Exploring Fourier Prior for Single Image Rain Removal

Xin Guo, **Xueyang Fu***, **Man Zhou**, **Zhen Huang**, **Jialun Peng** and **Zheng-Jun Zha**
University of Science and Technology of China, China
willing@mail.ustc.edu.cn, xyfu@ustc.edu.cn, {manman, hz13, pjl}@mail.ustc.edu.cn, zhazj@ustc.edu.cn

Abstract

Deep convolutional neural networks (CNNs) have become dominant in the task of single image rain removal. Most of current CNN methods, however, suffer from the problem of overfitting on one single synthetic dataset as they neglect the intrinsic prior of the physical properties of rain streaks. To address this issue, we propose a simple but effective prior - Fourier prior to improve the generalization ability of an image rain removal model. The Fourier prior is a kind of property of rainy images. It is based on a key observation of us - replacing the Fourier amplitude of rainy images with that of clean images greatly suppresses the synthetic and real-world rain streaks. This means the amplitude contains most of the rain streak information and the phase keeps the similar structures of the background. So it is natural for single image rain removal to process the amplitude and phase information of the rainy images separately. In this paper, we develop a two-stage model where the first stage restores the amplitude of rainy images to clean rain streaks, and the second stage restores the phase information to refine fine-grained background structures. Extensive experiments on synthetic rainy data demonstrate the power of Fourier prior. Moreover, when trained on synthetic data, a robust generalization ability to real-world images can also be obtained. The code will be publicly available at https://github.com/willinglucky/Exploring-Fourier-Prior-for-Single-Image-Rain-Removal.

1 Introduction

Outdoor computer vision systems are increasingly widely used, such as objection tracking, autonomous driving, and pedestrian detection. Still, the performance of these tasks is severely impaired under various rain conditions [Comaniciu *et al.*, 2003; He *et al.*, 2017]. Hence, image deraining is an important pre-processing task in the computer vision community and has drawn much research attention in recent years [Fu *et al.*, 2017a; Jiang *et al.*, 2020; Wang *et al.*, 2020].

Earlier, with the success of the dark channel prior [He *et al.*, 2010] in the dehazing task, exploiting prior knowledge to obtain clean images in the area of image deraining once rose. By exploring the physical properties of the rain layer and background layer, various priors are introduced to separate rain streaks from clean images. The representative methods achieved respectable deraining results proving the advantage of priors via dictionary learning [Kang *et al.*, 2011], discriminative sparse coding [Luo *et al.*, 2015], and Gaussian mixture models [Li *et al.*, 2016]. However, they depend on time-consuming iterative computations and subjective prior in specific scenarios, which lead to efficiency issues and poor adaptability in diverse rainy images from different sources.

Recently, CNNs have achieved significant success in single image rain removal. By end-to-end training, CNNs learn mapping rules from paired rainy/clean images. Researchers mainly focus on the design of complex networks for better deraining performance [Fu *et al.*, 2017a; Ren *et al.*, 2019; Jiang *et al.*, 2020; Jiang *et al.*, 2020; Chen *et al.*, 2021]. For example, the network in [Jiang *et al.*, 2020] consists of a multi-scale structure and Conv-LSTM modules with dense connection between each unit. Despite of the impressive capability to restore clean images from rainy images, these methods neglect the intrinsic prior of the physical properties of rain streaks, making them easily fall into the problem of overfitting on one single synthetic dataset [Wang *et al.*, 2020].

In this paper, we introduce a novel prior for image deraining. The motivation comes from a well-known property of the Fourier transformation: the Fourier phase spectrum preserves high-level semantics, while the amplitude spectrum contains low-level features [Oppenheim *et al.*, 1979]. Fig. 1 shows the results of exchanging the Fourier amplitude and phase spectrum of paired rainy/clean images. (a) and (c) are reconstructed with the phase of rainy image and the amplitude of ground truth, and (b) and (d) are obtained with the amplitude of rainy image and the phase of ground truth. From these images, three interesting phenomena arise. *First*, rain streaks in (a) and (c) are greatly suppressed, meaning most of the rain streaks information is preserved in the amplitude spectrum of rainy images. *Second*, comparing the ground truth and (a)/(c), we can see that the fine-grained structures of clean images are well preserved after replacing the phase spectrum with which of the rainy images. This indicates that the phase of rainy images keeps the similar background structures as the ground

