

# Towards Scale-Free Rain Streak Removal via Self-Supervised Fractal Band Learning

Wenhan Yang<sup>1</sup>, Shiqi Wang<sup>1</sup>, Dejie Xu<sup>2</sup>, Xiaodong Wang<sup>3</sup>, Jiaying Liu<sup>2\*</sup>

<sup>1</sup>City University of Hong Kong <sup>2</sup>Peking University <sup>3</sup>Beijing Institute of Electronic Engineering  
wyang34@cityu.edu.hk shiqwang@cityu.edu.hk dejia@pku.edu.cn  
casic\_wangxiaodong@126.com liujiaying@pku.edu.cn

## Abstract

Data-driven rain streak removal methods, which most of rely on synthesized paired data, usually come across the generalization problem when being applied in real cases. In this paper, we propose a novel deep-learning based rain streak removal method injected with self-supervision to improve the ability to remove rain streaks in various scales. To realize this goal, we made efforts in two aspects. First, considering that rain streak removal is highly correlated with texture characteristics, we create a fractal band learning (FBL) network based on frequency band recovery. It integrates commonly seen band feature operations with neural modules and effectively improves the capacity to capture discriminative features for deraining. Second, to further improve the generalization ability of FBL for rain streaks in various scales, we add cross-scale self-supervision to regularize the network training. The constraint forces the extracted features of inputs in different scales to be equivalent after rescaling. Therefore, FBL can offer similar responses based on solely image content without the interleave of scale and is capable to remove rain streaks in various scales. Extensive experiments in quantitative and qualitative evaluations demonstrate the superiority of our FBL for rain streak removal, especially for the real cases where very large rain streaks exist, and prove the effectiveness of its each component. Our code will be public available at: <https://github.com/flywh/AAAI-2020-FBL-SS>.

## Introduction

Bad weather conditions bring about a series of visibility degradations, *e.g.* occluding background scenes, altering the object content and changing contrast and color of images, *etc.* Due to detail loss and signal distortion, these degradations cause visual unpleasure and result in the failure of many outdoor computer vision applications, which built on taking high quality clean video frames as their input. As one

\*Corresponding author. This work is partially supported by National Natural Science Foundation of China under contract No.61772043, in part by Beijing Natural Science Foundation under contract No.L182002, in part by the Hong Kong RGC Early Career Scheme under Grant 9048122 (CityU 21211018), and in part by Hong Kong ITF UICP under Grant 9440203.  
Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

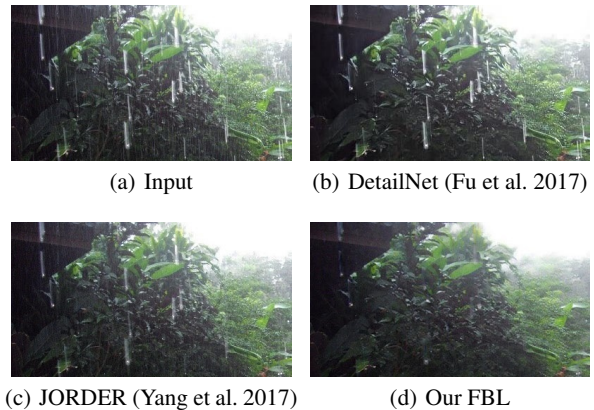


Figure 1: Visual comparison of our fractal band learning (FBL) with state-of-the-art approaches. In addition to removing small rain streaks as other methods, our FBL also removes most of the larger rain streaks.

of the most common degradations in rain frames, rain streaks cause severe intensity changes and light fluctuations, and hence obstruct and blur the background scene. An example of a rain streak image can be observed in Fig. 1(a).

As many degradation factors cause information loss, it is highly ill-posed to address single image rain removal problem. Previous methods (Kang, Lin, and Fu 2012; Huang et al. 2014; Sun, Fan, and Wang 2014; Luo, Xu, and Ji 2015) address the single image rain removal problem by separating rain streaks and background images, namely the corresponding rain-free versions, from their mixed versions. Various models are developed to extract rain streaks and background images based on their texture appearance patterns, such as frequency domain representation (Kang, Lin, and Fu 2012), sparse representation (Luo, Xu, and Ji 2015), and Gaussian mixture model (Li et al. 2016).

Recently, the rising of deep learning gives rise to deep-network based approaches. In (Fu et al. 2017), the image detail layer without background interference is regarded as the input, which directly reduces the mapping range from input to output and makes the learning process easier. In (Yang

et al. 2017b), a deep network is designed to jointly detect and remove rain streaks for heavy rain removal. Successive works (Zhang and Patel 2018; Li et al. 2018b) make great efforts to make networks more effective and efficient.

These methods achieve good performance in some cases. However, they still neglect some important issues:

- The degradations of rain scenes in real-world are very complex. Existing rain models often neglect the diversity of rain scales. A model trained with streaks of a scale is hardly to be generalized to handle streaks of a different one. Some previous works (Li, Cheong, and Tan 2017; Yang et al. 2019) make preliminary attempts. However, they either use a few times parameters of a single-scale model (Li, Cheong, and Tan 2017) or are restricted to process the images of a certain scale whose magnitude is the order of 2 compared to that in the training phase (Yang et al. 2019). However, a more economic and flexible framework that is easily to be generalized to provide an improved scale-robust rain streak removal is still absent.
- Recent deep-learning based methods take pure feed-forward CNN, ResNet or DenseNet *etc.* as backbone methods. However, there is no theoretic or conceptual connection between these neural models and traditional theories. This blank makes us away from injecting task-driven image priors into a model and developing novel competitive backbones.
- Models in previous data-driven methods are seldom designed to capture frequency band dependency of images explicitly. Intuitively, rain removal is a signal separation problem, where the features in the texture or frequency band domain play an important part. Thus, the properties of traditional band theory provides meaningful guidance to design a deraining network and facilitate a more effective automatic feature learning.

