormer (Jiang et al. 2022), spatial-domain-based methods can partially mitigate the impact of rain perturbations, but they often result in the loss of details, especially the structural information that aligns with the direction of the rain streaks. Compared to the wavelet-based DAWN (Jiang et al. 2023)), our FM-

RNet is still more effective in eliminating more rain perturbation while preserving finer background details in the derained images.

5 Conclusion

In this study, a novel frequency mutual revision network (FMRNet) is devised to remove rain perturbation in both the spatial and frequency spaces. More specifically, we investigate the potential relationships among rain-free and residue components at the frequency domain, where the mutual relations among low-frequency rain residue, rain residue and background are fully explored, so as to better separate the rain layer from clean background while preserving structural textures of degraded images. Meanwhile, we construct a multi-level mutuality fusion module (MMFM) to characterize the mutual relations, which help accurate rain estimation and background restoration. The effectiveness and efficiency of our proposed FMRNet is extensively validated through experiments on both synthetic and real datasets.

Although this study specializes in eliminating the aliasing artifacts and preserving structure, one limitation is that it behaves poor generalization to the hybrid degradation, like heavy rain veiling effect. The specific framework and decomposition pattern may be required to investigate and explore the intrinsic mutual relation of these complex degradation scenarios. In the future, we would explore how to solve these limitations.

Acknowledgments

References

Bae, S. 2019. Object Detection Based on Region Decomposition and Assembly. In *AAAI*, 8094–8101.

Barnum, P. C.; Narasimhan, S.; and Kanade, T. 2010. Analysis of rain and snow in frequency space. *IJCV*, 86(2-3): 256.

Chen, Y.-L.; and Hsu, C.-T. 2013. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *CVPR*, 1968–1975.

Deng, S.; Wei, M.; Wang, J.; Feng, Y.; Liang, L.; Xie, H.; Wang, F. L.; and Wang, M. 2020. Detail-recovery Image Deraining via Context Aggregation Networks. In *CVPR*, 14548–14557.

Fu, X.; Huang, J.; Ding, X.; Liao, Y.; and Paisley, J. 2017a. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.*, 26(6): 2944–2956.

Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017b. Removing rain from single images via a deep detail network. In *CVPR*, 3855–3863.

Garg, K.; and Nayar, S. K. 2005. When does a camera see rain? In *ICCV*, volume 2, 1067–1074.

Huang, H.; He, R.; Sun, Z.; and Tan, T. 2017. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In *ICCV*, 1689–1697.

Huang, H.; Yu, A.; Chai, Z.; He, R.; and Tan, T. 2021. Selective Wavelet Attention Learning for Single Image Deraining. *International Journal of Computer Vision*, 129(4): 1282–1300.

Huang, W.; Liang, C.; Yu, Y.; Wang, Z.; Ruan, W.; and Hu, R. 2018. Video-Based Person Re-Identification via Self Paced Weighting. In *AAAI*, 2273–2280.

Jiang, K.; Liu, W.; Wang, Z.; Zhong, X.; Jiang, J.; and Lin, C.-W. 2023. Dawn: Direction-aware attention wavelet network for image deraining. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7065–7074.

Jiang, K.; Wang, Z.; Chen, C.; Wang, Z.; Cui, L.; and Lin, C.-W. 2022. Magic ELF: Image Deraining Meets Association Learning and Transformer. In *Proceedings of the 30th ACM International Conference on Multimedia*, 827–836.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Han, Z.; Lu, T.; Huang, B.; and Jiang, J. 2020. Decomposition makes better rain removal: An improved attention-guided deraining network. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10): 3981–3995.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-Scale Progressive Fusion Network for Single Image Deraining. In *CVPR*, 8343–8352.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Wang, G.; Han, Z.; Jiang, J.; and Xiong, Z. 2021a. Multi-scale hybrid fusion network for single image deraining. *IEEE Transactions on Neural Networks and Learning Systems*.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Wang, Z.; Wang, X.; Jiang, J.; and Lin, C.-W. 2021b. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Transactions on Image Processing*, 30: 7404–7418.

Kang, L.-W.; Lin, C.-W.; and Fu, Y.-H. 2012. Automatic single-frame-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4): 3888–3901.

Kui, J.; Zhongyuan, W.; Zheng, W.; Peng, Y.; Junjun, J.; Jinsheng, X.; and Chia-Wen, L. 2022. DANet: Image Deraining via Dynamic Association Learning. In *IJCAI*.