*Corresponding author

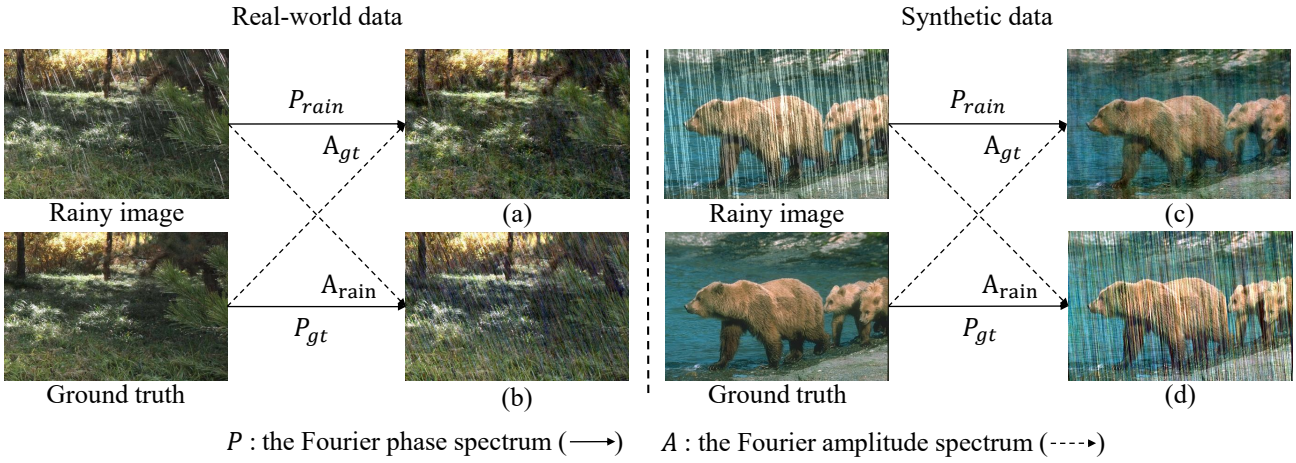

Figure 1: Examples of Fourier reconstructions. (a) and (c) are reconstructed with the phase of rainy images and the amplitude of ground truth. In contrast, (b) and (d) are reconstructed with the phase of ground truth and the amplitude of rainy images. There are three observations. First, it is obvious that rain streaks are diminished in (a) and (c), compared with (b) and (d). This means *the amplitude of rainy images contains most rain streaks information*. Second, similarities between ground truth and (a), (c) indicate that *the phase of rainy images keeps background structures similar to ground truth*. Third, the homogeneous banded degradation appears over the whole image in (b) and (d). This shows that *the amplitude of rainy images shares the same physical characteristic across the synthetic and real-world domain.*

truth. *Third*, (b) and (d) share a homogeneous banded degradation over the whole image, showing that the amplitude of synthetic and real-world rainy images have similar physical characteristics. Naturally, processing the amplitude spectrum of rainy images independently has the potential of effective rain removal and generalization to real-world datasets. Besides, the phase spectrum of rainy images can also be used to refine structural details of the background. In this way, the Fourier prior is achieved by learning the transformation of the amplitude and phase spectrum separately.

Following the Fourier prior, we design a simple but effective two-stage architecture by processing the amplitude and phase spectrum of rainy images sequentially. At the first stage, the network learns the mapping of the amplitude spectrum from the rainy domain to the clean domain. And the second stage learns the phase transformation and refinement of images, with the phase spectrum of the rainy images as part of input. To facilitate the transformation of the Fourier spectrum, we introduce a Fourier amplitude residual block (FARB) and a Fourier phase residual block (FPRB). FARB and FPRB alter the amplitude and phase spectrum respectively in the residual branch, avoiding destruction in the mainstream. Finally, the Fourier-prior-based architecture (FPNet) is built with two UNet-like networks [Ronneberger *et al.*, 2015] and FARB and FPRB as basic units. More sophisticated architecture may bring in larger performance gains, but that is not the priority of this paper. Compared with recent deep-learning-based methods, better performance reflects the superiority of FPNet and Fourier prior. More importantly, our proposed FPNet shows great generalizability on real-world rainy datasets when only trained on synthetic datasets.

The main contributions of this work are as follows:

- We propose a novel prior - the Fourier prior, which contains the physical properties of rainy images across different domains.
- We implement a simple but effective two-stage deraining network, where two effective residual blocks are introduced to facilitate the Fourier amplitude and phase spectrum transformations.
- Experiments demonstrate that our method outperforms the state-of-the-art methods on synthetic datasets and possesses better generalizability to real-world datasets.

2 Related Works

2.1 Single Image Deraining

Generally, there are two kinds of single image deraining: model-driven and data-driven.

Model-driven approaches separate the rain layer from the background layer by exploring physical attributes and prior knowledge of rain streaks. For example, by using dictionary learning to analyze morphology, rain streaks are extracted in [Kang *et al.*, 2011]. to remove rain streaks in high frequency regions. In [Luo *et al.*, 2015], based on image patches, a discriminative sparse coding is proposed to obtain rain-free content. And a Gaussian mixture models based patch prior [Li *et al.*, 2016] is introduced to adapt to rain marks in multiple directions and scales. But efficient issues and poor adaptability limit these method .