Considering these limitations of existing works, we aim to design a deep learning architecture inspired by frequency band recovery theory and is capable of effectively restoring clean images from inputs which contain rain streaks in different scales that may not appear in the training set. To realize that, we first briefly illustrate the connections between frequency band recovery and deep learning. Then, a band learning network is constructed to integrate most of common frequency band feature operations, *i.e.* *band refinement*, *expansion*, and *fusion* for rain streak removal. Wiring them together in a unified network facilitates the band-constrained automatic feature learning, leading to superior modeling capacity. Our network can be further augmented into a fractal band learning (FBL) network by being stacked in a fractal form. That is, the low-order constructed modules are used as the basic unit of a high-order one, which is beneficial to capturing potential hierarchical dependency among band features. To further improve the generalization ability of FBL to remove rain streaks in different scales, a cross-scale self-supervision is utilized to regularize network training. FBL forces consistency of the extracted features of inputs in different scales after zooming. Therefore, FBL can acquire cross-scale feature representation given input images in different scales. With the learned band feature representations

and the power of cross-scale self-supervision, our FBL is superior to previous deraining networks and is capable to remove rain streaks in different scales. Our contributions are summarized as follows,

- A cross-scale self-supervision constraint is proposed to regularize the training of a deraining network. The constraint guides the network to extract more scale-robust features and deal with rain images containing rain streaks in different scales. It can significantly benefit the rain removal on real rain images.
- We provide a novel understanding of the link between frequency band recovery and deep learning. The conceptual links between common band operations and learned neural modules are revealed.
- Based on the above understanding, we develop an FBL network based on frequency band structures, and the joint consideration in spatial and frequency domains. It performs band operations progressively and is further augmented by being stacked in a fractal form. Extensive experiments demonstrate the superiority of FBL for rain streak removal objectively and subjectively.

## Related Works

### Single Image Rain Removal

Single image deraining is highly ill-posed and is usually addressed via a signal separation or texture classification route. A series of non-deep learning-based methods (Kang, Lin, and Fu 2012; Chen and Hsu 2013; Kim et al. 2013; Luo, Xu, and Ji 2015; Li et al. 2016) are built based on statistical models. In (Fu et al. 2017), a deep network is taken to map the extracted texture images to the negative residues. In (Yang et al. 2017b), a deep joint rain detection and removal method was proposed to recurrently remove rain streaks and accumulation. In (Li, Cheong, and Tan 2017), to treat the rain streaks differently, several parallel sub-networks are trained with rain streaks in different scales. Zhang *et al.* (Zhang and Patel 2018) integrated rain-level estimation and rain streak extraction in a multi-path dense network. Later works focus on proposing advanced deep networks (Li et al. 2018a; Wang et al. 2019), exploiting more effective priors (Hu et al. 2019; Chang, Yan, and Zhong 2017; Zhu et al. 2017), or collecting more comprehensive and real datasets (Li et al. 2019; Wang et al. 2019) to facilitate rain removal. Following these works, our work also focuses on deep-learning based single image rain streak removal. Differently, we construct a deep network motivated by band recovery theory. The domain shift in streak scales between training and testing phases is addressed by network training with self-supervision.

### Deep Learning-Based Image Processing

The rise of the ResNet and DenseNet advances the development of low-level image processing tasks. It has been observed that refining features progressively like ResNet (He et al. 2016) or concatenating and fusing features from different levels like DenseNet (Huang et al. 2017) leads to better representations of pixels and their contexts for low-level visions. In (Yang et al. 2017a), the connection between ResNet

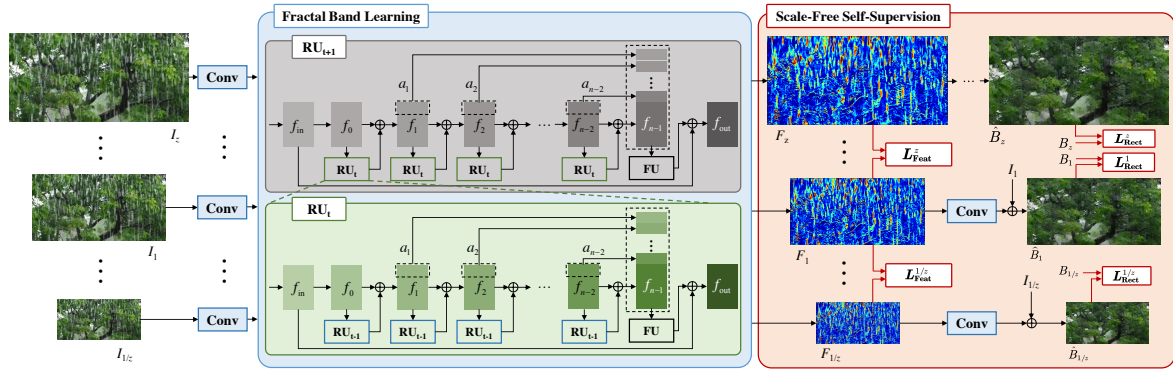


Figure 2: The framework of our proposed fractal band learning (FBL) for rain removal. It progressively performs learned band feature operations progressively and augmented by being stacked in a fractal form. It is trained with both image-level reconstruction constraint and feature-level cross-scale self-supervision loss.