Lai, W.-S.; Huang, J.-B.; Ahuja, N.; and Yang, M.-H. 2017. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, 624–632.

Li, R.; Cheong, L.-F.; and Tan, R. T. 2017. Single image deraining using scale-aware multi-stage recurrent network. *arXiv preprint arXiv:1712.06830*.

Li, S.; Araujo, I. B.; Ren, W.; Wang, Z.; Tokuda, E. K.; Junior, R. H.; Cesar-Junior, R.; Zhang, J.; Guo, X.; and Cao, X. 2019. Single image deraining: A comprehensive benchmark analysis. In *CVPR*, 3838–3847.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, 254–269.

Li, Y.; Tan, R. T.; Guo, X.; Lu, J.; and Brown, M. S. 2016. Rain streak removal using layer priors. In *CVPR*, 2736–2744.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust Recovery of Subspace Structures by Low-Rank Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 171–184.

Liu, H.; Jiang, B.; Song, Y.; Huang, W.; and Yang, C. 2020. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In *ECCV*, 725–741.

Liu, L.; Liu, B.; Huang, H.; and Bovik, A. C. 2014. No-reference image quality assessment based on spatial and spectral entropies. *SPIC*, 29(8): 856–863.

Luo, Y.; Xu, Y.; and Ji, H. 2015. Removing Rain from a Single Image via Discriminative Sparse Coding. In *ICCV*, 3397–3405.

Mallat, S. 1996. Wavelets for a vision. *Proceedings of the IEEE*, 84(4): 604–614.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7): 674–693.

Mehta, S.; Rastegari, M.; Shapiro, L.; and Hajishirzi, H. 2019. Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network. In *CVPR*, 9190–9200.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.*, 20(3): 209–212.

Nanba, Y.; Miyata, H.; and Han, X.-H. 2022. Dual heterogeneous complementary networks for single image deraining.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 568–577.

Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; and Meng, D. 2019. Progressive image deraining networks: a better and simpler baseline. In *CVPR*, 3937–3946.

Szu, H. H.; Telfer, B. A.; and Kadambe, S. L. 1992. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31(9): 1907–1916.

Teichmann, M.; Weber, M.; Zoellner, M.; Cipolla, R.; and Urtasun, R. 2018. Multinet: Real-time joint semantic reasoning for autonomous driving. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, 1013–1020. IEEE.

Wang, C.; Xing, X.; Wu, Y.; Su, Z.; and Chen, J. 2020. DCSFN: Deep Cross-scale Fusion Network for Single Image Rain Removal. In *ACM Multimedia*, 1643–1651.

Wang, Q.; Jiang, K.; Wang, Z.; Ren, W.; Zhang, J.; and Lin, C.-W. 2023. Multi-Scale Fusion and Decomposition Network for Single Image Deraining. *IEEE Transactions on Image Processing*, 33: 191–204.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P.; et al. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4): 600–612.

Yang, H.-H.; Yang, C.-H. H.; and Wang, Y.-C. F. 2020. Wavelet channel attention module with a fusion network for single image deraining. In *ICIP*, 883–887. IEEE.

Yang, W.; Liu, J.; Yang, S.; and Guo, Z. 2019. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE Trans. Image Process.*, 28(6): 2948–2961.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017. Deep joint rain detection and removal from a single image. In *CVPR*, 1357–1366.

Yu, Y.; Zhan, F.; Lu, S.; Pan, J.; Ma, F.; Xie, X.; and Miao, C. 2021. WaveFill: A Wavelet-based Generation Network for Image Inpainting. In *CVPR*, 14114–14123.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-Stage Progressive Image Restoration. In *CVPR*.

Zhang, H.; and Patel, V. M. 2017. Convolutional sparse and low-rank coding-based rain streak removal. In *WACV*, 1259–1267.

Zhang, H.; and Patel, V. M. 2018. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, 695–704.

Zhang, H.; Sindagi, V.; and Patel, V. M. 2020. Image deraining using a conditional generative adversarial network. *IEEE Trans. Circuits Syst. Video Technol.*, 30(11): 3943–3956.

Zhong, X.; Tu, S.; Ma, X.; Jiang, K.; Huang, W.; and Wang, Z. 2022. Rainy WCity: A real rainfall dataset with diverse conditions for semantic driving scene understanding. In *International Joint Conference on Artificial Intelligence*, 1743–1749.

Zou, W.; Wang, Y.; Fu, X.; and Cao, Y. 2022. Dreaming to prune image deraining networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6023–6032.