With the rise of deep learning and CNNs, data-driven methods have dominated the single image rain removal task in recent years. To use properties of rain layers, a multi-branch network using multi-task learning (DerainNet) is proposed [Yang *et al.*, 2017]. In [Fu *et al.*, 2017a], the authors present a three-layer network on high-frequency parts to simplify the learning processing. In [Zhang and Patel, 2018], a density-aware multi-stream CNN (DIDMDN) is introduced, which can automatically recognize the rain-density information. With consideration of the correlations between different stages of rain removal, a recurrent context aggregation network (RESCAN) is proposed in [Li *et al.*, 2018]. Unlike the

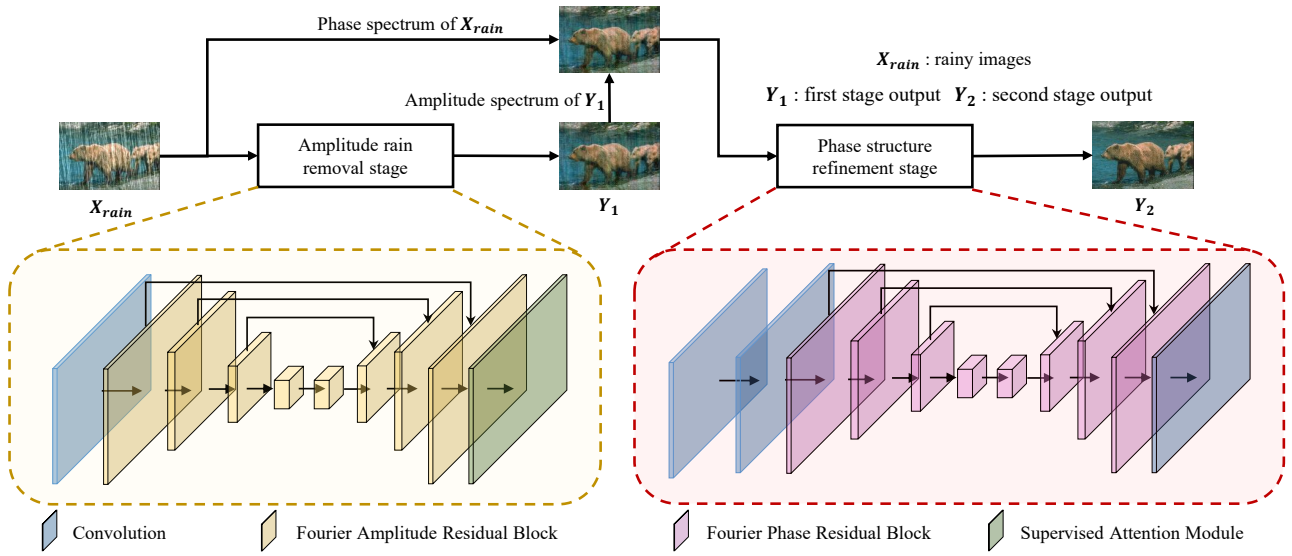

Figure 2: Overview of the proposed Fourier prior network (FPNet), which consists of the amplitude rain removal stage and phase structure refinement stage. The first stage restores the amplitude of rainy images to clean rain streaks, and the second stage restores the phase information to refine fine-grained background structures. In detail, there are 8 Fourier Amplitude Residual Blocks (FARBs) in the first stage and 8 Fourier Phase Residual Blocks (FPRBs) in the second stage. The images reconstructed with the amplitude of clean images and the phase of rainy images are used as the supervision signal of the first stage. And the clean images are used as the supervision signal of the second stage.

multi-scale network using cycle spinning (UMRL) [Yasarla and Patel, 2019], a progressive network (PReNet) is well-behaved with a simple architecture [Ren *et al.*, 2019]. The authors designed a semi-supervised transfer learning framework (SEMI) to tackle the domain adaptation in [Wei *et al.*, 2019]. By digging the correlations of rain streaks in an image, a multi-scale progressive fusion network (MSPFN) is introduced in [Jiang *et al.*, 2020]. In [Zamir *et al.*, 2021], a multi-stage progressive network (MPRNet) is proposed by further using features across each stage. On the basis of MPRNet, half instance normalization is introduced to be an important component of the network (HINet) in [Chen *et al.*, 2021]. Unlike the data-driven methods mentioned above, our approach introduces a novel prior that can improve the generalization ability of the network while taking advantage of the powerful mapping capabilities of CNNs.