and traditional band filter recovery is presented. In (Ledig et al. 2017), ResNet is used as the generator of the generative adversarial network. Afterwards, several improved models are built, *e.g.* multi-path ResNet (Yang et al. 2017b), multi-stream densely connected de-raining network (Zhang and Patel 2018), edge-preserving densely connected network for dehazing (Zhang et al. 2018a), MemNet capturing the human persistent memory for restoration (Tai et al. 2017a), combination of ResNet and DenseNet for super-resolution (Zhang et al. 2018b), *etc.* In our work, we summarize previous works from the view of band recovery theory. Based on our new understanding, we develop a new effective backbone for rain streak removal.

## Fractal Band Learning

### From Band Recovery to Deep Learning

In (Yang et al. 2017a), a systematic paradigm is provided to connect ResNet (He et al. 2016) and band recovery. The signal is reconstructed progressively with two steps: 1) residual block  $F_{RB}$  first generates a new band signal  $f_t$  based on a previous band estimator  $f_{t-1}$ . A summation is used to combine the new band signal (residue)  $F_{RB}(f_{t-1})$  and the input (previous band estimators)  $f_t$ :

$$f_t = F_{RB}(f_{t-1}) + f_{t-1}, \quad (1)$$

where  $F_{RB}(\cdot)$  denotes the process of residual block. This analysis only covers parts of band recovery characteristics. When we review the classical band filter representations, *e.g.* wavelet (Edwards 1992), steerable filter (Freeman and Adelson 1991), ringlet (Do and Vetterli 2005), and the related processing methods based on them (Song et al. 2016; Singh and Ahuja 2015), four critical properties are revealed:

- **Band refinement.** The given band features extracted from the source domain, *e.g.* a rain image, are transformed into the target domain, *e.g.* a rain-free image. As the analysis above, ResNet works this way.
- **Band expansion.** One band feature is split into several band features, and the total band feature number is increased. The popular DenseNet (Huang et al. 2017) pro-

vides this capacity. However, it will rapidly increase the number of used parameters. Thus, in our work, we choose a more parameter-economic way – directly forwarding parts of features to a certain layer, where all forwarded features are concatenated as a broader one.

- **Band fusion.** Several band features are combined into a more compact one. This can be achieved effectively by  $1 \times 1$  convolutions.
- **Hierarchical dependency** among different bands. Many of hand-crafted frequency band features are organized into a hierarchical structure. High-order bands usually have intrinsically potential connections to some low-order bands. We use fractal architectures, *i.e.* (Larsson, Maire, and Shakhnarovich 2017), to follow this rule.

In our work, we fully consider these properties and build our FBL based on their corresponding deep modules.

### Overview of Fractal Band Learning

The framework of our proposed FBL for rain streak removal is illustrated in Fig. 2:

- **Fractal Band Learning (FBL).** We hope to build a deep network based on band recovery theory to capture the hierarchical band dependency using end-to-end trainable components. Motivated by the basic elements of band operations, FBLs conduct *band feature refinement*, *expansion* and *fusion* operations in a band recovery paradigm, as shown in Fig. 2. To model the hierarchical band dependency, FBL is designed to have a fractal structure. The constructed refine unit ( $RU_t$ ) is used as the basic unit of a higher-order one ( $RU_{t+1}$ ). In  $RU_t$ , a band feature is refined progressively in a residual learning manner. The residue bands are generated by refine units (RU). Then, parts of each residue band are forwarded into the last layer, where they are concatenated with the output band feature of the penultimate layer and fused into a narrow band by a fusion unit (FU). On the top of Fig. 2, changes of the band feature dimension in the learning process are provided. The summation operations in dash boxes signify

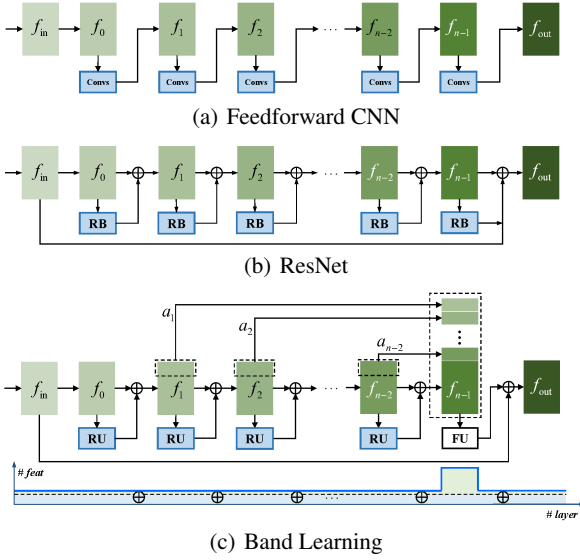


Figure 3: The framework of our band learning. The left side presents network structures. The bottom side illustrates band feature dimension changes during the learning process.

residual additions. From left to right, the band features are refined. Band features are expanded and fused in turn at the penultimate layer.

- **Cross-Scale Self-Supervision.** To handle scale variance of rain streaks, we impose additional constraints on the network such that extracted features tend to be scale-invariant for different scales. The basic idea is that the network should have similar responses when the input rainy images are with the same content but in different scales. In other words, the network should process rainy images in the same way no matter what scale the images are at. Therefore, for a given input image  $I_1$ , we may up-sample it into  $I_z$  or down-sample it into  $I_{1/z}$ , and then enforce the extracted feature  $F_z$  to be similar to  $F_1$  after down-sampling, or  $F_1$  to be similar to  $F_{1/z}$  after down-sampling as shown in Fig. 2. After training with this constraint, the extracted features are more scale-invariant and the model is better at removing rain streaks in different scales.

### Fractal Band Learning

Rain streak removal is intrinsically a texture-related signal separation problem. Thus, we seek to utilize the properties of traditional band recovery to guide the network design and facilitate a more effective automatic feature learning with band constraints. Four properties are of great importance: *band refinement*, *expansion*, *fusion* and *hierarchical dependency*. In this section, we build our FBL with these rules step by step.

**Feed-Forward CNN** The network structure used in SRCNN (Dong et al. 2015), as shown in Fig. 3 (a). The feature  $f_t$  is transformed through a chain of convolutions:

$$f_t = \sigma(W_t * f_{t-1} + b_t), \quad (2)$$

where  $*$  and  $\sigma$  denote the convolution operation and ReLU activation function, respectively.  $W_t$  and  $b_t$  are the weights and biases of the  $t$ -th convolutional layer, respectively. Even in this simplest way, stacking more layers usually lead to an improved network capacity.

However, increasing the layer number may face the problem of vanishing gradient in the learning. In other words, it is hard to pass tiny errors from very deep layers back to shallow layers, which causes the trained models to have difficulties in high frequency detail reconstruction.

**ResNet** To characterize tiny structures in signals, residual network is employed in VDSR (Kim, Lee, and Lee 2016) and DEGREE (Yang et al. 2017a). As shown in Fig. 3 (b), each residual block extracts a residue and adds it with the original signal:

$$f_t = F_{RB}(f_{t-1}) + f_{t-1}, \quad (3)$$

where  $F_{RB}(\cdot)$  denotes the feature transform of residual block. It can be set as stacked convolutions or more complicated modules. As discussed in (Yang et al. 2017a), this operation equals to *progressive band refinement*, which facilitates high frequency detail reconstruction.

**Band Learning** In ResNet, extracting a wide band feature, namely expanding the channel number of the whole network, is expensive. Thus, we propose a band learning to include the operations of *band expansion* and *fusion*. The features are refined by RU:

$$f_t = F_{RU}(f_{t-1}) + f_{t-1}, \quad (4)$$

where  $F_{RU}(\cdot)$  denotes the feature transform of RU, which usually consists of two or three stacked convolutions. Then, we perform *band expansion*.  $f_t$  is split into two band features  $f_t = [a_t, b_t]$ .  $a_t$  is forwarded to the penultimate layer by a skip connection. At the penultimate layer  $n - 1$ , all  $\{a_t | t = 1, 2, \dots, n - 2\}$  are concatenated with the last output feature  $f_{n-1}$ :

$$f_{\text{expansion}} = [a_1, a_2, a_3, \dots, a_{n-2}, f_{n-1}]. \quad (5)$$

After that, we conduct *band fusion* to transform  $f_{\text{expansion}}$  into a narrow one:

$$f_n = F_{FU}(f_{\text{expansion}}), \quad (6)$$

where  $F_{FU}(\cdot)$  denotes the feature transform of FU, which usually is set to a  $1 \times 1$  convolution. At last,  $f_n$  is added with the input feature  $f_{\text{in}}$  of the whole module to generate the output feature  $f_{\text{out}}$ :

$$f_{\text{out}} = f_{\text{in}} + f_n. \quad (7)$$

As observed from the bottom side of Fig. 3 (c), the information is updated regularly, expanded and fused at the end.

**Fractal Band Learning** So far, our module meets the following band properties – *refinement*, *expansion* and *fusion*. The potential *hierarchical structure* and *dependency* among band features are also important for signal modelling. Thus, we expand our band learning to fractal band learning. The architecture is presented in Fig. 2. The constructed module  $RU_t$  is used as RU of a high-order one  $RU_{t+1}$ . Our FBL performs band feature operations at different levels, and

the learned band features are extracted along the tree structure of FBL. Thus, it is capable of not only keeping global structures but also reconstructing the high-frequency details. From the view of band feature dimension changes (on the bottom of Fig. 3 (c)), the information is processed and flowed to the end. In Fig. 3 (a), the information is directly to the end. In Fig. 3 (c), the information is updated regularly, expanded and fused at different levels many times, thus a meaningful band feature is obtained.