2.2 Fourier Transform in Deep Learning

Some studies recently exploited Fourier transform on domain adaptation and domain generalization, demonstrating the importance of phase information. In [Yang *et al.*, 2020], the authors propose a phase consistency in source-target image translations, more competitive than the cycle consistency used for domain adaptation segmentation tasks. Similarly, in [Yang and Soatto, 2020], by replacing a small area in the centralized amplitude spectrum, images in the source domain become target-like. The work of [Xu *et al.*, 2021] explores the Fourier transform as a new type of data augmentation and achieves a remarkable performance. And their method is exchanging the Fourier amplitude and phase spectrum of different images from source and target domains. Different from the above work, we swap the amplitude and phase spectrum between paired clean/rainy images. By extensive observation, we summarize three characteristics and use them to conduct our network design.

3 Method

In this section, we will first introduce the definition of Fourier prior in Sec. 3.1. Then the network architecture and loss functions will be presented in Sec. 3.2 and Sec. 3.3.

3.1 Fourier Prior

Before introducing the Fourier prior, briefly reviewing the Fourier transformation is helpful for understanding our work. For an image $x \in R^{H\times W\times 1}$, its Fourier transformation $\mathcal{F}(x)$ is formulated as:

$$\mathcal{F}(x)(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w)e^{-j2\pi\left(\frac{h}{H}u+\frac{w}{W}v\right)}. \quad (1)$$

Accordingly, $\mathcal{F}^{-1}(x)$ is the inverse Fourier transformation. Both transformations can be implemented efficiently using the FFT algorithm in [Frigo and Johnson, 1998]. $\mathcal{A}(x)$ and $\mathcal{P}(x)$ are amplitude and phase spectrums expressed as:

$$\begin{aligned}\mathcal{A}(x)(u,v) &= \left[R^2(x)(u,v) + I^2(x)(u,v)\right]^{1/2},\\ \mathcal{P}(x)(u,v) &= \arctan\left[\frac{I(x)(u,v)}{R(x)(u,v)}\right],\end{aligned} \quad (2)$$

where $R(x)$ and $I(x)$ represent the real and imaginary part of $\mathcal{F}(x)$, respectively. For all images and feature maps in our method, the Fourier transformation is computed independently on each channel.

It is well-known that the phase component $\mathcal{P}$ preserves high-level semantics of the original signal, while the amplitude component $\mathcal{A}$ contains low-level statistics. Intuitively, in a rainy image, $\mathcal{P}$ and $\mathcal{A}$ would correspond to the image structures and rain streaks respectively. In order to evaluate this,

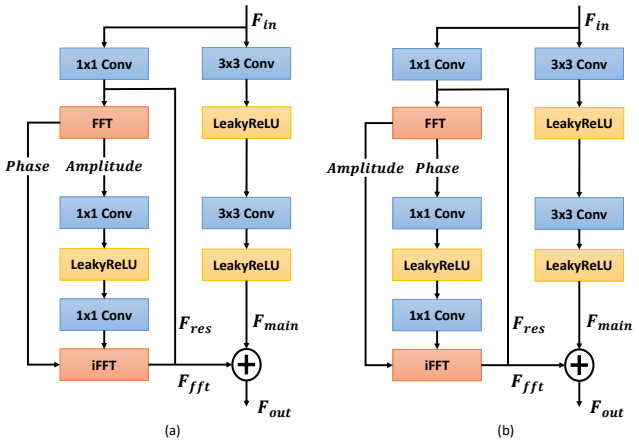

Figure 3: (a) Fourier Amplitude Residual Block; (b) Fourier Phase Residual Block.

we exchange the phase component $\mathcal{P}$ and amplitude component $\mathcal{A}$ of the rainy and clean images. Then, a series of common phenomena can be observed across synthetic and real-world datasets, as shown in Fig. 1. Let $\mathbf{X_{rain}}$ and $\mathbf{X_{clean}}$ represent the rainy image and clean image respectively. The rain streaks decrease in $\mathcal{F}^{-1}(\mathcal{A}(\mathbf{X_{clean}}), \mathcal{P}(\mathbf{X_{rain}}))$. Meanwhile $\mathcal{F}^{-1}(\mathcal{A}(\mathbf{X_{clean}}), \mathcal{P}(\mathbf{X_{rain}}))$ has the clean content similar to $\mathbf{X_{clean}}$. On the other hand, $\mathcal{F}^{-1}(\mathcal{A}(\mathbf{X_{rainy}}), \mathcal{P}(\mathbf{X_{clean}}))$ across synthetic and real-world domains have similar strip degradation which is evenly distributed in the whole image. From this, three important conclusions can be deduced: (1) the amplitude of rainy images contains most rain streaks information; (2) the phase of rainy images keeps fine-grained structures similar to ground truth; (3) the amplitude of rainy images shares the same physical characteristic across synthetic and real-world domains. So, the effective and generalizable deraining across synthetic and real-world domains can be performed on $\mathcal{A}$, while $\mathcal{P}$ retains the structural details of clean images. Based on these prior knowledges from Fourier transformation, it's natural and necessary to learn $\mathcal{A}(\mathbf{X_{rain}}) \rightarrow \mathcal{A}(\mathbf{X_{clean}})$ and $\mathcal{P}(\mathbf{X_{rain}}) \rightarrow \mathcal{P}(\mathbf{X_{clean}})$ separately in the image rain removal model.