### Cross-Scale Self-Supervision

For the input image  $I_1$  (the subscript denoting the relative scale of this image to the original one), besides constraining the generated result, which at the end is combined with  $I_1$  to approach  $B_1$ , we also enforce the rescaling version of the extracted features, *e.g.*  $P_z, P_1, P_{1/z}$ , to be similar. We first define three operations  $F_{\text{FBL}}(\cdot)$ ,  $F_{\text{up}}(\cdot)$ , and  $F_{\text{down}}(\cdot)$  to denote the processes of transforming features by FBL, up-sampling and down-sampling operations, respectively. Then, we rescale the input image  $I_1$  randomly with a scaling factor ( $z = 2, 3, 4$ ) for both down-sampling and up-sampling into a zooming-out image  $I_{1/z}$  or zooming-in one  $I_z$  as follows,

$$\begin{aligned} I_z &= F_{\text{up}}(I_1, z), \\ I_{1/z} &= F_{\text{down}}(I_1, z). \end{aligned} \quad (8)$$

Then, these input images  $I_1, I_{1/z}$ , and  $I_z$  are feed-forwarded into FBL to obtain the corresponding features  $P_1, P_{1/z}$ , and  $P_z$  as follows,

$$\begin{aligned} P_1 &= F_{\text{FBL}}(I_1), \\ P_{1/z} &= F_{\text{FBL}}(I_{1/z}), \\ P_z &= F_{\text{FBL}}(I_z). \end{aligned} \quad (9)$$

Subsequently, their consistency is ensured as follows,

$$L_{\text{Feat}}(P_1, P_z) = \|P_1 - F_{\text{down}}(P_z, z)\|, \quad (10)$$

$$L_{\text{Feat}}(P_1, P_{1/z}) = \|F_{\text{down}}(P_1, z) - P_{1/z}\|. \quad (11)$$

It is worth mentioning that, to avoid the resolution loss in the rescaling process, we always down-sample a feature map to a smaller one in the consistency measurement.

### Training Loss

Suppose we have a collection of paired rain and rain-free images  $\{y^i, x^i\}_{i=1, \dots, N}$ , where  $N$  is the total number of training samples. Then, we rescale each sample using Bilinear interpolation to get  $\{y_z^i, x_z^i\}_{i=1, \dots, N}$  and  $\{x_{1/z}^i, x^i\}_{i=1, \dots, N}$  with random scaling factor  $z = \{2, 3, 4\}$ . For convenience of representations, we at first only focus on the cases where  $z > 1$ . We use  $\Theta, \hat{x}_z^i, P_z^i$  to denote all the parameters in  $F_{\text{FBL}}$ , the output by FBL and its generated feature. We use  $F_{\text{Streak}}(\cdot)$  to represent the process of extracting negative streaks from the input images, namely  $F_{\text{Rect}}(F_{\text{FBL}}(\cdot))$ . We adopt the following loss function to

Table 1: PSNR results among different methods.

Baseline	Rain100L	Rain100H	Rain800
ID	23.13	13.78	20.54
DSC	24.16	15.66	25.57
LP	29.11	14.26	27.09
CNN	23.70	13.21	23.95
SRCNN	32.63	18.29	25.10
DetailNet	33.50	23.93	25.22
UGSM	28.83	14.06	23.12
JCAS	29.91	14.26	22.25
DID-MDN	28.27	13.85	22.55
ID-CGAN	23.39	16.86	23.81
JORDER	36.11	24.10	26.73
RESCAN	38.58	26.06	28.04
RWL	36.75	26.89	27.79
FBL	<b>40.32</b>	<b>30.47</b>	<b>28.68</b>

train  $F_{\text{FBL}}(\cdot)$  and  $F_{\text{Rect}}(\cdot)$ :

$$\begin{aligned} L(\Theta) &= \frac{1}{2N} \sum_{i=1}^N \sum_z (\lambda_z (w_1 L_{\text{Rect},z}^i(\Theta) + w_2 L_{\text{Feat},z}^i(\Theta))), \\ L_{\text{Rect},z}^i(\Theta) &= \|F_{\text{Down}}(F_{\text{Streak}}(y_z^i; \Theta), z) + y^i - x^i\|, \\ L_{\text{Feat},z}^i(\Theta) &= \|F_{\text{Down}}(F_{\text{FBL}}(y_z^i; \Theta), z) - F_{\text{FBL}}(y^i; \Theta)\|, \end{aligned}$$

where  $\lambda_z, w_1, w_2$ , and  $w_3$  are weighting parameters considering both relative area of images/features and importances of different terms. When  $z < 1$ , *i.e.*  $z = \{1/2, 1/3, 1/4\}$ , the two loss terms become:

$$\begin{aligned} L_{\text{Rect},z}^i(\Theta) &= \|F_{\text{Streak}}(y_z^i; \Theta) + F_{\text{Down}}(y^i - x^i, 1/z)\|, \\ L_{\text{Feat},z}^i(\Theta) &= \|F_{\text{Down}}(F_{\text{FBL}}(y_z^i; \Theta), 1/z) - F_{\text{FBL}}(y^i; \Theta)\|. \end{aligned}$$

In this case, up-sampling operations are not involved.

### Connections to FractalNet and MemNet

Compared to FractalNet (Larsson, Maire, and Shakhnarovich 2017) and MemNet (Tai et al. 2017b), our RFBL further provides a comprehensive understanding of the link between frequency band recovery and deep learning, our is closer to frequency band operations and different in the following aspects. 1) In each RU of RFBL, only parts of output features are fed into the last layer, which shrinks the dimension of the last layer and enables an increasing layer depth to model more complex mappings. The extracted features have larger variances and therefore be more informative to capture high-frequency information. 2) In MemNet, layers of different recursive units are connected by gate units. Comparatively, the output of our refine unit (RU) has a direct path to all layers of the next RU, thus the features in two adjacent RUs are tied closer. 3) FractalNet is designed for classification while ours is for low-level visions. 4) In FractalNet, the fractal unrolling doubles the channel dimension while the channel dimensions of RFBL are cyclicly changed at different steps of fractal unrolling. It is closer to some frequency transforms, *e.g.* wavelet transform. At each scale, a band can be decomposed into a fixed number of sub-bands and vice versa.

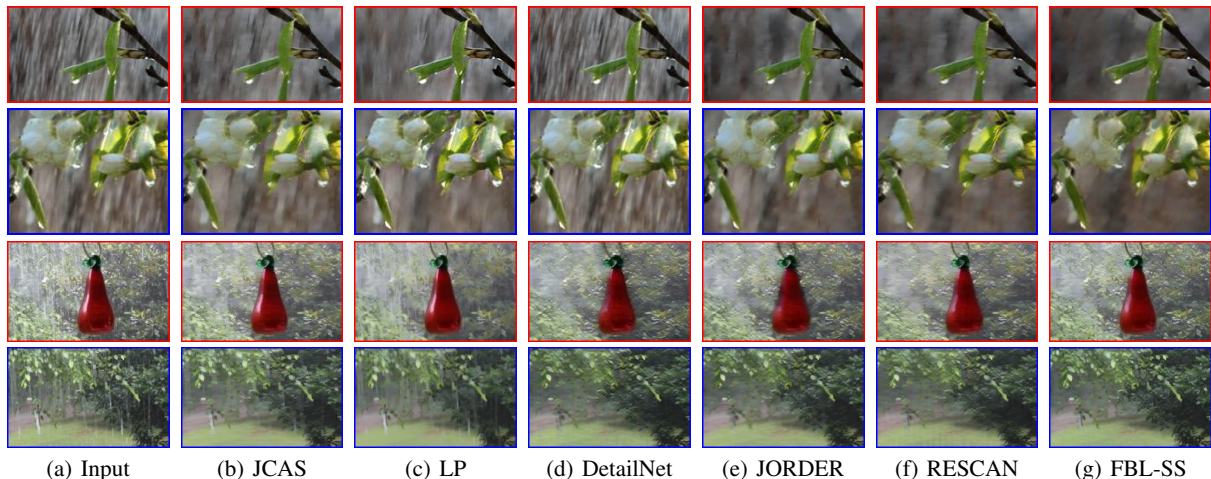


Figure 4: Visual comparison of different rain removal algorithms on real rain images. It is observed that, our FBL-SS successfully removes most rain streaks and better preserves texture details. Full-resolution results are provided in the supplementary material.

Table 2: SSIM results among different methods.

Baseline	<i>Rain100L</i>	<i>Rain100H</i>	<i>Rain800</i>
ID	0.7022	0.3968	0.6739
DSC	0.8728	0.5444	0.6521
LP	0.8786	0.4225	0.7801
CNN	0.8076	0.3712	0.6589
SRCNN	0.9357	0.6124	0.8232
DetailNet	0.9444	0.7251	0.8228
UGSM	0.8876	0.4454	0.7675
JCAS	0.9053	0.4837	0.7682
DID-MDN	0.8625	0.3748	0.7639
ID-CGAN	0.8275	0.4921	0.8072
JORDER	0.9705	0.7490	0.8683
RESCAN	0.9805	0.8128	0.8748
RWL	0.9754	0.8406	0.8795
FBL	<b>0.9863</b>	<b>0.8973</b>	<b>0.8875</b>

## Experiments

### Dataset

We compare our method with state-of-the-art methods on a few benchmark datasets: (1) *Rain100L* and *Rain100H* (Yang et al. 2017b), which are synthesized datasets with only one type of rain streaks and with three to five layers of rain streaks, respectively; (2) *Rain100H-S2* and *Rain100-S3* proposed in (Yang et al. 2019), synthesized with  $s$  rain streaks ( $s \in \{2, 3, 4, 5\}$ ) with different shapes and directions. The streak sizes are twice and three times as large as those in *Rain100H*, used for evaluating the performance when training and testing streaks have different sizes. (3) *Rain800*, a collection of diversified synthesized rain images from randomly selected outdoor images, which is split into testing set of 100 image and training set of 700 images.

### Baseline Methods

We compare our method with state-of-the-art methods: ID (Kang, Lin, and Fu 2012), DSC (Luo, Xu, and Ji 2015), LP (Li et al. 2016), CNN-based rain drop removal (Eigen,

Table 3: The ablation analysis for our FBL. R denotes refinement. E&F denotes band expansion and fusion. F denotes fractal structure. The evaluations are performed on *Rain100H*.

Methods	R	E&F	F	MS	PSNR	SSIM
Forward CNN	×	×	×	×	27.63	0.8518
ResNet	✓	×	×	×	28.35	0.8709
F-ResNet	✓	×	✓	×	28.83	0.8807
SBL	✓	✓	×	×	28.46	0.8734
FBL	✓	✓	✓	×	<b>29.02</b>	<b>0.8921</b>

Table 4: PSNR and SSIM results when training and testing streak sizes are different. FBL-SS: with self-supervision.