3.2 Network Architecture

Fourier Prior Network

To exploit the Fourier prior, we design a simple but effective two-stage architecture - the Fourier prior network (FPNet), as shown in Fig. 2. Our FPNet consists of the amplitude rain removal stage and phase structure refinement stage. The first stage restores the amplitude of rainy images to clean rain streaks, and the second stage restores the phase information to refine fine-grained background structures. $\mathcal{F}^{-1}(\mathcal{A}(\mathbf{X_{clean}}), \mathcal{P}(\mathbf{X_{rain}}))$ are used as the supervision signal of the first stage. And $\mathbf{X_{clean}}$ are used as the supervision signal of the second stage. Specifically, we do not directly use the first stage output $\mathbf{Y_1}$ as the input of the second stage. Instead, we use $\mathcal{F}^{-1}(\mathcal{A}(\mathbf{Y_1}), \mathcal{P}(\mathbf{X_{rain}}))$ to avoid the loss of $\mathcal{P}(\mathbf{X_{rain}})$ since the phase information will be altered when passing through the network [Yang *et al.*, 2020]. By now, the Fourier prior has been embedded into our network.

The first stage consists of 8 Fourier Amplitude Residual Blocks (FARBs) and the second stage consists of 8 Fourier Phase Residual Blocks (FPRBs). And there are three downsamplings and three upsamplings in each stage. Furthermore, the cross-stage feature fusion and supervised attention module are introduced from [Zamir *et al.*, 2021] for the stable optimization procedure.

Fourier Amplitude and Phase Residual Block

To facilitate the transformation of the amplitude spectrum, we propose the Fourier amplitude residual block (FARB). Since the convolution is conducted in the spatial domain, it is non-trivial to directly operate in the amplitude spectrum. To avoid destructing convolutional operation in spatial features, the amplitude spectrum is refined in the residual branch. As shown in Fig. 3(a), before performing the fast Fourier transform (FFT), the input features $F_{in} \in R^{H \times W \times C_{in}}$ is fed into one 1×1 convolution layer to output $F_{res} \in R^{H \times W \times C_{out}}$. $\mathcal{A}(F_{res}) \in R^{H \times W \times C_{out}}$ and $\mathcal{P}(F_{res}) \in R^{H \times W \times C_{out}}$ is calculated by $\mathcal{F}(F_{res})$ in Eq. 2. Then, $\mathcal{A}(F_{res})$ is fed into two 1×1 convolution layers to output $\hat{\mathcal{A}}(F_{res})$. Subsequently, $F_{fft} = \mathcal{F}^{-1}(\hat{\mathcal{A}}(F_{res}), \mathcal{P}(F_{res}))$ can be calculated by using the inverse fast Fourier transform(iFFT) algorithm. Finally, the amplitude residual block can be expressed as:

$$F_{out} = F_{main} + F_{fft} + F_{res}, \tag{3}$$

where $F_{out} \in R^{H \times W \times C_{out}}$ is the output feature and $F_{main} \in R^{H \times W \times C_{out}}$ is the main-branch feature.

The Fourier phase residual block (FPRB) is similar to FARB, just swapping the operations between $\mathcal{A}$ and $\mathcal{P}$, as shown in Fig. 3(b).

3.3 Loss Function

The Mean Square Error (MSE) produces the over-smoothed image because of its squared penalty. Thus, Mean Absolute Error (MAE) is adopted to balance rain removal and detail preservation in our method.

Let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ represent the output of the first and second stage respectively. For the first stage, we minimize the following loss functions:

$$\mathcal{L}_1 = \left\| \mathbf{Y}_1 - \mathcal{F}^{-1}(\mathcal{A}(\mathbf{X}_{clean}), \mathcal{P}(\mathbf{X}_{rain})) \right\|_1, \tag{4}$$

$$\mathcal{L}_{amp} = \left\| \mathcal{A}(\mathbf{Y}_1) - \mathcal{A}(\mathbf{X}_{clean}) \right\|_1, \tag{5}$$

where $\mathcal{L}_1$ and $\mathcal{L}_{amp}$ are performed in the space and frequency domain, respectively. For the second stage, we refine the phase spectrum and approximate clean images by using loss functions similiar to the first stage as follows:

$$\mathcal{L}_2 = \left\| \mathbf{Y}_2 - \mathbf{X}_{clean} \right\|_1, \tag{6}$$

$$\mathcal{L}_{fft} = \left\| \mathcal{F}(\mathbf{Y}_2) - \mathcal{F}(\mathbf{X}_{clean}) \right\|_1. \tag{7}$$

The overall loss function used to train FPNet is formulated as follows:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \lambda_{amp}\mathcal{L}_{amp} + \lambda_{fft}\mathcal{L}_{fft}, \tag{8}$$

where λ_{amp} and λ_{fft} are trade-off parameters, both set to 0.05 empirically.