Baseline	<i>Rain100H-S2</i>		<i>Rain100H-S3</i>	
	PSNR	SSIM	PSNR	SSIM
DetailNet	15.98	0.6328	15.03	0.6410
JORDER	16.11	0.6464	15.21	0.6483
DID-MDN	14.36	0.5383	14.70	0.6142
RESCAN	17.13	0.6965	16.31	0.6787
JCAS	14.27	0.6350	14.07	0.5779
UGSM	14.03	0.6260	13.82	0.5569
FBL	18.51	0.7355	16.32	0.6860
R-CNN	19.45	0.7668	17.32	0.7481
RWL	20.83	0.8050	18.28	0.7824
FBL-SS	<b>22.25</b>	<b>0.8063</b>	<b>20.85</b>	<b>0.8177</b>

Krishnan, and Fergus 2013), DetailNet (Fu et al. 2017), DGSM (Deng et al. 2018), JCAS (Gu et al. 2017), DID-MDN (Zhang and Patel 2018), JORDER (Yang et al. 2017b), RESCAN (Li et al. 2018b), and RWL (Yang et al. 2019). DetailNet, RESCAN, and JORDER are retrained with the online available codes. Other methods are directly evaluated with the online available codes. RWL is implemented by ourselves.

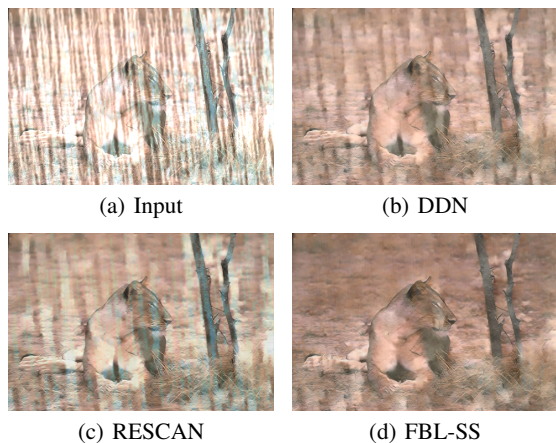


Figure 5: Evaluation results when training and testing streak scales are different.

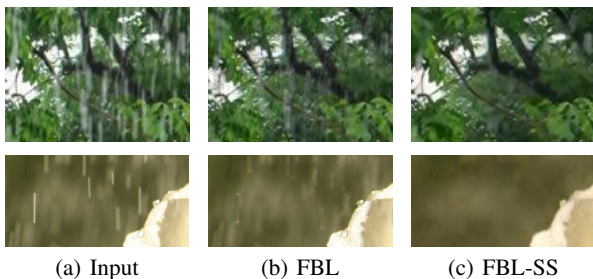


Figure 6: Ablation study for self-supervision. FBL: without self-supervision. FBL-SS: with self-supervision.

### Quantitative Evaluation

Tables 1 and 2 show the results of different methods on *Rain12*, *Rain100L* and *Rain100H*. As observed, our method considerably outperforms other methods in terms of both PSNR and SSIM. The PSNR of our FBL gains over RESCAN more than 3dB on *Rain100H* and almost 1dB on *Rain800*. Such a large gain demonstrates the effectiveness of proposed FBL on synthesized heavy rain images.

### Qualitative Evaluation

Fig. 4 shows the results of synthesized and real images. As observed, our method significantly outperforms previous state-of-the-art methods. Our FBL is better at removing large rain streaks (the first and second panels) and is superior in detail preservation (the third and fourth panels).

### Evaluation on Streak Size Mismatch

To prove the effectiveness of our FBL with self-supervision to handle the streak size mismatch problem between training and testing phases, we construct two testing sets, *Rain100H-S2* and *Rain100H-S3*. The streak sizes of these two sets are twice and three times of those in *Rain100H*, respectively. The testing results are provided in Table 4. It is observed that, FBL achieves superior performance than RWL, another

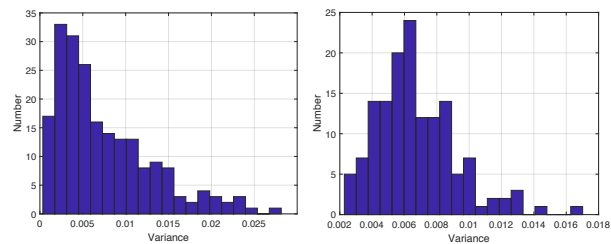


Figure 7: Comparison of extracted features in statistical distribution of DenseNet (left panel) and our FBL (right panel).

er state-of-the-art method that also aims to handle arbitrary-size rain streaks removal. The gain is more than 1.3 dB on *Rain100H-S2* and 2.5 dB on *Rain100H-S3*. Some visual results are provided in Fig. 5.

### Ablation Study for Network Structures

We compare six versions of our network: forward CNN, residual network (ResNet), fractal residual network (F-ResNet), sequential band learning (SBL), and fractal band learning (FBL). Their compositions and results are presented in Table 3. For a fair comparison, we make sure the parameters of these methods are almost the same. It is observed that, each component contributes to the final performance.

### Ablation Study for Self-Supervision

The cross-scale self-supervision provides the capacity of more “scale-robust” rain removal. We perform an ablation analysis on real images in Fig. 6. It is observed that, the results generated by FBL with self-supervision (c) obviously outperforms the one without it (b).

### Effectiveness in Frequency Information Extraction

We compare the penultimate features of RFBL and DenseNet with almost the same parameter number (193,283 VS. 195,283) in Fig. 7. We calculate the variances of normalized features. The feature variances of RFBL mainly distribute in [0.005, 0.01] and those of DenseNet distribute in [0, 0.005]. It is shown that, features of our RFBL are more informative.