Methods	Test100		Rain100H		Rain100L		Test2800		Test1200		Averages	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DerainNet [Yang *et al.*, 2017]	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 [Wei *et al.*, 2019]	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 [Zhang and Patel, 2018]	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 [Yasarla and Patel, 2019]	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 [Li *et al.*, 2018]	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 [Ren *et al.*, 2019]	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 [Jiang *et al.*, 2020]	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903
MPRNet [Zamir *et al.*, 2021]	30.27	0.897	30.41	0.890	36.40	0.965	33.64	0.938	32.91	0.916	32.73	0.921
HINet [Chen *et al.*, 2021]	30.29	0.906	30.65	0.894	37.28	0.970	33.91	0.941	33.05	0.919	33.03	0.926
FPNet (Ours)	**30.86**	**0.915**	**30.89**	**0.897**	**37.96**	**0.972**	**34.07**	**0.943**	**33.08**	**0.924**	**33.37**	**0.930**

Table 1: Comparison of average PSNR and SSIM values on synthetic benchmark datasets. The best results are **boldfaced**. Our FPNet not only attains the best performance on each dataset, but also achieves 0.34dB improvement compared to the recent best method HINet

4 Experiment

4.1 Datasets and Settings

Synthetic Data

As in [Jiang *et al.*, 2020; Zamir *et al.*, 2021], we train our model on $13,712$ clean-rain image pairs (for simplicity, denoted as Rain13k in the following) gathered from multiple synthetic datasets. With this single trained model, we perform evaluation on various test sets, including Rain100H [Yang *et al.*, 2017], Rain100L [Yang *et al.*, 2017], Test100 [Zhang *et al.*, 2019], Test2800 [Fu *et al.*, 2017b], and Test1200 [Zhang and Patel, 2018].

Real-world Data

To test the performance of our method in real scenarios, we conduct experiments on the testing samples of recent public real-world rainy dataset (for simplicity, denoted as RS in the following) [Quan *et al.*, 2021], which contains nearly 150 rainy/clean image pairs for training and 100 pairs for testing. All the images collected use a DSLR camera to capture rain images. Rain streaks are generated by spraying water, which is widely used to mimic rainy scenes in film industry.

Implementation Details

The networks are trained with Adam optimizer. The learning rate is set to 2×10^{-4} by default, and decreased to 1×10^{-8} with cosine annealing strategy. our models are trained on 256×256 patches with a batch size of 64 for 8×10^5 iterations. Random flipping and rotation are applied as data augmentation. PSNR and SSIM are used for the quantitative evaluation. Note that as the human visual system is sensitive to the Y channel of a color image in YCbCr space, PSNR and SSIM are computed based on this luminance (Y) channel [Yang *et al.*, 2017].

4.2 Evaluations on Synthetic Data

Table 1 reports the quantitative results with recent state-of-the-art methods. The proposed FPNet not only achives the best performance on each dataset, but also achieves 0.34dB improvement compared to the recent best method HINet [Chen *et al.*, 2021] when averaged across all datasets. Fig. 4 illustrates that FPNet preserves more fine-grained structures, with powerful deraining performance. This phenomenon shows that keeping the phase information unchanged as the second stage input assists in detail recovery.

Methods	RS	
	PSNR↑	SSIM↑
MSPFN [Jiang *et al.*, 2020]	24.76	0.691
MPRNet [Zamir *et al.*, 2021]	25.07	0.700
HINet [Chen *et al.*, 2021]	24.71	0.693
FPNet (Ours)	**25.34**	**0.707**

Table 2: Comparison of average PSNR and SSIM values on RS without training on corresponding real-world images. Our method achieves 0.27dB improvement compared to the second-best method. This illustrates the generalizability of our method directly.

4.3 Generalization on Real-world Data

To test our method's generalizability, the RS dataset is evaluated by using the model trained on Rain13k *without retraining*. As shown in Fig. 5, more real rain streaks are removed by our method while retaining reliable content information and details. At the same time, the results reported in Table 2 prove the favorable generalizability of FPNet, achieving 0.27dB improvement compared with the second-best method (MPRNet). This illustrates the Fourier prior can help the network learn the intrinsic characteristics of rainy images across real-world and synthetic domains.

4.4 Ablation Study

Fourier Prior

To validate the effectiveness of the Fourier prior, we design three variants of networks: Model 1 represents performing $\mathcal{P}(\mathbf{X_{rain}}) \rightarrow \mathcal{P}(\mathbf{X_{clean}})$ first and $\mathcal{A}(\mathbf{X_{rain}}) \rightarrow \mathcal{A}(\mathbf{X_{clean}})$ second; Model 2 represents performing $\mathbf{X_{rain}} \rightarrow \mathbf{X_{clean}}$ at both stages; Model 3 represents performing $\mathcal{A}(\mathbf{X_{rain}}) \rightarrow \mathcal{A}(\mathbf{X_{clean}})$ first and $\mathcal{P}(\mathbf{X_{rain}}) \rightarrow \mathcal{P}(\mathbf{X_{clean}})$ second, but using Y_1 as the input of the second stage.

Methods	Rain13K		RS	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Model 1	31.34	0.912	24.65	0.683
Model 2	32.65	0.919	24.69	0.685
Model 3	33.13	0.927	25.22	0.703
FPNet	**33.37**	**0.930**	**25.34**	**0.707**

Table 3: Evaluation on the Fourier prior. The results show that our Fourier prior is very effective for improving generalizability and performance of image deraining.

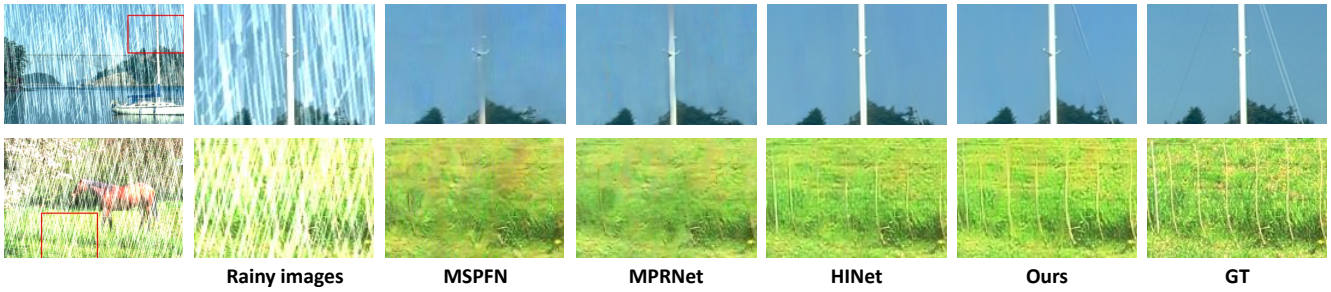

Figure 4: Visual comparisons on synthetic datasets. This shows that our method has a powerful ability to preserve details and restore content.

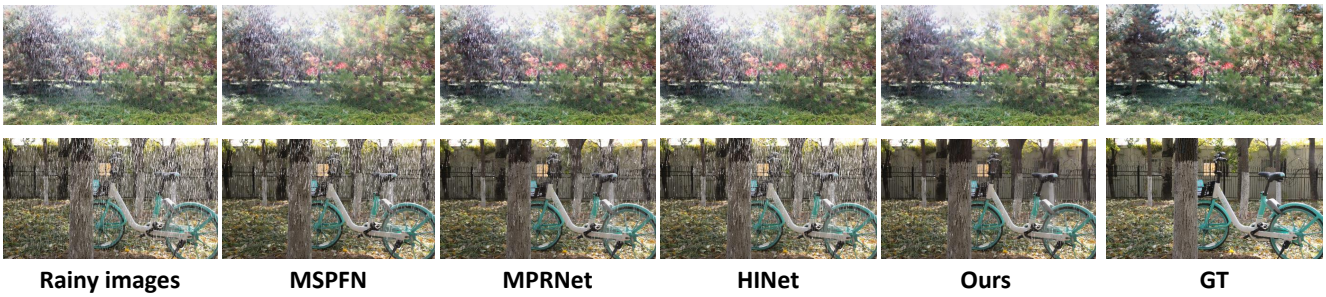

Figure 5: Visual comparisons on real-world datasets. The model is only trained on Rain13k without retraining on real rainy images. This shows that the Fourier prior has a powerful generalization ability across real-world and synthetic domains.

Methods	Rain13K PSNR↑	Rain13K SSIM↑	RS PSNR↑	RS SSIM↑
Baseline	32.93	0.924	25.03	0.697
Baseline + FARB	33.12	0.925	25.17	0.704
Baseline + FPRB	33.05	0.924	25.09	0.701
FPNet	**33.37**	**0.930**	**25.34**	**0.707**

Table 4: Ablation study on the effect of FARB and FPRB.

Methods	MACs(G)↓	Time(ms)↓
MSPFN [Jiang *et al.*, 2020]	604.70	117.3
MPRNet [Zamir *et al.*, 2021]	141.28	28.3
HINet [Chen *et al.*, 2021]	170.71	19.8
FPNet (Ours)	**126.19**	**17.5**

Table 5: Comparison on MACs and runtime. MACs and Time are estimated with the input size of $3 \times 256 \times 256$ the Nvidia 3090 GPU.