### Conclusion and Discussions

In this paper, we design a fractal band learning network trained with self-supervision for scale-robust rain streak removal. A fractal band learning network is built to perform frequency band feature operations, offering more superior deraining performance. The learned features extracted by our FBL are constrained to be consistent after rescaling operations, which improves the robustness of the method in handling streaks in different scales. It is observed from the second row of Fig. 6 that, the edges and textures are a little blurred. In the future, there might be two directions to address the problem: 1) applying adversarial constraint on the result background; 2) developing more effective feature-level constraints to additionally measure the background blurriness and texture authenticity.

## References

- Chang, Y.; Yan, L.; and Zhong, S. 2017. Transformed low-rank model for line pattern noise removal. In *ICCV*.
- Chen, Y.-L., and Hsu, C.-T. 2013. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *ICCV*.
- Deng, L.-J.; Huang, T.-Z.; Zhao, X.-L.; and Jiang, T.-X. 2018. A directional global sparse model for single image rain removal. *Applied Mathematical Modelling*.
- Do, M. N., and Vetterli, M. 2005. The contourlet transform: an efficient directional multiresolution image representation. *TIP*.
- Dong, C.; Loy, C.; He, K.; and Tang, X. 2015. Image super-resolution using deep convolutional networks. *TPAMI*.
- Edwards, T. 1992. Discrete wavelet transforms: Theory and implementation.
- Eigen, D.; Krishnan, D.; and Fergus, R. 2013. Restoring an image taken through a window covered with dirt or rain. In *ICCV*.
- Freeman, W. T., and Adelson, E. H. 1991. The design and use of steerable filters. *TPAMI*.
- Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017. Removing rain from single images via a deep detail network. In *CVPR*.
- Gu, S.; Meng, D.; Zuo, W.; and Zhang, L. 2017. Joint convolutional analysis and synthesis sparse representation for single image layer separation. In *ICCV*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*.
- Hu, X.; Fu, C.-W.; Zhu, L.; and Heng, P.-A. 2019. Depth-attentional features for single-image rain removal. In *CVPR*.
- Huang, D.-A.; Kang, L.-W.; Wang, Y.-C. F.; and Lin, C.-W. 2014. Self-learning based image decomposition with applications to single image denoising. *TMM*.
- Huang, G.; Liu, Z.; v. d. Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *CVPR*.
- Kang, L. W.; Lin, C. W.; and Fu, Y. H. 2012. Automatic single-image-based rain streaks removal via image decomposition. *TIP*.
- Kim, J. H.; Lee, C.; Sim, J. Y.; and Kim, C. S. 2013. Single-image deraining using an adaptive nonlocal means filter. In *ICIP*.
- Kim, J.; Lee, J. K.; and Lee, K. M. 2016. Accurate image super-resolution using very deep convolutional networks. In *CVPR*.
- Larsson, G.; Maire, M.; and Shakhnarovich, G. 2017. Fractalnet: Ultra-deep neural networks without residuals. In *ICLR*.
- Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; and Shi, W. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*.
- Li, Y.; Tan, R. T.; Guo, X.; Lu, J.; and Brown, M. S. 2016. Rain streak removal using layer priors. In *CVPR*.
- Li, G.; He, X.; Zhang, W.; Chang, H.; Dong, L.; and Lin, L. 2018a. Non-locally enhanced encoder-decoder network for single image de-raining. In *ACMMM*.
- Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018b. Rescan: Recurrent squeeze-and-excitation context aggregation net. In *ECCV*.
- 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*.
- Li, R.; Cheong, L.-F.; and Tan, R. T. 2017. Single Image Deraining using Scale-Aware Multi-Stage Recurrent Network. *ArXiv e-prints*.
- Luo, Y.; Xu, Y.; and Ji, H. 2015. Removing rain from a single image via discriminative sparse coding. In *ICCV*, 3397–3405.
- Singh, A., and Ahuja, N. 2015. Super-resolution using sub-band self-similarity. In Cremers, D.; Reid, I.; Saito, H.; and Yang, M.-H., eds., *ACCV*.
- Song, S.; Li, Y.; Liu, J.; and Quo, Z. 2016. Joint sub-band based neighbor embedding for image super-resolution. In *ICASSP*.
- Sun, S.-H.; Fan, S.-P.; and Wang, Y.-C. F. 2014. Exploiting image structural similarity for single image rain removal. In *ICIP*.
- Tai, Y.; Yang, J.; Liu, X.; and Xu, C. 2017a. Memnet: A persistent memory network for image restoration. In *ICCV*.
- Tai, Y.; Yang, J.; Liu, X.; and Xu, C. 2017b. Memnet: A persistent memory network for image restoration. In *ICCV*.
- Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *CVPR*.
- Yang, W.; Feng, J.; Yang, J.; Zhao, F.; Liu, J.; Guo, Z.; and Yan, S. 2017a. Deep edge guided recurrent residual learning for image super-resolution. *TIP*.
- Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017b. Deep joint rain detection and removal from a single image. In *CVPR*.
- Yang, W.; Liu, J.; Yang, S.; and Guo, Z. 2019. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *TIP*.
- Zhang, H., and Patel, V. M. 2018. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*.
- Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; and Fu, Y. 2018a. Residual dense network for image super-resolution. In *CVPR*.
- Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; and Fu, Y. 2018b. Residual dense network for image super-resolution. In *CVPR*.
- Zhu, L.; Fu, C.; Lischinski, D.; and Heng, P. 2017. Joint bi-layer optimization for single-image rain streak removal. In *ICCV*.