As shown in Table 3, Model 1 presents the worst performance. It's hard for Model 1 to learn the transformation since $\mathcal{F}^{-1}(\mathcal{A}(\mathbf{X_{rain}}), \mathcal{P}(\mathbf{X_{clean}}))$ is full of degradation across the whole image, which means the first sub-network is not fully utilized. Second, Model 3 surpasses Model 2 with 0.53dB on RS dataset. This verifies the effectiveness of the proposed architecture. Third, FPNet shows better performance than Model 3, meaning that keeping $\mathcal{P}(\mathbf{X_{rain}})$ invariant play an significant role in the Fourier prior. All of the above results reflects the superiority of FPNet and Fourier prior.

FARB & FPRB

To verify the effectiveness of FARB and FPRB, we ablate the usage of FARB and FPRB by utilizing the regular residual block. Table 4 shows our proposed FARB and FPRB significantly outperform the regular residual block. The results indicate that both modules are the key ingredients of FPNet.

Complexity

To validate the efficiency of our method, we test the MACs and inference time using 100 images with a size of $3 \times 256 \times 256$ on one Nvidia 3090 GPU. Comparisons on average MACs and run-time are shown in Table 5. The results show that our FPNet is able to achieve efficient calculation consumption and inference time.

5 Conclusion

In this paper, we present the Fourier prior of rainy images that are effective for improving generalizability and performance and designed Fourier prior network for single image deraining. The experiments on synthetic and real-world rainy images illustrate our method can restore clean images from rainy images with contents and details well preserved. But if the image is degraded severely by heavy rain, the phase spectrum cannot preserve the original image structure information well. Despite this limitation, the effectiveness of our method suggests that physical priors are still worth exploring, and more attention can be paid to the frequency domain for single image rain removal.

Acknowledgements

This work was supported by the National Key R&D Program of China under Grant 2020AAA0105702, the National Natural Science Foundation of China (NSFC) under Grants U19B2038 and 61901433, the University Synergy Innovation Program of Anhui Province under Grants GXXT-2019-025, the Fundamental Research Funds for the Central Universities under Grant WK2100000024.

References

[Chen *et al.*, 2021] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *CVPR*, pages 182–192, 2021.

[Comaniciu *et al.*, 2003] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Transactions on pattern analysis and machine intelligence*, 25(5):564–577, 2003.

[Frigo and Johnson, 1998] Matteo Frigo and Steven G Johnson. Fftw: An adaptive software architecture for the fft. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, volume 3, pages 1381–1384. IEEE, 1998.

[Fu *et al.*, 2017a] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017.

[Fu *et al.*, 2017b] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *CVPR*, pages 3855–3863, 2017.

[He *et al.*, 2010] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[He *et al.*, 2017] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[Jiang *et al.*, 2020] 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 *CVPR*, pages 8346–8355, 2020.

[Kang *et al.*, 2011] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing*, 21(4):1742–1755, 2011.

[Li *et al.*, 2016] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *CVPR*, pages 2736–2744, 2016.

[Li *et al.*, 2018] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, pages 254–269, 2018.

[Luo *et al.*, 2015] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *ICCV*, pages 3397–3405, 2015.

[Oppenheim *et al.*, 1979] A Oppenheim, Jae Lim, Gary Kopec, and SC Pohlig. Phase in speech and pictures. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 632–637. IEEE, 1979.

[Quan *et al.*, 2021] Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing raindrops and rain streaks in one go. In *CVPR*, pages 9147–9156, 2021.

[Ren *et al.*, 2019] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, pages 3937–3946, 2019.

[Ronneberger *et al.*, 2015] 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*, pages 234–241. Springer, 2015.

[Wang *et al.*, 2020] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3103–3112, 2020.

[Wei *et al.*, 2019] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *CVPR*, pages 3877–3886, 2019.

[Xu *et al.*, 2021] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *CVPR*, pages 14383–14392, 2021.

[Yang and Soatto, 2020] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *CVPR*, pages 4085–4095, 2020.

[Yang *et al.*, 2017] 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 *CVPR*, pages 1357–1366, 2017.

[Yang *et al.*, 2020] Yanchao Yang, Dong Lao, Ganesh Sundaramoorthi, and Stefano Soatto. Phase consistent ecological domain adaptation. In *CVPR*, pages 9011–9020, 2020.

[Yasarla and Patel, 2019] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, pages 8405–8414, 2019.

[Zamir *et al.*, 2021] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021.

[Zhang and Patel, 2018] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018.

[Zhang *et al.*, 2019] 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, 2019.